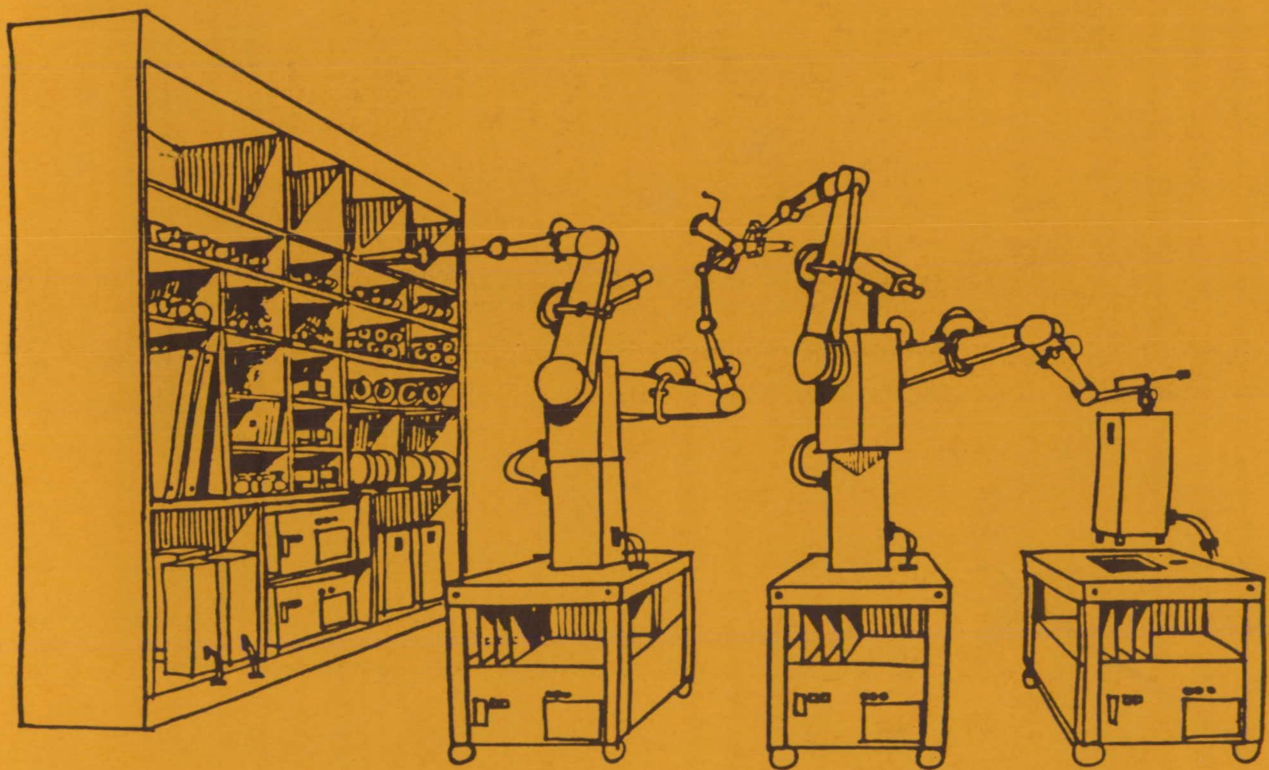
NASA Conference Publication 2255

Advanced Automation for Space Missions



Proceedings of the 1980 NASA/ASEE
Summer Study held at the
University of Santa Clara
Santa Clara, California
June 23-August 29, 1980

NASA

NASA Conference Publication 2255

Advanced Automation for Space Missions

Edited by
Robert A. Freitas, Jr.
Space Initiative/XRI
Santa Clara, California

William P. Gilbreath
NASA Ames Research Center
Moffett Field, California

Proceedings of the 1980 NASA/ASEE Summer Study sponsored by the National Aeronautics and Space Administration and the American Society for Engineering Education held at the University of Santa Clara
Santa Clara, California
June 23-August 29, 1980

NASA
National Aeronautics and Space Administration

Scientific and Technical Information Branch

1982

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402

The painting was created by Mr. Rick Guidice. It captures the spirit of the space missions described in this study. In the center of the picture are human beings who, we believe, will continue to play a controlling role in future space missions. To the right of the circle are two space systems representing a partially automated Space Manufacturing Facility which would eventually utilize nonterrestrial resources. In the upper-right corner is Saturn attended by its largest natural satellite Titan, the proposed destination of our advanced space-exploration mission. The upper-left corner depicts the deepest reaches of the Cosmos that humans someday may explore. At center left is the Earth, which is under intensive study by an intelligent Earth-sensing information system that is able to obtain and deliver data in a far more effective manner than present-day methods. In the lower left corner, a lunar manufacturing facility rises from the surface of the Moon. Someday, such a factory might replicate itself, or at least produce most of its own components, so that the number of facilities could grow very rapidly from a single seed.

PREFACE

Since the late 1950s NASA has devoted itself to the acquisition and communication of information about the Earth, the planets, the stars, and the Universe. It has launched an impressive series of spectacularly successful exploration missions and numerous Earth-orbiting satellites that have added to an immense, growing pool of useful knowledge about terrestrial resources, weather and climatic patterns, global cartography, and the oceans. Each mission has made use of some level of automation or machine (artificial) intelligence (AI).

Mission complexity has increased enormously as instrumentation and scientific objectives have become more sophisticated. In the next two decades there is little doubt that NASA will shift its major focus from exploration to an increased emphasis on utilization of the space environment, including public service and industrial activities. The present study was sponsored by NASA because of an increasing realization that advanced automatic and robotic devices, using machine intelligence, will play a major role in all future space missions. Such systems will complement human activity in space, accomplishing tasks that people cannot do or that are too dangerous, too laborious, or too expensive. The opportunity to develop the powerful new merger of human intellect and machine intelligence is a result of the growing capacity of machines to accomplish significant tasks. Indeed, the growth in capability of onboard machine intelligence will make many missions technically or economically feasible. This study has investigated some of the ways this capacity may be used as well as a number of research and development efforts necessary in the years ahead if the promise of AI is to be fully realized.

There is a great deal of theoretical research (and in some cases practical development) in progress at several institutions in the United States and throughout the world. These research and development programs are necessary for the eventual success of the applications described elsewhere in this report. They are also a part of the rationale which has led to NASA's current strong interest in the potential of machine intelligence in space. However, even a vigorous research effort does not necessarily imply an applications development process adaptable to future NASA needs. Furthermore, the technology-transfer problem is aggravated by the relative scarcity of qualified workers in the AI field. NASA may begin to alleviate this manpower crisis by directly supporting artificial intelligence and robotics research in colleges and universities throughout the United States.

During the first 2 weeks of the study the group was introduced to the status of work in artificial intelligence by a series of lectures given by scientists from SRI International. A number of NASA program engineers participating in the study reviewed agency interests in relevant mission areas. To reduce the problem of automation-technology assessment to manageable proportions, the summer study group divided into four mission teams that could select single missions for concentrated attention in order to illustrate fully the potential for advanced automation. The task divisions among the teams guaranteed that all major classes of possible future NASA missions were considered, including public service, space utilization, and interplanetary exploration.

The teams spent the major part of the summer elaborating their missions with particular emphasis on the special role of machine intelligence and robotics technology. In addition to mission scenarios, the study produced two other significant outputs: Advanced Automation Technology Assessment to identify technology needs for mission capabilities representative of NASA programs in the 2000-2010 time frame; and an Epilogue which assures that an evolutionary NASA space program scenario with coordinated developmental initiatives is undertaken in the next 20 years to help establish an aggressive, multidisciplinary program of space exploration and utilization early in the 21st century. The Epilogue provides project/mission recommendations in each of the task team areas, an integrated project/mission scenario, and a series of recommended NASA planning options that include a consistent space-program strategy, technological development priorities, and updates to the OAST Space Systems Technology Model.

ACKNOWLEDGMENTS

The participants would like to thank all the visitors from NASA Headquarters, who gave tirelessly of their time and assistance in these endeavors, the Office of Aeronautics and Space Technology and other NASA offices for their support, and especially the NASA Administrator, Dr. Robert A. Frosch, for his personal attendance on two separate occasions at our deliberations. We would also like to thank various NASA Centers, SRI International, Lockheed Missiles & Space Company, Martin-Marietta Corporation, Battelle Memorial Institute, Science Applications, Inc., and a number of visitors from NASA/JPL for the many invaluable seminars, personal contacts, demonstrations, and facilities graciously and effectively provided during the course of the Summer Study. We also thank our hosts, University of Santa Clara for its lovely physical facilities and Ames Research Center for providing the administrative support. Special thanks goes to the Technical Information Division at Ames for continuing graphics assistance and particularly to Susie Rydquist for the final editing and preparation of this report.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This document is the final report of a study on the feasibility of using machine intelligence, including automation and robotics, in future space missions. The 10-week study was conducted during the summer of 1980 by 18 educators from universities throughout the United States who worked with 15 NASA program engineers. The specific study objectives were to identify and analyze several representative missions that would require extensive applications of machine intelligence, and then to identify technologies that must be developed to accomplish these types of missions.

The study was sponsored jointly by NASA, through the Office of Aeronautics and Space Technology (OAST) and the Office of University Affairs, and by the American Society for Engineering Education (ASEE) as part of their continuing program of summer study faculty fellowships. Co-hosts for the study were the NASA-Ames Research Center and the University of Santa Clara, where the study was carried out. Project co-directors were James E. Long of the Jet Propulsion Laboratory and Timothy J. Healy of the University of Santa Clara.

The study was sponsored by NASA because of an increasing realization of the major role that advanced automatic and robotic devices, using machine intelligence, must play in future space missions (fig. 1.1). Such systems will complement human activity in space by accomplishing tasks that people cannot do or that are otherwise too dangerous, too laborious, or too expensive. The opportunity to develop a powerful new merger of human intellect and machine intelligence is a result of the growing capacity of machines to accomplish significant tasks. The study has investigated some of the ways this capacity may be used as well as a number of research and development efforts necessary in the years ahead if the promise of advanced automation is to be fully realized.

1.1 Survey of Artificial Intelligence

Many of the concepts and technologies considered in this study for possible use in future space missions are elements of a diverse field of research known as "artificial intelligence" or simply AI. The term has no universally accepted definition or list of component subdisciplines, but is commonly understood to refer to the study of thinking and perceiving as general information processing functions – the science of intelligence. Although, in the words of one researcher, "It is completely unimportant to the theory of AI who is doing the thinking, man or computer" (Nilsson, 1974), the historical development of the field has followed largely an empirical and engineering approach. In the past few decades, computer systems have been programmed to prove theorems, diagnose diseases, assemble mechanical equipment using a robot hand, play games such as chess and backgammon, solve differential equations, analyze the structure of complex organic molecules from mass-spectrogram data, pilot vehicles across terrain of limited complexity, analyze electronic circuits, understand simple human speech and natural language text, and even write computer programs according to formal specifications – all of which are analogous to human mental activities usually said to require some measure of "intelligence." If a general theory of intelligence eventually emerges from the AI field, it could help guide the design of intelligent machines as well as illuminate various aspects of rational behavior as it occurs in humans and other animals.

AI researchers are the first to admit that the development of a general theory of intelligence remains more a goal for the future than an accomplishment of the present. In the meantime, work is progressing in a number of more limited subdisciplines. The following seven topical research areas include most elements normally considered to be a part of the field.

1.1.1 Planning and Problem Solving

All of artificial intelligence involves aspects of planning and problem solving, a rather generic category. This includes planning and organization in the program development phase as well as the dynamic planning required during an actual mission. Problem solving implies a wide range of tasks including decisionmaking, optimization, dynamic resource allocation, and many other calculations or logical operations that arise throughout a mission.

1.1.2 Perception

Perception is the process of obtaining data from one or more sensors, and analyzing or processing these data to facilitate subsequent decisions or actions. One simple example is a visual perception system that views a scene, determines whether or not a specified round object is in the scene, and if so, initiates a signal which causes an automatic arm to move the object out of the scene. Perception may be

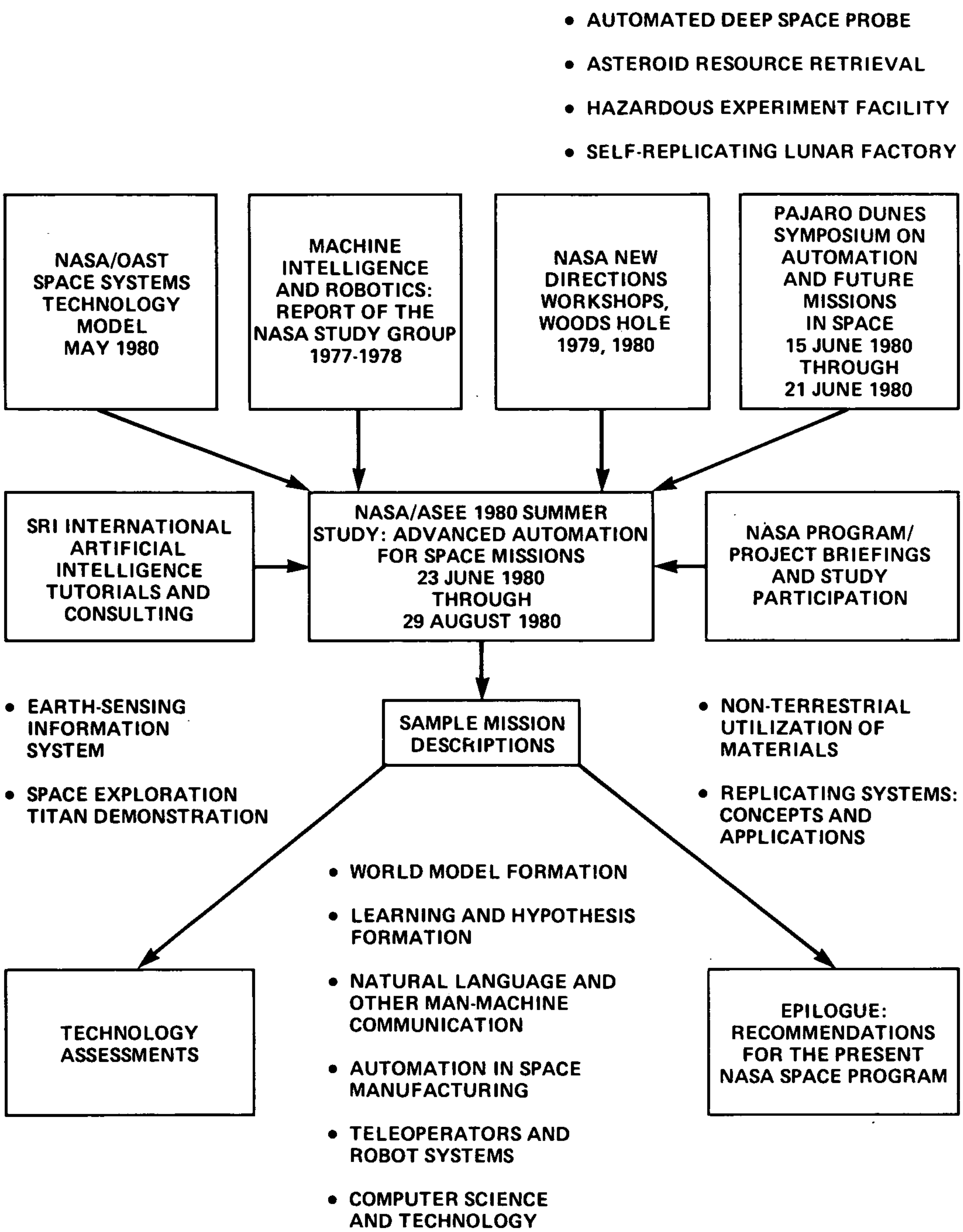


Figure 1.1.– Overview of NASA/ASEE 1980 Summer Study on Advanced Automation for Space Missions.

electromagnetic (visual, infrared, X-ray, microwave), aural, tactile, chemical; the possibilities are virtually unlimited.

The basic problem in perception is to extract from a large amount of sensed data some feature or characteristic that permits object identification. If viewed scenes can contain only two possible object classes, say, round or square, then the problem of deciding which is present may be relatively simple. But if thousands of characteristics are important in the scene, the task of creating a perceptual model of sufficient richness to permit unambiguous identification may be formidable indeed.

1.1.3 Natural Languages

One of the most difficult problems in the evolution of the digital computer has been communication between machine and human operator. The operator would like to use an everyday language – a natural language – to gain

access to the computer system. But proficiency in communication between human beings, and between machines and people, requires (1) mutual intimate familiarity with contextual understanding, (2) a very large base of data, (3) linguistic inferential capability, and (4) broad utilization of jointly accepted models and symbols. The process, very complex and detailed, demands expensive computer hardware and software to achieve accurate and efficient translation between machine and human languages. Extensive research is now in progress in the AI field to better understand the fundamentals of human language and to improve the quality of communication between man and machine.

1.1.4 Expert Systems

Scientific expertise typically develops in human beings over many years of trial and error in some chosen field. So-called "expert systems" permit such individual expertise to be stored in a computer and made available to others who have not had equivalent experience. Successful programs have been developed in fields as diverse as mineral exploration, mathematical problem solving, and medical diagnosis. To generate such a system, a scientific expert consults with software specialists who ask many questions in the chosen field. Gradually, over a period of many months, the team builds a computer-based interactive dialogue system which, to some extent, makes the expert's experience available to eventual users. The system not only stores scientific expertise but also permits ready access to the knowledge base through a programmed capacity for logic and inference.

Typically, a user interrogates the expert system via a computer terminal, typing in, for example, statements about apparent symptoms in a medical case. The system may then inquire about other conditions or symptoms, request that specific tests be performed, or suggest some preliminary diagnosis, thus attaching a probability or level of confidence to its conclusion and supplying an explanation upon demand. Therefore, user and system interact and gradually approach an answer to some question, whether on a diagnosis of an illness, the location of a mineral deposit, or the solution to a problem in mathematics.

1.1.5 Automation, Teleoperation, and Robotics

Automatic devices are those that operate without direct human control. NASA has used many such machines for years for diverse purposes including antenna deployment, midflight course changes, and re-entry parachute release.

Teleoperation implies a human operator in control of a mechanical system remotely. Executive signals are transmitted from controller to device over hard wires if the distance is small, as in the case of a set of master-slave arms in an isolation room (e.g., "P4" biohazard facility, radioisotope handling, etc.). Or, control signals may travel millions of kilometers over a radiowave link to a planet light-hours away.

Robotic devices have the capacity to manipulate or control other devices. They may be mobile, able to move to some distant physical location where an action must be taken. Robots can be either automatic or teleoperated.

1.1.6 Distributed Data Management

Large amounts of data are involved in the operation of automatic and robotic devices. This may include control information that specifies the next action to be taken in some sequence of operations, archived data that are being transmitted from one memory bank to another, or sensed or measured data that give the status of a geographical area, the position of an actuator, or the speed of a spacecraft. The field of distributed data management is concerned with ways of organizing such data transmission and distribution so that it is accomplished rapidly, efficiently, and in a manner which best supports overall system operation, and with ways of optimizing cooperation among independent but mutually interacting databases.

1.1.7 Cognition and Learning

In this study, cognition and learning refer to the development of a machine intelligence capable of dealing with new facts, unexpected events, and contradictory information in novel situations. Many potential applications of advanced automation require a level of adaptive flexibility that is unavailable with present technology. Today's automatic computer-controlled machines handle new data by a method or approach programmed into them when they were developed. Tomorrow's more sophisticated tools may need the ability to learn, even understand, in the sense of changing their mode of operation as they encounter new situations.

1.1.8 Research and Development in Artificial Intelligence

At present, there is a great deal of AI theoretical research (and in some cases practical development) in progress at several institutions in the United States and throughout the world. Much of the early work in the field was accomplished at five major centers: Carnegie-Mellon University, Edinburgh University, MIT, SRI International, and Stanford University. Today, however, the list of active sites is much longer and includes, in the United States alone, such schools as the University of Illinois, the University of Massachusetts, Yale University, the University of Southern California, Texas University, the University of California at Berkeley, etc. Corporations with ongoing work include Bolt Beranek and Newman, General Motors, IBM,

Lockheed, Rand, Schlumberger, Texas Instruments, and Xerox-PARC. Other institutions in this country have shown increasing interest in the field. International activity is concentrated in Great Britain, Japan, and the Soviet Union, with some work under way in Canada, France, Italy, West Germany, and Sweden.

These research and development programs are necessary for the eventual success of the applications described elsewhere in this report. They are also a part of the environment which has led to NASA's current strong interest in the potential of machine intelligence in space. However, even a vigorous research effort does not necessarily imply an applications development process adaptable to future NASA needs. The technology transfer problem is further aggravated by the relative scarcity of qualified workers in the AI field. NASA may begin to alleviate this manpower crisis by directly supporting artificial intelligence and robotics research in colleges and universities throughout the United States.

1.2 History of NASA Automation Activities

Since its inception in the late 1950s, NASA has been primarily devoted to the acquisition and communication of information about the Earth, the planets, the stars, and the Universe. To this end, it has launched an impressive string of spectacularly successful exploration missions including the manned Mercury, Gemini, and Apollo vehicles and the unmanned Surveyor, Mariner, Pioneer, Viking, and Voyager spacecraft to the Moon and beyond. Numerous Earth-orbiting NASA satellites have added to an immense, growing fund of useful knowledge about terrestrial resources, weather and climatic patterns, global cartography, and the oceans. Each mission has made use of some level of automation or machine intelligence.

Mission complexity has increased enormously as instrumentation and scientific objectives have become more sophisticated and have led to new problems. The Mariner 4 mission to Mars in 1965 returned about 10^6 bits of information and was considered a tremendous success. When Viking revisited the planet only a decade later, roughly 10^{10} bits were returned with greatly increased flexibility in data acquisition. Even now, the amount of data made available by NASA missions is more than scientists can easily sift through in times on the order of a decade or less. The situation can only become more intractable as mission sophistication continues to increase in the future, if traditional data acquisition and handling techniques are retained.

A 1978 JPL study suggested that NASA could save from $500 million to $5 billion per annum by the year 2000 AD if the technology of machine intelligence is vigorously researched, developed, and implemented in future space missions. According to a special NASA Study Group: "Because of the enormous current and expected advances in machine intelligence and computer science, it seems possible that NASA could achieve orders-of-magnitude improvement in mission effectiveness at reduced cost by the 1990s [and] that the efficiency of NASA activities in bits of information per dollar and in new data-acquisition opportunities would be very high" (Sagan, 1980). Modern computer systems, appropriately programmed, should be capable of extracting relevant useful data and returning only the output desired, thus permitting faster analysis more responsive to user needs.

During the next two decades there is little doubt that NASA will shift its major focus from exploration to an increased emphasis on utilization of the space environment, including public service and industrial activities. Current NASA planning for this eventuality envisions the construction of large orbital energy collection and transmission facilities and space stations operated either in Earth or lunar orbit or on the surface of the Moon. The first steps toward space industrialization already have been taken by NASA's Skylab astronauts who in 1973 performed a number of successful material-processing experiments. Progress will resume when the Space Shuttle delivers the first Spacelab pallet into orbit, and this line of experimentation continues.

Economy is perhaps the most important reason why robotic devices and teleoperated machines will play a decisive role in space industrialization. A conservative estimate of the cost of safely maintaining a human crew in orbit, including launch and recovery, is approximately $2 million per year per person (Heer, 1979). Since previous NASA mission data indicate that astronauts can perform only 1 or 2 hr of zero-gravity extravehicular activities (EVA) per day, the cost per astronaut is on the order of $10,000/hr as compared to about $10-100/hr for ground-based workers. This suggests that in the near term there is a tremendous premium attached to keeping human beings on the ground or in control centers in orbit, and in sending teleoperated machines or robots (which are expected to require less-costly maintenance) into space physically to perform most of the materials-handling jobs required for space industrialization.

In summary, the objective of NASA's space automation program is to enable affordable missions to fully explore and utilize space. The near-term technology emphasis at OAST includes:

- Increasing operational productivity
- Reducing cost of energy
- Reducing cost of information
- Enabling affordable growth in system scale
- Enabling more cost-effective high performance missions (planetary, etc.)
- Reducing cost of space transportation

The growth in capability of onboard machine intelligence will make possible many missions (OAST, 1980) technically or economically infeasible without it. The startling success of the recent Viking and Voyager robot explorers has demonstrated the tremendous potential of spacecraft controllers even when computer memory alone is augmented. Earlier spacecraft computers were limited to carrying out activity sequences entirely predetermined by programmed instructions; the advanced Viking and Voyager machines could be reprogrammed remotely to enable them to perform wholly different missions than originally planned – a flexibility that ultimately yielded more and better data than ever before.

1.2.1 NASA Study Group on Machine Intelligence and Robotics (1977-78)

Recognizing the tremendous potential for advanced automation in future space mission planning and development, and suspecting that NASA might not be utilizing fully the most recent results in modern computer science and robotics research, Stanley Sadin at NASA Headquarters requested Ewald Heer at the Jet Propulsion Laboratory (JPL) to organize the NASA Study Group on Machine Intelligence and Robotics, chaired by Carl Sagan. The Study Group was composed of many leading researchers from almost all major centers in the fields of artificial intelligence, computer science, and autonomous systems in the United States. It included NASA personnel, scientists who worked on previous NASA missions, and experts in computer science who had little or no prior contact with NASA. The Study Group met as a full working group or as subcommittees between June 1977 and December 1978, and devoted about 2500 man-hours to an examination of the influence of current machine intelligence and robotics research on the full range of space agency activities, and recommended ways that these subjects could assist NASA in future missions (Sagan, 1980).

After visiting a number of NASA Centers and facilities over a two-year period, the Study Group reached four major conclusions:

- NASA is 5 to 15 years behind the leading edge in computer science and technology.
- Technology decisions are, to a great degree, dictated by specific mission goals, thus powerfully impeding NASA utilization of modern computer science and automation techniques. Unlike its pioneering work in other areas of science and technology, NASA's use of computer science has been conservative and unimaginative.
- The overall importance of machine intelligence and robotics for NASA has not been widely appreciated within the agency, and NASA has made no serious effort to attract bright, young scientists in these fields.
- The advances and developments in machine intelligence and robotics needed to make future space missions economical and feasible will not happen without a major long-term commitment and centralized, coordinated support.

The Study Group recommended that NASA should adopt a policy of vigorous and imaginative research in computer science, machine intelligence, and robotics; that NASA should introduce advanced computer science technology into its Earth orbital and planetary missions, and should emphasize research programs with a multimission focus; and that mission objectives should be designed flexibly to take best advantage of existing and likely future technological opportunities.

The Study Group concluded its deliberations by further recommending that (a) the space agency establish a focus for computer science and technology at NASA Headquarters to coordinate R&D activities; (b) computer scientists should be added to the agency advisory structure; (c) a task group should be formed to examine the desirability, feasibility, and general specification of an all-digital, text handling, intelligent communication system for the transfer of information between NASA Centers; and (d) close liaison should be maintained between NASA and the Defense Mapping Agency's (DMA) Pilot Digital Operations Project because of the similarity of interests.

1.2.2 Woods Hole New Directions Workshop (1979)

Soon after the NASA Study Group on Machine Intelligence and Robotics completed its work, the NASA Advisory Council (NAC) convened a New Directions Workshop at Woods Hole in June, 1979. The NAC, a senior group of scientists, engineers, sociologists, economists, and authors chaired by William Nierenberg (Director, Scripps Institute of Oceanography), had become concerned that people in the space program "might have lost some of their creative vitality and prophetic vision of the future" (Bekey and Naugle, 1980). Before setting off for Woods Hole, 30 workshop members assembled at NASA Headquarters for briefings on the agency's current program and long-range plans, the projected capabilities of the Space Transportation System, and various interesting concepts that had not yet found their way into formal NASA planning. The Workshop members then divided themselves into eight working groups, one of which, the Telefactors Working Group, was charged with examining possible future applications of very advanced automation technologies in space mission planning and implementation.

The Telefactors Working Group recognized that the cost of traditional space operations, even if transportation becomes relatively inexpensive, makes many proposed

large-scale enterprises so expensive that they are not likely to gain approval in any currently foreseeable funding environment. Long delays between large investments and significant returns make the financial burden still less attractive. The crux of these difficulties is the apparent need to carry fully manufactured machinery and equipment to generate useful output such as oxygen, water, or solar cells *in situ*. The Group decided to see if the feasibility of certain large-scale projects could be enhanced by using machines or machine systems that are able to reproduce themselves from energy and material resources already available in space. Such devices might be able to create a rapidly increasing number of identical self-replicating factories which could then produce the desired finished machinery or products. The theoretical and conceptual framework for self-reproducing automata, pioneered by von Neumann three decades ago, already exists, though it had never been translated into actual engineering designs or technological models.

The difference in output between linear and exponentiating systems could be phenomenal. To demonstrate the power of the self-replication technique in large-scale enterprises, the Telefactors Working Group assumed a sample task involving the manufacture of 10^6 tons of solar cells on the Moon for use in solar power satellites. A goal of 500 GW generating capacity – to be produced by entirely self-contained machinery, naturally occurring lunar materials, and sunlight for energy – was established. From an initial investment estimated at $1 billion, to place a 100-ton payload on the surface of the Moon, a nonreplicating or "linear" system would require 6000 years to make the 10^6 tons of solar cells needed – clearly an impractical project – whereas, a self-replicating or "exponentiating" system needs less than 20 years to produce the same 10^6 tons of cells (fig. 1.2, Bekey, 1980).

The Working Group concluded that replicating machine systems offer the tantalizing possibility in the near future that NASA could undertake surprisingly ambitious projects in space exploration and extraterrestrial resource utilization without the need for unreasonable funding requests from either public or private sources. In practice, this approach might not require building totally autonomous self-replicating automata, but only a largely automated system of diverse components that could be integrated into a production system able to grow exponentially to reach any desired goal. Such systems for large-scale space use would necessarily come as the end result of a long R&D process in advanced automation, robotics, and machine intelligence, with developments at each incremental stage finding wide use both on Earth and in space in virtually every sphere of technology.

The Telefactors Working Group, believing that robotics, computer science, and the concept of replicating systems could be of immense importance to the future of the space program, recommended that NASA should proceed with studies to answer fundamental questions and to determine the most appropriate development course to follow.

1.2.3 Pajaro Dunes Symposium on Automation and Future Missions in Space (1980)

Because of the burgeoning interest in machine intelligence and robotics within NASA, the decision was made in September 1979 to fund an automation feasibility study to be conducted the following year as one of the annual joint NASA/ASEE Summer Study programs. To help provide the Summer Study with a set of futuristic goals and possibilities, an interactive symposium was organized by Robert Cannon at the request of Robert Frosch to take place the week before the opening of the summer session. During 15-22 June 1980, 23 scientists, professors, NASA personnel, and science fiction authors gathered at Pajaro Dunes near Monterey, California, to consider two specific questions: (1) What goals involving self-replicating telefactors might NASA possibly pursue during the next 25, 50, or 100 years, and (2) what are the critical machine intelligence and robotics technology areas that need to be developed? (Proceedings of the Pajaro Dunes Workshop, June 1980, unpublished).

A large number of highly imaginative missions were discussed, including automatic preparation of space colonies, an automated meteor defense system for the Earth, terrestrial climate modification and planetary terraforming, space manufacturing and solar power satellites, a geostationary orbiting pinhole camera to permit high-resolution solar imaging, lunar colonies, a Sun Diver probe capable of penetrating and examining the solar photosphere, advanced planetary surface exploration, and so forth. However, Workshop participants selected four missions they regarded as most significant to NASA's future and to the development of advanced automation technology:

- Mission I – Very Deep Space Probe – highly automated for Solar System exploration and eventually to be extended to include interstellar missions capable of searching for Earthlike planets elsewhere in the Galaxy
- Mission II – Asteroid Resource Retrieval – includes asteroids, jovian satellites and lunar materials that will use mass drivers, nuclear pulse rockets, etc., for propulsion
- Mission III – Hazardous Experiment ("Hot Lab") Facility – an unmanned scientific laboratory in geostationary orbit with isolation necessary to safely handle such dangerous substances as toxic chemicals, high explosives, energetic radioisotopes, and genetically engineered biomaterials
- Mission IV – Self-Replicating Lunar Factory – an automated unmanned (or nearly so) manufacturing facility consisting of perhaps 100 tons of the proper set of

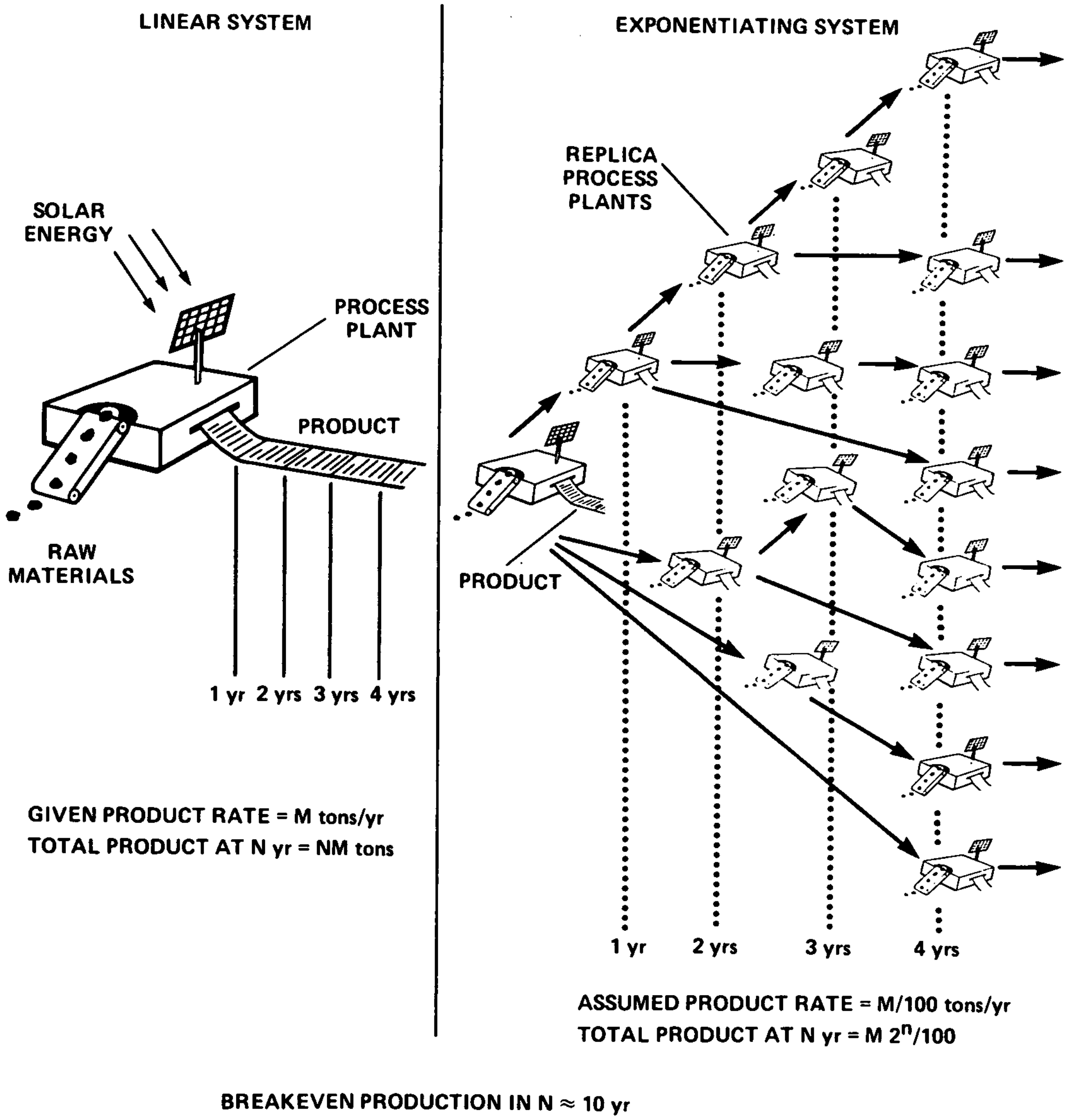


Figure 1.2.– Comparison of linear and exponentiating (self-replicating) systems in production capability.

machines, tools, and teleoperated mechanisms to permit both production of useful output and reproduction to make more factories.

Mission IV appears to have generated the most excitement among Workshop participants, in part because it has not yet been extensively studied by NASA (or elsewhere) and the engineering problems are largely unexplored. A number of important issues were raised and concepts defined, and there was a general consensus that virtually every field of automation technology would require further development for the self-replicating factory to become a reality.

Six important robotics and machine intelligence technology categories were identified as most critical by Workshop participants:

(1) Machine vision capabilities, especially in the areas of depth perception, multispectral analysis, modeling, texture and feature, and human interface

(2) Multisensor integration, including all nonvision sensing such as force, touch, proximity, ranging, acoustics, electromagnetic wave, chemical, etc.

(3) Locomotion technology to be used in exploration, extraction processes and beneficiation, with wheeled, tracked, or legged devices under teleoperated or autonomous control

(4) Manipulators, useful in handling materials both internal and external to the machine, general purpose and special purpose, teleoperated or fully automatic

(5) Reasoning or intelligence, including logical deductions, plausible inference, planning and plan execution, real-world modeling, and diagnosis and repair in case of malfunction

(6) Man-machine interface, including teleoperator control, kinesthetic feedback during manipulation or locomotion, computer-enhanced sensor data processing, and supervision of autonomous systems.

1.3 Summer Study on Advanced Automation for Space Missions

Immediately following the conclusion of the Pajaro Dunes Symposium, the present summer study was convened on 23 June 1980 and completed its formal work (roughly 10,000 man-hours) on 29 August 1980. During the first two weeks of the study the group was introduced to the status of work in artificial intelligence by a series of lectures given by scientists from SRI International. A number of NASA program engineers participating in the study reviewed agency interests in relevant mission areas.

Study members then focused their work by selecting four space missions which appeared to have great potential for the use of machine intelligence and high relevance to future NASA program goals. There was no assumption that these specific missions would ever be carried out. The four teams and the missions they chose to examine were:

(a) Terrestrial applications – an intelligent Earth-sensing information system

(b) Space exploration – Titan demonstration of a general-purpose exploratory system

(c) Nonterrestrial utilization of materials – automated space manufacturing facility

(d) Replicating Systems – self-replicating lunar factory and demonstration.

The teams spent the major part of the summer elaborating their missions (summarized below), with particular emphasis on the special role that machine intelligence and robotics technology would play in these missions.

The study has produced three significant outputs, outlined briefly in the remainder of this chapter, as follows: Mission Scenarios, Advanced Automation Technology Assessment, and an Epilogue.

1.3.1 Mission Scenarios

Over the last few years literally hundreds of mission opportunities beyond the 10-year time frame have been developed by the NASA Office of Aeronautics and Space Technology and assembled into a comprehensive Space Systems Technology Model (OAST, 1980). To reduce the problem of automation technology assessment to manageable proportions, the summer study group formed four mission teams that could select single missions for concentrated attention in order to illustrate fully the potential of advanced automation. The task divisions among the teams guaranteed that all major classes of possible future NASA missions were considered, including public service, space utilization, and interplanetary exploration. A fifth group, the Space Facilities and Operations Team, consisted largely of NASA and industry personnel whose duty it was to ensure that all mission scenarios were technically feasible within the constraints of current or projected NASA launch- and ground-operations support capabilities.

(a) Terrestrial Applications Team. The Terrestrial Applications Team elected to examine a sophisticated, highly intelligent information processing and delivery system for data obtained from Earth-sensing satellites. Such a system can play an immediate and practical role in assisting people to manage local resources, and, in a broader sense, could provide continuous global monitoring that is useful in the management of the individual and collective activities of man. The mission scenario presented in chapter 2 includes basic systems descriptions and hardware requirements, a discussion of "world model" structures, and a suggested developmental timeline.

(b) Space Exploration Team. The Space Exploration Team developed the concept of a general-purpose, interstellar-capable, automated exploratory vehicle that can (1) operate in complex unknown environments with little or no *a priori* knowledge, (2) adapt system behavior by learning to enhance effectiveness and survivability, (3) independently formulate new scientific hypotheses by a process called abduction, (4) explore with a wide variety of sensory and effector-actuator systems, (5) coordinate distributed functions, and (6) exchange information with Earth via an entirely new form of man-machine interactive system. A demonstration mission to Titan was examined in some detail and is presented in chapter 3, including mission operational stages, hardware specifications, sensing and modeling functions, and machine intelligence and other advanced technology requirements.

(c) Nonterrestrial Utilization of Materials Team. The Nonterrestrial Utilization of Materials Team considered options for a permanent, growing, and highly automated space manufacturing capability based on the utilization of ever-increasing fractions of extraterrestrial materials. The major focus was the initiation and evolutionary growth of a general-purpose Space Manufacturing Facility (SMF) in low Earth orbit. The mission scenarios in chapter 4 include surveys of solar system resources and various manufacturing processes especially applicable to space, a description of several basic industrial "starting kits" capable, eventually, of evolving to complete independence of Earth materials resupply, and discussions of the rationales for and implications of such ventures.

(d) Replicating Systems Concepts Team. The Replicating Systems Concepts Team proposed the design and construction of an automated, multiproduct, remotely controlled or autonomous, and reprogrammable lunar manufacturing facility able to construct duplicates (in addition to productive output) that would be capable of further replication. The team reviewed the extensive theoretical basis for self-reproducing automata and examined the engineering feasibility of replicating systems generally. The mission scenarios presented in chapter 5 include designs that illustrate two distinct approaches – a replication model and a growth model – with representative numerical values for critical subsystem parameters. Possible development and demonstration programs are suggested, the complex issue of closure discussed, and the many applications and implications of replicating systems are considered at length.

1.3.2 Advanced Automation Technology Assessment

A principal goal of the summer study was to identify advanced automation technology needs for mission capabilities representative of desired NASA programs in the 2000-2010 period. Six general classes of technology requirements derived during the mission definition phase of the study were identified as having maximum importance – autonomous "world model" information systems, learning and hypothesis formation, man-machine communication, space manufacturing, teleoperators and robot systems, and computer science and technology.

Technology needs were individually assessed by considering the following sequence of questions in each case:

(1) What is the current state of the relevant technology?

(2) What are the specific technological goals to be achieved?

(3) What developments are required to achieve these goals?

After mission definition was completed (the first seven weeks), summer study personnel reorganized into formal technology assessment teams with assignments based on interest and expertise. During this phase of the study, participants focused their attention on the evaluation of advanced automation technologies required to attain the desired mission capabilities. The results of this activity are presented in chapter 6 of this report.

1.3.3 Epilogue

The purpose of the epilogue (chapter 7) is to present carefully targeted recommendations for the present NASA space program. An evolutionary NASA space program scenario was assumed by the study group, based on relevant planning documents and other information sources. The major premise was that coordinated developmental initiatives would be undertaken by the agency in the next 20 years to establish a basis for an aggressive, multidisciplinary program of space exploration and utilization early in the twenty-first century. The epilogue includes a space facilities and programs overview, specific goals for growth scenario, and a series of recommended NASA planning options including a consistent space program strategy, technological development priorities, and updates to the OAST Space Systems Technology Model.

1.4 REFERENCES

Bekey, Ivan; and Naugle, John E.: Just Over the Horizon in Space. *Astronautics and Aeronautics,* vol. 18, May 1980, pp. 64-76.

Heer, Ewald: Prospects for Robotics in Space. *Robotics Age,* vol. 1, Winter 1979, pp. 20-28.

Nilsson, Nils J.: Artificial Intelligence. *Information Processing 74,* vol. 4, Proceedings of IFIP Congress 74, 1974, pp. 778-801.

Office of Aeronautics and Space Technology (OAST): *NASA Space Systems Technology Model: Volume III – Opportunity Systems/Programs and Technologies,* NASA, Washington, D.C., May 1980.

Sagan, Carl, Chmn.: *Machine Intelligence and Robotics: Report of the NASA Study Group,* NASA TM-82329.

CHAPTER 2

TERRESTRIAL APPLICATIONS: AN INTELLIGENT EARTH-SENSING INFORMATION SYSTEM

2.1 Introduction

The Terrestrial Applications Team was charged with identifying a sample near-Earth NASA mission that could be implemented during the next two or three decades and that would require intensive application of artificial intelligence and robotics technologies. The team initially considered a long list of missions that included the design and automated fabrication of a satellite solar power station, weather sensing and prediction, crop assessment, large communication satellites, and disaster monitoring. As the catalogue of possible tasks evolved, it became clear that artificial intelligence would be most useful when applied to missions that generate data at very high rates – such as the NASA applications (Landsat) satellites which provide imaging data of the Earth. The team focused on the development of an integrated, user-oriented, Earth-sensing information system (fig. 2.1) incorporating a maximum of artificial intelligence capability for two primary reasons.

First, substantial economic benefits may accrue from the effective use of an integrated, intelligent remote Earth-sensing system. For example, a reduction in weather damage to crops, the location of mineral deposits and earthquake faults, and more efficient means of surveying large tracts of land may save time, money, and even human lives. With superior definition of models for weather forecasting, climate and oceanic processes may eventually make possible more precise meteorological prediction and ultimately even weather control and global climate modification (*Outlook for Space,* 1976). Such an intelligent sensing system can play a dominant role in the activity of understanding the Earth as a dynamic physical entity, and can provide a major

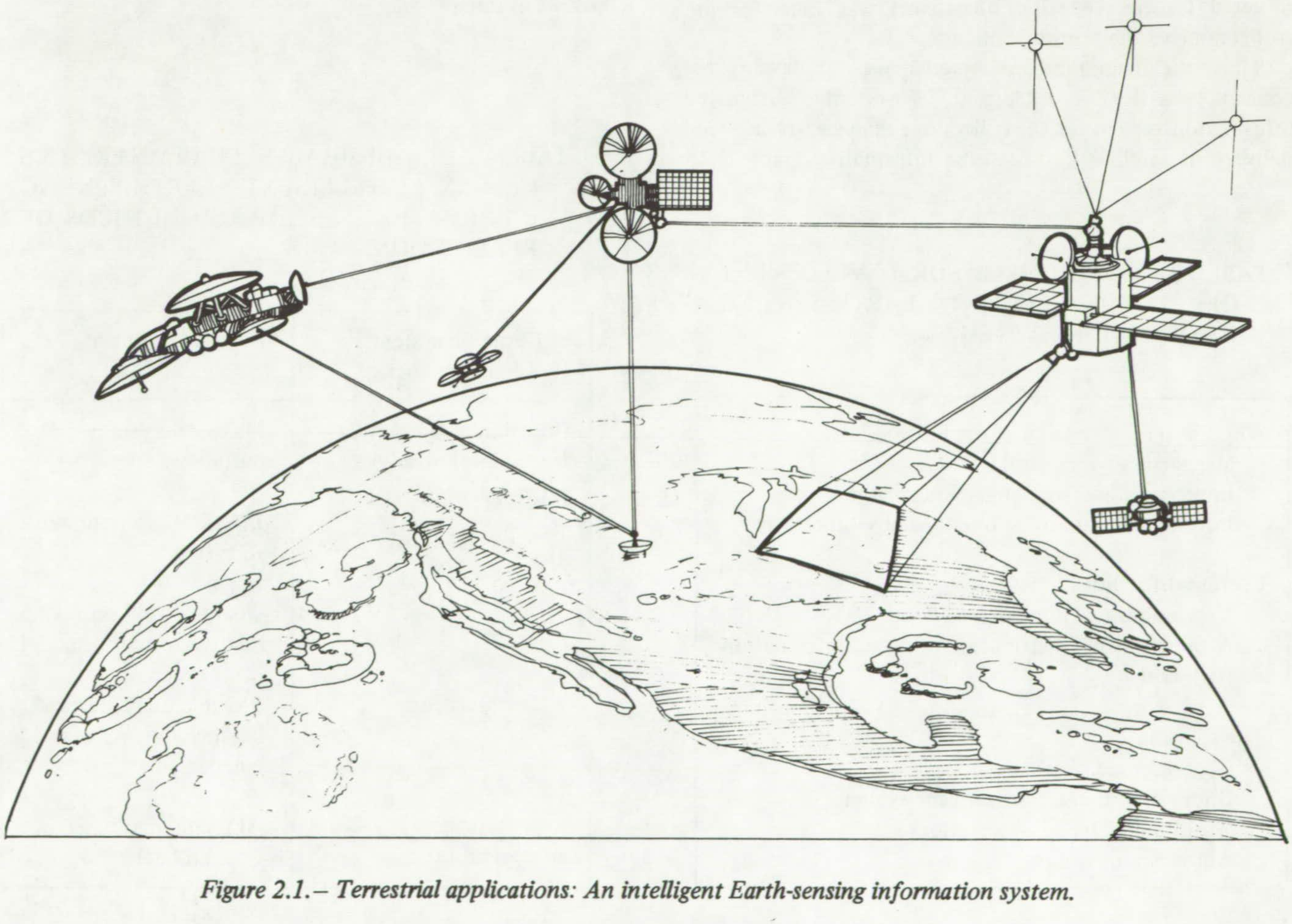


Figure 2.1.– Terrestrial applications: An intelligent Earth-sensing information system.

part of the ongoing monitoring of Earth useful in effective management of the individual and collective activities of man.

Second, NASA currently is obtaining and storing data from Earth-sensing satellites at a rate far out of proportion to the present or expected utilization of that data. The potential utility of collected data is not being realized because the raw data are not accessible in a timely and convenient manner, and because most potential users do not have the resources to extract useful information from the raw files. The current philosophy of data collection and storage had its origin in the early days of space research when sensors were sent into space, turned on, and all results transmitted back and stored. While this appears to maximize the utilization of the space vehicle, it has proven to be a false economy – the vast majority of uncategorized, generally unorganized data have never been and possibly never will be analyzed. The data format, its raw condition (digital conversions of analog sensor readings), and the complete lack of cross-referencing of contents make the data extremely difficult to find, interpret, or use. The tremendous volume of information already amassed and the expected increases in future rates of collection due to improved sensor technology make the philosophy of unorganized data acquisition obsolete. An alternate philosophy of goal-oriented data collection (information is gathered to meet specific objectives) was taken as the cornerstone of the proposed mission.

Thus, the main mission objective was to develop the concept of a flexible, intelligent, user-oriented automated information system for the collection, analysis, storage, and delivery of satellite Earth-sensing information (table 2.1). Within rational cost bounds, the system should maximize the utilization of this information for the following purposes: scientific, managerial, commercial, and humanitarian. In addition, the collection and storage of data having little or no utility should be minimized, and the costs of acquiring, interpreting, and storing Earth resource information must also be reduced.

Inexpensive data delivery can be accomplished by a system operating with relatively little human intervention. Price reduction requires that images be processed without costly manual procedures, and that the physical satellite system be managed so as to obtain a maximum of useful data for the given configuration of orbits and sensors. It seems possible to design and construct, by the year 2000, a largely autonomous system that can directly interface with individual users in natural language, accept requests for information, and provide answers based on satellite observations coupled with a resident theoretical model of the state of the world. Such a system should be able to achieve sophisticated data interpretation at modest cost through advances in machine hardware and artificial intelligence techniques. Table 2.2 lists several desirable system characteristics and suggested methods for their achievement. The key to the proposed system is a sophisticated world model (section 2.3) that enables the system to perceive both the present state of the world and how that state changes in time.

TABLE 2.1.– RATIONALE FOR DEVELOPMENT OF AN INTELLIGENT EARTH-SENSING INFORMATION SYSTEM.

Why use remote sensing of the Earth
Management – control
Improved understanding – knowledge
Information cannot be obtained any other way
Current difficulties
Vast amount of unorganized data
Acquisition and distribution of useful information
High cost
The solution
Goal-oriented observation
Direct user interaction with the system
World model-based observations
Autonomous system

TABLE 2.2.– DESIRABLE CHARACTERISTICS OF AN INTELLIGENT EARTH-SENSING INFORMATION SYSTEM AND METHODS OF REALIZATION.

Desirable system characteristics	Methods of realization
Cost minimization relative to level of service provided	Maximize system autonomy
	Interface users directly to system
	Goal-oriented observing relative to world model
Wide utilization	Flexible user interfacing including natural language
Automatic data interpretation	AI techniques based on world model

2.1.1 Relationship to NEEDS

NASA has instituted the NASA End-to-End Data System (NEEDS) program, the goal of which is to improve the effectiveness and efficiency of the agency's data and information management methodology. NEEDS began with Phase I which addressed some very near-term data handling and processing problems. Phase II, initiated in 1978, concentrated on complete subsystems development to permit nearly real-time data management. Future Phase III will concentrate on low-cost communications and data distribution, and Phase IV will deal with integration of modular subsystems and systems techniques. NEEDS is a complex program which evolved on a problem-by-problem basis to accommodate everincreasing demands placed on it by the changing nature of the space program. A summary of Phases I and II projects appears in table 2.3.

TABLE 2.3.– NASA END-TO-END DATA SYSTEM (NEEDS) PROGRAM.

Phase I
1. Synthetic aperture radar processor
2. Multispectral data processor
3. Digital data systems
4. Multipurpose user oriented software techniques
5. Resource effective data system definitions
Phase II
1. Systems analysis and integration
2. Modular data transport systems
3. Information adaptive systems
4. Database management systems
5. Archival mass memory
6. Massively parallel processor

The intelligent Earth-sensing information system proposed in this report, and also the Titan mission described in chapter 3, will also place significantly increased demands on the present NEEDS program. This is not surprising as both may be viewed as parts of a natural evolution of present or near-term planned NASA missions. Moreover, with these new proposals as goals, the NEEDS program can implement Phases III and IV in a comprehensive fashion rather than on a problem-by-problem basis. The Earth-sensing mission demands are at the far extremes of present NEEDS activities, particularly in planning, scheduling, and control of satellite systems.

2.2 System Description

The system and mission goals described in this chapter are best summarized as an attempt to propose a flexible, ongoing tool of tremendous utility and sophistication. Most of the details presented are offered solely to illustrate one of many possible alternative approaches. The intent was not to prepare an encyclopedic discussion of design specifics but rather to indicate the general nature of probable solutions and provide sufficient subsystem details to permit preliminary technology assessment of the basic concept.

The Intelligent Earth-Sensing Information System (IESIS) has the following major features:

- An intelligent satellite system that gathers data in a goal-directed manner, based on specific requests for observation and on prior knowledge contained in a detailed self-correcting world model (section 2.3). The world model eliminates the processing and storage of redundant information.
- A user-oriented interface that permits natural language requests to be satisfied without human intervention from information retrieved from the system library or from observations made by a member satellite within the system.
- A medium-level onboard decisionmaking capability which optimizes sensor utilization without compromising users' requests.
- A library of stored information that provides a complete detailed set of all significant Earth features and resources, adjustable for seasonal and other identifiable variations. These features and their characteristics are accessible through a comprehensive cross-referencing scheme.

IESIS has five major components: (1) System/user interface, (2) uplink, (3) satellite sensing and processing, (4) downlink, and (5) on-ground processing. The basic system is illustrated in figure 2.2. The user connects to the on-ground processor via a communication link and interactively defines his needs with the assistance of the system, accessing the database or directing IESIS to collect, process, and deliver information as required. The link might be a standard telephone line, and the entire transaction could occur in keyed natural language. Often, the user request should be answerable entirely from information already available in the system database, in which case IESIS appears to function much like any other interactive question-answering system (fig. 2.3).

Frequently, however, requests will require satellite observation data not yet available, in which case appropriate observations are scheduled by the on-ground processor. These instructions are uplinked via geosynchronous

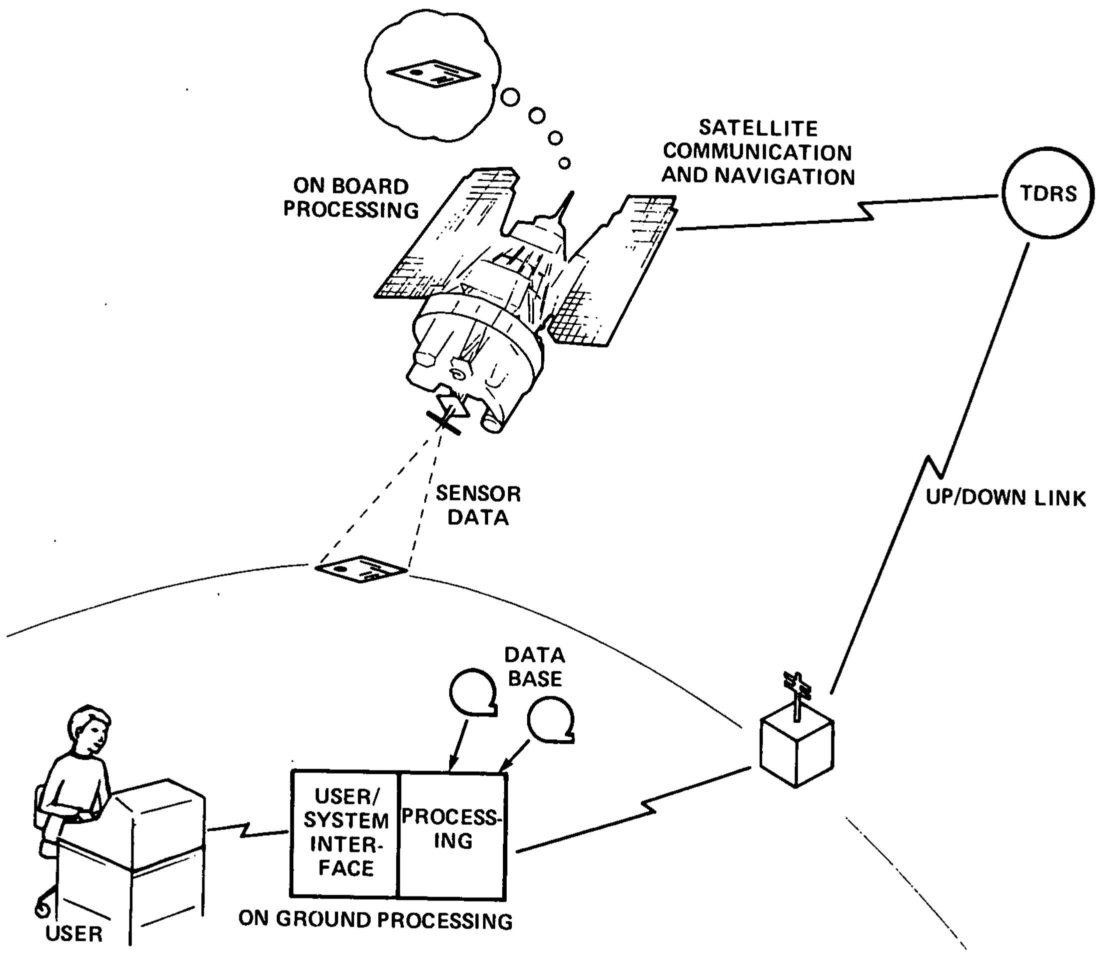


Figure 2.2.– The basic intelligent Earth-sensing information system.

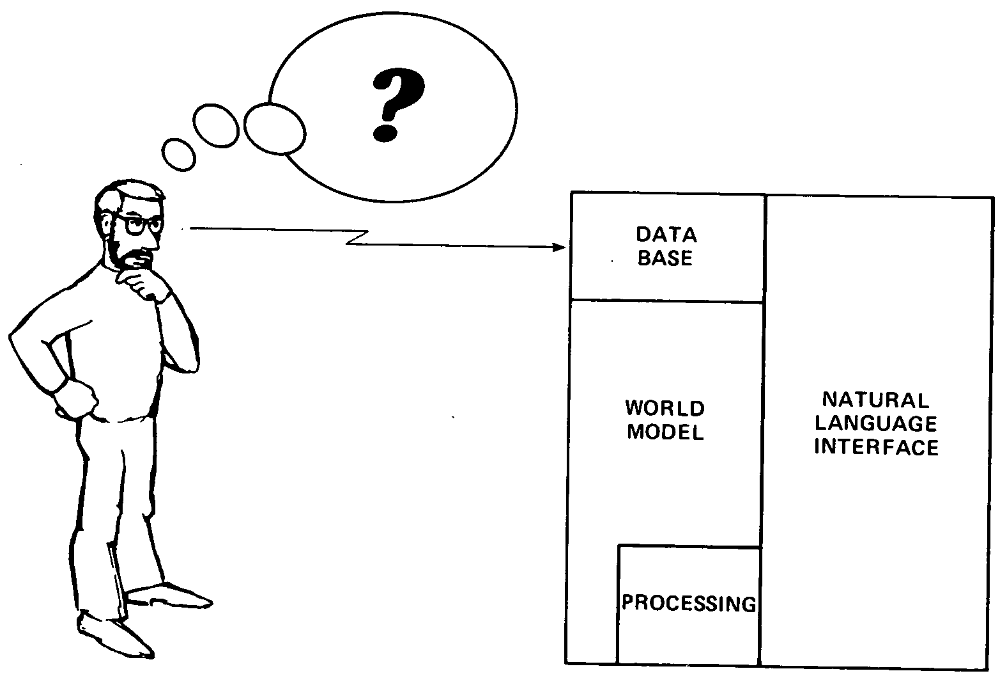


Figure 2.3.– IESIS response to user request for information already available in the system database.

communication satellites to orbital IESIS components that acquire and process the desired data and return them to the ground processor through a downlink (section 2.4.2). The ground processing unit further refines the information obtained, if necessary, then delivers it to the requester (fig. 2.4).

IESIS nominally operates with only one human in the loop – the user. This high degree of autonomy enables the system to be cost-effective and capable of rapid response. The sequence of operations during any user/IESIS interaction is outlined in figure 2.5.

IESIS has two basic modes of operation called background and foreground. The background mode of operation performs continuous goal-oriented observations of Earth and abstracts from these useful information for storage in a readily accessible, cross-referenced database. Background mode builds a broad scientific knowledge base that provides useful historical data at low cost for theory verification and testing. The foreground mode allows individual users to request that observations be taken and processed in nonstandard ways. The system must be sufficiently intelligent to help "naive users" obtain the information they want in an optimal or near-optimal fashion without restricting or unduly burdening the more sophisticated user.

2.2.1 Background Mode

In background mode IESIS continuously observes the Earth and gathers information to update the world model and to identify anomalies (sensor readings differing significantly from the expected). The system uses its world model to eliminate transmission of duplicate data and to implement basic principles of management by exception. The IESIS world model describes the topography and environment of Earth and can predict what a member satellite should record during its next observation period. During that period, the system collects data for "features" (e.g., lakes, forests, coastlines) and identifies all anomalies. Feature information is summarized to specify feature status without describing every pixel observed. For instance, if the height of a lake is known at its inlet and outlet, then the lake height at all points and flow rate can be determined. Only two pieces of data need be stored and transmitted by IESIS, rather than complete data for each pixel of the lake.

Anomalies are of two types. The first consists of transient normal events occurring at random, which are not to be permanently included in the world model. Examples are ships on an ocean, cars on a road, an iceberg, or a forest fire. IESIS should be capable of identifying such events by

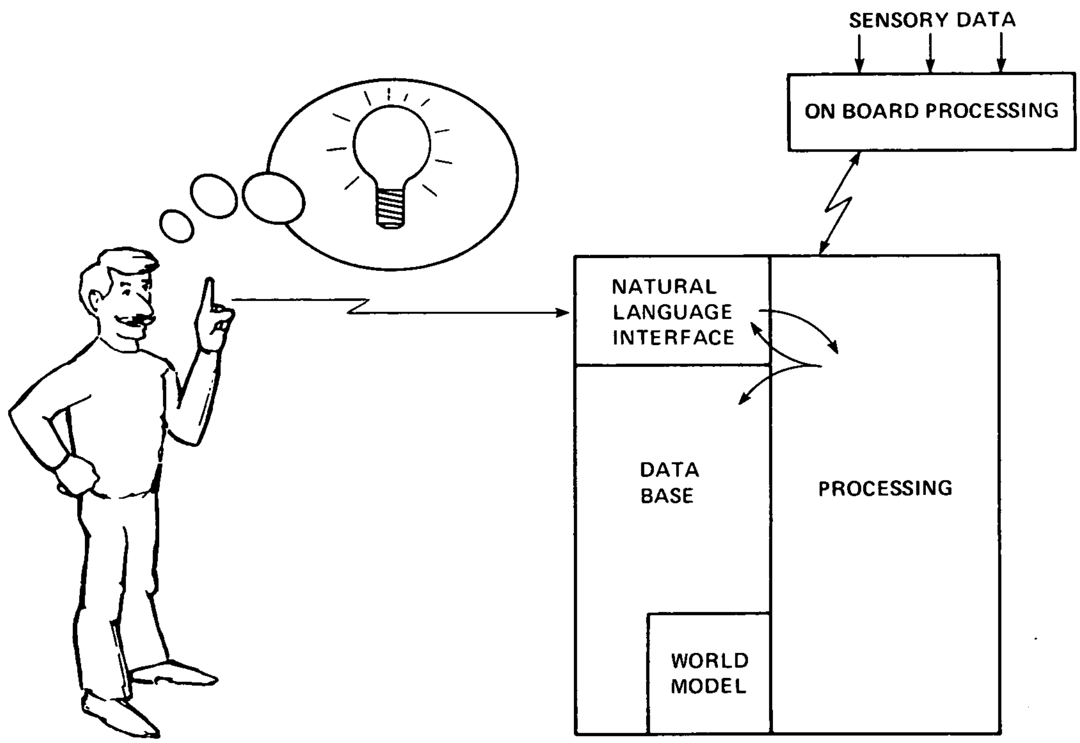


Figure 2.4.– IESIS response to user request for information that requires satellite observation.

DATA ON DEMAND

USER

LOGON

SYSTEM/USER INTERFACE
- NATURAL LANGUAGE QUERY-RESPONSE

ON GROUND PROCESSING
- KNOWS ALL SATELLITES
- KNOWS ALL USER REQUESTS
- KNOWS ALL PRIORITIES
- BUILDS PLAN FOR IMPLEMENTATION
- GENERATES COST ESTIMATE PRIOR TO EXECUTION OF PLAN

UPLINK

SATELLITE SENSING & PROCESSING
- RAW DATA COLLECTION
- MAKES DECISION ON DATA COLLECTION
- ON BOARD PROCESSING
- DOWNLINK
  - IMAGE

DOWNLINK

ON GROUND PROCESSING
- PICTURE PROCESSING
- UPDATE WORLD MODEL
- ARCHIVE
- DELIVER RESPONSE TO USER

USER

Figure 2.5.– Typical data processing steps in any user/IESIS interaction.

their signatures. Observations of these anomalies may produce a sample count of the observation type, trigger an alarm, or generate a report of the incident automatically sent to persons who should be apprised of the situation. The second class of anomaly consists of unexpected, wholly novel events. Upon observation of such an anomaly (e.g., the eruption of Mount Saint Helens' volcano), all sensor data are returned to Earth for analysis, identification, and possibly, action. The expected anomaly file is updated to include the identity of the phenomenon together with directions on actions to take upon re-observation, if any.

Processed sensor readings for features encountered during an observation are archived. Archival data collected on the basis of features and their properties then may be used to improve world model accuracy or to build detailed models of particular features (e.g., Lake Erie) or types of features (e.g., fresh water lakes). Individually observed data points lose informational value over time and can be reduced to models such as a Fourier time series to retain more valuable long-term trends once sufficiently detailed surveys accumulate. While the importance of this aspect of data reduction will grow over time, the majority of data reduction is associated with the world model in the process of eliminating the storage of redundant data. The world model will enable individual features as well as groups of features to be studied and summarized easily.

2.2.2 Foreground Mode

The IESIS foreground mode allows individual users to make natural language requests for particular data to be collected and processed for their own purposes. The fulfillment of this request becomes a goal of the system. IESIS determines the appropriate sensor algorithms and requested data are acquired the next time the requisite sensors are within view of the feature or area to be observed. The system must ascertain that conditions specified by the requester are satisfied during observation (e.g., absence of cloud cover, proper sun angle). If they are not, IESIS informs the user and reschedules the run. Nonstandard data processing may be performed on sensor data with output in any user-specified format including terminal printout, photograph/hard copy, and so forth. IESIS must have default processing/output modes as well as a choice of several optimal preprogrammed methodologies. An unsupported user-written software library similar to that maintained by IBM also could be provided.

2.2.3 *User/System Interface*

The user/system interface illustrated in figure 2.6 has two basic functions. The first is to process data previously collected and stored on the database to provide desired output. The second is to schedule satellites to make new observations, process data obtained by system routines or user-written procedures, and to notify users when information is available for delivery in an appropriate format.

Individuals communicate with IESIS through a generally accessible information net. After valid identity is established, the user indicates via a natural language interface whether data retrieval or satellite scheduling is desired. If this choice is not already known, the system may be interrogated regarding relative costs in time or money of using the most recent file data or the next expected observations. Users requiring data retrieval may interact with IESIS to determine the type of information needed and to develop a carefully tailored retrieval and processing scheme. The user then ends the session or enters additional requests. Upon termination, system files and customer billing records are updated.

If a user desires satellite scheduling, the optimum method for obtaining required information including decisions about appropriate sets of sensors, observation frequencies, sites, and so on, is developed interactively. Customers are provided initial estimates as to the probable time of completion and expected costs for data collection, and may revise or cancel a request on the basis of these preliminary appraisals. Each such interaction updates the request schedule so that observing satellites can perform appropriate observations and deliver derived data to the correct user after processing. Individuals may access the system to cancel or modify a previous request at any time up to the actual taking of data. Further, IESIS can accept requests for time- or event-oriented measurements such as photography of a particular area every month, or for observations contingent on specific events or conditions recognizable by the system such as volcanic eruptions or forest fires.

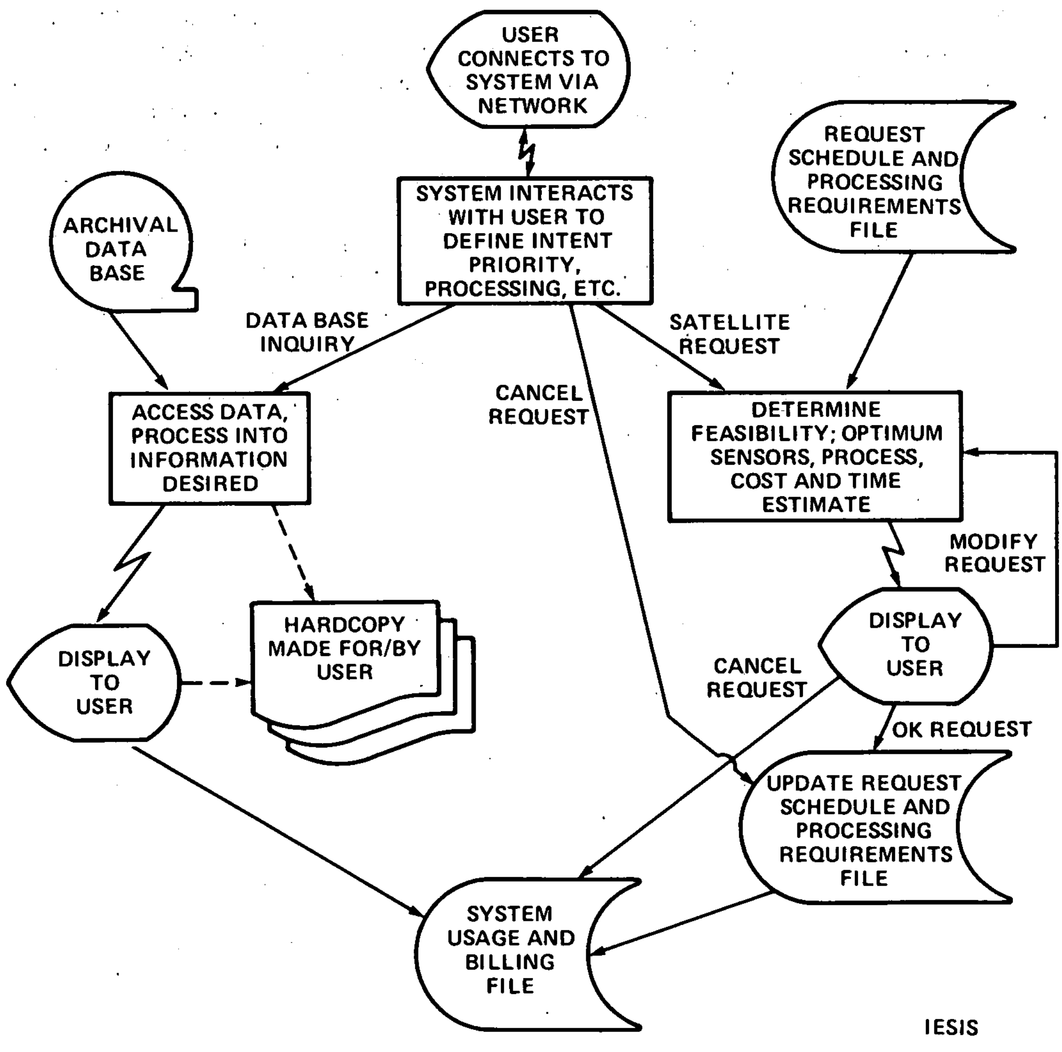


Figure 2.6.– User/system interface.

2.2.4 *Natural Language Interface*

The primary reason for a natural language interface is to allow the largest possible number of novice users to access IESIS directly without need for "interpreters." The principal advantage for more knowledgeable customers is convenience, though many customers will likely require a more formal and precise interface language in addition to the natural language capability.

There are major problems with current natural language interfaces that require careful consideration. The two primary difficulties are:

(1) The machine has only a very rudimentary notion of the subject of conversation during sentence interpretation even if it is quite competent to answer questions posed in a formal query language.

(2) The machine is not a social being and thus is deaf to most of the subtle information content in sentences generated by people.

Such flaws exist in all current natural language systems. For the anticipated "naive" user these obstacles must be minimized or IESIS will prove extremely uncomfortable and time-consuming. These flaws could possibly lead to system avoidance by less-sophisticated users, thus defeating one of the major mission goals.

Perhaps the simplest way to overcome such problems in natural language systems is to restrict the domain of discourse to a small set of possible concepts keyed to known individual human differences. Ultimately, the following may be the best approach for the "naive" user: Each would have a personal identification code known to IESIS, permitting the system to adjust its language to a compatible dialect. Knowledge of customer category could enable IESIS to employ reasonable default assumptions in restricting the domain of discourse (and thus the vocabulary) to a manageable subset of the overall system domain.

It is important that IESIS be able to communicate at the appropriate level of complexity and brevity. To accomplish this requires a system capability of modeling individual users. Some initial work in this area has been done (Rich, 1979), but no known current technique offers the level of performance necessary for IESIS. Natural language interfacing is one area that requires considerable advancement before it can hope to meet the IESIS system requirements: domain model, user model, dialogue model, reasonable default assumptions and common world knowledge, and explanatory capabilities.

2.2.5 *Artificial Intelligence Problem-Solving*

Clearly IESIS presents the usual difficulties in problem-solving typically involved in AI question-answering systems. But there is another new and important dimension added – effective combination of a world model database, world model theory, and a potentially resource-limited observational capability. The power of the problem-solvers and planners, and their capacity to execute plans in a dynamic and only partially known environment, will be instrumental in achieving a high-quality information delivery service at minimum cost.

Two specific areas where the quality of problem solving affects overall system efficiency and cost were considered for illustrative purposes. The first is communication link capacity. Given the goal of answering a large number of information requests, an intelligent planner able to isolate a parsimonious set of observations can considerably reduce ground link and intrasatellite link volumes. This set of observations is determined by consideration of individual requests, the ensemble of all current requests, and a prediction of expected requests.

A second area of concern is the number of satellites. If the system can employ a very sophisticated theory of observation, then it may be possible to shift most data-taking tasks to lower resolution. This system would allow data-taking by orbiters at higher altitudes having greater fields of view; thus a smaller total number of satellites would achieve the same frequency of coverage.

A major IESIS goal is to perform appropriate automatic data interpretation. System success in this activity demands a high-level capability to understand relationships between sensor readings and the actual state of the world as defined by human-oriented descriptors. This is precisely the problem in visual perception, an active current area in the field of artificial intelligence. Section 2.2.6 further discusses several aspects of the perception problem for an Earth-sensing system, and section 2.2.7 describes the need for flexibility and adaptability in IESIS. In both areas – perception, and system flexibility and adaptability – there is a tremendous need for development of fully autonomous techniques far more powerful than those presently available.

2.2.6 *Theory of Observation*

While the number of distinguishable states of the world of human interest (at a particular level of resolution and description) is extremely large, this figure is still dwarfed by the vast number of distinguishable ways the world may appear to rudimentary sensors. Machine sensors and human eyes see entirely different things when minor changes in the world state occur. For instance, in hilly terrain at low sun angle, satellites sensor readings vary rapidly as the shadows progress, but most of what is of interest to human beings is invariant.

To extract interesting information from sensors it is necessary first to understand the nature of the sensor as a transducer so that a mathematical inversion process can be performed on the readings. This involves computation of the electromagnetic reality at the image sensor location

(i.e., at the satellite, but before the sensor transduces visual photons into a signal). This extracted information is called "satellite local reality." If satellite local reality is not obtained, then an interpretation correlation problem arises when different sensors are used at various times to observe the same thing.

Once satellite local reality has been determined, it must be mapped onto an "object local reality" where "object" refers to the surface or volume under observation. For example: suppose that the state of the relatively clear atmosphere and a power spectrum at several chosen frequencies for a particular image pixel element of the Earth's surface are accurately known. The theory of observation must be able to use the world model (in this case the atmospheric component of the model) to determine which parts of the satellite local power spectrum are generated from a ground effect, and which are atmospheric phenomena. Reflection, refraction, absorption and emission cannot, in general, be clearly separated. Thus, the attempt to translate satellite local to object local descriptions requires some additional information.

The predictive aspects of the world model are important here. The model can predict much of the satellite local power spectrum, so inversion to object local spectrum need not be performed at all except in specific (and hopefully, relatively infrequent) cases where observations deviate significantly from prediction. These deviations are called anomalies. When an anomaly occurs, there may be simple alternative world states IESIS can hypothesize in an attempt to find an explanation for the anomaly. In this mode of action the model is altered and a new prediction for the observation generated. If a reasonable world model alteration leads to a predicted image that matches that actually observed, then the altered model may represent an adequate estimate of current reality. Other specific observations should be designed with the objective of testing the new hypothesis.

If an anomaly cannot be disposed of by hypothesizing new world states, alternative mechanisms are needed. It is desirable to proceed as far as possible without explicit human intervention. One alternative is to automatically schedule other observational configurations (different satellites, sensors, or lighting conditions) to gather enough satellite local information and permit clear computation of object local signature. Once this is accomplished the final interpretation must be made which involves mapping object local signatures into a state description suitable for incorporation into the world state component of the world model (see section 2.3).

The role of a predictive model in efficient image gathering is essential. Even when it is theoretically possible to clearly map satellite local reality into a high-level description of object local reality without such a model, model use can significantly increase the efficiency of the observation and the speed of the interpretation process. The following is a partial list of the many ways world models may be helpful:

- Extending the range of possible viewing conditions under which usable information can be gathered (e.g., compensating for variations in sun angle)
- Predicting when certain types of observations are impossible because of unfavorable viewing conditions which can be known prior to the time of observation
- Computing the least costly set of sensors (e.g., fewest sensors for shortest time, or use of sensors which at the particular time have no other demands on them) needed to determine a particular fact about the world
- Avoiding taking certain new observations by deriving at least some responses to requests from information already in the world model database.

IESIS is not oriented toward the storage of information as images. Rather, images are processed in real time (or almost real time) and only extracted information is stored in the world model. In such a context, the emphasis shifts from finding observation strategies which yield absolutely unique sensor signatures for identifying the condition of the world, to observing only what is necessary to identify the state of the world in the context of the world model. The theory of observation can be viewed as part of the world model, and represents a large part of the knowledge necessary to connect sensor-encoded information with more human-oriented descriptions of reality contained in the world model database.

2.2.7 System Flexibility

It is very difficult to anticipate the entire range of users to whom Earth-sensing systems may be applicable. IESIS must be flexible enough to allow a scale-up of total system throughput to accommodate a growing number of customers. Similarly, it is unlikely that the mission system will have available, by the year 2000, the ultimate in sensor technology. Almost certainly, a rapid evolution of ideas and technology will occur after a short period of system use. If the system is not to be the seed of its own rapid obsolescence, it is imperative that it be flexible enough to accommodate new modes of observation including new equipment and new processing procedures.

The general philosophy of providing a flexible information system for the sophisticated user virtually demands that the user be able to specify new algorithms for controlling the data accumulation and data-analysis processes. User-defined data collection control becomes important for the advanced user when observation scheduling must be sensitive to dynamic events in real time – where it would be impractical or impossible to use the standard system-scheduling mechanisms. Such individuals may require specialized data interpretation processing for a variety of

reasons including the needs for recognition categories not defined as standard categories or for specialized data displays.

The argument can be made that, besides the general desirability of system flexibility, it is important to give advanced users a flexible, complex, and adaptive tool. Very likely some of the most innovative and important IESIS applications will arise through the efforts of such individuals.

At present, Landsat data customers represent a relatively small but sophisticated population comprised mainly of engineers and scientists. IESIS is intended to reach a much broader spectrum of potential users, the majority of whom are "naive" with regard to computer technology. It is imperative that a reasonable model of this target population be generated as the system is implemented. Such knowledge is necessary for detailed design of user/system interfaces, ensuring insertion of user-related elements into the world model, and for determining signature analysis and pattern recognition techniques required to answer probable user questions.

2.2.8 Data Archiving and Compaction

The traditional NASA information gathering philosophy has been to collect as much raw data as possible from each mission and then allow university, industrial, and government researchers to complete the analysis. In the early days of space exploration this strategy was reasonable, based on spacecraft investment, insofar as it maximized return. But today, advancing satellite technology has greatly expanded the number of sensors flown and available data rates. The resulting torrential flow of information has overwhelmed the capacity of the system – only 0.05% of data collected from space has ever been analyzed. The great unused bulk of observations must be stored even though much of it is of marginal quality (e.g., obscured by clouds) and probably never will be analyzed.

NASA should consider revising its philosophy of data collection to: (1) Make use of knowledge gained from previous sensing missions to reduce redundancy, (2) adopt a goal-oriented approach to Earth-sensing and other observational missions, (3) begin to identify and dispose of poor-quality data, (4) condense information as it ages and becomes less useful, and (5) present data with full indexing and cross-referencing to maximize their utility to the consumer.

Knowledge and experience combined with artificial intelligence techniques can eliminate redundancy. For example, it is extremely inefficient for an Earth-sensing satellite to "rediscover" a lake, road, or city on every orbital pass. The truly important aspects of the object are its fundamental attributes – area, temperature, color, texture, etc. – many of which are either constant or predictable. The use of a world model to eliminate continual rediscovery of such features could greatly reduce the extraordinary redundancy of most visual imagery.

All object attributes studied must reflect worthwhile goals. One goal should be the assembly of a historical record of Earth features. Others may include more specific user-defined objectives. This new emphasis on goal-directed observations does not preclude data utilization by the scientific community in the investigation and verification of new theories; quite the contrary, it should actually enhance this activity by enabling researchers quickly and easily to direct IESIS to collect and process data under closely controlled conditions.

Many time-oriented observations lose some of their value with age. After an extended period of time, long-term trends are much more useful than individual data values. In the proposed IESIS system only the long-term trends are retained – original data are eventually discarded. Thus, the system processes all data immediately for specific goals and, at a later time, integrates trend information into a more compact world model representation such as a long-term temperature gradient.

As data are collected in orbit, Earth features and their processed image characteristics must be fully indexed to sort features and characteristics and to analyze them by group. The data then may be manipulated from within a fully cross-referenced base. For instance, area type can be called out and summed to obtain the total rye acreage in a given state. This cross-referencing feature is critical to the effectiveness of the Earth-sensing information system. The ability to automatically cross-reference and access data by content and feature allows rapid aggregation and correlation, and may promote new research as rapid access to useful scientific data becomes routine. The proposed database is organized using geographic location as the primary key (similar to the World Reference System used for Landsat data) with individual features also keyed. Feature keys greatly simplify the generation of inverted files listing, say, all lakes, deserts, wetlands, forests, or housetop areas – obvious widespread applications. The detailed mechanisms of records layouts, file structures, and database languages are beyond the scope of the present study and are recommended for future investigation.

The expense of storing data is a very significant part of computer system cost. This consists of direct charges for storage media as well as the costs of transferring data to and from local and remote storage devices. Data compaction and compression produce cost savings by reducing storage and transmission requirements. In addition, these data reduction methods enable more efficient information retrieval and more economical transmission of large quantities of data over computer networks.

There are no standard definitions for compaction and compression, so the following usages are adopted for the purposes of the present study:

Compaction of data – any technique that reduces the size of physical data representation while preserving the relevant information.

Compression of data – the application of some function to elements of the database.

If x is a specified element of the database then the compression of x is y, where $y = f(x)$. Usually f is invertible, which means that the original information may be recovered whole from the compressed data.

Compaction techniques other than compression involve elimination of information deemed superfluous, in order to decrease overall storage requirements. One such method is abstraction. Abstraction is accomplished by processing data over important common image features (in the case of photographic information) by using, for instance, a world model. After abstraction it is not possible to recover the original image.

Mission data such as are received daily or are already stored in various NASA facilities (e.g., the EROS data center at Sioux Falls, South Dakota) and slated for compaction may be classified within two broad categories – continuous and noncontinuous – depending on timing and event characteristics. Classes of continuous data include:

- Periodic data – When the same event appears again and again, only one copy need be saved.
- Trendless data – If the data are random continuous, a sample should be taken to check for trend. If no trend is found, the data may be represented by a histogram updated regularly as more information accumulates. Only histogram parameters need be saved.
- Data with trend – If a trend is detected, multiple regression and curve-fitting are best to record the feature. It may be possible to correlate variation in one time series with that in another (e.g., how sea level is affected by temperature or pressure). Data compression is achieved in this case by fitting polynomial segments, possibly straight lines, to the data.
- Data with turning points – Turning or "inflection" points, such as where an upward trend suddenly shifts downward in a time series graph, may require different models to be fitted to the two parts of the series. Only the parameters of fitted polynomials and a few statistical abstractions (e.g., maxima and minima, mean, variance, and several others) in any particular range need be saved.

Suggested classification and processing techniques for noncontinuous numerical data are summarized in figure 2.7.

2.3 The World Model

The world model is a crucial element in the achievement of specific goals. Without a sophisticated model two serious problems are encountered with remote Earth-sensing data, particularly images:

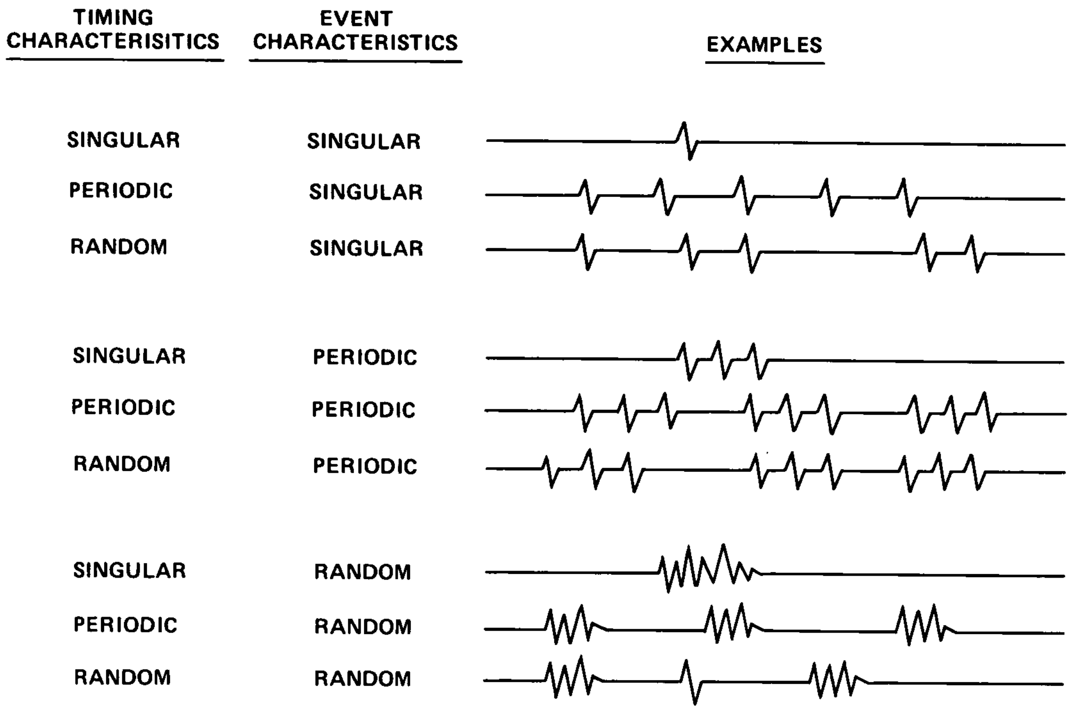
- It is very difficult, if not impossible in many instances, to accurately separate interesting from noninteresting observations.
- It is difficult to comprehend raw sensor data in terms readily understandable by human beings.

The first of these leads to the collection and retention of great volumes of data, simply because there is no practical way to perform an appropriate selection of the useful subset of information applicable to a user request. The second problem results in gross underutilization even of potentially useful data. The lack of a world model in present-day spacecraft makes necessary a voluminous and costly stream of highly redundant data which must be transmitted and collected on the ground before any useful information is retrieved, leaving a huge reservoir of unprocessed data in expensive storage facilities. It is the world model which transforms IESIS from a collection of remote cameras into an entity able to perceive the planet in a manner interesting and informative to humans. This world model is a compact representation of persistent spatial and temporal characteristics of the Earth (its land, oceans, and atmosphere), and algorithms for use of the model.

2.3.1 World Model Structure

The IESIS world model has two separate components. The first is the state component, which defines the physical status of the world to a predetermined level of accuracy and completeness at a specified time. Second is the theory component that allows derivation of the following information from the state component: (1) Values of parameters of the world state not explicitly stored in the state component, and (2) a forecast of the time evolution of the state of the world. The theory component gives the world model a predictive capability, in that the model can predict facts about the world not explicitly retained in the database.

The disparity between predicted information and reality generally increases with increasing computational distance separating the starting information and the derived result, increasing time in the case of forecasting, and certain other factors. The world model requires a continual influx of new observations to remain temporally current. A major research goal for efficient IESIS operation is to develop the AI capability to construct an effective real-time world model which can act as a database for answering questions



Timing	Event		
	Singular	Periodic	Random
Singular	Save event and location	Save one event, location, period, and number	Save data
Periodic	Save single event and its period	Save one event, location, short period, long period, number in short period, number in long period	Sample data and save statistical information (period, number)
Random	Save one event Study distribution of waiting times, fit Poisson process and save parameters of Poisson process with statistical information	Save one event, period, number, and information on distribution of waiting times	Same data and save statistical information

Figure 2.7.– Classification of and suggested processing techniques for noncontinuous numerical mission data.

about the state of the world and as a predictive mechanism for controlling observation satellites and interpreting observations.

The world model database must contain state component information about the expected character of points on the Earth. This includes land use (crop type, urban type, etc.) and ground topography. The world model theory component must predict some rather ubiquitous changes that occur, such as alterations in foliage color and foliage loss for certain vegetation areas as a function of season and precipitation history; ice formation and melting with the seasons; and variations in appearance of rivers from flow rate changes due to watershed runoff. As with the ensemble of observing satellites, it should be possible initially to set the Earth-sensing system into operation with a limited world-modeling capability, and later expand this model as the technology progresses.

The key element in handling, processing, and storing data in the proposed satellite system is the use of a world model to abstract useful information from images and thus accomplish a large reduction in required data transmission and storage. The model gives the satellite "experience" by which to judge new observations. Data compression has been investigated in video imagery and, in some cases, compressions of 20:1 have been accomplished, but not at acceptable fidelity (Graham, 1967). It has been found that methods using feature extraction without benefit of a world model are capable of compression ratios in excess of 100:1 (Chien and Peterson, 1977). A world model permits still greater data compression by interpreting features and their established properties in the full context of the known land, sea, and atmospheric environment of the Earth.

The simplest world model is a flat land map. Figure 2.8 illustrates such a map of Mildura City in Victoria, Australia, representing an area encompassing 16 features in 5 distinct land types: (1) river, (2) lake, (3) forest, (4) cropland, and (5) city area. Each feature is termed a "niche" – a surface feature on land or ocean possessing a relatively clear boundary, common features within that boundary, and whose location does not change rapidly with time. On land, niches are closely spaced, whereas in the oceans and atmosphere they may be quite large. Table 2.4 gives general characteristics of land, ocean, and atmospheric world models.

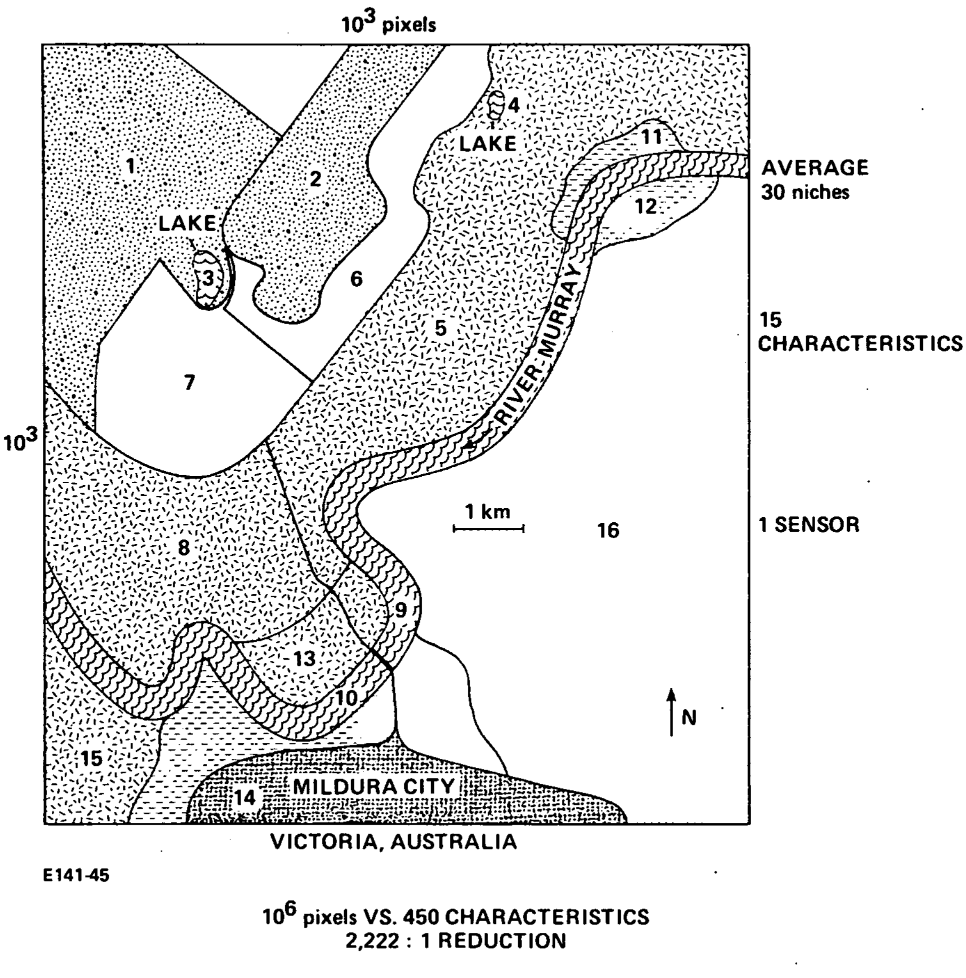


Figure 2.8.– Sample niche features map of Mildura City, Victoria, Australia.

TABLE 2.4.– WORLD MODEL CHARACTERISTICS OF LAND, OCEAN, AND ATMOSPHERIC WORLD MODELS.

Land Extensive, detailed contour maps exist Relatively static Sharp boundaries Small feature size Widely populated
Oceans Gross maps exist Constant altitude Wide boundaries Large feature size Sparsely populated
Atmosphere 3-dimensional Elementary model exists Rapidly varying Boundaries difficult to identify Large shifting patterns Some small feature sizes Substantially unpopulated

Niches possess a high degree of spatial redundancy. A large lake remains the same lake on each satellite overpass, and parts of the lake are very similar to other parts, yet are very distinct from the surrounding land. The large redundancy permits very substantial data reduction by processing each sensor reading across a single niche and extracting desired characteristics of that niche. These characteristics can be combined across sensors to produce yet further compaction. Despite the large reduction in physical data, no useful information is lost.

The likely scale of data compaction is illustrated by considering a 10 km × 10 km land region at a pixel resolution of 10 m. This scale gives an image measuring 1000 × 1000 pixels. If just one observational wavelength is involved, and 8 bits are used to represent intensity at each of the 10^6 pixels, the resulting image is 8×10^6 bits. A limited examination of aerial photographs suggests such a region will possess an average of 30 niches. Each must be fully described in terms of characteristics important to it such as area, average sensor value, variance, higher moments, two-dimensional sensor intensity gradient, and texture. If 15 characteristics are sufficient to describe most niches, then only 3600 bits (15 characteristics × 30 niches × 8 bits/niche-characteristic) are needed to replace the 8,000,000 bits of the full image. A reduction of 2222:1 is immediately accomplished.

Further reduction can be achieved by combining data across the approximately 20 sensor wavelengths proposed for the satellites, and also across the 15 characteristics. This reduced data can be used, for example, in signature identification, specification of niche status, or for classifying an anomaly. Near-maximum reduction occurs when imagery is required to answer a sophisticated question posing a choice from 256 ($= 2^8$) alternatives, an answer requiring just 8 bits – the 10^6 pixel elements over 20 sensors demand one 8-bit transmission and subsequent storage for a reduction of 20,000,000:1 ([10^6 elements × 20 sensors × 8 bits/sensor-element] / [8 answer bits]). If the question requires a "yes" or "no" answer only 1 bit must be transmitted and the absolute maximum data reduction in this simplified example is achieved – 160,000,000:1 – as summarized in table 2.5.

IESIS also is capable of discovery. Novel occurrences can be detected by the satellite system when searching for anomalous features in the imagery as compared to the world model. Many of these anomalies may be boundary changes in the existing map, e.g., overflow of the River Murray as shown with dark lines in figure 2.9. Others will involve unusual values of sensory characteristics of the niche (e.g., blight on a corn field). Prompt identification of these anomalies allows real-time management action in response to the "abnormal" occurrence. Figure 2.10 illustrates a hypothetical set of readouts from an intelligent satellite scan over Mildura.

The most efficient use of a world model requires placement of a simplified model in memory onboard the satellite system (to accomplish direct data reduction) and retention of a more sophisticated model in the ground operations facility. The latter serves as a master Earth model for use in updating, calibrating, and further processing transmitted data. (Estimated memory requirements are given in section 2.3.2.) Another very important feature of the ground-based world model is that it allows full cross-indexing of

TABLE 2.5.– DATA REDUCTION USING WORLD MODEL IN A SIMPLIFIED EXAMPLE.

Transmission/storage task	Total bits	Net reduction
10^6 image pixels; 20 sensors 8 bits each	1.6×10^8	1:1
30 niches, 20 sensors, 15 characteristics at 8 bits	7.2×10^4	2,222:1
256-choice answer	8	20,000,000:1
Yes or no answer required	1	160,000,000:1

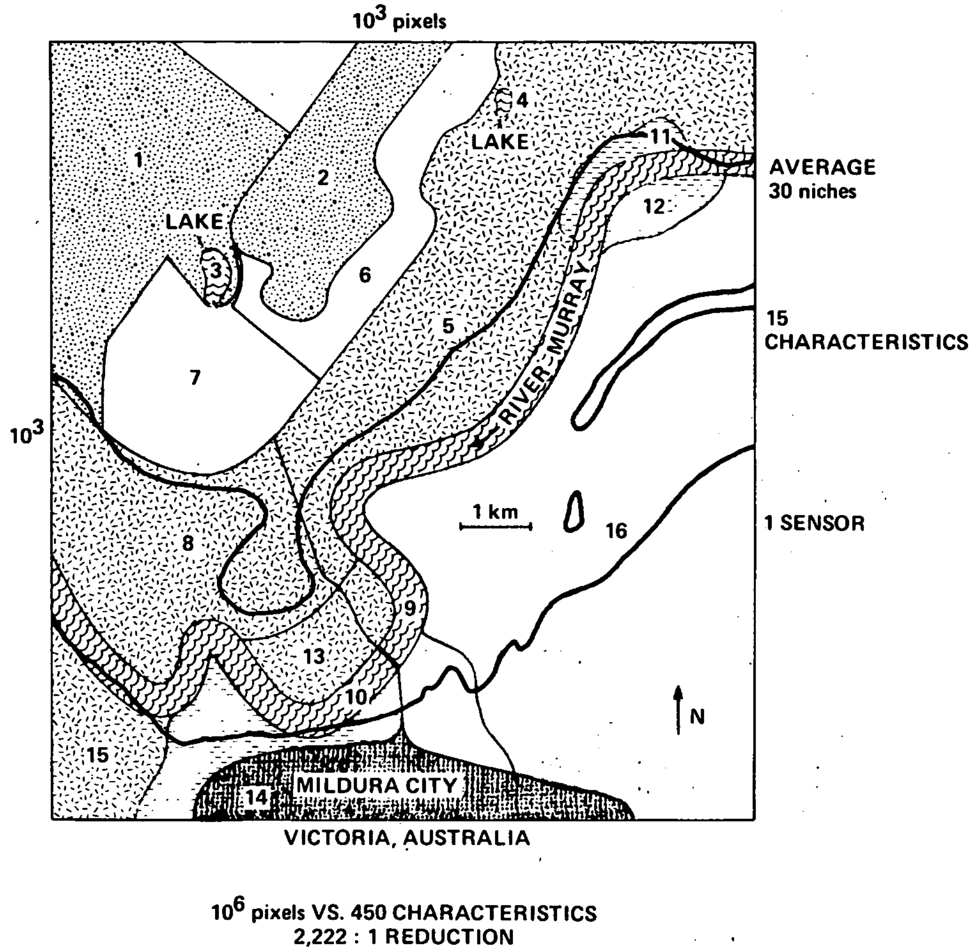


Figure 2.9.– Sample niche features map of Mildura City; alteration in boundaries of the River Murray due to flooding.

terrestrial niches, sensor characteristics, and subsidiary characteristics. The ground model thus constitutes a full working library. A land agent, for instance, easily could retrieve moisture content on all corn acreage in southwest Iowa from the master files.

In addition to a physical database memory onboard and on the ground, the world model requires extensive artificial intelligence software including expert systems which use the database in controlling sensing, image location and rectification, data processing, labeling, anomaly search, and decisionmaking on board the satellite and on the ground. A set of expert systems is needed to handle overall coordination, observation scheduling and user-generated processing tasks (fig. 2.11).

2.3.2 Onboard Memory

Present predictions for onboard memory in the year 2000 for Earth-sensing satellites are on the order of 10^{14} bits (Opportunity for Space Exploration to Year 2000. Address delivered by A. Adelman at Goddard Space Flight Center, 1980). In this section, these projections are compared with the storage capacity required for IESIS abstracting and onboard processing. The following estimates emphasize image processing because this is the type of data transmitted at the highest rates – on the order of 650 Mb/sec for SAR and 320 Mb/sec for the multilinear array (Nagler and Sherry, 1978).

The bits stored aboard the satellite or satellites for use in immediate processing tasks presumably are substantially less than the number stored on the ground in IESIS. The on-ground world model is a continually updated version of the Earth model. Future feasibility experiments plus new developments in computer science and technology will dictate the specific allocations of memory required in space and on the ground. Hence, the following are only crude estimates of the storage requirements under a range of plausible assumptions.

For the present work it is assumed that a correlation is to be performed between the incoming image and its image description stored in the onboard world model. Significant

PRINT OUT

IMAGE COORDINATES: E141-45 S30-06

LOCAL NAME: SEC 192, MILDURA, VICTORIA, AUSTRALIA

9 · 12 · 74 21:08 VMT

NICHES	CLASS	NAME
1	LAKE	GEORGE
2	RIVER	CHARLES
3	FOREST	STIRLING
4	GLEN	GREEN
5	GRASSLAND	LOW
.	.	.
.	.	.

CHANNEL #

AVERAGE SENSOR READING:	1 BLUE	2 GREEN	3 YELLOW	4 RED	5 NEAR/R	6 FAR/R	T_{COLOR}	T_{IR}
	85 ± 7	35 ± 3	7 ± 1	1 ± 1	6 ± 1	8 ± 1	5960 ± 38	310 ± 28
	63 ± 4	60 ± 2	30 ± 3	16 ± 5	5 ± 1	1.6 ± 5	4500 ± 26	305 ± 6
	7 ± 1	58 ± 7	25 ± 2	12 ± 2	8 ± 2	3 ± 1	3106 ± 22	316 ± 5
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.

H2 H3

WEIGHTED AVERAGE OVER CHANNEL:		
	COLOR TEMP	4302 ± 20
	GROUND TEMP	307 ± 02
	MOISTURE	0.06
	.	.
	.	.

H4

WOMBAT COUNT	$506/mi^2$
CLOUD COVER	15%
MURRAY IN FLOOD	

H5

HIERARCHY
- H1 ALL DATA, ALL SENSORS
- H2 NICHE AVERAGE, ALL SENSORS
- H3 NICHE CHARACTERISTIC
- H4 NEIGHBORHOOD CHARACTERISTIC
- H5 SPECIAL REQUEST

Figure 2.10.– Hypothetical IESIS user printout following scan of Mildura City.

features of the incoming image are contrasted and identified with distinguishable features (niches) catalogued in the model. The number of bits N needed to store this information on some or all features is $N = ni$, where n is the number of niches in all or part of the world model and i is the number of bits required to describe a niche. Examples that estimate N are given below.

Niches. The niche is a broad distinguishable geographical region having some common features across its surface when viewed from space. Niches are easily separable, somewhat permanent features (large rivers, canals, lakes, major highways, cultivated areas or forests) recognizable within a predetermined orbital swath. Since niches consist of common features, data acquired across them have very high redundancy. A large amount of data reduction is obtained by describing the entire niche with a limited number of common values abstracted from the whole set of niche data on record.

It is estimated that a total of $i = 122$ bits is required for individual niche identification, as follows: Location of centroid, 6 bytes, 48 bits; maximum and minimum of horizontal, 2 bytes, 16 bits; maximum and minimum vertical, 2 bytes, 16 bits; orientation, 2 bytes, 16 bits; abstract shape among 1000 choices, 10 bits; naming the niche, 2 bytes, 16 bits. In addition, s sensors are assumed scanning

GROUND

- CHECK LOCATION OF SATELLITES
- PREPARE NEW WORLD MODEL DATA, OVERALL MEASUREMENT ROUTINE FROM EXPECTED CLOUD COVER, WEATHER, SUN POSITION, ETC., FOR EACH SATELLITE IN UPCOMING MEASURMENT
- TRANSMIT DATA AND INSTRUCTIONS
- RECEIVE SATELLITE DATA
- PROCESS UNIDENTIFIED ANOMALIES AND TAKE APPROPRIATE ACTION
- UPDATE WORLD MODEL, HYPOTHESIZE SEASONAL AND TIME DEPENDENT PATTERNS, DETERMINE NEW CORRELATIONS AND TRENDS
- INTERFACE WITH USER AND DATA ARCHIVAL SYSTEMS AND NATURAL LANGUAGE INTERFACE
- COORDINATE AND MAINTAIN ALL SYSTEM COMPONENTS
- DETERMINE SYSTEM STATUS AND PERFORMANCE

SATELLITES

- DETERMINE LOCATION FROM AUTONOMOUS NAVIGATION SYSTEM USING POSITION SENSITIVE ANGULAR MEASUREMENTS TO CELESTIAL OBJECTS, NICHE LOCATIONS AND MEASUREMENTS TO KNOWN BEACONS AND THE GPS
- CALL OUT NICHES BY STRIP ALONG OVER PASS
- APPLY TEMPORAL, SENSORAL, SUN AND OTHER CORRECTIONS
- FROM NICHE TYPES AND BOUNDARIES DETERMINE OPTIMUM SENSOR COMBINATIONS, ADJUSTS MEASUREMENT STRATEGY AND ALGORITHMS
- APPLY OTHER PREPROCESSING
- PROCESS DATA
- CHECK RESULTS
- USE RESULTS TO DETERMINE NEW LOCATION
- APPLY POST PROCESSING PROCEDURES
- DETERMINE TRANSMISSION ROUTING TO GROUND

- PREPARE FOR SPECIFIC SATELLITE ORBIT
- CHANGE ALGORITHM
- CHANGE MEASUREMENT STRATEGY
- CHECK RESULT
- CHANGE MODEL
- CHANGE DATABASE

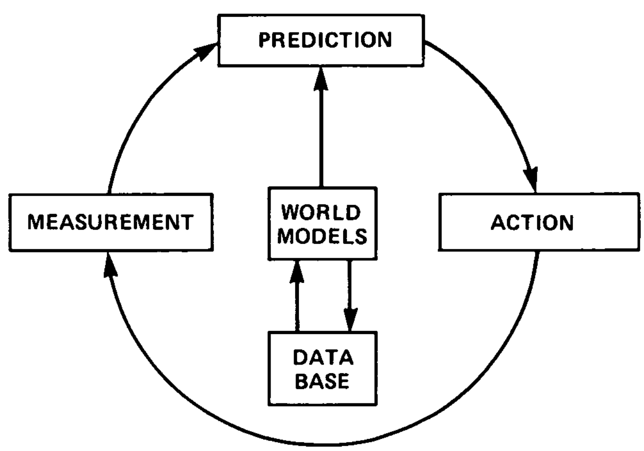


Figure 2.11.– Functions of IESIS expert systems.

various bands covering a niche having k characteristics coded in memory. Thus, 8 sk bits are needed for niche sensing, and 20 sensors having 15 characteristics require 2400 bits. It is also conceivable that 10 other characteristics might be detailed yielding an additional 80 bits. Consequently, without using boundary analysis in the image correlations, the total information i required for each niche is approximately 2600 bits, more than an order of magnitude larger than the information needed simply to identify the niche.

Thus, it is most efficient if the onboard world model simply consists of descriptions of those niches the IESIS satellite(s) will encounter along an orbital swath on the next revolution. The model is uploaded from the ground prior to each pass. If swath width in kilometers is W, repeat time (in days) for the satellite to cover the entire equatorial

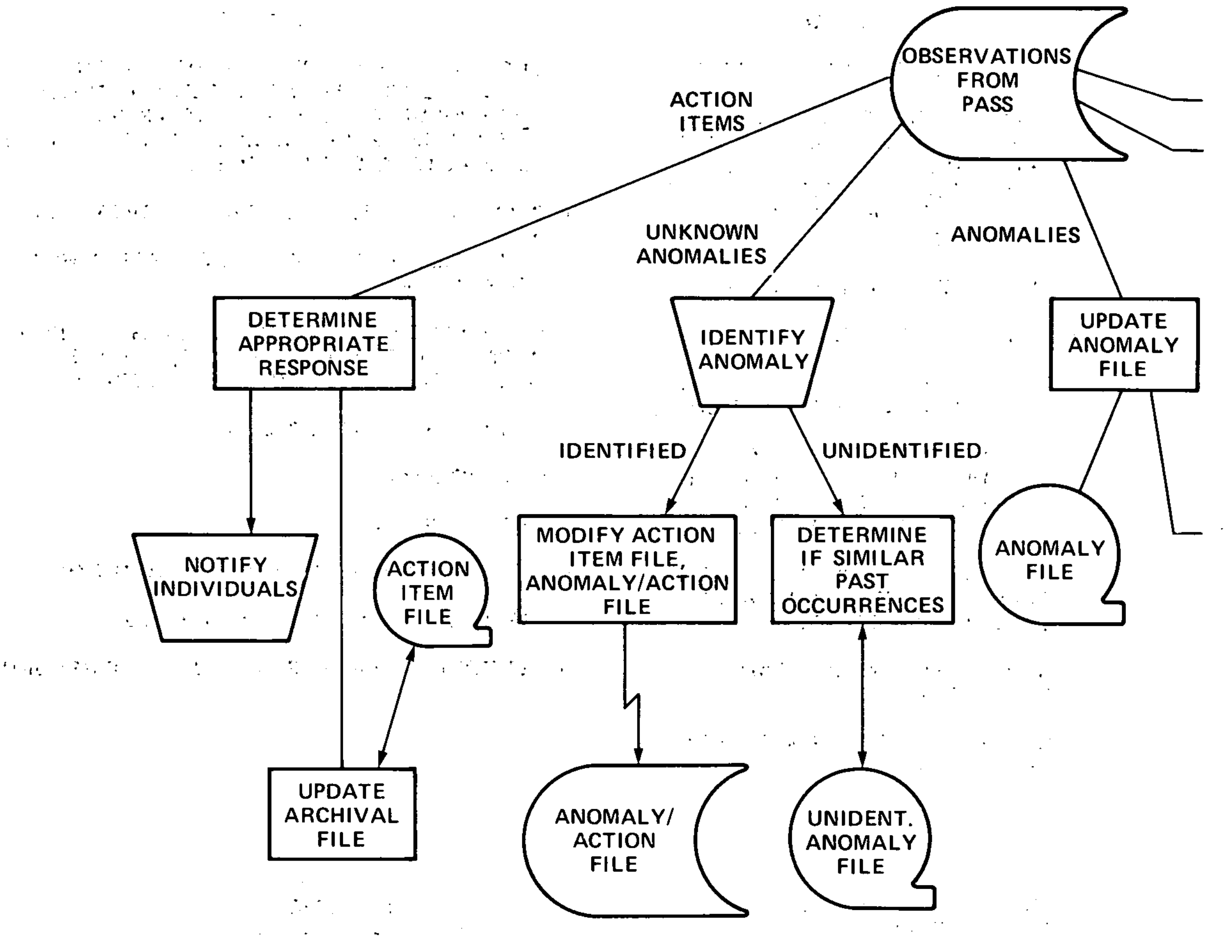


Figure 2.12.– Ground-based world model data processing operations.

region is T, and the number of passes per day (generally 15-16) is f, then, taking the Earth's circumference roughly as 4×10^4 km, $T = (4\times10^4)/fW$. If n denotes the number of niches in one pass for a given swath width, then $n = (4\times10^4)W/\text{niche area}$.

For example, if swath width is 330 km and the satellite executes 15 rev/day, then T is on the order of 8 days. If the area of a representative niche is roughly (for simplicity) 3.3 km^2, then $n = 4\times10^6$ niches and $N = ni = (4\times10^6)(2600) = 1.04\times10^{10}$ bits.

Niche at each pixel. A gross theoretical estimation is obtained by assuming that each pixel element is associated with 8 bits of information of t different types. Thus, each pixel element plays the role of a niche. If A is the resolution of each pixel in km^2 and E is the approximate surface area of Earth (5.15×10^8 km^2), then the total number of bits required to describe pixel elements covering the entire planet is $8tE/A$. If pixel resolution is 10 m, then $A = 10^{-4}$ km^2/pixel which gives

$$(5.15\times10^8 \text{ km}^2)/(10^{-4} \text{ km}^2/\text{pixel}) = 5.15\times10^{12} \text{ pixels}$$

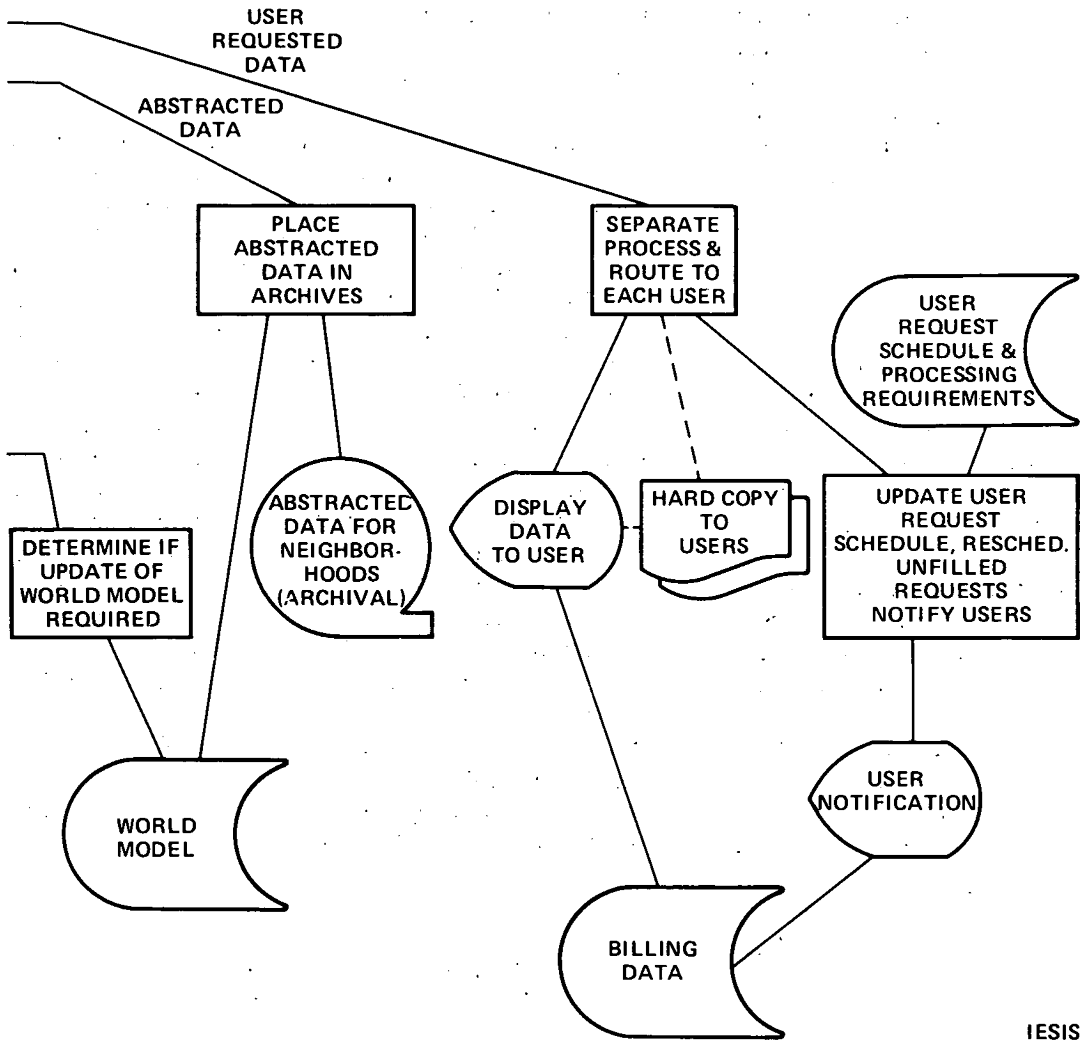
for all of Earth. The total information required to store this gross world model is of the order

$$8tE/A = 8t(5.15\times10^8)/10^{-4} = 4.12\times10^{13}\ t \text{ bits.}$$

It appears that a memory capacity onboard a satellite of the order 10^{11} bits is sufficient to allow system operation on an orbital swath basis. Presumably between 10^{14}-10^{15} bits are needed for ground memory. It is estimated that by the year 2000 roughly 10^{14} bits of in-space memory will be available, so these requirements do not appear particularly stringent (Opportunity for Space Exploration to Year 2000. Address delivered by A. Adelman at Goddard Space Flight Center, 1980).

2.3.3 Ground-Based World Model Processing

Processing at the central ground facility can be broken into paths as shown in figure 2.12. Action items which the system should generate in response to observation of an anomaly are carried out, appropriate alarms sounded and individuals notified. Anomalies that the system cannot



identify undergo close human scrutiny – if subsequently identified, the expected anomaly/action file is updated; if not, the event is catalogued and correlated with other similar but also unidentified anomalies. Identified anomalies keyed to time of occurrence, intensity, and other important parameters are archived and used to update the world model if necessary. Data abstracted from observed features are archived and used for world model updating. Information gathered in response to user requests is transmitted to customers for their exclusive use and is not automatically stored by the system.

2.4 Autonomous Satellites

This section focuses on individual satellites and the collective IESIS system, as suggested in figure 2.13. To satisfy overall IESIS goals, all satellites must be equipped with an appropriate ensemble of sensors and reside in orbits providing sufficiently frequent observation opportunities for all points of interest. Just as important, however, is the ability of each device to accept brief high-level instructions to guide its observations, and to perform massive onboard processing for abstraction of high-level information.

2.4.1 Onboard Processing

The observing satellite receives the schedule for its next pass from the uplink (section 2.4.2) and performs the requisite processing according to this instruction set. Terrestrial cloud cover and weather conditions are obtained from one of the IESIS geostationary satellites and navigational data are transmitted from a global positioning system already in place. Each satellite adjusts its attitude as required, turns detectors on and off, modifies sensor resolution, and takes both active and passive observations of the Earth. Data then are processed by comparison to predicted observables as derived from the world model.

Anomalies are identified if possible, catalogued, and appropriate action taken. Anomalies that cannot reliably be identified by IESIS are placed on file for transmission to a human analyst for interpretation and action. All sensor data are abstracted and summarized by feature and placed in the abstracted observation file for archiving and world model updating. Any user-requested processing is then performed and a user file established. These processes are shown schematically in figure 2.14.

2.4.2 Uplink and Downlink

The uplink sends detailed observational and processing schedules to IESIS satellites. A possible component of the uplink package is a set of sensor values expected to be observed during the next set of observations. This expectation is generated by using the world model as the standard against which to compare sensor observations so that anomalies may be identified. The uplink is illustrated in figure 2.15. Information needed during the next observation period is transmitted from the central processing facility via ground station and geosynchronous communications satellite up to the observation satellite as discussed in section 2.2.

The downlink reverses the uplink process by consolidating data gathered by satellite and returning them to the ground in one transmission. The downlink also uses the geosynchronous communications link to transmit data to a ground station which relays the information to the central processing facility. The process of downlink consolidation is illustrated in figure 2.16.

2.4.3 Image Processing

A primary IESIS requirement is the necessity to perform rapid and massive data reduction aboard the satellite in the sequence suggested in table 2.6. The focus is on image data acquired at high rates – presently 120 Mb/sec for SIR, 85 Mb/sec for the Thematic Mapper, and in excess of 600 Mb/sec for SAR systems (Nagler and Sherry, 1978). Such rates may arise in each of perhaps 20 sensors in certain extreme cases of IESIS operations, which requires that the spacecraft carry onboard high-speed processors.

IESIS high-speed processors might evolve from faster serial logic devices, e.g., those which may be developed from Josephson tunnel technology. However, an alternative approach is the use of parallel logic to perform many operations simultaneously. By executing thousands of computations at once an intrinsic speed advantage equal to the number of individual processors is theoretically possible. In practice, this hypothetical limit may not be attainable due to pragmatic technological restrictions on each individual element in an array of thousands of processors. In spite of this, computing speeds within two orders of magnitude of the theoretical limit have already been obtained (Schaefer, 1980). Data handling using thousands of active elements simultaneously is called "parallel processing."

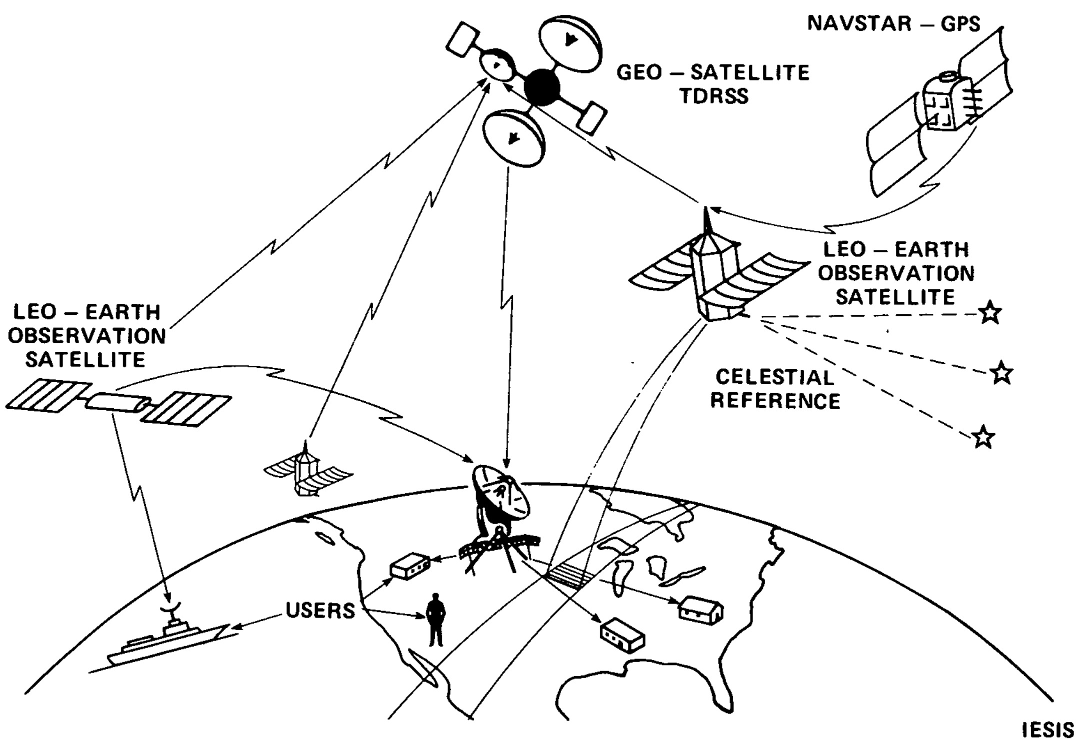


Figure 2.13.– IESIS satellite system.

Preprocessing functions including extremely rapid rectification of images, formatting, noise removal, imbalance, and radiometric correcting can be performed best by manipulating data in parallel rather than serially. Projections for serial onboard image processing range up to 10^9 operations per second (Opportunity for Space Exploration to Year 2000. Address delivered by A. Adelman at Goddard Space Flight Center, 1980), whereas, those for parallel processors extend up to 10^{11} operations per second per pixel (assuming 100 clock cycles per operation). Another capability expected to be using parallel processing is a 1000 X 1000 parallel input array operating at up to 100 MHz clock rates (James Strong, personal communication, 1980). In one proposed 1000 X 1000 input array, as shown in figure 2.17, threshold photosensors on a wafer lead directly to massively parallel processing elements which in turn are connected (also through the wafers) to memories. Processing rates expected for massively parallel

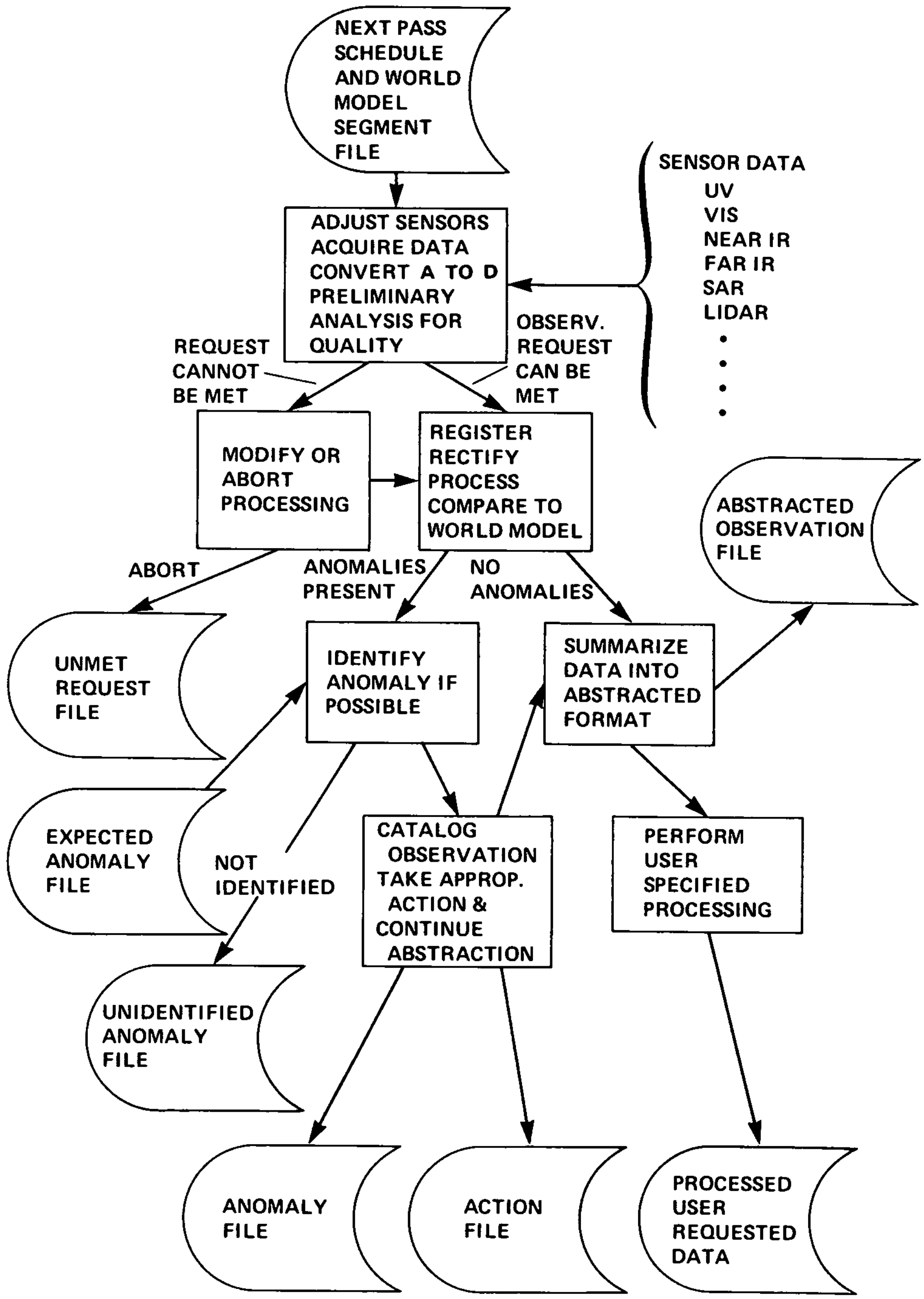


Figure 2.14.– Onboard processing flowchart.

TABLE 2.6.– ONBOARD DATA REDUCTION PROCESSES.

Preprocessing
Name image region
Format image data
Apply sensor corrections
Processing
Cross-correlate image and map
Geometric correction
Resample data
Obtain niche boundary
Generate niche mask
Process niche data
Combine data across sensors
Detect and characterize anomalies
Postprocessing
Attach niche labels
User tags
Assign priorities
Update data base

processors (MPP) are tabulated in table 2.7. Further information and technical descriptions are available in Gilmore et al. (1979).

TABLE 2.7.– ESTIMATED SPEED OF COMPUTATIONS USING MASSIVELY PARALLEL PROCESSOR.

Function	10^8 Hz clock
Area	6 μsec
Average	50 μsec
Variance	0.11 msec
Slope	1.4 μsec
Fourier transform	11.7 msec
Histogram	1.5 msec
Classification	0.6 msec
Matching two images	81.8 msec
Resampling	0.05 msec

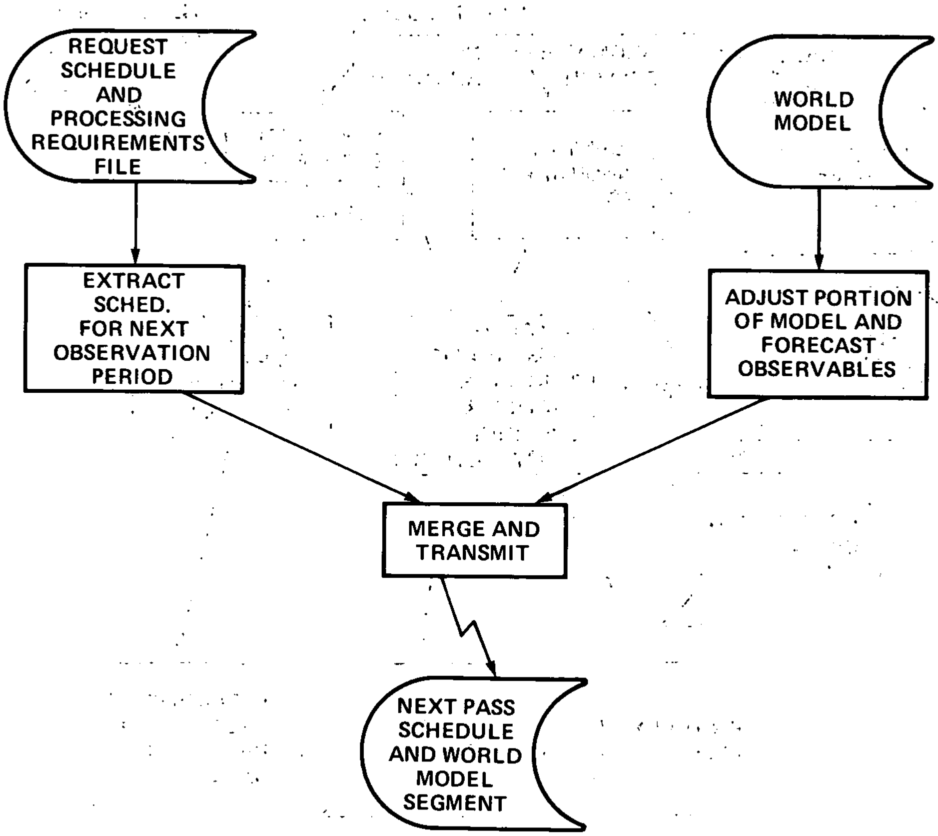


Figure 2.15.– IESIS uplink processing.

2.4.4 Sensor Requirements

IESIS sensor requirements are dictated by the specific terrestrial environment that must be scanned to fulfill an Earth resource mission. Full utilization of satellite capabilities and operating time demands useful sensing during daylight and nighttime passes of an orbit and during cloud cover.

Daylight operation involves observation by sunlight filtered through atmosphere twice before detection by satellite sensors. Most filtered sunlight lies in the visible region extending somewhat into the near-UV and further into the near-infrared. An optimum match between the daylight Earth environment and the satellite passive scanning system must include a visible, near-infrared, and some ultraviolet detection capability. A properly chosen 3-dimensional color space obtained using a red, green, and blue filter set yields color discrimination roughly comparable to that of the human eye. Detection at the chlorophyll absorption region near 650 nm gives useful discrimination for vegetation, while for water detection the low reflectivity region near 850 nm is useful (Schappell and Tietz, 1979). A pair of UV, five visible, and three near-IR bands should provide sufficiently broad color space (10 dimensions) to allow very widespread signature analysis of important terrestrial features such as crops, rivers, lakes, clouds, forests, and snow covers.

The nighttime environment may be scanned passively for thermal radiation at a temperature near 300 K. The blackbody emissions of the cool Earth peak in the far-IR near 10 μm. Four wideband far-IR sensors would allow accurate temperature and signature definition of nighttime features, although not to the same precision and resolution as with daytime sensing.

All-weather capability requires active microwave scanning of the Earth. The Synthetic Aperture Radar (SAR) operating at 1–10 GHz (Nagler and Sherry, 1978; OAST, 1980) is capable of essentially all-weather observation at good resolution. The SAR system also would augment nighttime passive measurement in the far-IR.

Altitude sensing provides useful information about terrestrial resources such as crop height, reservoir levels, or mountain snow cover. Height is recorded from differential altitude measurements performed at a boundary, e.g., by comparing the heights of crop tops to nearby level ground. A differential altitude measurement system is possible using rapid Q-switched LIDAR. Absolute altitude measurements can be obtained by LIDAR or microwave altimeter. Differential velocity measurements at a boundary (e.g., a river bank) can be taken by Doppler shift analysis.

Undoubtedly there will be requirements for additional specialized optical, infrared, and microwave sensor bands to detect important surface and atmospheric components such as ozone, water, water vapor, and carbon dioxide (Golovsko

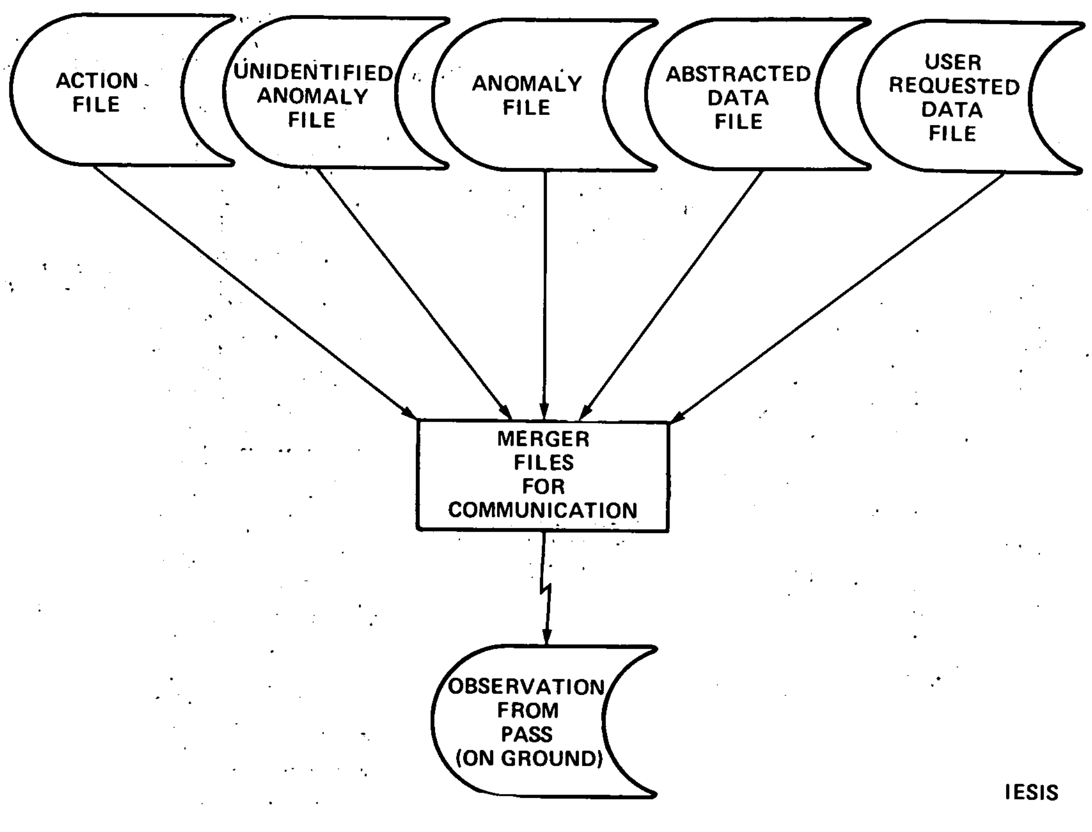


Figure 2.16.– IESIS downlink processing.

and Pakhomov, 1978). Somewhat artibrarily, six specialized sensor bands have been allocated to these uses for purposes of the present study.

Sensors are configured into a wide-angle medium-resolution sensor array and additionally into at least two arrays of narrow-angle, high-resolution sensors capable of independent accurate aiming over at least the swath width of the wider-angle array. Narrow arrays allow independent coverage of terrestrial features needed to satisfy conflicting requests during a particular orbital pass. They also may be used to obtain stereoscopic imagery, say, of cloud tops, by setting one array to view forward and the other aft during a pass and later combining the two image streams appropriately. Table 2.8 summarizes one possible sensor set.

2.4.5 *Data Rate Estimation*

The bit rate generated by one optical or IR sensor is given approximately by $8SBVW/A$, where S is the number of sensor sets, B is bands per sensor set, V is orbital velocity, W is swath width, and A is pixel resolution in km^2 for an 8-bit pixel. For two sets of narrow sensors and one set of wide sensors the total bit rate generated at a 7 km/sec orbital speed is $8(2)(10)(7 \text{ km/sec})(110 \text{ km})/(5\times10^{-3} \text{ km})^2 + 8(1)(20)(7 \text{ km/sec})(330 \text{ km})/(15\times10^{-3} \text{ km})^2 = 6.6\times10^9$ bits/sec. Of the remaining sensors the SAR will produce the maximum data rate by far. Today's SAR apparatus generates data at 0.65×10^9 bits/sec (Nagler and Sherry, 1978). Representative Doppler LIDAR, Doppler radar, and laser altimeters return data at the rate of several tens of kilobits per second. Thus, onboard computing capability requirements must be sufficient to handle data rates near 7×10^9 bits/sec. This is roughly an order of magnitude higher than that used in present Landsat orbiters.

2.4.6 *Satellite Requirements*

A summary of required measurement rates has been provided by Nagler and Sherry (1978) for a wide range of environmental and resource assessments. The necessary frequency of observation generally is lowest for land-based features, higher for ocean observation, and highest for atmospheric and weather assessments, with considerable overlap in the requirements. Table 2.9 indicates the frequencies of Earth observation and attendant swath widths believed reasonable for the IESIS system.

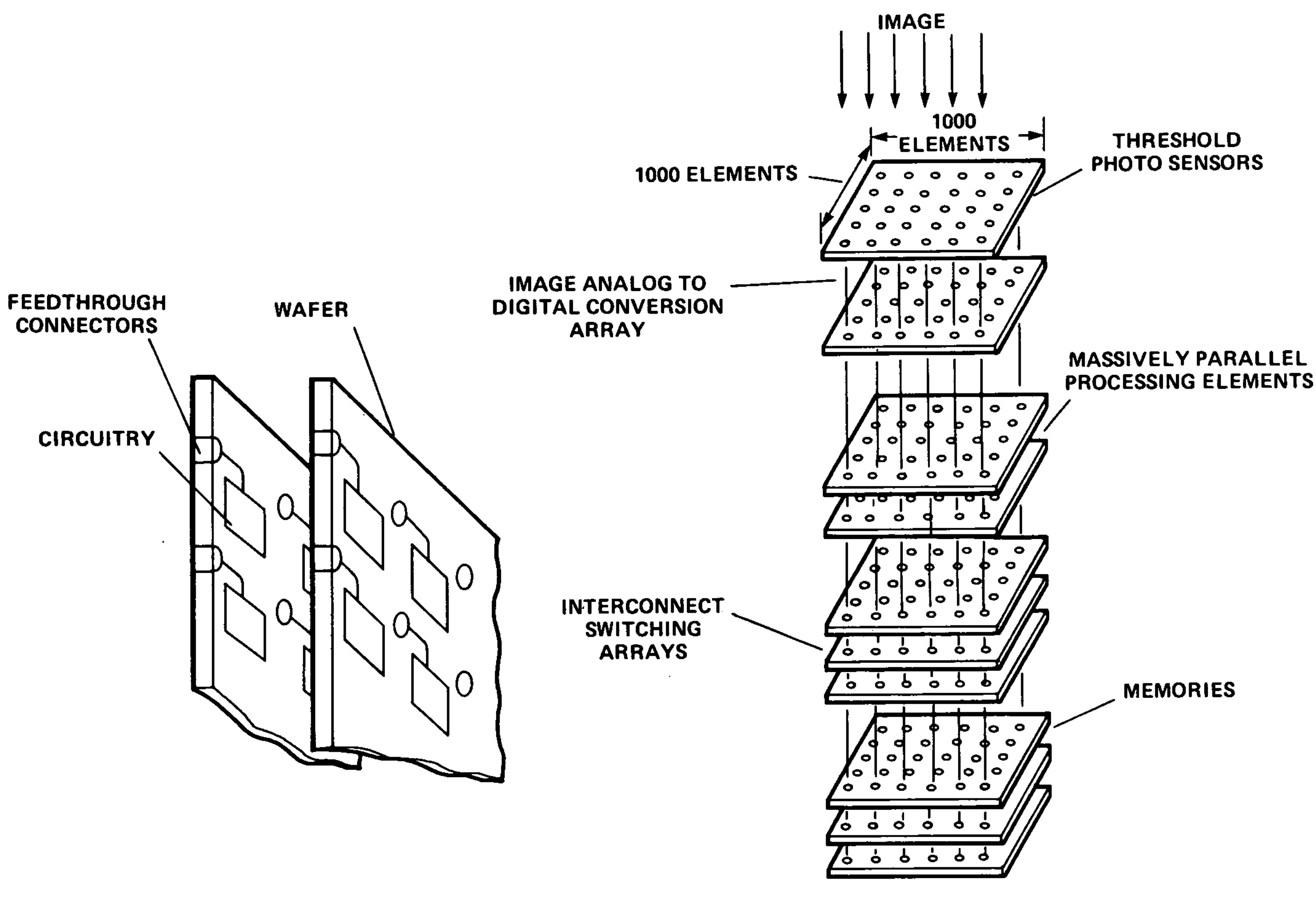


Figure 2.17.– A projected advanced massively parallel processing system.

TABLE 2.8.– SENSOR CONFIGURATION OF AN EARTH OBSERVING SATELLITE.

Configuration and array of a possible set of active and passive sensors in an early mission. All sensor bands to be operated simultaneously if desirable.

Configuration

1 set of wide angle sensors comprising the full array of sensors to scan 330 km swath at 15 m resolution, or as limited by individual sensor.

2 sets of narrow angle UV, visible and near IR sensors, capable of accurate aiming, to cover 110 km swath at 5 m resolution; $\sim 7\times 10^9$ bits/sec.

Sensor Array

10 bands – UV, visible and near IR (daylight)
4 bands – far IR (night)
6 specialized bands (atmospheric composition)
SAR (all weather)
LIDAR
Differential height
Differential velocity
Altimeter

TABLE 2.9.– POSTULATED OBSERVATION FREQUENCY AND SWATH WIDTHS.

Niche features observed	Observation frequency, per day	Maximum swath width, km
Land	0.5	350
Ocean	3	700
Atmosphere	12	1400

The IESIS satellite program is envisioned as developing in a long-term sequence carrying well into the next century. A detailed world model of land features already exists as contour maps covering a significant portion of the continents. Land features have sharp boundaries and vary only slowly over time. Oceans have wider geographic features that vary seasonally. The atmosphere requires three-dimensional modeling of rapidly varying phenomena. An obvious difference between land and ocean or atmosphere from a user standpoint is the large human population on land and its virtual absence elsewhere. Table 2.4 summarizes the characteristics of world models of the land, oceans, and atmosphere.

The logical deployment sequence of user-oriented resource satellites begins with a set of basic land-observing satellites whose world model already can be rather fully detailed. Since the satellites will spend about 75% of their time over the ocean it is natural to include ocean-sensing capability with as much ocean modeling as is feasible at the time of design and launch. Atmospheric sensing and rudimentary modeling should be included, both for understanding the state of the atmosphere and also as a necessary part of the interpretation process for sensor readings of land and ocean observations.

To assure long life for these sophisticated satellites, reasonably high orbits are required. Atmospheric path distortion and sun angle introduce errors and complications into the interpretation process for imaging data. Path distortion causes reddening and other wavelength-dependent absorptions, and Rayleigh and Mie scattering are especially sensitive to particle size in the atmosphere and to sun angles.

The use of sun-synchronous satellites simplifies the situation considerably, a rational initial constraint which could be removed at some later time when more sophisticated modeling becomes available. It appears reasonable to have a set of sun-synchronous satellites operating continuously so that each Earth ground point is covered at the equator every 2 days by at least one satellite of the set. Sun-synchrony produces roughly the same sun-angle conditions over an observed land point for a particular satellite and helps to standardize image interpretation for that satellite at that point. An orbit near present-day Landsat altitudes (920 km, nominal) will support a long-lived satellite. If altitude is adjusted to a 14-1/8 rev/day rate, the ground trace of a particular satellite repeats every 8 days. Four such satellites could cover the Earth with the desired 2-day period.

The swath width required of a satellite for 8-day coverage at 14-1/8 rev/day is about 350 km. However, in order to take account of partial cloud cover the team proposes six satellites in sun-synchronous orbits. If these are placed substantially uniformly about the Earth's circumference the local viewing times for each satellite are spaced about 2 hr apart. Bunching may be desirable if there are reasons to pick a particular local viewing time. Polar conditions can be monitored by a seventh polar satellite, which may also act as a spare if one of the sun-synchronous satellites is disabled.

To relay data to the continental United States, two geostationary satellites are required. These satellites are also used to monitor global conditions, particularly cloud cover. Global cover information is compiled by IESIS to prepare each satellite for the tasks it can most usefully perform during its upcoming orbit, by enabling modifications in sensors and processing to optimize the information obtained from each series of observations.

Ocean coverage of a particular ocean point three times per day with a 700-km swath width requires 12 satellites,

and atmospheric coverage at a rate of 12 times per day with a 1400 km swath requires 24 satellites, each with an 8-day repeat cycle (assuming the same sun-synchronous orbital parameters given above for land-observing satellites).

The technology available in the year 2000 (hypothetical IESIS deployment date) will, of course, dictate the actual satellite configuration employed. Still, the initial set of satellites should emphasize land observation with more sophisticated oceanic and atmospheric satellites phased in as the ability to model these systems develops. Figure 2.18 summarizes the basically land-observing system described earlier.

2.5 Time Phasing

The intelligent Earth-sensing information system proposed herein is an evolutionary system which considerably extends both planned and existing NASA missions. The time-phasing chart in figure 2.19 is oriented to development of the various components culminating in an operational system by the year 2000. Little attention was given to sensor development as this technology is driven by the various demands of other users and by general progress in this technical area (Breckenridge and Husson, 1979). Most attention was directed to software and artificial intelligence development as these lie at the heart of IESIS, although advanced hardware technology R&D also is required to achieve high packing densities, large wafers, fault-tolerant designs, advanced cooling techniques, advanced interconnections, more logic functions between array elements, advanced data output, and parallel input from sensors to buffer memory. Some of the major points are as follows:

- Research into automatic mapping and world model development should begin early. A world model for use onboard should be ground-demonstrated by mid-1987 and a Shuttle demonstration of the world model/sensor operation completed by 1990 to meet the projected IESIS deployment date (2000 AD).
- Parallel processor development should be given high priority. The Massively Parallel Processor (a 128 × 128 array processor) will be operational in 1982. A 1000 × 1000 (or perhaps a 10,000 × 100) array for parallel processing should be developed by 1990 and should be flown on a Shuttle test satellite by 1995 to meet the 2000 AD deadline.
- Natural language user interfacing with the data system should be operational by 1990.
- Development of a model of the user population should begin immediately and be phased with natural language and world model development. The prospective ground demonstration of the world model using direct data from an advanced Landsat can be made available at some point to selected users on an experimental basis. The information on prospective selected users can form a preliminary user model.
- Signature analysis, data handling, and security will require continuing development and algorithm refinement. By 1995, software should be flown onboard both experimentally and as an initial phase-in on the autonomous satellites.
- A large world model encompassing terrestrial, oceanographic, and atmospheric components and a satellite system scheduler/controller should be ready by the year 2000.
- A gradual phase-out of Landsat D orbiters and phase-in of more autonomous "smart" satellites should begin. By the year 2000, fully autonomous satellites carrying world models should be available for long-term operation and initiation of the complete IESIS program.

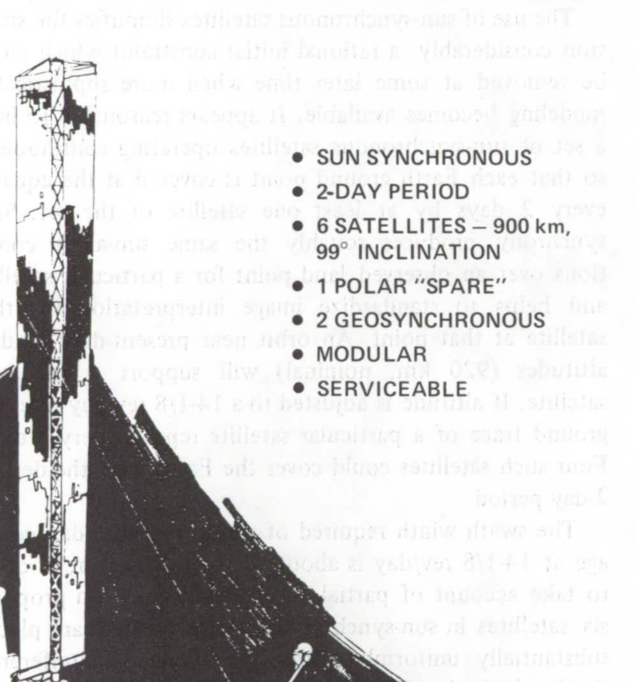


Figure 2.18.– Orbit characteristics: Possible initial IESIS satellite configuration.

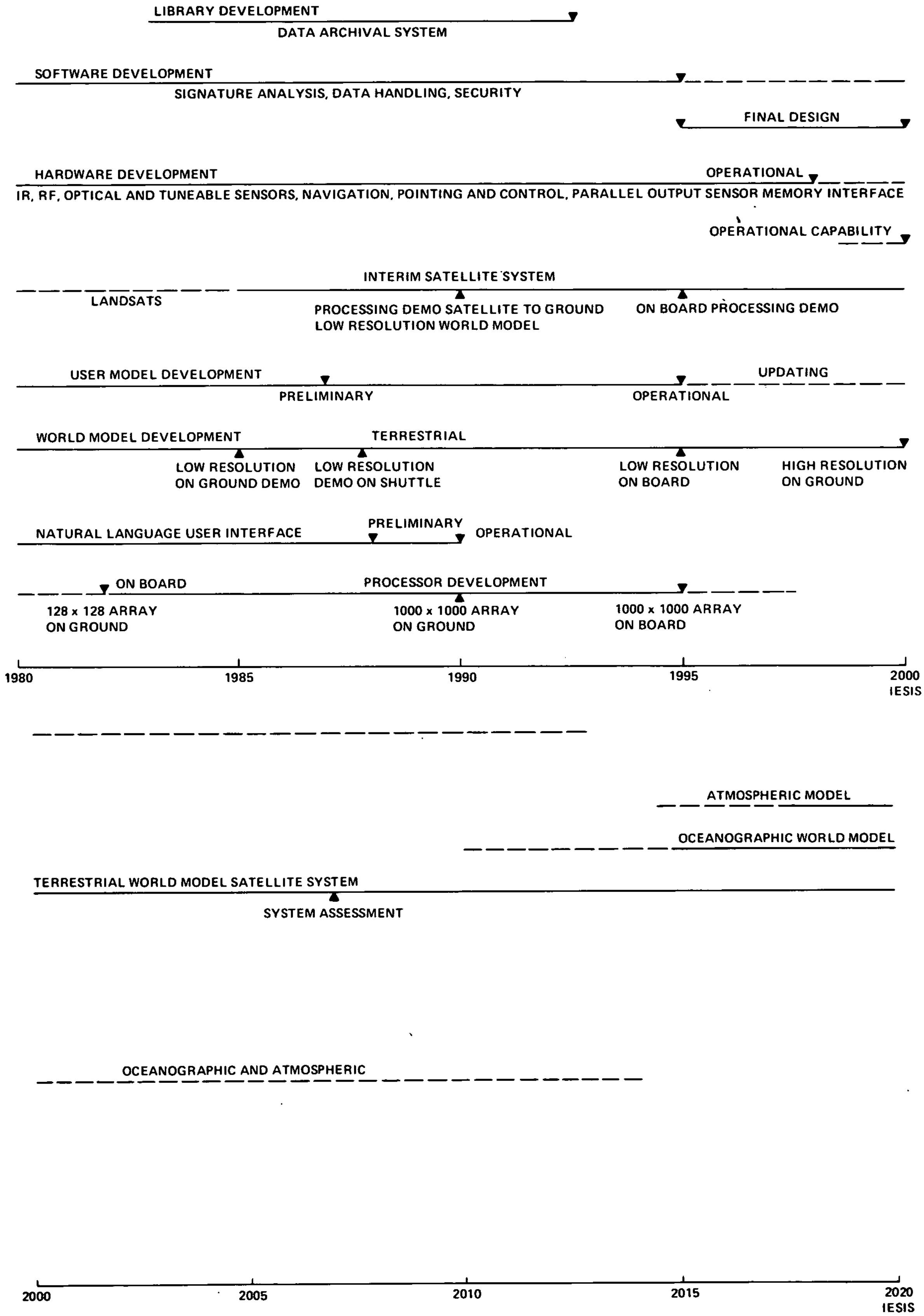


Figure 2.19.– IESIS time phasing chart, 1980-2020.

2.6 References

Breckenridge, R. A.; and Husson, C.: Smart Sensors in Spacecraft: Impacts and Trends, in Remote Sensing of Earth from Space: Role of "Smart Sensors," Roger A. Breckenridge and Martin Summerfield, eds., Progress in Astronautics and Aeronautics, AIAA, vol. 67, pp. 3-12, 1979.

Chien, R. T.; and Peterson, L. J.: Image Compression and Reconstruction Using Feature Extraction, in Proceedings of the 5th International Joint Conference on Artificial Intelligence, MIT, 22-25 August 1977, Cambridge, p. 658, 1977.

Gilmore, P. A.; Batcher, K. E.; Davis, M. H.; Lott, R. W.; and Buckley, J. T.: Massively Parallel Processor. Phase I Final Report, prepared by Goodyear Aerospace Corporation for Goddard Space Flight Center, July 1979.

Golovko, V. A.; and Pakhomov, L. A.: Simultaneous Determination of Atmospheric Temperature, Water Vapor Profiles and Thickness from Infrared Interferometer Measurements on Board "Meteor 28." In Remote Sounding of the Atmosphere from Space (Advances in Space Exploration, vol. 4), COSPAR, Pergamon Press, New York, pp. 247-250, 1978.

Graham, Donald Norman: Image Transmission by Two-Dimensional Contour Coding, Proc. IEEE, vol. 55, no. 3, pp. 336-346, March 1967.

Nagler, R. G.; and Sherry, E. J.: Global Services Satellite Circa 1995: Volume II – Environmental and Resource Management Measurement Services, Jet Propulsion Laboratory, Pasadena, California, 30 September 1978.

Office of Aeronautics and Space Technology (OAST): NASA Space Systems Technology Model: Volume I – NASA System/Program Descriptions and Technology Needs, NASA TM-81041, May 1980.

Outlook for Space: NASA SP-386, 1976.

Rich, E.: Building and Exploiting User Models. In Proceedings of the 6th International Joint Conference on Artificial Intelligence, 20-23 August 1979, Tokyo, pp. 720-722, 1979.

Schaefer, D. H.: NASA End-to-End Data System – Massively Parallel Processor, Goddard Space Flight Center, Greenbelt, Maryland, 30 May 1980.

Schappell, R. T.; and Tietz, J. C.: Landmark Identification and Tracking Experiments. In Remote Sensing of Earth from Space: Role of "Smart" Sensors, Roger A. Breckenridge and Martin Summerfield, eds., Progress in Astronautics and Aeronautics, AIAA, vol. 67, pp. 134-151, 1979.

CHAPTER 3

SPACE EXPLORATION: THE INTERSTELLAR GOAL AND TITAN DEMONSTRATION

3.1 Introduction

The small Pioneer 10 spacecraft, launched from Earth on March 2, 1972, represents mankind's first physical extension into interstellar space. Having traversed the Asteroid Belt and given scientists their first good look at Jupiter and its satellites, the vehicle now rushes toward the edge of the Solar System at a speed of about 3 AU/yr. The exact moment of penetration into extrasolar space is unpredictable because the boundary of our System is not precisely known, and because the spacecraft's ability to transmit useful data will likely degrade by the time of passage (circa 1986) that it will be unable to report transit of the heliosphere when this occurs.

Several other unmanned vehicles will also eventually exit the Solar System. However, as Pioneer 10 none of these were designed specifically as interstellar probes, and comparatively little work has yet been accomplished with the aim of developing such craft. Still less effort has been directed toward the ultimate goal of manned interstellar exploration.

3.1.1 Automated Interstellar Space Exploration

The most extensive study of interstellar space exploration to date has been Project Daedalus, an analysis conducted by a team of 13 people working in their spare time under the auspices of the British Interplanetary Society from 1973 to 1978 (Martin, 1978). The focus was a feasibility study of a simple interstellar mission using only present technology and reasonable extrapolation of foreseeable near-future capabilities.

The proposed Daedalus starship structure, communications systems, and much of the payload were designed entirely within today's capabilities. Other components, including the machine intelligence controller and adaptive repair systems, require a technology which Project members expected would become available within the next several decades. For example, the propulsion system was designed as a nuclear-powered, pulse-fusion rocket engine burning an exotic deuterium/helium-3 fuel mixture, able to propel the vessel to velocities in excess of 12% of the speed of light. Planetary exploration and nonterrestrial materials utilization were viewed as prerequisites to the Daedalus mission, to acquire useful experience and because the best source of helium-3 propellant is the atmosphere of the gas giant Jupiter (to be mined using floating balloon "aerostat" extraction facilities). This ambitious interstellar flyby was thought possible by the end of the next century, when a solar-system-wide human culture might be wealthy enough to afford such an undertaking. The target selected for the first flight was Barnard's star, a red dwarf (M5) sun 5.9 light years away in the constellation Ophiuchus.

The central conclusions of the Project Daedalus study may be summarized roughly as follows: (1) Exploration missions to other stars are technologically feasible; (2) a great deal could be learned about the origin, extent, and physics of the Galaxy, as well as the formation and evolution of stellar and planetary systems, by missions of this kind; (3) the necessary prerequisite achievements in interplanetary exploration and the accomplishments of the first interstellar missions would contribute significantly to the search for extraterrestrial intelligence (SETI); (4) a funding commitment over 75-80 years is required, including 20 years for vehicle design, manufacture and checkout, 30 years of flight time, and 6-9 years for transmitting useful information back to Earth; and (5) the prospects for manned interstellar flight are not very promising using current or immediately foreseeable human technology.

A more recent study (Cassenti, 1980) concludes on a more optimistic note: "We are like 19th Century individuals trying to imagine how to get to the Moon. Travel to the stars is extremely difficult and definitely expensive, but we did get to the Moon and we can get to the stars." Cassenti supports the Project Daedalus judgment that only vehicles capable of achieving more than 10% of the speed of light should be examined and that the preferred propulsion system now is "a version of the nuclear pulse rocket for unmanned exploration and combinations of the nuclear pulse rocket and the laser-powered ramjet for propelling manned interstellar vehicles."

Even more imaginative and longer-range interstellar missions of galactic exploration have been considered by Robert A. Freitas Jr., a participant in the present study (Freitas, 1980a, 1980b; Valdes and Freitas, 1980). He concludes that self-reproducing interstellar probes are the preferred method of exploration, even given assumptions of a generation time of about 1000 years and a 10-fold improvement in current human space manufacturing technology. He envisions "active programs lasting about

10,000 years and involving searches of 1,000,000 target stars to distances of about 1000 light years in the Galactic Disk ..." and states that interstellar probes will be superior to beacon signals in the search for extraterrestrial intelligence.

The Space Exploration Team was charged with defining a challenging mission for the next century which could be a technology driver in the development of machine intelligence and robotics. Interstellar exploration was early identified as the ultimate goal, where this would focus on an investigation of planetary systems in the solar neighborhood discovered through SETI operations or by searches with large apodized visual telescopes (Black, 1980) in Earth orbit. Though previous studies of interstellar exploration missions are few, even these clearly suggest the need for high levels of automation.

The Team defined a general concept of space exploration centered on the notion of an autonomous extrasolar exploratory machine system. This system incorporates advanced machine intelligence and robotics techniques and combines the heretofore separate and manpower-intensive phases of reconnaissance, exploration, and intensive study into a single, integrated mission. Such an automatic scientific investigation system should be useful in the exploration of distant bodies in the Solar System, such as Jupiter and its satellites; Saturn and its rings; Uranus, Neptune, Pluto and their moons; and perhaps comets and asteroids as well. It may provide tremendous economies in time, manpower, and resources. Interstellar exploration seems virtually impossible without this system, which is itself a magnificent technology driver because the level of machine intelligence required far outstrips the state of the art (see section 3.3).

This report cannot review the entire gamut of reasons for human interest in the physical exploration of the Solar System and the Universe. Recent space research programs have stimulated large numbers of people from various scientific disciplines to join in the challenge of interplanetary exploration. Astronomers and geologists have participated since they represent the sciences traditionally most involved in the observation and classification of planetological and celestial phenomena. During the last two decades researchers from other physical sciences and the biological sciences have become interested in investigating how the laws of nature operate in the cosmos, using the techniques of radio astronomy and space exploration including direct biological samplings of other planets. Interest in the outer Solar System and deep space will likely remain high among natural scientists.

It is assumed that these reasons, coupled with the seemingly basic need of human beings to satisfy their inherent curiosity when confronted by new environments, are sufficient to motivate the economical exploration programs that advanced machine intelligence systems will make possible. Appendix 3A includes a summary of the ideas of the team's student member, Timothy Seaman, whose feelings may be representative of those of the generation of young Americans most likely to receive the first major benefits from mankind's more ambitious future ventures into space.

Although interstellar exploration was identified as the ultimate goal, detailed mission analyses are not provided. The determination of technological, economic and political feasibility for such complex, expensive, and extraordinarily long-duration undertakings must wait until advanced machine-intelligence capabilities of the type required for an extrasolar voyage have been successfully demonstrated in planetary missions conducted entirely within the Solar System. Accordingly, the major emphasis of the present study is a Titan Demonstration Mission (fig. 3.1) conceptualized to require the evolution of equipment and machine intelligence capabilities which subsequently may be applied to autonomous interstellar operations.

3.1.2 The Titan Demonstration Mission

The demonstration mission concept leads ultimately to development of a deep space-system incorporating advanced machine intelligence technology capable of condensing NASA's current three investigatory phases – reconnaissance, exploration, and intensive study – into a single, integrated, autonomous exploratory system. This should yield significant economies in time and resources over present methods (table 3.1).

TABLE 3.1.– SPACE EXPLORATION: THE INTERSTELLAR GOAL AND TITAN DEMONSTRATION

Goal: Evolution of capability for autonomous investigation of unknown domain.

Approach: Integrate previously separate investigation steps into single mission.

- Advanced propulsion capability
- Global scale investigation by remote sensing
  - Advanced sensors
  - Machine intelligence for information extraction and plan follow-up
  - Limited number of in situ exploration vehicles
  - Autonomous hypothesis formation to classify information and develop new theories

The Space Exploration Team proposes a general-purpose robot explorer craft that could be sent to Titan, largest of Saturn's moons, as a technology demonstration experiment and major planetary mission able to utilize the knowledge and experience gained from previous NASA efforts. Titan was chosen in part because it lies far enough from Earth to preclude direct intensive study of the planet from terrestrial

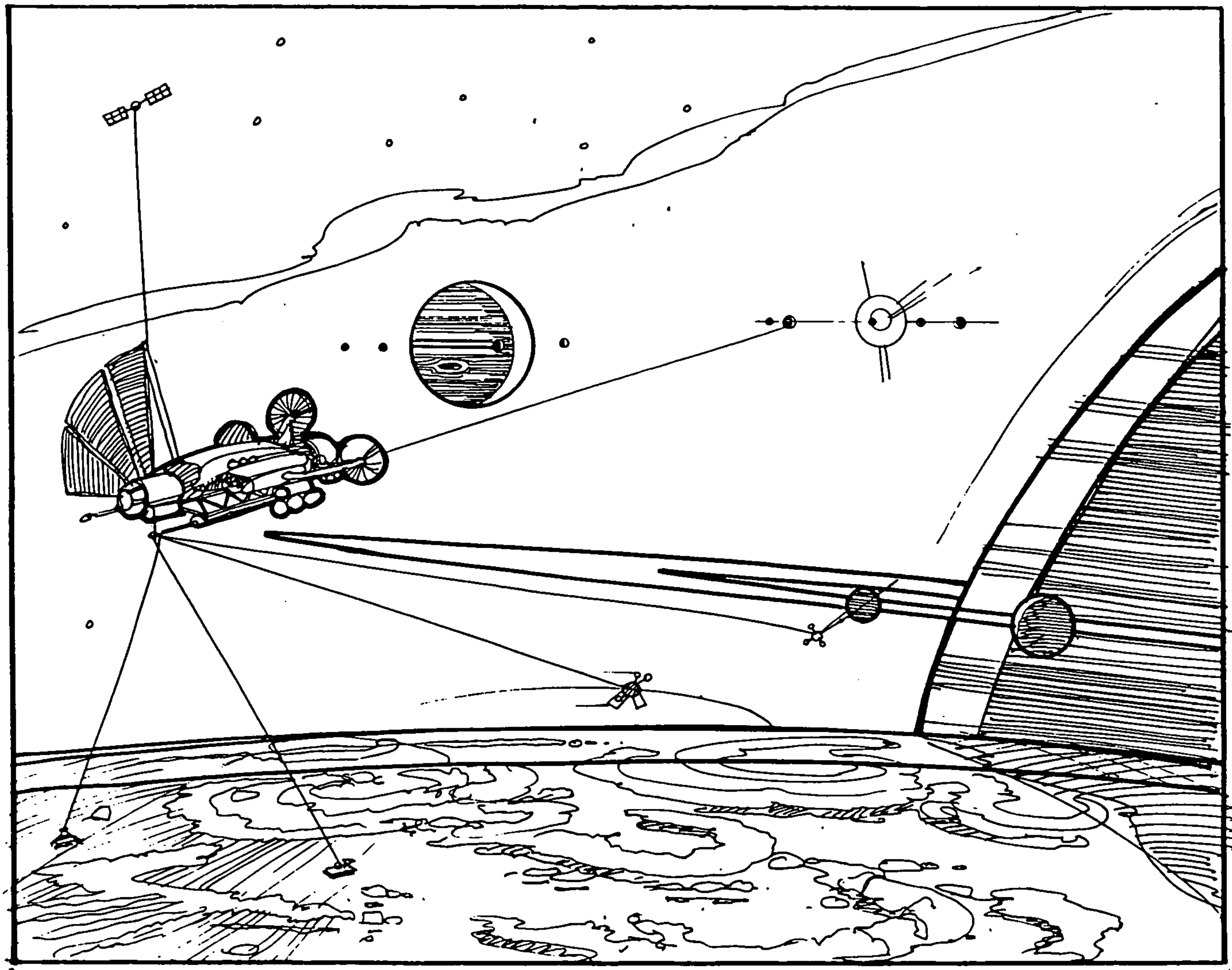


Figure 3.1.– Titan Demonstration Mission.

observation facilities or easy teleoperator control, yet is near enough for system monitoring and human intervention as part of a developmental process in the demonstration of a fully autonomous exploration technology. Such capability must include independent operation from launch in Low Earth Orbit (LEO); spiral Earth escape; navigation; propulsion system control; interplanetary flight to Saturn followed by rendezvous with Titan; orbit establishment; deployment of components for investigation and communication; lander site determinations; and subsequent monitoring and control of atmospheric and surface exploration and intensive study. The target launch date for the Titan Demonstration Mission was taken as 2000 AD with 5 years on-site. Knowledge gained from the Titan exercise could then be applied to the design of follow-on exploration missions to other planetary systems.

A number of specific criteria were decisive in the selection of Titan as a premier demonstration site for the autonomous exploration system concept:

(1) Titan is one of the few bodies in the Solar System where the physical and atmospheric conditions are partially unknown and interesting, but also still lie within acceptable tolerance ranges for equipment survivability.

(2) Titan, 9.54 AU distant from the Sun, is far enough from Earth to preclude intensive study using terrestrially based, scientific, experimental, and observational equipment, to deny easy teleoperator operations, and to require fully autonomous systems functioning while still being close enough for monitoring and intervention by humans as the demonstration experiment evolves.

(3) The existence of a heavy atmosphere provides a good test for system flexibility since atmospheric modeling

is crucial in understanding surface conditions and evaluating the possibility of life. Thus, smart multispectral correlation systems development is essential.

(4) The shrouded surface provides an unknown environment in which to test imaging systems without bias.

(5) Titan is better capable of capturing and holding the public interest than other bodies for some of the same reasons that it has received increasing scientific attention; for instance, the fact that it holds a faint hope for lifeforms (past, present, or future) and requires the full NASA array of equipment including the manned Shuttle. The Saturnian moon already has been popularized by Carl Sagan in his PBS television series "Cosmos" with a visually striking simulated Saturn ring penetration and Titan landing, and Voyager I vastly increased our scientific knowledge of Titan during its encounter with the planet in November 1980.

(6) Precursor missions will provide enough knowledge of Titan and the Saturn environment to allow verification by Earth-based scientists of the atmospheric and surface models sent back by hypothesis-formation modules operating aboard the Titan spacecraft.

(7) A partial knowledge of the Titan environment permits equipment and experiment economies over later missions wherein many more contingencies and hypotheses must be anticipated.

A Titan Demonstration Mission in the year 2000 AD would benefit from two types of heritage (fig. 3.2). The first, knowledge heritage, allows the use of spacecraft com-

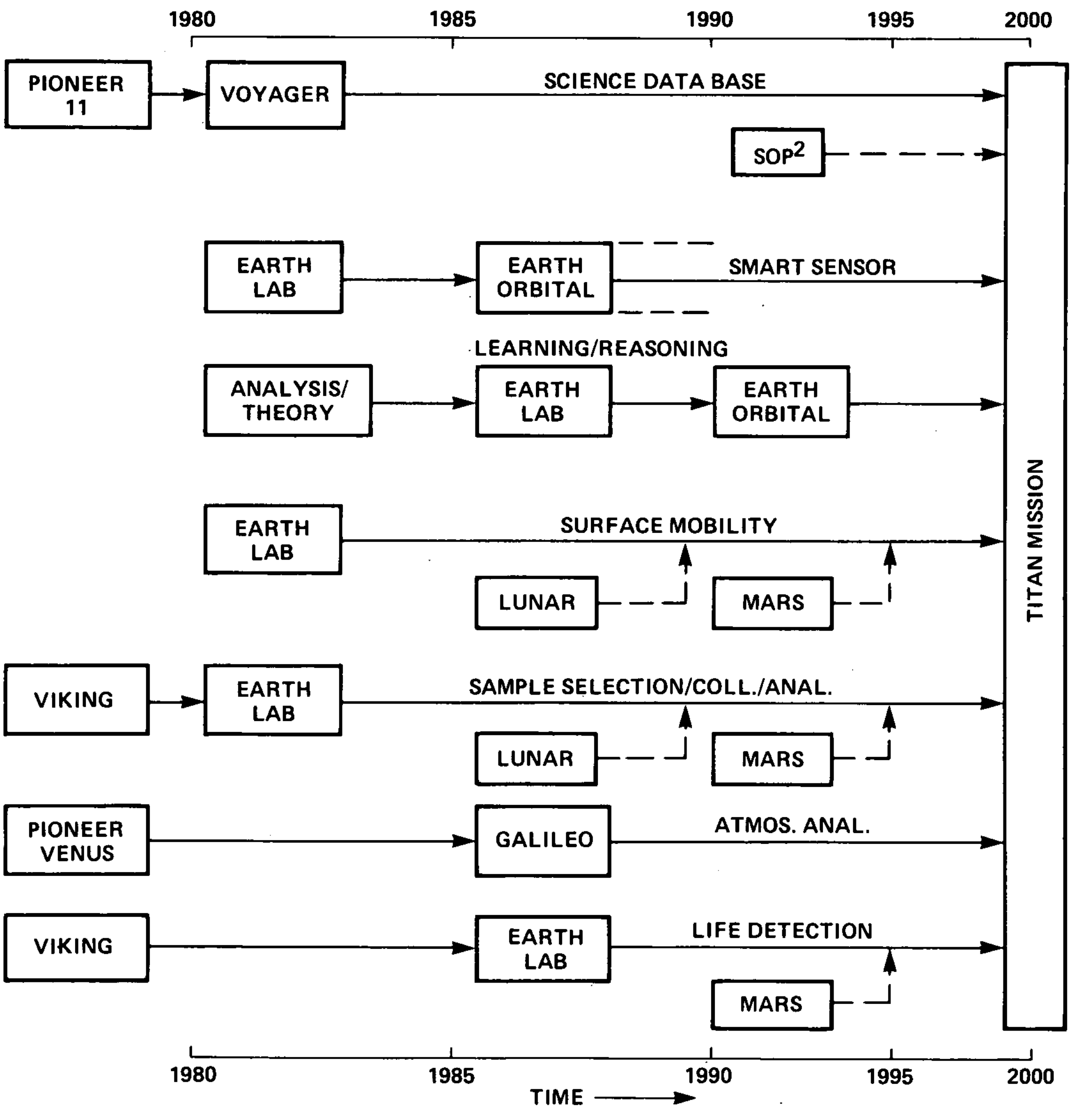


Figure 3.2.– Prior mission contributions to desired Titan mission capabilities.

ponents which need not be designed to cope with wholly unknown alien environments. The experience gained during the Pioneer 11 and Voyager encounters with Saturn and its moons has provided essential prior scientific and engineering data on Titan and its surroundings. The second, equipment heritage, permits investigative techniques developed for earlier missions to be adapted in modified form for the Demonstration. Many pre-Titan spacecraft operations address the same basic objectives in planetary exploration and provide a useful remote-sensing technology base to carry them out. For example, the Viking, Pioneer Venus, and Galileo missions furnish techniques for in-situ atmospheric analysis, and valuable experience with surface analyses searching for microbial life was gained during the Viking mission to Mars.

A number of planned or opportunity missions currently under consideration by NASA offer further possibilities for technology development in directions useful for the Titan Demonstration – such as the proposed lunar and Mars missions employing autonomous surface roving vehicles and advanced methods for sample selection, collection and analysis, and the VOIR (Venus Orbiting Imaging Radar) system for the development of a planetary radar mapping capability. Since the global characteristics of Titan are included within the scope of the Demonstration, opportunities for knowledge and for equipment heritage exist with respect to the proposed Saturn Orbiter Dual Probe (SOP^2) spacecraft.

In summary, the proposed Titan technology demonstration experiment and major space exploration mission utilizes the knowledge and experience gained in previous NASA operations. In turn, the Demonstation itself serves as the verifying mission for an autonomous space exploration capability which is the ultimate goal.

Figure 3.3 shows the relationships between the research areas of the four Study Teams and the Titan and interstellar

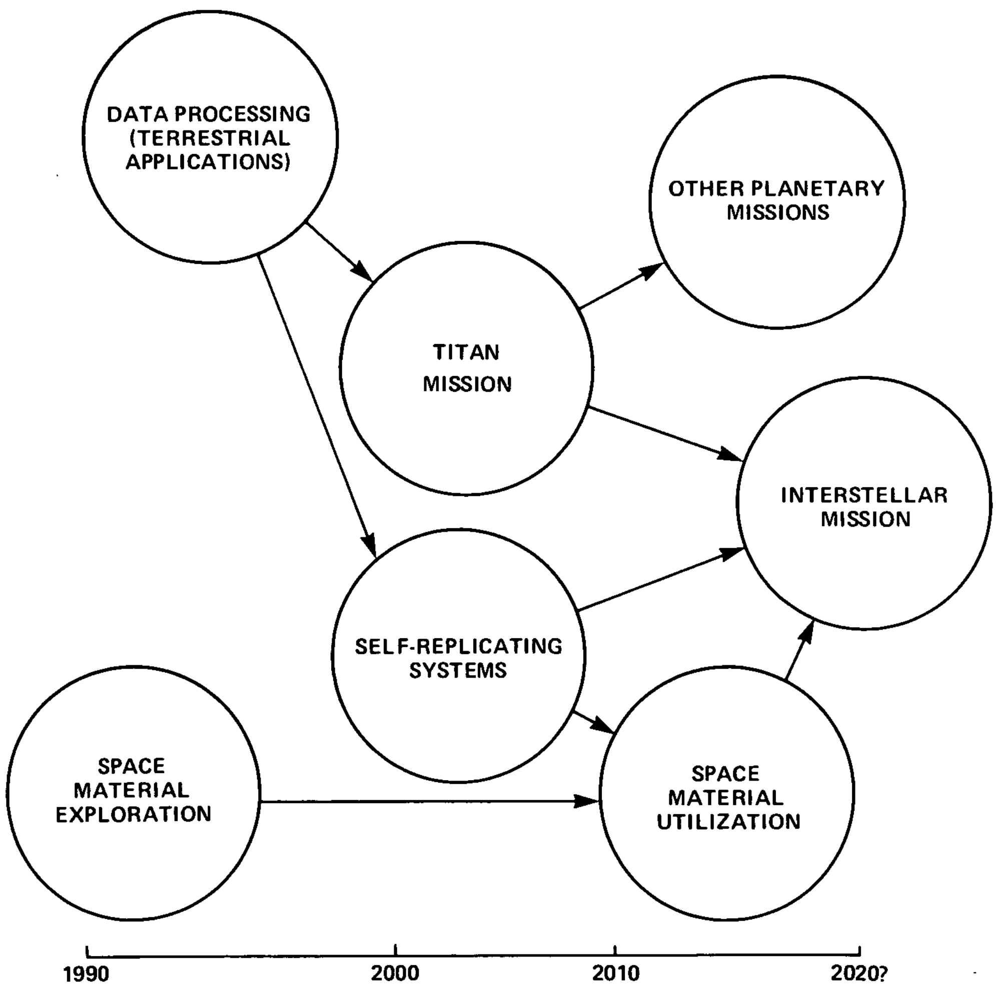


Figure 3.3.– Relationships between space exploration and other 1980 study areas.

mission concepts addressed in this chapter. Of particular interest is the question, "How soon after the Titan mission will extraterrestrial materials be utilized to facilitate interstellar exploration missions?" A Delphi poll was conducted using all Study participants (considered the best sample of experts immediately available to consider the question) and the results were: Median year 2028 AD, with the 14 estimates ranging from 1995 AD through 2100 AD.

3.2 Titan Demonstration Mission Definition

The Titan Demonstration Mission as envisaged by the Space Exploration Team encompasses a continuum of scientific investigative activities culminating in a fully autonomous extrasolar exploratory capability. The primary focus is on condensing into a single extended mission NASA's present sequential approach of reconnaissance, exploration and intensive study. In the past, interplanetary discovery has required Earth-launch of consecutive exploratory devices designed on the basis of data gathered by precursor craft. This approach assumes a broad range of sophisticated sensing equipment but little capability for onboard processing. Analysis of acquired data typically has been relegated to earthbound scientists who make judgments to determine the best next course of action, a procedure which incurs considerable time delays in return transmission of data as well as in ground-based control of distant spacecraft. An even more dramatic delay problem emerges with respect to the deployment of subsequent exploratory devices. In the case of Mars, for example, an initial reconnaissance vehicle (Mariner 4) was dispatched in 1964 but it was more than 10 years later (in 1975) before Viking 1 could be launched to attempt a Martian landing and a more intensive planetary investigation.

Mars, of course, is one of Earth's closest neighbors. Time delays in data transmission and control functions reach a maximum of 21 min in each direction, and travel time from Earth to Mars is approximately 1 year. In the outer Solar System the delay for one-way data transmission and control is measured in hours or days, while at interstellar distances, delay is measured in years with travel times of decades or more. As exploration goals are extended into the farthest reaches of space, development of nontraditional techniques and systems requiring a lesser dependency on Earth-based operations and possessing far greater autonomy become increasingly desirable and necessary. It is in this spirit that the Titan Demonstration Mission is proposed – anticipation of the potential for advanced machine intelligence eventually to permit fully autonomous exploration of the interstellar domain, a capability born of earlier demonstrations within the closer context of the Solar System.

In order to maintain linkages with current and future NASA activities (e.g., Voyager, Saturn Orbiter Dual Probe) and between short- and long-term objectives, the initial Titan demonstration relies upon extensions of current artificial intelligence (AI) techniques where these are appropriate. For example, by the year 2000 a considerable amount of information about Titan's characteristics, including a basic atmospheric model, already may have been compiled. Assuming research and development progresses in both interacting simulation models and rule-based automated decisionmaking, then extensions of current AI knowledge-based systems will have the potential to contribute to the automatic maintenance of mission integrity to insure the survival of mission functions and components.

To the extent that new developments in machine intelligence technology move in the appropriate directions, the Titan mission might include demonstrations of autonomous onboard processing of mechanically acquired data in at least one sample of scientific investigation. This results in great compression of return information because only the "important" or "interesting" hypotheses about the target planet are transmitted back to Earth. Such a function presupposes a machine capacity both for hypothesis formation and for learning, neither of which is inherent in state-of-the-art AI technology (see section 3.3). Significant new research in machine intelligence is a clear prerequisite to successful completion of the proposed Titan Demonstration Mission (see table 3.2).

TABLE 3.2.– TITAN EXPLORATION MISSION DRIVERS

Technology

- A coordinated surrogate scientific community on and around Titan
- Long system life – 10 years or more reliable/redundant propulsion/energy
- Distributed decision and expert systems
- Self-monitor and repair ability
- Semi-autonomous subsystems
  Probes, Landers, Rovers, Satellites
- Data storage and reduction; information communication to Earth
- Integrated multisensor capability

Intelligence

- Overcome the intelligence barrier. Current AI capabilities and research will not achieve autonomous MI needs for space exploration
- MI for space exploration must be able to learn from and adapt to environment. To be able to formulate and verify hypotheses is essential, but may not be sufficient.

Goal: Full autonomic exploration system with human intervention option.

While Titan is too distant to explore efficiently using traditional methods it is still near enough to monitor the performance of automated functions and to take intervening action should the need arise. As exploration distances extend farther out into the Solar System, such intervention becomes increasingly difficult so the demand for greater mission autonomy and higher-level machine intelligence rapidly intensifies. An outline of operational mission stages integral to the full range of exploratory activity, from the Titan demonstration to interstellar exploration, is presented below. Each phase underscores a variety of machine capabilities, some unique and some overlapping, required if full autonomy is to be achieved. These capabilities represent the primary technology drivers for machine intelligence in future space exploration.

3.2.1 Titan Mission Operational Stages

A fully automated mission to Titan (and beyond) requires a very advanced machine intelligence as well as a system which is highly adaptive in its interactions with its surroundings. This latter aspect is even more significant in extrasolar missions because a sufficient operational knowledge base might not be available prior to an encounter with new planetary environments. The explorer must generate and use its own information regarding initially unspecified terrain, and this knowledge must evolve through the updating of databases and by the continual construction and revision of models. Such a machine system should be capable of considerably higher-order intelligent activities than can be implemented with state-of-the-art techniques in artificial intelligence and robotics.

The short-term mission objective is to encompass the tripartite staging of NASA missions within a single, fully automatic system capable of performing scientific investigation and analysis, the immediate objective being a complete and methodical account of Titan. Later, and as a longer-term goal, given the successful achievement of the short-term objective, a similar exploration of the outermost planets and bodies of the Solar System could be conducted with improved equipment, building on the systems operations knowledge gained at Titan.

The proposed exploration system must be capable of the following basic functions:

(1) Select interesting problems and sites.

(2) Plan and sequence mission stages, including deployment strategies for landers and probes.

(3) Navigate in space and on the ground by planning trajectories and categorizing regions of traversibility.

(4) Autonomously maintain precision pointing, thermal control, and communications links.

(5) Budget the energy requirements of onboard instrumentation.

(6) Diagnose malfunctions, correct detected faults, and service and maintain all systems.

(7) Determine data-taking tasks, set priorities, and sequence and coordinate sensor tasks.

(8) Control sensor deployment at all times.

(9) Handle and analyze all physical samples.

(10) Selectively organize and reduce data, correlate results from different sensors, and extract useful information.

(11) Generate and test scientific and operational hypotheses.

(12) Use, and possibly generate, criteria for discarding or adopting hypotheses with confidence.

One way to formalize the precise characteristics of a proposed mission is in terms of a series of prerequisite steps or stages which, in aggregate, capture the nature of the mission as a whole. The operational mission stages selected for the Titan demonstration analysis are: configuration, launch, interplanetary flight, search, encounter, orbit, site selection, descent, surface, and build. Each is discussed briefly below.

Configuration. This initial phase addresses considerations of size, weight, instrument specifications and other launch vehicle parameters, and usually depends on the equipment and tasks required for a specific mission. Questions concerning the precise nature of the investigation and experimentation traditionally are taken up at this point.

For deep-space exploration, spacecraft configurations must be general and flexible enough to handle a wide range of environments. Hardware and software impervious to extreme pressure, temperature, and chemical conditions and with long lifespans are required. Also, a diverse assortment of onboard sensors with broad capabilities is necessary to produce basic information via complementary and selective sensing to be used in scientific investigation and planning.

Launch. The focus of this stage depends to some extent on the perceived configuration of the mission vehicle. Issues related to propulsion and energy needs and appropriate launch sites (e.g., Low Earth Orbit vs vicinity of extraterrestrial resources utilized for the mission) are decided. The launch phase is conducted largely by Earth-based humans, but could benefit from machine intelligence capabilities (e.g., CAD/CAM/CAT) for testing, checkout, flight preparation, and launch support.

Interplanetary flight. Prior to Viking and Voyager, unmanned flyby and orbiter spacecraft were totally dependent upon Earth-based remote observation and direct human intervention to accomplish accurate navigation, stationkeeping, and rendezvous and docking maneuvers (Schappell, 1979). This underscores the control and communication time delay problem that limits efficient investigation of distant bodies such as Titan and even more dramatically constrains exploration of the interstellar realm. Some ground-based support for the initial Titan

Demonstration Mission may be appropriate in computing navigational corrections, but subsequent deep-space exploration requires a fully autonomous navigation system. Such systems also improve cost-effectiveness by reducing the amount of ground support necessary to accomplish the missions. Potential savings in equipment complexity, operational costs, and processing time will motivate the development of autonomous systems for near-Earth and deep-space vehicles.

Consider, for instance, the Viking mission, one of the most complex interplanetary operations attempted to date. The Mars landers were remotely operated robot laboratories equipped with comparatively highly automated instrumentation. Many spacecraft functions could be performed adaptively, accommodating to changing necessities during the mission. Even so, the operational system required major navigational changes to be specified 16 days before the indicated flight action. Several hundred people on Earth were involved in science data analysis, mission planning, spacecraft monitoring, data archiving, data distribution, command-sequence generation, and system simulation. An infusion of advanced machine intelligence could significantly reduce this major mission cost.

In addition to navigation, the spacecraft also must maintain attitude and configuration control, thermal control, and communications links. These functions involve the use of feedback loops and built-in test routines. One way to visualize a greatly improved system is to conceptualize a machine intelligence capable of sequentially modifying its activity as a result of experience in the environment, with an additional capability of internalizing or "learning" the relationship between environmental states and corrections to guide future modifications and coordinate them with anticipated states. Such goal-directed intelligent functioning is not possible with state-of-the-art AI technology. However, it is conceivable that a machine system could be provided with a capacity to represent its present state, some goal-state of equilibrium or stability and a means of noting and measuring any discrepancy between the two, and, finally, effectors or actuators for modifying the present state in accordance with the programmed goals.

Search. During the Search phase the system performs preliminary analyses while approaching the target body. The information acquired is integral in making decisions about subsequent activities as well as the point at which to begin preliminary analysis. The spacecraft must be able to employ appropriate sensing equipment to collect raw data and to modify sensor utilization as a result of feedback information. Inherent in this formulation is the capacity of the system to perform some analysis using the raw data it has collected and to make decisions about mission sequencing based on analysis results.

Complementary and concurrent sensing tasks are scheduled according to the time required for their completion, the point at which their output becomes important to ongoing model construction, and the relative importance of the results. Another significant factor is spacecraft-instrumentation power scheduling, assuming that the supply of energy is insufficient to allow all subsystems to operate simultaneously. Scientific tasks must be scheduled to take into account possible mission-control functions that might override them. Collection tasks producing data having multiple uses or particular utility in mission integrity operations (self-maintenance, survival, and optimization) have high priority. All operations are to be accomplished without benefit of direct human intervention.

For the initial Titan mission, one might attempt to automate all search functions by means of an onboard expert system that utilizes known information about the conditions on Titan and that is capable of examining and choosing from among preselected resident hypotheses (leading finally to some judgment as to what action to take based on probability calculations). However, such a system could be highly fallible because information gaps and inaccuracies in its available range of hypotheses might lead to serious misjudgments. In the case of the long-term objective – interstellar navigation – the consequences of an incomplete knowledge base are even more dramatic. The team concludes that expert systems of the current AI variety cannot satisfactorily perform the Search task.

One possible solution, and a potentially valuable technology driver, is an advanced type of expert system able to update and modify its own knowledge base as a result of experience – that is, as a result of the analytical actions which it performs on its own environment. On Earth the advent of such an advanced system would eliminate time-consuming and costly human analysis and reprogramming typical of state-of-the-art expert systems (which would be particularly inefficient in space applications where huge time delays often must be accommodated). Self-modification of advanced expert systems also prepares the exploration system to make autonomous decisions and corrections regarding its relationship with the environment.

An additional essential task *en route* to an unknown planetary system around another star is the determination of gross parameters such as sizes, masses, densities, orbital periods, rotational periods, axial tilts, and solar distances for each member planet and moon. A fully autonomous spacecraft would utilize these characteristics, determined by early data collection, in making onboard selections of appropriate bodies to explore.

Given the existence of specific atmospheric conditions determined by long-range remote sensing, logical hypotheses may be generated to predict the surface conditions of the chosen celestial body in terms of the possibility of life and the compatibility of the planet with spacecraft hardware and engineering. Decisions must then be made on the basis of preliminary analyses whether to proceed and establish orbit around the planet for further exploration, or

to choose another target. An intriguing alternative would be a system capable of redesigning or adapting its equipment to accommodate the relevant alien environmental conditions.

Encounter. The processing of image data is probably one of the most computationally demanding tasks performed during planetary exploration missions. In the Encounter phase, when the spacecraft controller must make a quick go/no-go decision on the question of orbital insertion, the data processing challenge includes speed as well as volume. The problems of distance and communications delays, coupled with the necessity of making rapid local decisions, virtually demand that image analysis during Encounter be accomplished by fully autonomous onboard processing systems.

One possibility is an imaging system capable of describing a planetary body much as an astronaut would. For example: "The surface is bluish with some brownish areas near the equator. There appear to be thin wispy clouds covering a 100 X 200 km area centered about 75° N and 30° W." The observation of "bluish" and "brownish" indicates the processor's ability to match raw data inputs to color concepts understood by humans. The identification of "wispy clouds" suggests the capability of matching data in a sequential region of the image to the known concept of "wispy." The ability to match regions, spectral data, and other features in an image to stored concepts in memory requires a reasonably high level of machine intelligence.

Another part of the description of the image observed by spacecraft sensors locates the "wispy" area at a given latitude and longitude. To do this, the processor must be able to establish the geometrical shape of the body encountered and to apply a coordinate system to it. Once this coordinate system is computed it forms the cartographic grid to which all surface features are mapped. While this is a well-understood mathematical procedure, the "number crunching" load is significant and must be executed very rapidly during the Encounter phase.

Orbit. When preliminary analysis suggests a reasonably benign environment warranting further investigation, orbit is established to conduct a more detailed study. The establishment and maintenance of orbital position, like most of the functions already mentioned, should be a fully autonomous process with characteristics similar to the autonomous interplanetary flight navigation system. Onboard automated decisionmakers determine an optimal orbit using information gathered during preliminary analyses, and orbital insertion is achieved.

Multisensor analysis is implemented concurrently with the establishment of orbital position, permitting a more comprehensive investigation of planetary characteristics than during Encounter. During Orbit phase a variety of sensors and sophisticated image processing techniques are employed to examine atmospheric and surface conditions. Analyses should be conducted both in the context of (1) pragmatic decisionmaking, including assessments of atmospheric pressure, density, and identifications of surface conditions to be utilized in judging which equipment to deploy, and of (2) scientific investigation, such as information acquisition for hypothesis generation.

For the Titan mission an advanced expert system may be used to form judgments about appropriate exploratory equipment for specific environmental conditions. For instance, when deploying probes or landers smart sensors might first assimilate data regarding atmospheric density and pressure. The advanced expert system could then make probability judgments as to how fast probes should fall and how much retrorocket energy is required for landing. Additional assessments could be made of surface conditions, such as whether the surface is composed of a solid, liquid, or gaseous base. This information supports subsequent decisions about necessary configurational requirements of landing craft (e.g., should it be a wheeled, walking, hovering, or floating vehicle?). The above machine intelligence applications could probably be developed on a relatively short-term basis, utilizing minimal extensions of state-of-the-art AI techniques.

In the deployment of such exploratory mechanisms as atmospheric and surface probes, balloons, and landers, intelligent coordination of autonomous orbit maintenance and control is crucial. Since deployment of onboard equipment alters the total mass and mass distribution of the orbiter, some simultaneous revision of the altitude control function, ideally based on "anticipatory information," is required. That is, the spacecraft must anticipate changes in its state prior to component deployment and be prepared to adapt to concomitant variations in its physical state (a specific example of the type of feedback system required to maintain mission integrity).

A much more serious problem for development in the area of machine intelligence is the scientific analysis of data and the autonomous formulation of hypotheses and theories. Current expert systems technology cannot generate and test unique hypotheses that have not been preprogrammed by a human operator. This limitation restricts an exploratory device based on state-of-the-art AI to data analysis, categorization, and classification in terms of existing structures of thought or taxonomies of knowledge. However, in alien environments, particularly those accessible in an interstellar mission, pre-formed scientific notions may not reasonably be applicable; on the contrary, they may serve only to distort higher-order understanding of incoming data. Thus, a major technology driver is the development of an advanced machine intelligence system capable of reorganizing rejected hypotheses, integrating that information with data acquired through sensory apparatus, generating new hypotheses which coordinate all

existing information, and, finally, testing these hypotheses in some systematic fashion. (See appendix 3B for a hypothetical illustration of this point.)

Lander site selection. During this phase some form of mobile surface device compatible with local environmental conditions is deployed according to planetary orbiter directives. This device performs in situ surface and geologic data acquisition, imaging, and representative physical sample collections. Its deployment requires the selection of appropriate landing sites, a major task for the autonomous exploration system controller.

Processed image data of planetary surface conditions permits a mapping of topographic surface characteristics with respect to terrain configuration – a cataloguing of mountains, craters, canyons, seas, rivers, and other features to be correlated with maps of temperature, moisture, cloud cover, and related observables. These maps become the basis for a determination of optimal landing locations. Site selection analysis also must include some judgments regarding areas of greatest "interest" for investigation, necessitating some means of detecting regions of the environment which are anomalous with respect to expectations based on prior preliminary analyses of the locale. Criteria for site selection, as for example geological significance or the possibility of lifeforms, are stored in memory. Imagery to be compared to this set of criteria could be obtained from a world model (see chapter 2) developed during the orbital phase, an application ripe for machine intelligence technology development.

Hazard avoidance at the landing site and terrain traversibility for mobile landers are additional considerations in the site-selection process. Some mechanism for self-preservation should be included so that an assessment of potential landing sites is made according to whether they pose a danger or are benign. Only then can adaptive action patterns be undertaken with some reasonable expectation of success.

Descent to surface. The descent to surface should be fully automated even in relatively near-future explorations of the Solar System. Autonomous feature-guided landing poses a unique challenge to image-processing technology. For instance, during a parachute descent the target landing site must be located and tracked by an image processor. As the assigned target is tracked, the lander parachute must be manipulated to steer toward the target much like a sports parachute. While the tracking task is not conceptually difficult, the processing speeds required do not exist in present-day computer hardware. As the surface draws closer, the potential landing site must be reexamined for obstacles hazardous to the craft. This presupposes some stored knowledge of precisely what could pose a hazard, as well as the ability to act upon that information. In the Descent phase, machine intelligence integral to the surface exploration system will require high-accuracy processing and ultra-high speed hardware.

On the surface. Once surface contact is achieved the most interesting and probably the most difficult image processing begins. Self-inspection for damage comes first, followed by verification of the lander's position. This may involve comparing the surrounding scene with possible projected scenes assembled from the world model, or the analysis could be based on tracking by the main orbiting spacecraft. Next is the planning, scheduling, and commencement of experiments. All conflicts must be comprehended and resolved. If one experiment calls for rock density measurements and no rocks are within reach of the lander's end-effectors, a decision must be made to schedule another experiment or to move the lander. Such operational decisions require intelligent scene analysis and concept/theory matching.

If preliminary analyses suggest that further investigation is warranted and safe, the lander system for image processing of the surrounding area is deployed. This accompanies the collection of local temperatures, pressures, and general ambient conditions data, as well as sample collection and analysis. To provide these functions the lander (an intelligent robotic device) is equipped with a wide variety of sensor and end-effector apparatus. Vision is especially important for obstacle avoidance and mobility. Stereo vision may prove an invaluable aid in successfully traversing three-dimensional spaces, and also an important safety feature for avoiding depth hazards.

Mobile lander data collection responsibilities require several specific machine intelligence capabilities including (1) pattern recognition to correlate visual images and to detect similarities and differences among data alternatives and (2) decisionmaking to determine whether a particular datum is worth collecting. While it is conceivable that minimal extensions of state-of-the-art expert systems might prove adequate to address the problem of datum "worth," still there remains a sizable gap between current capabilities in computer perception (pattern recognition) and capabilities needed for tasks integral to the proposed mission – another crucial technology driver.

While some of the Titan mission performance demands on robot manipulators are not as critical as on industrial assembly lines, still there are definite constraints. Spacecraft effectors must operate in completely unstructured environments unlike state-of-the-art factory robots which move only in small, comparatively well-defined work areas. Precision requirements are fairly modest for explorer manipulators when they are handling physical samples, but placement accuracy must be considerably improved whenever the system is responsible for joining closely fabricated pieces during instrument repair, component reconfiguration or construction. Manipulator supervision is supported primarily by visual sensing, though a wide variety of other sensor inputs may supplement optical techniques.

A potentially difficult image processing task is the coordination of manipulator movements with those of the

target object, better known as "hand-eye coordination." Image processing may be used accurately to find the position (in three-dimensional space) of the object to be manipulated as well as the grasping surfaces of the manipulator itself. Locating these surfaces might involve matching the received images with memorized models (or concepts) of the object and the end-effector, a tremendous challenge to present-day machine intelligence technology. An alternative method requires using pressure- and force-feedback, as well as proprioceptive information (sensory input designating body or effector orientation) to reduce image processing requirements.

Movement of the lander demands that a safe, obstacle-free path be found across the landscape. This may entail generating a contour map of the surface surrounding the lander (perhaps using high-resolution satellite/orbiter data) and derivation of a clear path from this map. State-of-the-art laser scanning techniques already have proven adequate to handle the task of topographic analysis for purposes of local wild-terrain locomotion. Hazards hidden from view along the intended itinerary must be identified *en route,* and the path ahead continually re-scanned and updated as in the case of a human walking through a rocky area.

An alternative (and more difficult) approach places greater reliance on autonomous lander processing systems. A planet model provides an apparently traversible path from the landing site to another location observable from the landing site (based on low-resolution data). This "fuzzy" trail is given to the lander controller which then must negotiate its own path from the first position to the second, must identify and work its way around such obstacles as gulleys, creeks, or rubble invisible in the low-resolution model. In addition, during each traverse the lander analyzes the surrounding scenery and searches for significant or unusual objects while also keeping track of its location. Thus, a great deal of image processing and map updating must be done that requires formidable onboard computing power, as well as advanced machine intelligence techniques.

Build. The Build phase actually lies in the domains of space manufacturing (chapter 4) and machine replication (chapter 5), but nevertheless, is worth mentioning here as an important prerequisite for extending the proposed mission to intensive Solar System and interstellar exploration. At some (yet undefined) point it becomes necessary to provide machines with mining, materials processing, construction, repair, and perhaps, even replicative capabilities in order to escape the enormous cost of building and launching burgeoning masses of exploration equipment from Earth (Freitas, 1980b). With respect to the Titan Demonstration Mission, a first step toward the ultimate goal of machine self-sufficiency would be an onboard provision for machine hardware components with the ability to make adaptive modifications to the system as a result of preliminary analyses of probe and landing craft needs.

3.2.2 Scientific Investigation: Remote Sensing and Automated Modeling

The concept of space exploration presented above suggests the potential capability of an interstellar spacecraft to develop complete detailed models of planets and moons in other solar systems and to return these to Earth as major scientific discoveries about the Galaxy. These models would include information about the planets' atmospheres, surfaces, subsurfaces, electromagnetic and gravitational fields, and any evidence of lifeforms.

Having first characterized the operational mission stages and identified the important machine intelligence requirements of each, the Space Exploration Team chose to consider at greater length one aspect of the Titan Demonstration system capacity to conduct useful scientific investigations: automated modeling of an unknown celestial body. This particular aspect of the scientific investigation capability was selected because it involves the full range of high-level machine intelligence required for autonomous space exploration, while simultaneously relating to the orbit-based world model deployment scheme contemplated by the Terrestrial Applications Team (see chapter 2).

In terms of the preceding discussion of the operational phases of space exploration missions, the task of creating such models is the first and foremost task of the Orbit stage. Detailed remote sensing is undertaken in the mission orbital phase to complete atmospheric modeling and to map various physical parameters of the surface. Perhaps as much as 90% of the total information gathered in the exploration of an unknown body can be collected by the orbiter.

A complete world model describes atmospheric and surface physical features and characterizes the processes which govern the dynamic states of the planet and its atmosphere. The job of constructing a world model may be broken down into two separate categories: building an atmospheric model and examining processes in the surface environment, described below. Since a great deal of work is under way at NASA and at various universities in the analysis of Landsat and weather satellite information, it can be anticipated that much of the groundwork in the techniques for assembling a planetary model will have been laid long before deployment of the Titan mission. Not only is the development of a terrestrial world model an essential precursor research program in pursuit of interstellar mission technical requirements, but it also provides valuable Earth resource information in the more immediate future. Creating and automatically modifying world models based on inputs from a variety of sensors is a machine intelligence technology in which research should be encouraged.

Atmospheric modeling. An accurate atmospheric model is essential to successful landing, scientific analysis, and the prediction of the possibility of indigenous life. The construction of an atmospheric model for Earth (including composition, structure and dynamics) has taken many years, an iterative process dictated by evolving technology plus the developing knowledge and expertise of investigators in a young field. To a large extent this emerging methodology has been driven by the measurability of accessible variables, which may or may not be optimal from a systems theoretical point of view. But given higher technology, observational freedom from Earth's atmosphere, and fresh unknown territory to explore, many more options become available with respect to what should be measured and in what order to define an atmosphere most efficiently and unambiguously. The process has not yet been adequately systematized to permit clear-cut rational choices.

Atmospheric modeling should begin early in the approach to an unknown planet since many mode-of-exploration decisions require information on the nature of the atmosphere. During the course of the mission the atmospheric model accumulates greater detail with continuous updating as higher sensor resolution is achieved and probes are deployed for direct measurements. The investigation of an atmosphere differs from studies of surface characteristics in that it involves the complex integration of many interrelated subhypotheses and measurements of numerous allied parameters. Studies of the surface are more a problem of deriving hypotheses from completed maps representing different measurements and then overlaying these maps as a final step.

Specific initial tasks related to atmospheric modeling include:

- Determination of the region of the spectrum in which most of the electromagnetic radiation is emitted.
- Determination of the sources of opacity for selection of optimum communications link frequency (for landers and probes) and for choosing wavelengths in which to perform infrared and millimeter radiometry.
- Search for unbroadened spectral lines above the atmosphere to provide information on the overall composition of the air.
- Observe where spectral lines interfere with blackbody temperature measurements and determine the wavelength(s) at which the atmosphere may be fully penetrated and planetary surface temperatures accurately recorded.
- Perform preliminary temperature and pressure measurements, to be updated once a comprehensive atmospheric model has been constructed.
- Begin atmospheric modeling with remote sensing at millimeter and infrared wavelengths.

Surface modeling. The best method for planetary surface structure hypothesis formation requires scanning of the body with sequentially increasing resolution in at least four distinct steps. The first step obtains global average values for temperature, surface structure, composition, etc., and establishes norms for keying future observations at higher resolution. Gross features such as lunar maria and highlands or the martian polar caps would appear in this type of survey.

The second observational phase exposes finer detail, identifying regions on the scale of the Tharsis Plain of Mars or the Caloris basin of Mercury. As the explorer approaches Titan, higher-resolution observations of the surface become possible and morphological changes can be observed in each succeeding frame. Recognition of features such as craters, mountains, rivers, and canyons may be accomplished by an advanced expert system which includes models of surface processes in its knowledge base, although present-day pattern recognition and vision systems will require significant refinement before this capability can be realized.

The third step is the recognition of sites with high potential for usefulness in the construction of world models. Such sites mainly include unusual features that are interesting because of their anomalous nature. Identification requires a stored concept of "usual," as for instance: "There is usually a sharp boundary between continents and oceans" and "Craters viewed from directly above usually are circular." An original supply of these simple concepts are programmed into the system by humans before the mission begins; however, additional and revised definitions of normality must be developed and refined as the mission study of a particular planetary body progresses, with self-developed concepts of "usualness" updated by the system as various stages and modes of multisensor investigation are completed. The recognition of that which is "unusual" is discussed at greater length below.

The fourth and final step includes detailed surveys at maximum resolution of selected sites and additional imaging of various undistinguished sites spaced along a grid to pick up interesting features missed by other searches at lower resolution.

Automated selection of interesting sites. It is desirable to minimize raw data storage in order to maximize the efficiency of onboard concurrent mission tasks and analyses. Some method must be found to deal with the information overload which might result from exhaustive exploratory surveys, particularly high-resolution topographic mapping.

Data preprocessing and compression are needed not only because of memory limitations but also to help reduce the complexity of information to be assimilated into world models. Without some way of narrowing the field of interest or of identifying "highlights," the task of converting multiple correlations of many detailed data sets into complete models is cumbersome and impractical. Simplification

also is needed to perform initial but fairly exhaustive searches for sites warranting further investigation (e.g., potentially interesting, safe for landing, etc.). This activity requires a high-level machine intelligence system able to make good choices of which high-resolution data to save and which to throw away.

One possible approach to the selection problem is for the first mapping system to earmark data anomalies for surveillance at higher-than-normal resolution in subsequent surveys. Anomalies are sought by making detailed comparisons of successive maps of the same region. Alternatively, scan data can be searched for locales in which the measured parameter deviates significantly from a predetermined norm. In practice, single-pixel measurements might be saved if values exceed specified thresholds. Data also may be saved in map regions where measured parameter gradients are as steep or steeper than the slope defined as a "significant" edge for that type of measurement. Effectiveness using norm comparisons depends upon appropriate thresholding, whereas the detection of anomalies by successive map matching may be a less subjective approach. (Measured quantities in either case might include rock types, textures, slopes, temperatures, gas concentrations, symmetries, and colors.)

Abridged maps of anomalies detected using the initial survey maps are then compared in a search for correlations that might identify interesting sites. Both the most common and the most unusual sites have high priority for further examination. Areas of interest are identified and ranked for intensive investigation according to the total number of different types of "edges" they contain. The degree of uniqueness of any given site is a criterion for prioritizing follow-up studies. The occurrence of more than one site exhibiting the same edges or of locations with similar correlations indicates a need for additional study. The distribution of correlated sites might suggest some common factor among them; for instance, latitude or regular temporal variations.

The above method of data analysis is one way of focusing on a few features or locations most useful in constructing world models. These high-value sites are identified by their discontinuous character as compared to their surroundings. The method should correctly report features that would be included in an eyewitness description of the celestial body given by a human observer. For example, the crisscrossing lines of Europa might be singled out, as well as the canyons and streambeds of Mars.

Such striking features are necessary but hardly sufficient to specify an entire planet. A complete view also must include: (1) Large-scale structure illuminated by lower-resolution mapping (e.g., the overall smooth surface of Europa), (2) the construction of models inferred from surface mapping data (e.g., that the cracks and smooth surface of Europa indicate a young, active crust), and (3) the incorporation of atmospheric modeling. A machine intelligence system which can quickly single out and characterize important features of celestial bodies is required. This system should first be tested with known bodies to verify its ability to rediscover what humans would consider to be significant.

Two approaches for model formation have already been presented. The first applies to atmospheric determination – specifically, a process of multiple iteration and revision starting with a distant view of the unknown world. The second applies to surface environment studies after the explorer system has entered orbit. Both methods correctly recognize that models are not effectively constructed by trying to answer all questions for every pixel. Hypothesis formation is followed by a process of testing and checks in each approach.

3.2.3 Titan Mission Components Concepts

The comprehensive exploration and intensive study of Titan will require an appropriate system of spacecraft components. In this section, preliminary technical specifications are provided for each candidate spacecraft function involved in the Titan Demonstration Mission. Of course, final system configuration is dependent upon progress in machine intelligence techniques and on advances in hardware technology that may occur. The technical level of the following specifications is compatible (though not presented in the same format) with the NASA Space Systems Technology Model. In each case, criteria of maximum cost-effectiveness and minimum equipment proliferation are applied.

Rather than discuss every detailed hardware requirement, the Space Exploration Team elected to focus primarily on aspects of the proposed mission which demand significant advances in current technology. Consequently, and also because the design is largely conceptual, the following quantitative and qualitative information would be ranked "level 3" (relatively low confidence) in the Space Systems Technology Model.

The general features of the Titan Demonstration Mission are given in table 3.3. Table 3.4 lists the candidate spacecraft system elements, including the typical number of each type, their operational locations once deployed, and

TABLE 3.3.– GENERAL FEATURES OF THE TITAN DEMONSTRATION MISSION

Status: Opportunity Mission (not in current NASA plans)
Lifetime: 10 years; includes 5 years at Titan
Launch/transfer vehicle: Shuttle/400 kW Nuclear Electric Propulsion (NEP)
Operational location: Titan, Saturn's largest satellite
Total mass: 13,000-17,000 kg
Total power: About 400 kW

mass and power requirements. The major mission accomplishments expected of each system component are shown in table 3.5.

surface mobility, and physical sample selection, collection, and analysis. Other candidate system elements have more specialized functions, the management of which can be assumed, at least in part, by advanced sensors and machine intelligence aboard the orbiter or landing craft.

TABLE 3.4.– CANDIDATE SPACECRAFT FOR THE TITAN DEMONSTRATION MISSION

Spacecraft type	Typical number	Operational location	Mass, kg	Power, kW
Nuclear electric propulsion	1	Earth to Titan orbit	10,000[a]	400
Main orbiting spacecraft	1	Circular polar Titan orbit at 600 km altitude	1,200	---[b]
Lander/Rover	2	Surface	1,800	1
Subsatellites	~3	One at a Lagrange point; others on 100 km tethers from NEP	300	0.3
Atmospheric probe	~6	Through Titan atmosphere to surface	200	0.1
Powered air vehicle	1	Atmosphere	1,000	10
Emplaced science	~6	Surface	50	0.1

[a]Does not include propellant.
[b]Uses NEP power.

The minimum duration of Titan operations is 1 year. While this would be barely sufficient to complete a nominal mission, it is a short time in comparison to seasonal changes in the Saturn system. (Saturn's solar orbital period is 29 years.) The most significant seasonal effects may be expected within about 5 years of the solar equinox of Saturn and Titan – which occurs in 1980, 1995, and 2010 AD. Hence, the preferred arrival dates are 2005 or 2010 AD, with a nominal mission duration of 5 years. Adding 5 more years for interplanetary flight, the preferred Earth-launch dates are 2000 or 2005 AD.

The success of the Titan Demonstration Mission depends on two essential elements – (1) the main orbiting spacecraft and (2) the lander/rover – and on the machine intelligence which they possess. High-level AI capabilities are needed by the main orbiter to coordinate other system components and to conduct an ambitious program of scientific investigation, and are required by the lander/rover to complete its tasks including safe and accurate landing,

Nuclear electric propulsion. The early phases of the mission, beginning with launch from Earth and continuing through Saturn arrival, require a high-performance propulsion system which can deliver the payload within a reasonable flight time (4 to 6 years). Low-thrust Nuclear Electric Propulsion (NEP) is the preferred technology for this purpose. The entire NEP system can be delivered to LEO, then be used for spiral escape from Earth, Earth-to-Saturn transfer, for Titan-rendezvous from a circular orbit around Saturn, and finally for spiral capture into Titan orbit and all subsequent spacecraft orbital adjustments. The main orbiter spacecraft and the NEP system share responsibilities for navigation, guidance, control and sequencing, system monitoring, and communications with Earth.

NEP technology has been studied for a long time but has no current planned application beyond possible cargo transport operations from LEO to Geosynchronous Earth Orbit (GEO). Planetary missions such as the proposed Titan

TABLE 3.5.– TITAN MISSION SPACECRAFT ACCOMPLISHMENTS

Spacecraft type	Possible accomplishments
Nuclear electric propulsion	Spiral escape from low Earth orbit; interplanetary transfer to Saturn; rendezvous with Titan; and spiral capture into 600 km circular polar orbit.
Main orbiting spacecraft	Automated mission operations during interplanetary and Titan phases: this includes interfacing with one supporting other spacecraft before deployments; deploying other spacecraft; communicating with other spacecraft and with Earth; studying Titan's atmosphere and surface using remote sensing techniques at both global characterization and intensive study levels; and selecting landing sites.
Lander/Rover	Lands at preselected site, avoids hazards; intensive study of Titan's surface; selects, collects and analyzes samples for composition, life, etc., explores several geologic regions.
Subsatellite	Lagrange point satellite monitors environment near Titan and is continuous communications relay; tethered satellite measure magnetosphere and upper atmosphere properties.
Atmospheric probe	Determines surface engineering properties and atmospheric structure at several locations/times.
Powered air vehicle	Intensive study of Titan's atmosphere; aerial surveys of surface; transport of surface samples or surface systems.
Emplaced science package	Deployed by long-range rover to form meteorological and seismological network. (Alternatives are penetrators or extended lifetime probes.)

Demonstration represent significant new possible applications. However, a NEP development program must be initiated in the 1980s to be operational in time for a Titan mission around the turn of the century.

The only major alternative propulsion technology is a chemical system using cryogenic liquids (the so-called Orbit Transfer Vehicle or OTV) for Earth escape, followed by gravity assists from Jupiter (in 1998) or from Earth and Venus, followed by aerocapture at Titan in the 2005-2010 time frame.

Main orbiting spacecraft. The principal vehicle for exploration in near-Titan space is an orbiter craft which remains with the NEP system. During the spiral capture process, the spatial structure of fields and particles around Titan can be measured. Following capture, the main spacecraft is parked in a circular polar orbit roughly 600 km above the surface of the body. Such an orbit has relatively little atmospheric drag and is highly desirable for close measurement and deployment of subsidiary system components into the atmosphere and to the surface of Titan.

During operations in near-Titan space, the main spacecraft must support a set of sophisticated remote-sensing instruments needed for global characterization and intensive study. In addition, it must continue to provide essential functions initiated during the interplanetary phases and support for deployed subcraft including navigation and communications with Earth. The estimated data collection volume is estimated at 10^{10}-10^{11} b/day, significantly greater than the 10^9 b/day characteristic of previous planetary missions. Most of this is accumulated from instruments aboard the main orbiter, with perhaps 10% supplied by subsatellites and surface vehicles. Assuming that all raw data are returned to Earth, the required downlink communications capability is 10^5-10^6 b/sec or 3-30 times the Voyager mission capacity from Saturn. However, significant amounts of data compression using advanced machine intelligence techniques should greatly reduce the transmission burden on the terrestrial downlink and also between elements of the Titan Mission.

The technologies developed in present and future planetary missions (especially Galileo, VOIR, and Earth-orbital) are generally applicable to this spacecraft. For instance, while in Titan orbit, the main orbiter is nadir-pointing much like VOIR and many Earth-sensing satellites. Major advancements are expected in the areas of machine intelligence and smart sensors, which suggests an increased capacity for data handling and communications as compared to previous planetary missions by the time of the Titan Demonstration.

Lander/rover. A lander/rover is needed to perform detailed surface and atmospheric measurements as well as the intensive level of study. Deployment of this spacecraft system is deferred until Titan's ground terrain has been fully mapped and an appropriate target site selected.

Atmospheric data are taken during the lander descent phase, and this continues as long as the vehicle remains operational on the planetary surface. Small rocket thrusters are used to guide the craft to a safe place free of large boulders, deep crevasses, or steep slopes. After a soft landing, the surroundings are characterized in preparation for site selection for sample collection. Physical samples are then acquired using extensible manipulators (scoops, drills, slings, etc.) and are immediately analyzed to determine chemical composition, layering effects, evidence for indigenous lifeforms, etc.

After this has been accomplished, the lander requires samples taken from a wider area to complete its preliminary investigations. The general solution to this problem is the rover, a vehicle deployed by the lander and used to explore the local neighborhood and to bring back samples. The simplest rover design might operate no more than 100 m from the lander and would remain almost totally dependent upon it. Such a machine is useful for collecting samples more free of contamination and more representative of the surface than those taken nearer the landing site. However, the Space Exploration Team prefers a more ambitious design, an autonomous rover able to operate up to 10 km from the lander. This larger-area capability permits the lander/rover system to return data which better contributes to an overall understanding of the geological structures of complex sites. Such advanced rovers already have been considered for lunar and martian applications.

It is also necessary to provide the capability of performing intensive studies at more than one surface landing site. This flexibility is possible by deploying multiple lander/rover teams which may be carried from site to site using powered air vehicles for very-long-distance transport. Physical samples could also be returned to stationary landers by similar means. Another possibility is a highly sophisticated long-range rover having a complete set of instruments and sample collection and analysis equipment, and designed for higher speeds, longer traverses (more than 100 km), and enhanced survivability over more difficult terrain with more challenging obstacles. Long-range rovers could visit any number of distinct geologic regions during their lifetimes and might be used to deploy a network of stationary science packages across the surface of the entire planet. The orbit of the main spacecraft is such as to permit regular contact with surface vehicles twice each Titan day (once each Earth week).

The lander/rover system needs extensive machine intelligence capability. Technology requirements are greatest for a long-range rover operating independently in the absence of continuous communications with the main orbiting spacecraft or with Earth. This capability is highly desirable, since without it the operational demands placed on other mission elements – such as the subsatellites for ground-to-orbit Titan uplink or powered air vehicles necessary for sample and system component transport – rapidly may become unmanageable.

A significant heritage may be expected from experience gained with the Viking landers and from any future martian or lunar missions, several of which might be approved and flown prior to the Titan Demonstration. One potential major difference is the unknown character of the surface including the possible existence of open liquids on Titan. If fluidic features are widespread it may be necessary to devise new methods of surface mobility and long-distance planetary exploration. New rover concepts for the reduction of machine intelligence requirements by decreased susceptibility to hazards should also be investigated.

Subsatellites. In addition to the main orbiter, subsatellites may be needed for certain specific purposes. One example is a free-flying spacecraft stationed at the L1 Lagrangian point between Titan and Saturn. This could be used to monitor the particle/field environment beyond Titan's magnetosphere, to observe the target atmosphere, and to communicate with mission elements located on the Saturn side of Titan. Another example is a tethered subsatellite system operating within 100 km of the main orbiter – such multiple devices can more easily distinguish spatial and temporal variations in particles and fields and probe the upper atmosphere (which would cause unacceptable drag on the main spacecraft if it attempted these measurements directly).

The subsatellite concept is new to planetary mission planning. However, these devices currently are projected for use on the Space Shuttle and also are under consideration in connection with manned and unmanned orbital platforms. This technology should become available by the time of the Titan Demonstration (e.g., the spin-stabilization of Mission relay subsatellites). There may also exist some commonality with previous planetary missions such as Pioneer 10/11 and Pioneer Venus.

Atmospheric probes. Several mission components must be sent into Titan's atmosphere at selected locations to make in situ measurements of the air and to carry small instrument packages to the surface. These probes are deployed by the main orbiter from its 600-km circular polar orbit, thus permitting considerable flexibility in choice of geographical entry points and timing. Atmospheric entry probes measure vertical profiles of the atmosphere at the time of deployment, and provide sufficient information to meet mission objectives at the "exploration" level. The Pioneer Venus, Galileo, and proposed Saturn Orbiter Dual Probe (SOP^2) missions all include atmospheric entry probes among their equipment.

One large entry probe and at least three small probes are necessary to fulfill the major objectives of Titan exploration. As in the Pioneer Venus mission, all probes measure atmospheric structure, pressure, temperature, etc., whereas

only the large probe takes more detailed data regarding composition, cloud structure, and planetary heat balance. (The large probe considered for the Titan Demonstration is roughly the same size and complexity as the device proposed for the SOP^2 mission.) Both types of probes also may serve as limited-purpose surface stations.

Powered air vehicles. Many options exist for intensive atmospheric investigation using still more sophisticated vehicles. A superpressure or passive hot-air (Montgolfier) balloon can be designed to float along an isobar for extended periods of time, providing a continuous record of wind speeds and other atmospheric data. Tethered balloons or kites could be used to sample the aerial environment surrounding a surface station. Powered air vehicles such as airplanes, helicopters, and dirigibles can study still larger regions of the atmosphere.

Of the options considered, the powered air vehicle – especially one having an inexhaustible energy supply for long-term operation – appears preferable. Such craft could be used to support extended surface operations, to conduct remote-sensing observations near the base, and even to help collect samples to be returned to the base site for detailed analysis. Regardless of whether the vehicle is an airplane or dirigible, it is highly unlikely that much previous experience will have been acquired with such systems in planetary missions. While the aerodynamic properties of fliers may match those of some Earth-based machines, control and propulsion requirements are likely to differ significantly. Control problems perhaps may be solved using a combination of smart sensors and an advanced machine intelligence capability, together with a satisfactory energy source such as a 10 kW nuclear-power generator to drive an efficient propeller. Titan's atmosphere possibly could be utilized for the production of propellants or buoyant gas.

Packaging the entire system and deploying it at Titan is an additional concern.

Surface science network. A scientific network should be established consisting of at least three permanent sites on the Titanian surface. The network collects seismographic and meteorological data needed to infer subsurface structure and global atmospheric circulation patterns. There are several ways to establish a network, such as (1) using long-range rovers to deploy stationary science packages, (2) deploying surface penetrators dropped from the main orbiter, and (3) extending the lifetime of the atmospheric probes (also dispatched from the main orbiter).

The network concept emphasizes long-term observation – as much as 5 years or more – on Titan's surface. Assuming network stations communicate directly to the main orbiting spacecraft, data must be stored for about a week following collection before uplinking. Each station must be able to function in an extremely cold thermal environment (about 100 K) with internal parts maintained at reasonable operating temperatures not below 220 K. Stations must be well-coupled to the planetary surface for seismometric purposes but must not thaw crustal ices. One solution is the radiation of excess heat up into the atmosphere.

All of the above components are relatively simple systems, mostly achievable using current or foreseeable aeronautical technology.

3.2.4 Machine Intelligence and Automation Requirements

In outlining the operational mission stages for a Titan demonstration and for the exploration of deep space, a number of automation technology drivers were identified in each of two general categories of system functions:

(1) Mission integrity, including self-maintenance, survival of the craft, and optimal sequencing of scientific study tasks.

(2) Scientific investigation, including data processing and the methodical formation of hypotheses and theories.

Both categories impose considerable strain on current AI technology for development in several overlapping areas of machine intelligence. These requirements represent research needs in domains of present concern in the AI community, as well as new research directions which have not yet been taken.

Success in mission integrity (fig. 3.4) requires the application of sophisticated new machine intelligence techniques in computer perception and pattern recognition for imaging and low-level classification of data. This also presupposes the utilization of a variety of remote- and near-sensing equipment. Onboard processing of collected data serves to coordinate the distributed systems and planning activity in terms of reasoning, action synthesis, and manipulation. More capable remote sensing is the key to efficient exploration, making more selective and efficient use of highly complex equipment for atmospheric and planetary surface monitoring.

With respect to reasoning, automated decisionmaking emerges as an important research area. Within this field, development might depart from current expert systems with advancements coming in the form of interacting simulation models of the processes which structure given domains and hypothesis formulating logics. New research directions lie in the areas of alternative computer logics, self-constructing knowledge bases, and self-learning systems.

A need has been identified regarding action synthesis, or procedural sequencing, for representing the relationship between predefined goal states and the current state, and for reducing the discrepancy between the two through automated implementation of subgoals and tasks. Such a system implies the utilization of a sequential informational feedback loop. A more difficult problem is simultaneous

coordination through anticipation, or prediction of the most appropriate action patterns followed by implementation of such action before a large discrepancy occurs. Complementary to the above capability is the capacity for automated construction of unprogrammed goal states as the result of environmental feedback. These latter two technology drivers fall under the general heading of automated learning and are not part of current research interests in the AI community at large.

Another broad technology requirement within the category of mission integrity is manipulation. A fully autonomous system should be capable of self-maintenance and repair, as well as sample collection for data analysis and utilization in decisionmaking processes. The former task presupposes some initial ability for self-diagnosis, while both tasks require a variety of effector capabilities for dealing with a wide range of situational demands. Here, advances in robotics with respect to hand-eye coordination

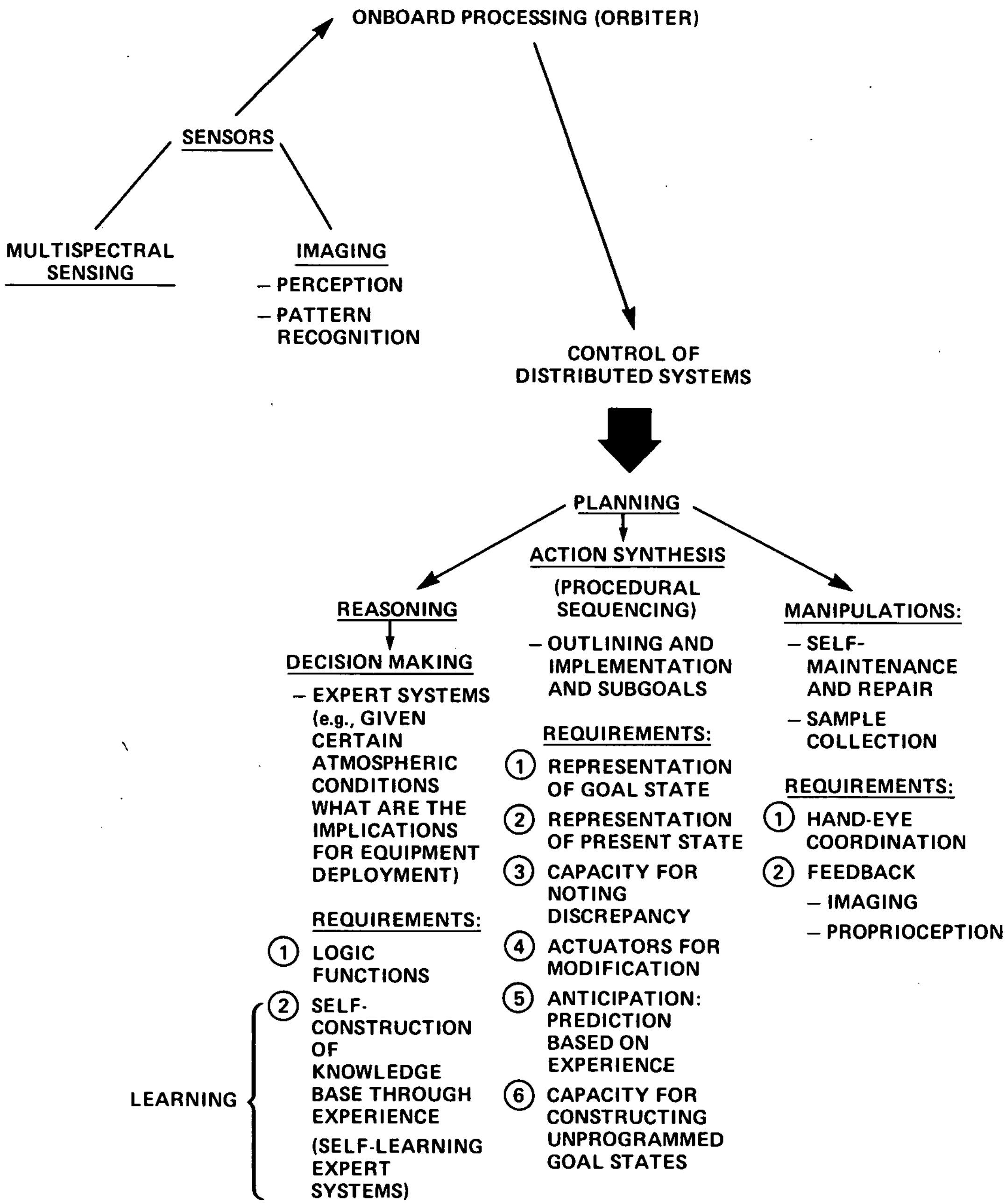


Figure 3.4.– Mission integrity.

and force/proprioceptive feedback systems emerge as significant.

The technology drivers identified for the scientific investigation category of mission functions (fig. 3.5) overlap to some degree those outlined for mission integrity. Automated intelligent planning is perceived as a general requirement in terms of defining scientific goals (both preprogrammed and self-generated) and for the definition of appropriate subgoals. Advanced decisionmaking also is an essential prerequisite for implementing scientific research and for conducting experiments. Decisions such as whether or not an experiment should be carried out, or where and when it should be conducted, probably could be accomplished (as with mission integrity) through extensions of current expert systems technology.

Reduction of collected sensory data to informational categories is yet another significant technology driver. A number of requirements emerge, starting with the ability to describe data at the simplest perceptual level. A higher-order task is the addition of data descriptions to a knowledge base for purposes of classification. This classification may be accomplished in terms of given categories of knowledge requiring some low-level hypothesis generation and testing. More advanced is the necessary capability for reorganizing old categories into new schemes or structures as a consequence of active information acquisition. Underlying this form of classificatory activity is again the self-learning process of hypothesis formation and testing. Each of the aforementioned tasks require varying levels of research and development to transform them into fully realized capabilities.

Finally, a requirement exists within the area of communication – transmitting acquired information back to human users. Here the emphasis is on automated selection processes in which an advanced decisionmaking system determines what information and which hypotheses are appropriate and sufficiently interesting to report. The obvious need to communicate with human beings in this case underscores the need for further developments in the field of natural language interfaces.

A scenario illustrating the great complexity of data processing and high-level hypothesis formation capability required for scientific investigation by an autonomous exploration system is presented in appendix 3C.

3.3 Machine Intelligence in Space Exploration Missions

The advanced machine intelligence requirements for general-purpose space exploration systems can be summarized largely in terms of two tasks: (1) Learn new environments, and (2) formulate new hypotheses about them.

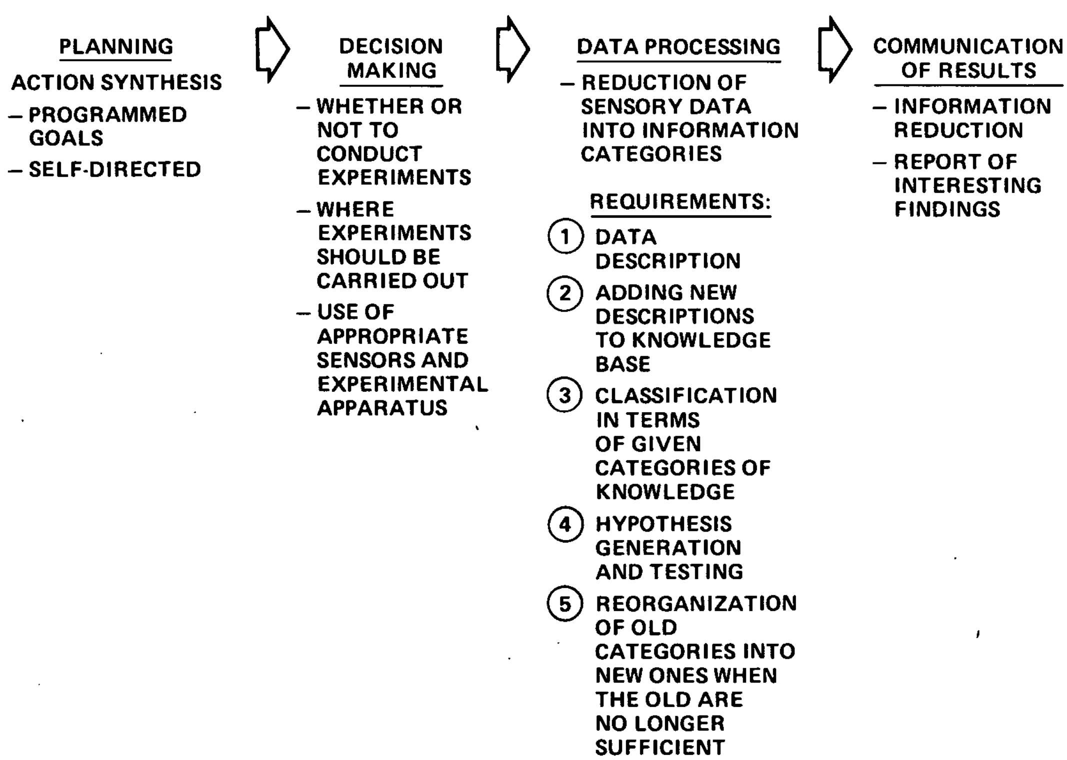


Figure 3.5.– Scientific investigation.

Hypothesis formation and learning have emerged as central problems in machine intelligence, representing perhaps the primary technological prerequisites for automated deep space exploration.

The Titan, outer planet, and interstellar missions discussed by the Space Exploration Team require a machine intelligence system able to autonomously conduct intensive studies of extraterrestrial objects. The artificial intelligence capacity supporting these missions must be adequate to the goal of producing scientific knowledge regarding previously unknown objects. Since the production of scientific knowledge is a high-level intelligence capability, the AI needs of the missions may be defined as "advanced-intelligence machine intelligence," or, more briefly, "advanced machine intelligence."

3.3.1 A Working Definition of Intelligence

Before an advanced machine intelligence system can be developed and implemented, the concept must be precisely defined and translated into operational terms. One way of doing this is to specify the patterns of inference which constitute the high-level intelligence – the design goal for advanced AI systems. Optimally, designers would have at their disposal an ideal definition of "intelligence" stating the necessary and sufficient conditions for achieving their goal. Such a definition, in addition to precisely stating what intelligence is, also would provide a set of criteria with which to decide the question: "Does entity X possess intelligence?" Unfortunately, no generally accepted ideal definition of intelligence is yet available.

However, a working definition sufficient for the purposes of the present investigation can be formulated. This inquiry addresses the general question of the characteristics of an advanced machine intelligence system needed for autonomous space exploration missions. As such, the investigation should address two questions in particular: "What intelligence capabilities must be designed into space exploration systems?" and "By what criteria will it be determined whether or not the final system actually possesses the high-level intelligence required for the mission?"

American Pragmatism, the major school of American philosophy, developed an acount of intelligence that contains the key to a useful working definition (Davis, 1972; Dewey, 1929, 1938; Fann, 1970; Mead, 1934, 1938; Miller, 1973; Peirce, 1960, 1966; and Thayer, 1968). The major figures of this school – John Dewey, William James, George Herbert Mead, and Charles Sanders Peirce – claimed that an entity's intelligence consists of its ability to reduce the complexity and variety of the world to patterns of order sufficient to support successful action by that entity. For example, human beings have reduced their welter of sensations to patterns of order, e.g., in comparative distinctions between nutrients and non-nutrients, chemical qualitative analysis schemes, and abstract aesthetic concepts. These patterns are, in turn, the bases of human actions including (following the above examples) satisfaction of the need for food, identification of an unknown chemical compound, and the creation of a work of art.

The Pragmatists further claimed that these action-related patterns of order exhaust an entity's knowledge. In other words, all knowledge is action-related – indeed, according to Peirce, "to have a belief is to be prepared to act in a certain way." This view is summarized in the fundamental Pragmatist principle that intelligence is always displayed in action and can be detected only in action. In this view intelligence is a dynamic process, rather than a static state, having at least two dimensions. First, unless an entity has a continuing history of action its intelligence is not displayed, cannot be detected, and therefore cannot be presumed to exist. Second, since a given pattern of order is linked to a related type of action, the success or failure of a particular action reflects on the "correctness" of the underlying pattern of order. An entity can have a continuing history of successful activity only if it can modify or replace those patterns of order which lead to failure. Therefore, an entity's intelligence is far more than merely the possession of a fixed stock of knowledge – even when this knowledge consists of action-related patterns of order. Rather, intelligence is the ability to preserve a high ratio of successful to unsuccessful outcomes.

The Pragmatists' account of intelligence can be summarized by this definition: Intelligence is the ability to formulate and revise patterns of order, as evidenced by the eventual emergence of successful over unsuccessful actions. There may well be aspects of intelligence that escape the definition, but nevertheless it provides a useful framework for the present investigation. This is because it focuses on capabilities which must be designed into advanced machine intelligence systems required for autonomous space exploration, as well as on the criteria with which to test for the presence of these capabilities.

A working definition of "advanced machine intelligence" in the context of autonomous scientific investigation of extraterrestrial objects can be formulated by utilizing the above general definition. The Pragmatists held that intelligence is a matter of degree and that among biological entities the question is never intelligence versus nonintelligence, but rather the level thereof. The actions by which biological entities display intelligence range from the amoeba's avoidance of toxic materials to the human's acquisition of scientific knowledge. The patterns of order underlying this spectrum of activity are characterized by a wide range of complexity paralleling that of the related actions. Machine intelligence also admits of degrees. Applying the Pragmatists' general definition is primarily a matter of specifying the level of capabilities with which the investigation is concerned.

In particular, application of the general definition to AI in space applications requires interpreting "actions" to

mean "scientific investigation and mission survival" (the two most complex sets of tasks facing an autonomous exploratory system) and "patterns of order" to mean "the complex abstractive and conceptual structures related to scientific investigation and mission survival" (e.g., hierarchical schemes and terrain maps, respectively). Hence the working definition of advanced machine intelligence in the context of the present study may be summarized as follows:

> Advanced machine intelligence is the ability of a machine system to autonomously formulate and to revise the patterns of order required for it to conduct scientific investigations and to survive, as evidenced by continued systemic survival and investigatory behavior despite any environmental challenges it may encounter.

This working definition provides ready answers to the capabilities and criteria issues raised earlier. These responses may be restated from the above definition as follows: (1) An advanced machine intelligence system for autonomous space exploration must possess the capability to utilize already formulated patterns of order and to devise new or revise existing patterns of order; and (2) the criteria by which to determine whether a system actually possesses intelligence is its observed ability to self-correct unsuccessful actions and eventually to act successfully in situations novel to the system.

3.3.2 A Systems Approach to Intelligence

Systems analysis may be used to translate the above definition into practice. Stated in general terms, the design goal is to achieve a machine intelligence capability to autonomously conduct scientific investigations and ensure mission survival. "Intelligence" can be an omnibus term which refers to a broad range of abilities including "knowing," "emoting," "fantasizing," etc. However, only rational cognition such as "knowing" is immediately relevant to machine intelligence for space exploration.

Of course, "knowing" is itself an omnibus term having a range of usages differing somewhat in meaning. In the present context it refers to the rational dimensions of intelligence, the processes of acquiring justified, though possibly fallible, statements about the world and its constituents. Among those dimensions are (1) identifying things and processes, (2) problem-solving, and (3) planning, since the outcomes of each of these processes are statements about the world selected from among a number of alternatives and justified on some basis. The essence of "knowing" in the context of a given environment is the ability to organize and thereby reduce the complexity and variety of perceived events, entities, and processes in the surroundings – a broad general class of rational activity required for machine-intelligent space exploration systems.

A "classification scheme" is any distinction or set of distinctions which can be used to divide events, entities, or processes into separate classes. By this measure taxonomies, analytical identification procedures, scientific laws and theories (e.g., "$F = ma$" names, hence, distinguishes forces and masses), decision criteria (e.g., go/no-go configurations in a given context), and concepts (e.g., "true" divides all statements into two separate classes) all are examples of classification schemes. Thus, a scheme is any statement, theory, model, formula, taxonomy, concept, categorization, classification, or other representational or linguistic structure which identifies the recurring characteristics of particular environments.

Tasks by which knowing is accomplished may be divided into two distinct types: (1) Utilization of preformulated fixed classification schemes, and (2) generation of new classification schemes or revision of old ones by formulating new components for the schemes. These two task types differ fundamentally both in the characteristics of the tasks and in the types of inference which underlie them.

When preformulated, fixed classification schemes are used, outcomes include identifications, classifications, and descriptions of events, entities, and processes occurring in the environment. These outcomes take the form of statements of the following general types:

- "X is an entity of type A."
- "Y is an instance of process B."
- "Z is a class-C event."

In each case, perceived constituents of the environment are matched with the general classes of constituents into which the classification schemes divide the world. The pattern of inference underlying this type of task is the analytic comparison of actual environmental constituents with "known" assertions about general environmental characteristics. Thus, an important aspect of the utilization of classification schemes is the confrontation of these schemes with the facts of experience. Knowing of this type cannot be successful – indeed, cannot even continue – if the actual state of affairs in the environment and that postulated by the classification schemes differ significantly. So, while the utilization of preformulated classes is an important type of knowing activity, the actual knowing of a given environment is deficient if the schemes are incomplete or incorrect. Knowing can be complete only when new classification schemes can be formulated and incorrect ones revised.

The creation and revision of classification schemes is the second major type of task involved in knowing. The outcomes of this task are either new classification schemes or new parts for preformulated ones. This task can, in turn, be divided into subtasks – the invention of new or revised classification schemes and the testing of these schemes for completeness and correctness prior to general use. Quite different types of inference underlie these two subtasks. Testing new or revised classification schemes requires analytic comparison of the claims made by these schemes with the facts of the world, exactly the same kind of process

involved in the utilization of classification schemes. However, the invention of new or revised schemes demands completely different types of inference. Two patterns of inference comprise this advanced activity – "induction" (included in all standard accounts of inference) and "abduction" (first described by Peirce, 1960, 1966; see also Burks, 1946; Fann, 1970; and Frankfurt, 1958) – as discussed at length below.

The systems approach leads to two important conclusions about machine intelligence (MI). First, MI involves the ability to utilize existing knowledge structures and to invent new ones. Second, although the utilization and invention of classification schemes require the formation of hypotheses, the inference for formulating hypotheses which apply existing classification schemes are logically distinct from the inferences used in formulating hypotheses which invent new or revised classification schemes (see fig. 3.6).

These conclusions have implications for machine intelligence systems designed for autonomous deep space exploration. If classification schemes applicable to the Earth were complete and correct for all extraterrestrial bodies, then an autonomous system utilizing these schemes via analytic inferences alone could successfully complete the knowing process. However, it is probably true that at least some of the available classification schemes are either incomplete or incorrect in the extraterrestrial context and, in any case, the most prudent design philosophy for a space exploration system would be to assume that gaps do exist. Under the assumption that novelty will be encountered in space, an autonomous exploratory system may successfully complete the knowing process only if it can utilize preformulated classification schemes and also invent new or revised ones, that is, only if it can make inferences of the inductive and abductive types in addition to inferences of the analytic type.

3.3.3 *Patterns of Inference for Hypothesis Formation*

Analytic, inductive, and abductive inferences will now be characterized in terms of the information inputs and outputs of each. An existence argument for abductive inference, which also establishes its centrality to scientific investigation, is offered, and the process involved in abduction is characterized in some detail. Finally, the requisite state of development for each of the three basic inferential types is contrasted with AI state of the art in the context of autonomous scientific investigation, the ultimate goal.

Analytic inferences are logical patterns by which existing scientific classification schemes (principles, laws, theories, and concepts) are applied to information about the events and processes of the world for the purpose of producing identifications and descriptions of these events and processes as well as predictions and explanations about them (Alexander, 1963; Harré, 1960; Hempel, 1965, 1966;

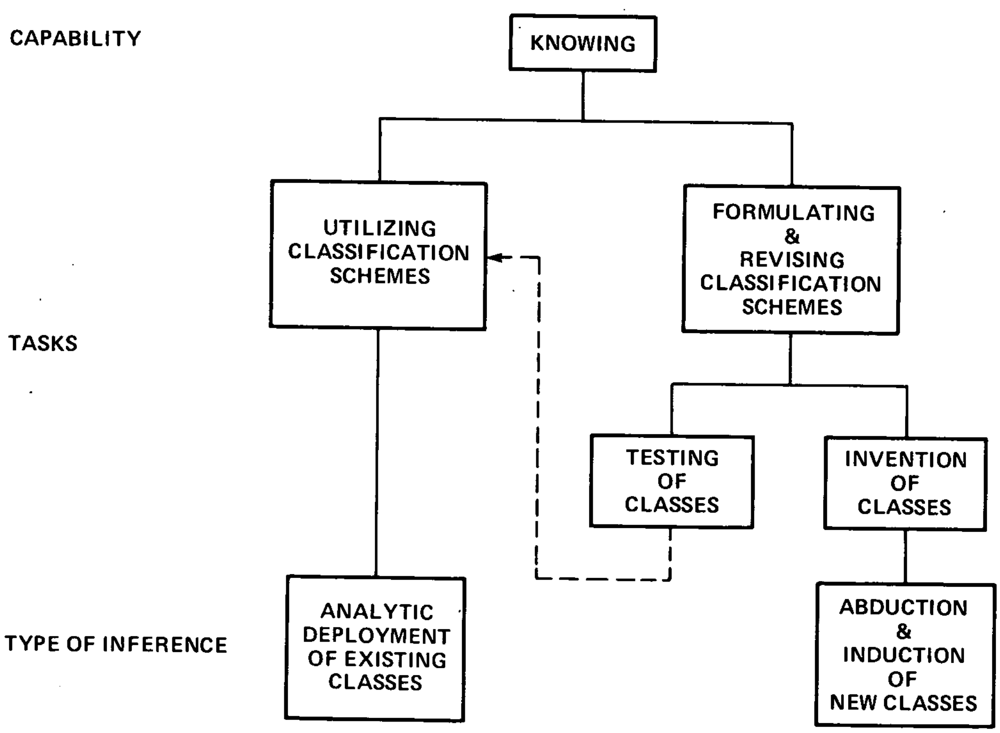


Figure 3.6.– Systems graph for machine intelligence.

Popper, 1963; Wisdom, 1952). This information itself is produced by applying current scientific classification schemes to raw data in an attempt to structure and interpret it. The reasoning is deduction, whether formal deductive logic or other deductivist analytical procedures. Models play an important, though indirect, role in analytic inference (Hanson, 1958; Kuhn, 1970; Toulmin, 1960). The quantitative and symbolic information and the identifications, descriptions, predictions, and explanations which are the outputs of analytic inferences are derived from detailed knowledge such as equations, formulas, laws, and theories. However, standing behind this detailed knowledge is a fundamental model of the "deep structure" of the world which, in effect, provides a rationale for applying that particular kind of detailed knowledge to that specific data. For instance, the kinetic-molecular theory of gases is one such fundamental model whose scientific function is to provide a rationale for searching and then applying a particular kind of detailed knowledge about gases. Figure 3.7 shows the input/output structure of analytic inference.

Inductive inferences are logical patterns for moving from quantitative or symbolic information about a restricted portion of a domain to universal statements about the entire domain (Cohen, 1970; Good, 1977; Hilpinen, 1968; Horton, 1973; Lehrer, 1957, 1970; Rescher, 1961; Salmon, 1967; Skyrms, 1966). There are two somewhat different aspects of inductive inference: Inductive generalization and abstraction. In inductive generalization, some finite set of measurements of an independent variable and its dependent variable are generalized into a mathematical function which holds for all possible values of those variables. Alternatively, in abstraction, some finite set of symbolic representations of just a few members of some domain is the basis for inferring some abstractive characteristic common to all members of the domain. Examples of abstraction include moving from a set of white objects to the concept of "white," and inferring from the information that all observed ravens are black; the principle that being black is a defining characteristic of ravens. As was the case with analytic inferences, models play an important though indirect role (Hanson, 1958; Kuhn, 1970; Toulmin, 1960). These models serve to restrict the range of mathematical functions or abstractive concepts that can characterize a domain, hence, they focus the inductive inference from information to generalization. For instance, we know that Robert Boyle was guided in the processing of pressure and volume data by a model of gases that required volume to decrease while pressure increased (Toulmin, 1961). Figure 3.8 suggests the input/output structure of inductive inference.

Abductive inferences are logical patterns for moving from an input set that includes:

- some theoretical structure T consisting of models, theories, laws, concepts, classification schemes, or some combination of these,
- some prediction P derived from T by means of an analytic inference, and
- some set of quantitative or symbolic data D which contradict P (D = not-P),

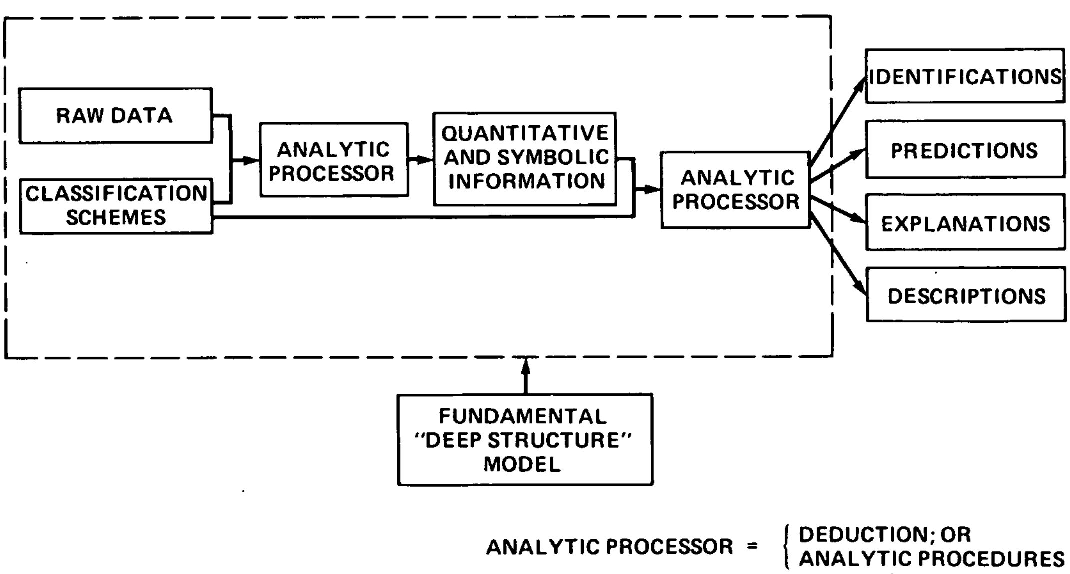


Figure 3.7.– Analytic inference.

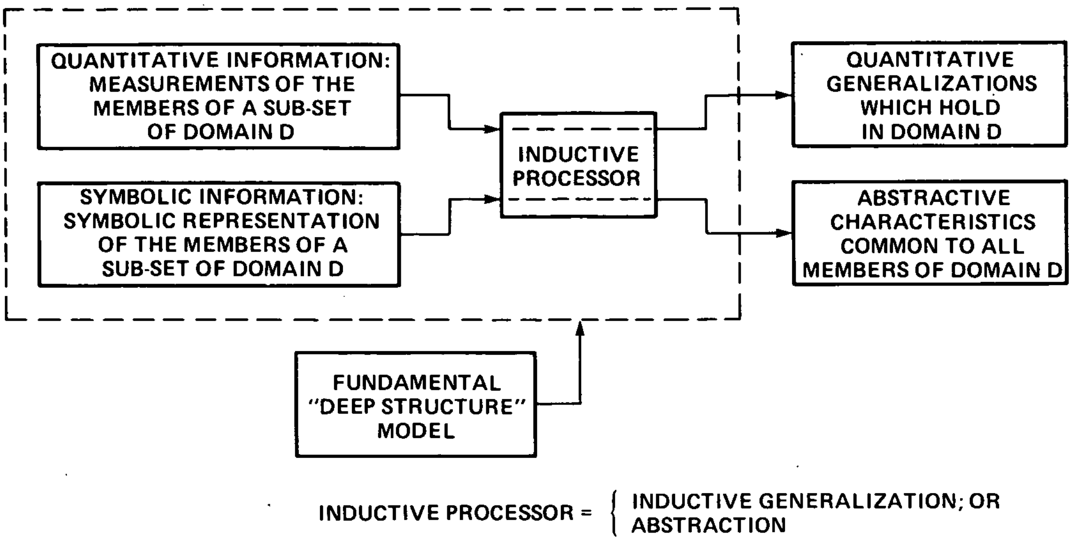


Figure 3.8.– Inductive inference.

to an output set that includes:

- a new or revised theoretical structure T^*,
- a prediction P^* derived from T^*, and
- a set of quantitative or symbolic data D^* which both agrees with P^* and is the representation of D in T^*;

that is, $D^* = P^*$ and D^* is the mapping of D onto T^* (Burks, 1946; Davis, 1972; Dewey, 1929, 1938; Fann, 1970; Frankfurt, 1958; Gravander, 1975, 1978; Hanson, 1958, 1961, 1965, 1967, 1969; Kuhn, 1957, 1970, 1977; Lakatos, 1970, 1976, 1977; Mead, 1934, 1938; Miller, 1973; Peirce, 1960, 1966; Simon, 1965; Toulmin, 1960, 1961, 1972; Van Duijn, 1961). Models play a far more important role than in just analytic and inductive inferences: In abduction, fundamental models of processes structuring the world enter directly into the inference. Such models sometimes are the component of a theoretical structure replaced or modified by the inference. Of course, not every replacement or revision of the theoretical structure involves model modification. Those abductive inferences which revise or replace such components as laws or generalizations take the model to be a premise of the inference. The input/output morphology of abductive inference is shown in figure 3.9.

Probably there exists a family of abductive inference species. However, all members of this family must bear many resemblances to one another. Two such family characteristics are particularly important. First, the logical impetus behind the transition from T to T^* is the ability of T^* to explain data which T cannot. Second, the attainment of explanation involves a re-representation of information – i.e., T fails to explain D and T^* explains D^*, where D^* is not D but rather the representation of D in T^*. As Lakatos (1976) notes, "discovery" is a process in which a theory stated in language L fails to explain a fact; therefore, it cannot adequately be represented in L, so a theory stated in L' must be found to explain it and allow its representation in L'.

Virtually all standard accounts of scientific investigation include analytic and inductive inferences as important components of the logic of science. Abductive inferences are not as widely accepted or understood. Nevertheless, numerous detailed analyses of actual scientific discoveries have demonstrated that there are inferences in these discoveries that are neither analytic nor inductive in nature (Gravander, 1975; Hanson, 1958; Kuhn, 1957; Lakatos, 1977; McMullin, 1978). Examination of these scientific discoveries establishes that the researcher involved in the discovery possessed a determinate set of initial information, including some existing theory and data contradicting a prediction of the theory, and that there is a detailed inference which takes this initial information as its premise and provides the discovery as a conclusion. Whether it is possible to prove that the scientists in question actually followed this inference step by step is irrelevant insofar as the present investigation is concerned. The analyses reported demonstrate the existence of a family of nonanalytic and noninductive inferences which produce new or revised theoretical structures as output. This demonstration constitutes an existence argument for abductive inference.

It cannot be emphasized too strongly that the analysis of actual scientific discoveries is valid only as an existence argument for abduction, not as a research program for mechanizing it. Investigations into the logical process

underlying abductive inference certainly is a first step toward mechanizing the invention of new or revised scientific laws and concepts. But these inferences cannot be demonstrated to be the inference which the scientist followed to the new notion, rather, only an inference having this new notion as its conclusion. Thus, it is not at all clear that a theory of abduction adequate for machine intelligence applications must await a full understanding of human cognition. Quite the contrary; the preferred approach is to attempt to develop a theory of abductive inference on the basis of a direct logical analysis, retreating to the more fundamental problem of human cognition only if the techniques of logical analysis fail.

To consider what might be expected from a direct attack on the logic of abduction a brief characterization of inferential steps constituting such inference is presented below. This characterization takes the viewpoint of some unspecified knower "X," either a scientist or an abductive machine intelligence system.

(1) X is surprised while using theoretical structure T by some occurrence, 0, because 0 is not among X's set of expectations that are based on T.

(2) X represents 0 by a determinate set of data, D.

(3) X demonstrates that D is more than simply unexpected; it is anomalous in the sense that T predicts not-D.

(4) X traces not-D back to those components $[T_1, T_2, \ldots]$ of its total theoretical structure which entered directly into T's prediction of not-D.

(5) X determines which element, T_j, in $[T_1, T_2, \ldots]$ is the most likely "villain" behind X's misexpectation.

(6) X attempts to reformulate T_j in such a way that when the new T_j^* is substituted for T_j in a revised T^*, 0 can be represented by D^* which, in turn, is predicted by T^*. (If successful, the next step is (9) below.)

(7) If not successful, X repeats steps (5) and (6) above with the remaining elements of $[T_1, T_2, \ldots]$ in order of decreasing likelihood until all possibilities are exhausted. (If successful, the next step is (9) below.)

(8) If still not successful, X repeats steps (5) and (6) with the remaining elements of T, in order of increasing theoretical content and scope, the last component tried being the fundamental "deep structure" model itself.

(9) X makes all adjustments in T^* necessitated by the adoption of T_j^*, including generating a new set of expectations 0^*.

(10) X uses T^* until the next "surprising" occurrence.

This characterization of abduction, though not as detailed and precise as that which would result from further investigation, is precise enough to suggest three key problems standing in the way of mechanized abductive inference. First, how should 0 best be represented as data so that later re-representation is facilitated, and how should these re-representations be performed? Second, is the initial selection of "villains" best achieved by parallel search, hierarchical serial search, or some other technique? Third, can the formulation of the T_j^* replacement of T_j be captured in a stepwise inference in which preceding steps uniquely constrain the selection of the next succeeding step, or must some other technique be used? Note that all of these may be addressed on logical grounds, independent of the broader questions of human cognition.

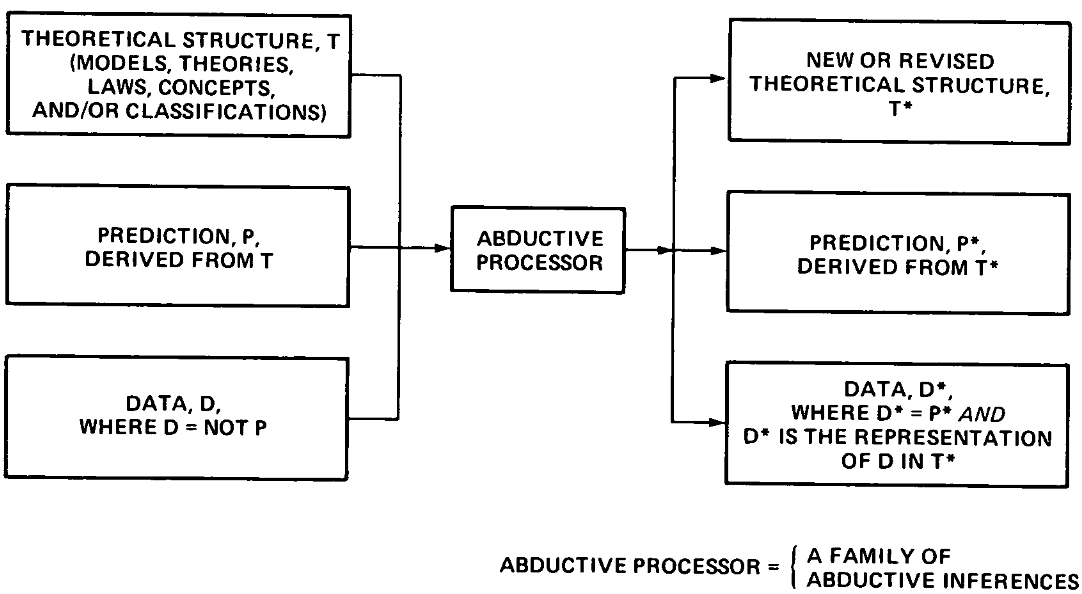


Figure 3.9. – Abductive inference.

Finally, it is instructive to contrast state-of-the-art AI treatments of analytic, inductive and abductive inference with the optimal treatment required to achieve working machine intelligence systems with highly advanced capabilities. (See also chapter 6.) First, with respect to analytic inference, current AI research is not addressing the central problem of supporting the detailed knowledge in the classification schemes with fundamental models. Second, although some preliminary work has been done in mechanizing inductive inference (Hajek and Havranek, 1978), this work also has not adequately addressed the basic problem of connecting fundamental models to the generalizing process. Third, only tentative steps have been taken in the development of mechanized abductive inference (Hayes-Roth, 1980), and even these efforts are not grounded on a mature theory of abduction for machine intelligence.

3.3.4 The Inference Needs of Autonomous Space Exploration Systems

For an autonomous space exploration system to undertake knowing and learning tasks, it must be capable of mechanically formulating hypotheses using all three of the distinct logical patterns of inference, as follows:

- Analytic inference – needed by the explorer system to process raw data and to identify, describe, predict, and explain events and processes in terms of existing knowledge structures.
- Inductive inference – necessary to formulate quantitative generalizations and to abstract the common features of events and processes, both of which amount to the invention of new knowledge structures.
- Abductive inference – needed by the system to formulate hypotheses about new scientific laws, theories, models, concepts, principles, and classifications. The formulation of this type of hypothesis is the key to the ability to invent a full range of novel knowledge structures required for successful and comprehensive scientific investigation.

Although the three patterns of inferences are distinct and independent, they can be ordered by difficulty and complexity. This ordering is the same as comparing their ability to support the invention of new knowledge structures. Analytic inference is at the low end. An automated system that performs only this type of inference could probably undertake reconnaissance missions successfully. Next is inductive inference. A machine system able to perform this type as well as analytic inference could successfully undertake missions combining reconnaissance and exploration, provided the planet explored is represented well enough by the fundamental models with which the system would be preprogrammed. But if the processes underlying the phenomena of the new world are not well-represented by the fundamental models, automated combined reconnaissance and exploration missions will require abductive inference. Abduction is at the top of both orderings. It is the most difficult as well as the heart of knowledge invention. An automated system capable of abductive reasoning could successfully undertake missions combining reconnaissance, exploration, and intensive study.

3.3.5 Cognitive Processes in Intelligent Activity

One significant technology driver in fully autonomous space exploration is the capacity for learning and the need for adaptive forms of machine intelligence in future space missions (fig. 3.10). However, a review of the literature (Arden, 1980; Boden, 1977; Raphael, 1976) and personal consultations with experts in the field of AI indicate that theoretical and technological research in this area has not seriously been pursued for many years.

For this reason it is useful to approach the goal of adaptive intelligence from the perspective of a related field of study in which it has already received considerable attention: Cognitive psychology. Clearly, descriptions of human thought processes leading to intelligent behavior cannot serve as a direct template for machine intelligence programming – it is a recognized philosophy of the AI community that software need not exactly mimic human processes to achieve an intelligent outcome. Rather, the objective is to describe some aspects of human cognition in hopes of bridging the gap between present limitations in the AI field and the level of machine intelligence likely to be needed in future space exploration missions.

Perception and pattern recognition. The most fundamental kinds of intelligence are perception and the related activity of pattern recognition. Each has been the subject of much study by cognitive and physiological psychologists. For example, evidence from Sperling (1960) suggests that perceptual input is held briefly in a sensory buffer register, thus, permitting the activation of control processes to encode the data in terms of meaningful categories. Stimuli presented to the human sensorium arrive in conscious awareness first as some perceptual-level description, then later with some useful label attached. Exactly how these processes work remains unknown, in part because perception occurs below the subject's level of awareness. Progress to date provides only partially integrated theories of perceptual data handling, yet these are sufficiently well-developed to deserve a brief review in the context of the present study.

A definition of perception at the descriptive level, popular in the psychological literature, holds that sensory processing is essentially inferential or interpretive, based on

raw sensory cues available in the environment, and produces and subsequently tests interpretations about what the world looks like. The percept is the phenomenological result of the interpretation. In this view, perception is a subconscious, "hard-wired" constructive process involving the formation of a hypothesis, a test of that hypothesis, and a consequent decision as to whether the hypothesis accurately encompasses the sensory information. The literature of psychology contains much evidence to support such a description as a reasonable characterization of human perception (Neisser, 1967; Rock, 1975), and the AI community has accepted, in principle, a similar view (Arden, 1980). However, the techniques and operations typically employed to achieve computer pattern-sensing generally fail to properly incorporate the notion of perception and recognition as active constructive processes.

Cognitive psychological theory has largely emphasized two general approaches in characterizing pattern recognition schemes – template matching and feature extraction theory. Each has a different focus of attention with respect to the three major aspects of recognition called "description," "representation," and "matching" (of new images against stored representations).

Template matching theorists propose that a literal copy of perceived stimuli stored in memory is matched against new incoming stimuli. Although this view has been criticized as too simplistic and naive (Klatsky, 1975; Neisser, 1967), updated versions of the hypothesis still hold sway. For instance, one modification retains the notion that literal copies are stored in memory but suggests that new percepts are "normalized" before matching. In this view, some precomparison processing takes place in which edges are smoothed out, oriented in the appropriate plane, and centered with respect to the surrounding field. In addition, image context helps in the normalizing process by reducing the number of possible patterns the stimulus might match (Klatsky, 1975). In the field of AI technology, the Massively Parallel Processor or "MPP" (an imaging system currently under development at Goddard Space Flight Center) uses visual data-handling techniques with characteristics

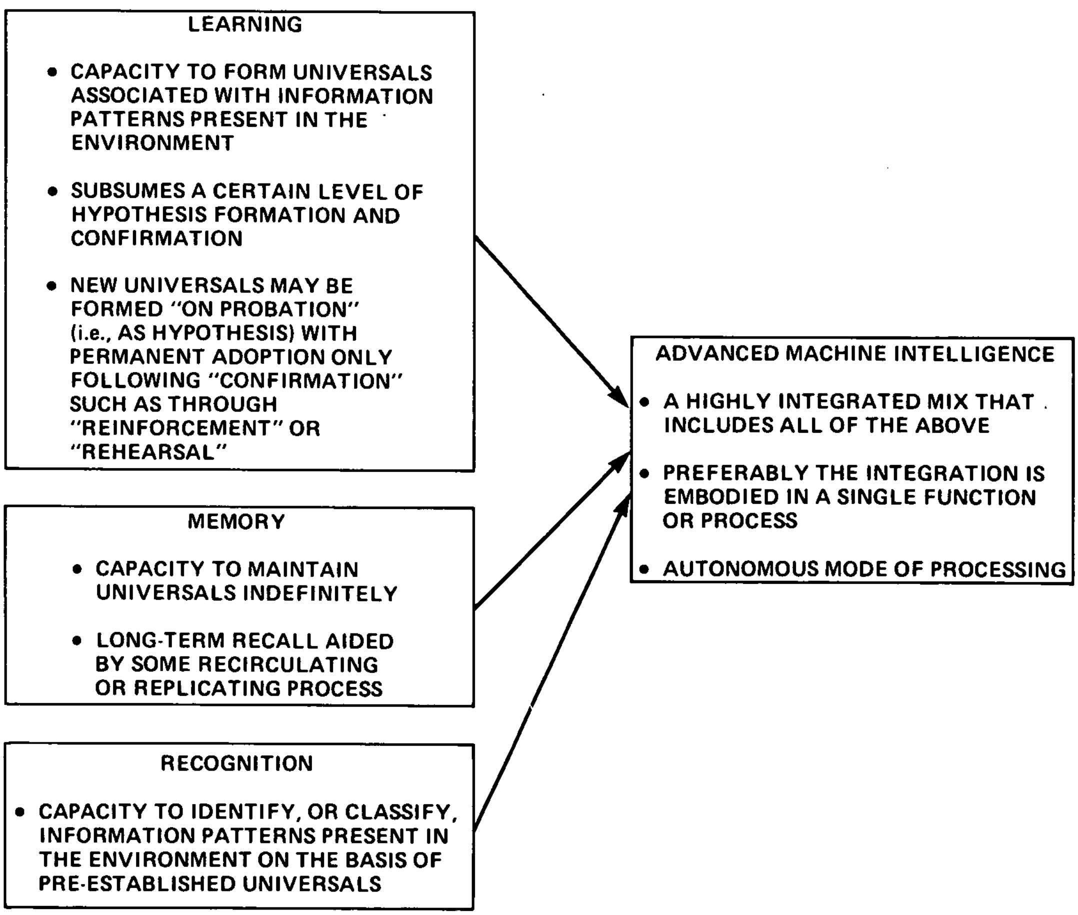


Figure 3.10.– Adaptive machine intelligence for advanced space exploration.

remarkably similar to those described in the normalizing and template matching theories. Given information on its sensory perspective and images stored in its memory, the MPP performs precomparison processing to orient incoming images for compatibility with stored images.

Another hypothesis of perception with similar assumptions is feature detection or feature extraction theory. According to this formulation a pattern may be characterized as a configuration of elements or features which can be broken down into constituent subcomponents and put back together again. Recognition is a comparison process between lists of stored features (which when combined, constitute a pattern) and features extracted from incoming stimuli (Klatsky, 1975). An early theoretical AI model of the feature detection hypothesis was Pandemonium (Selfridge, 1966). This system performs a hierarchical comparison of low-level through higher-order features until the incoming pattern is recognized. More recent scene analysis paradigms have grown from similar assumptions that the raw scene must be "segmented" into regions, or edges of regions, out of which desired objects may be constructed (Arden, 1980; Barrow, private communication, 1980). Scene-analysis models developed on the basis of higher-order features of greater complexity than those proposed by Selfridge have achieved moderate success in limited environments. The major problem is that the system can only deal with familiar or expected input data. All categories within which items are recognized, must be explicitly defined by the programmer in terms of their subcomponents. This eliminates the possibility of recognition processes in novel environments.

Reviewed together, template matching and feature detection reflect the processes modeled by most AI imaging and pattern recognition research. Hence, current AI systems are incapable of handling new category construction and other advanced perceptual tasks which might be required in future space missions. This limitation suggests that an alternative approach to the problem of automated pattern recognition may be needed.

Despite abundant research supporting the existence of feature detectors in humans (Hubel and Wiesel, 1966; Lettvin et al., 1959), other evidence suggests that feature and template theory do not provide a complete explanation of recognition. The above approaches are regarded today as unsophisticated in their conception of how events are mentally represented, and erroneous in ignoring the problem of how representations are achieved. Experiments conducted by Franks and Bransford (1971) indicate that the human mental representation used for feature comparison may be prototypical and holistic rather than literal and elemental. That is, what is actually stored in memory is the product of an active construction, developed over time. In this view the cognitive system extracts and stores the converging "essences" of items to which it is exposed, and this abstraction is then utilized in the recognition process. The emphasis is on conceptual representational construction and conceptually driven (top-down) processing, rather than matching and data-driven (bottom-up) processing. The advantage of a prototype approach to perception is that minor distortions or transformations within a limited range will not interfere with the recognition process.

The prototype approach may be considered in terms of two different aspects – the abstract analogical nature of representation and category or concept construction. With respect to machine intelligence, perhaps the closest approximation to the notion of prototypical representation is illustrated by Minsky's "frame" concept. A frame in Minsky's formulation is a data structure for representing a stereotyped situation (Minsky, 1975) and corresponds in many ways to the psychological notion of schema (Bartlett, 1961). Though not really analogical in nature, the frame conception contributes to scene analysis by permitting the system to access data in a top-down fashion and to utilize generalized information without relying on simplistic features. The frames, however, must be described within the system by a programmer and are relatively static. There is no capability for frame reorganization as a result of experience.

Consider now the second aspect of the prototype approach, the construction of abstract categorical representations. Category construction may be viewed as a brand of concept formation. Experimental evidence suggests that the formation of new conceptual categories is the result of a hypothesis generation and testing process (Levine, 1975) in which recursive operations are evoked which infer hypotheses about how a number of particulars are related and then test those hypotheses against feedback information from the environment. Some additional evidence suggests that a number of these hypotheses may be tested simultaneously (Bruner et al., 1956). The result is considered an abstract analogical representation capturing an essence which subsumes all the particulars. Since the hypothesis theory of concept formation typically has been considered in the context of conscious processes, it may seem somewhat far afield of perceptual processing. However, since perception itself has been described as an unconscious inferential process, it may be the case that similar underlying logical operations are at work in the formation of higher-order concepts, prototypes, and in perceptual construction. The precise nature of acquisition, how an "elegant" hypothesis is formed, is not clearly specified in any of these theories. (See section 3.3.3.)

Only a minimal amount of work has been done on AI approaches to the formation of new conceptual structures. A classic attempt was Winston's concept formation program in which a machine was taught through example to acquire new concepts (e.g., the architectural concept of "arch"). Using informational feedback from the programmer as to whether a particular example illustrated the concept or not, and by assessing the essential similarities and differences

among the examples it was shown, Winston's software created structural descriptions of the essentials of the concept in the form of a semantic network.

The function of such a program may appropriately be defined as concept learning. However, the programming techniques appear more closely wedded to the notion of concepts as feature lists rather than as prototypical, analogical structures. This "feature view" has theoretical limits in the domains of human and artificial intelligence since a number of abstract categories can be identified in which constituent members have a few or no structural features in common but whose relationship is either more functional in nature or salient "in more broadly specifiable terms" (Boden, 1977; Rosch and Mervis, 1975). Salience for Winston's program relates only "to categorizations made by its human teacher for human purposes" (Boden, 1977). It is difficult to see how a program with a feature list assumption could move beyond predefined categories to handle the construction of new abstract concepts. This is a significant constraint on state-of-the-art AI technology in terms of future space missions requiring autonomous exploration in novel environments where "there is no guarantee that categorizations previously found useful would still be salient" (Boden, 1977).

Genetic epistemology. One final consideration with respect to intelligent activity comes from Jean Piaget's work on genetic epistemology. This topic is relevant to the issues addressed in this chapter because genetic epistemology offers one of the most comprehensive views of intelligence to be found in the literature today. Piaget's conceptions of the underlying processes of "natural" intelligence encompass the behavioral and cognitive activities of humans and animals. Moreover, the processes are sufficiently general possibly to be captured in a nonliving artifact which would then serve as an effective realization of non-natural intelligence (Piaget, 1970).

How can intelligence be characterized in terms of structures and processes so that it might be embodied in a computer system? One important assumption of Piaget's theory is that any account of the evolution of cognitive activity and intelligence must include the nonteleological aspects of adaptation and purpose. The process of equilibration, a regulative function which propels the subject toward more inclusive and stable interactions with its environment, is basic to the theory. The deterministic result of equilibrium is seen as a characteristic structuring of the relations between subject and environment (Piaget, 1963).

There are two processes that subjects must coordinate in order to achieve a state of equilibrium: Assimilation and accommodation. Assimilation, exhibited by all organisms, is the functional aspect of structure formation by which subjects, acting on their environment, modify it in terms of existing structures (Piaget, 1970). Each organism possesses a set of generalized behavior patterns, or action schemes, which support its repetitive modification of its environment for the purpose of producing an expanded set of interactions. Accommodation is the modification of the assimilatory cycle itself as a result of the subject's interactions with its surroundings (Piaget, 1963). Accommodation involves the transformation of existing structures in response to continuous environmental stimulation. The result is the construction of new categories of experience which then become part of the organism's general behavioral repertoire.

For Piaget, these "schemes" are the basic units for structuring knowledge (Rosenberg, 1980), the means by which all overt behavioral and cognitive activity is organized. The notion of "scheme" defined by Piaget has certain similarities to Minsky's "frames" as the basic units of knowledge representation. Both notions imply a top-down processing schedule for intelligent activity. However, the two notions differ dramatically in terms of their dynamics. The frame permits a kind of assimilatory activity (organization of particulars within its structure) but the structure itself is relatively static – there seems to be no possibility for reorganization of the structure (the frames) in response to experience. Alternatively, the scheme emphasizes both assimilative and accommodative processes. Accommodation in this case is the restructuring of available schemes into new higher-order schemes which subsume all previous particulars while simultaneously permitting the inclusion of new ones. Again the primary gap between the level of intelligence available with current AI approaches and that which characterizes more advanced intelligent activity appears in the domain of emergent change. Transforming present knowledge structures into new higher-order schemes is a prerequisite for fully intelligent activity, and this capability is absent from state-of-the-art AI techniques.

While the utilization of a genetic epistemological framework has not yet received much study by researchers in the AI field, it has attracted some recent attention in other quarters. For instance, Rosenberg (1980) suggests a number of ways to blend Piaget's theory and current AI methodology to their mutual benefit. Perhaps this represents the beginning of a recognition of the need for comprehensive formulations of natural intelligence to be incorporated into the development of a theory of intelligence in nonhuman artifacts.

3.4 Technology Drivers for Automated Space Exploration

The most important single technology driver for automated space-exploration missions of the future is advanced machine intelligence, especially a sophisticated MI system able to learn new environments and to generate scientific hypotheses using analytic, inductive, and abductive reasoning. Within the AI field the most powerful technology driver is the demonstrable need for an abductive inferential capability useful for inferring new successful knowledge

structures from failed ones. Required machine intelligence technologies include:

- Autonomous processing (essentially no programming)
- Autonomous "dynamic" memory
- Autonomous error-correction
- Inherently parallel processing
- Abductive/dialectic logical capabilities
- General capacity for acquisition and recognition of patterns
- Universal "Turing Machine" computability.

Numerous other supporting technologies also are essential for the staging of autonomous space exploration missions, including low-thrust propulsion systems; general-purpose surface exploration vehicles able to function on both solid and fluid surfaces; reconfigurable sensor nets and smart sensors; flexible, adaptive general-purpose robot manipulators; and distributed intelligence/database systems.

3.5 References

Alexander, Peter: A Preface to the Logic of Science. Sheed & Ward, London, 1963.

Arden, Bruce W., ed.: What Can Be Automated? The Computer Science and Engineering Research Study (COSERS), MIT Press, 1980. 934 pp.

Bartlett, Frederic Charles: Remembering: A Study in Experimental and Social Psychology. The Cambridge Univ. Press, 1932. Revised Edition, 1961.

Black, D. C., ed., Project Orion: A Design Study of a System for Detecting Extrasolar Planets, NASA SP-436, 1980.

Boden, Margaret: Artificial Intelligence and Natural Man. Basic Books, Inc., 1977.

Bruner, G. S.; Goodnow, J. J.; and Austin, G. A.: A Study of Thinking. Wiley, New York, 1956.

Burks, Arthur: Peirce's Theory of Abduction. Philosophy of Science, vol. 13, 1946, pp. 301-306.

Cassenti, B. N.: A Comparison of Interstellar Propulsion Systems, AIAA paper 80-1229, AIAA/SAE/ASME 16th Joint Propulsion Conf., 30 June-2 July, 1980. Hartford, Connecticut.

Cohen, Laurence Jonathan: The Implications of Induction. Methuen Press, London, 1970.

Davis, William H.: Peirce's Epistemology. Martinus Nijhoff, The Hague, 1972.

Dewey, John: Experience and Nature. W. W. Norton and Co., New York, 1929. Revised Ed.

Dewey, John: Logic: The Theory of Inquiry. Henry Holt and Co., New York, 1938.

Fann, K. T.: Peirce's Theory of Abduction. Martinus Nijhoff, The Hague, 1970.

Frankfurt, Harry G.: Peirce's Notion of Abduction. J. of Philosophy, vol. 55, July 1958, pp. 593-597.

Franks, Jeffrey J.; and Bransford, John D.: Abstraction of Visual Patterns. J. of Exper. Psych., vol. 90, no. 1, 1971, pp. 65-74.

Freitas, Robert A., Jr.: Interstellar Probes: A New Approach to SETI. J. British Interplanetary Soc., vol. 33, March 1980a, pp. 95-100.

Freitas, Robert A., Jr.: A Self-Reproducing Interstellar Probe. J. British Interplanetary Soc., vol. 33, July 1980b, pp. 251-264.

Good, I. J.: Rationality, Evidence, and Induction in Scientific Inference. E. Elcock and D. Michil, eds., Machine Intelligence, Halstead Press, New York, 1977, pp. 171-174.

Gravander, Jerry Wallace: Newton's New Theory About Light and Color and the Hypothetico-Deductive Account of Scientific Method: Scientific Practice contra Philosophic Doctrine. Ph.D. dissertation, Univ. of Texas, 1975. Univ. Microfilms, Ann Arbor, Michigan, 1975.

Gravander, Jerry Wallace: Mead's Logic of Discovery. Michael P. Jones et al., eds., The Individual and Society, The Southwestern J. Philosophy Press, Norman, Oklahoma, 1978, pp. 187-207.

Hajek, Peter; and Havranek, Tomas: Mechanizing Hypothesis Formation: Mathematical Foundations for a General Theory. Springer-Verlag, Berlin, 1978.

Hanson, Norwood Russell: Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science. Cambridge University Press, Cambridge, 1958.

Hanson, Norwood Russell: Is There a Logic of Discovery? In H. Feigl and G. Maxwell, eds., Current Issues in the Philosophy of Science, Holt Rinehart and Winston, New York, 1961, pp. 20-35.

Hansen, Norwood Russell: The Idea of a Logic of Discovery. Dialogue, vol. 4, 1965, pp. 48-61.

Hanson, Norwood Russell: An Anatomy of Discovery, The Journal of Philosophy, vol. 64, 1967, pp. 321-352.

Hanson, Norwood Russell: Perception and Discovery: An Introduction to Scientific Inquiry. Freeman, San Francisco, 1969.

Harré, Romano: An Introduction to the Logic of the Sciences. Macmillan, London, 1960.

Hayes-Roth, Frederick, Theory-Driven Learning: Proofs and Refutations as a Basis for Concept Discovery, paper delivered at the Workshop on Machine Learning, Carnegie-Mellon University, July, 1980.

Hempel, Carl Gustav: Aspects of Scientific Explanation. Free Press, New York, 1965.

Hempel, Carl Gustav: Philosophy of Natural Science. Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

Hilpinen, Risto: Rules of Acceptance and Inductive Logic. North-Holland, Amsterdam, 1968.

Horton, Mary: In Defense of Francis Bacon: A Criticism of the Critics of the Inductive Method. Studies in the History and Philosophy of Science, vol. 4, 1973, pp. 241-278.

Hubel, D. H.; and Wiesel, T. N.: Receptive Fields, Binocular Interaction, Functional Architecture in Cats' Cortex. L. Uhr, ed., Pattern Recognition. John Wiley and Sons, Inc., New York, 1966, pp. 262-277.

Klatsky, Roberta L.: Human Memory: Structures and Processes. W. H. Freeman and Co., San Francisco, 1975.

Kuhn, Thomas Samuel: The Copernican Revolution: Planetary Astronomy in the Development of Western Thought. Harvard Univ. Press, Cambridge, 1957.

Kuhn, Thomas Samuel: The Structure of Scientific Revolutions. (Second ed.), University of Chicago Press, Chicago, 1970.

Kuhn, Thomas Samuel: The Essential Tension: Selected Studies in Scientific Tradition and Change. Univ. of Chicago Press, 1977.

Lakatos, Imre: Falsification and the Methodology of Scientific Research Programmes. In Imre Lakatos and Alan Musgrave, eds., Criticism and the Growth of Knowledge. University Press, Cambridge, 1970, pp. 91-196.

Lakatos, Imre: The Changing Logic of Scientific Discovery, Cambridge Univ. Press, Cambridge, 1977.

Lakatos, Imre: Proofs and Refutations: The Logic of Mathematical Discovery, Cambridge Univ. Press, Cambridge, 1976.

Lehrer, Keith: Induction, Rational Acceptance, and Minimally Inconsistent Sets. G. Maxwell and R. L. Anderson, Jr., eds., Induction, Probability, and Confirmation, University of Minnesota Press, Minneapolis, 1957, pp. 295-323.

Lehrer, Keith: Induction, Reason and Consistency. British Journal for the Philosophy of Science, vol. 21, 1970, pp. 103-114.

Lettvin, J. W. *et al.*: What the Frog's Eye Tells the Frog's Brain. Proceedings of the IRE, vol. 47, Nov. 1959, pp. 1940-1951.

Levine, Marvin J.: A Cognitive Theory of Learning. Lawrence Erlbaum Associates, Hillsdale, New Jersey; Halsted Press, 1975.

Martin, A. R., ed.: Project Daedalus: The Final Report on the BIS Starship Study. British Interplanetary Soc. Suppl., British Interplanetary Soc., London, 1978.

McMullin, Ernan: The Conception of Science in Galileo's Work. R. Butts and J. Pitt, eds., New Perspectives on Galileo, Reidel, Dordrecht, 1978, pp. 209-257.

Mead, George Herbert: The Philosophy of the Act. Charles W. Morris, John M. Brewster, Albert M. Dunham, and David L. Miller, eds., Univ. of Chicago Press, Chicago, 1938.

Mead, George Herbert: Mind, Self, and Society. Charles W. Morris, ed., Univ. of Chicago Press, Chicago, 1934.

Miller, David L.: George Herbert Mead: Self, Language, and the World. Univ. of Texas Press, Austin, 1973.

Minsky, M.: A Framework for Representing Knowledge. Patrick Henry Winston, ed., The Psychology of Computer Vision, McGraw Hill, 1975.

Neisser, Ulric: Cognitive Psychology. Appleton-Century-Crofts, New York, 1967.

Peirce, Charles Santiago Sanders: Collected Papers, Volumes 1-6. Charles Hartshorne and Paul Weiss, eds., The Belknap Press of Harvard University Press, Cambridge, 1960.

Peirce, Charles Santiago Sanders: Collected Papers, Volumes 7-8. Arthur W. Burks, ed., The Belknap Press of Harvard University Press, Cambridge, 1966.

Piaget, Jean: The Psychology of Intelligence. Littlefield, Adams and Co., New Jersey, 1960.

Piaget, Jean: Structuralism. Basic Books, Inc., New York, 1970.

Popper, Karl Raimund: Conjectures and Refutations. Routledge and Kegan Paul, London, 1963.

Raphael, B.: The Thinking Computer: Mind Inside Matter. W. H. Freeman and Co., San Francisco, 1976.

Rescher, Nicholas: Rules of Inference and Problems in the Analyses of Inductive Reasoning. Synthese, vol. 13, 1961, pp. 242-251.

Rock, I.: An Introduction to Perception. Macmillan, New York, 1975.

Rosch, E.; and Mervis, C.: Family Resemblances: Studies in the Internal Structure of Categories. Cognitive Psychology, vol. 7, 1975, pp. 573-605.

Rosenberg, J.: Piaget and Artificial Intelligence. Published in Proceedings of the First Annual Conference on Artificial Intelligence, 1980.

Salmon, Wesley C.: The Foundations of Scientific Inference. Univ. of Pittsburg Press, Pittsburg, 1967.

Schappell, Roger T., et al.: Application of Advanced Technology to Space Automation, NASA CR-158350, 1979.

Selfridge, O. G.: Pandemonium: A Paradigm for Learning. L. Uhr, ed., Pattern Recognition. John Wiley and Sons, New York, 1966, pp. 339-348.

Simon, Herbert A.: The Logic of Rational Decision. The British J. Philosophy Sci., vol. 16, 1965, pp. 169-186.

Skyrms, Brian: Choice and Chance: An Introduction to Inductive Logic. Dickenson, Belmont, California, 1966.
Sperling, George: The Information Available in Brief Visual Presentations. Psychological Monographs, vol. 74, no. 11, 1960, pp. 1-29.
Thayer, Horace Standish: Meaning and Action: A Critical History of Pragmatism. The Bobbs-Merrill Company, Inc., Indianapolis, 1968.
Toulmin, Stephen Edelston: The Philosophy of Science. Harper and Row, New York, 1960.
Toulmin, Stephen Edelston: Foresight and Understanding: An Enquiry into the Aims of Science. Indiana Univ. Press, Bloomington, 1961.
Toulmin, Stephen Edelston: Human Understanding, Volume I. Clarendon Press, Oxford, 1972.
Valdes, Francisco; and Freitas, Robert A., Jr.: Comparison of Reproducing and Non-reproducing Starprobe Strategies for Galactic Exploration. J. British Interplanetary Soc., vol. 33, Nov. 1980, pp. 402-408.
Van Duijn, P.: A Model for Theory Finding in Science. Synthese, vol. 13, 1961, pp. 61-67.
Wisdom, John Oulton: Foundations of Inference in Natural Science. Methuen, London, 1952.

APPENDIX 3A

WHY INTERSTELLAR SPACE EXPLORATION?

The first question a skeptic today might ask is: "Why an interstellar mission?" (fig. 3-11). Twenty years ago many people similarly inquired "Why go to the Moon?" Besides political reasons, there were other goals when the Apollo Program began. For instance, scientists had high hopes for a better understanding of the Earth, the Moon, and the Universe. Yet, although the Solar System is many worlds with countless strange phenomena, still its scientific treasures are miniscule in comparison to those of the Galaxy. Interplanetary space travel is no longer a dream, but a reality – the new dream is interstellar space travel.

Mankind cannot survive forever tied to the cradle of the Earth. In perhaps six billion years our Sun will burn itself out, exhausted of its thermonuclear fuel. But Earth should become uninhabitable long before that. Nuclear war, asteroid collisions, or innumerable other planet-scale disasters could wipe out much of terrestrial life including mankind. The human species remains at risk until humanity extends itself beyond its homeworld. As a young person eventually must leave his parents' home to seek his own path, so must mankind extend its grasp far beyond its ancestral birthplace. Interstellar travel offers the hope of ultimate long-term perpetuation of human life.

We, as a species, possess a deep instinct to survive. Adventure and risk attract many people. It is possible to imagine a manned interstellar mission with all of the above in mind and more, and to dream of life afresh on an alien world with room to grow and a chance for countless new beginnings. Is this really so different from the early settlers who crossed the Atlantic in search of a "New World?"

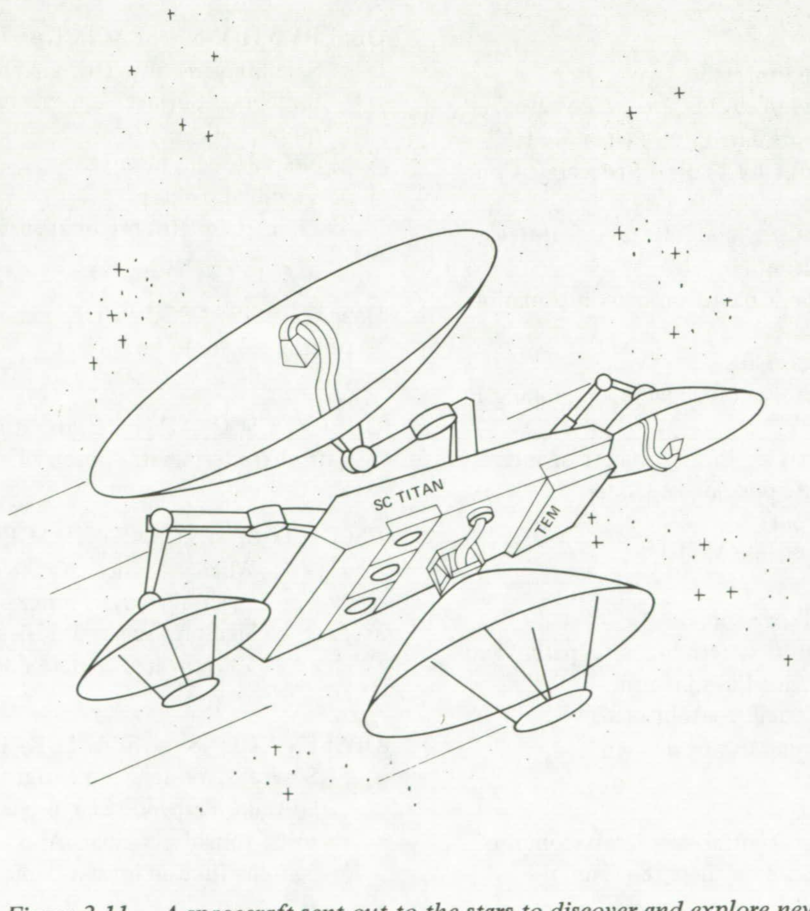


Figure 3.11.– A spacecraft sent out to the stars to discover and explore new worlds.

APPENDIX 3B

EXCERPT FROM HYPOTHETICAL TITAN MISSION LOG

1. Central Spacecraft Computer.
   a. Monitors progress of all operations.
   b. Initializes all tasks.
   c. Keeps log and communicates to Earth.
   d. Makes and tests hypotheses when anomalies from predefined "Planet Model" are found.

2. Spacecraft Imager.
   a. Records images in "snapshot" fashion (on retina-like array of detectors).
   b. Finds features asked for by central spacecraft computer.
   c. Notes anomalies in images from predefined "Planet Model."
   d. Is capable of describing images in terms of predetermined concepts.
   e. Updates "Planet Model" with new information based on image input.

3. Spacecraft Control Image Processor.
   a. Processes data scanned by visible, IR, microwave, and other image sensors. Puts data into "Planet Model."
   b. Performs tests asked for by Central Spacecraft Computer on this data.
   c. Identifies surface features and matches features to concepts stored in "Planet Model."
   d. Updates "Planet Model" based on new information.

4. Lander Central Image Processor.
   a. Main lander vision processor capable of looking in any direction.
   b. Performs scene analysis to locate objects of interest on surface and to locate position of lander.
   c. Retinal-type sensor input.
   d. Adds surface data to "Planet Model."

5. Lander Guidance Image Processor.
   a. Processes image data to determine safe path from present location to assigned destination.
   b. Updates "Planetary Model" contour map.
   c. Finds obstacles on ground during descent.

6. Central Lander Computer.
   a. Handles requests from central spacecraft computer.
   b. Plans lander actions based on these requests.
   c. Assigns tasks to Lander Central Image Processor and Lander Guidance Image Processor.

HYPOTHETICAL SHIP'S LOG – SPACESHIP TITAN July 4, 2010

REPORT: CENTRAL SPACECRAFT COMPUTER (CSCC)

9:00 am GMT – Have Titan in view on spacecraft imager. Based on size of disk at 300 mm focal length, we are 504300 km from satellite. This agrees ±100 km with microwave (5680 GHz) ranging system.

9:10 am GMT – Zoomed to 3000 mm focal length. Approximately 5:30 am local zero meridian time.

OBSERVATIONS – SPACECRAFT IMAGER.

1. Satellite generally dark in appearance with some light blotches. Surface appears smoother than Earth's moon.
2. No polar caps observed.
3. Terminator sharp.
4. Limb at equator bright at surface.

HYPOTHESIS: CSCC – This indicates high-density model of atmosphere to be correct.

QUERY: CSCC – High-density atmosphere does not agree with sharp terminator observed.

CSCC TO SPACECRAFT IMAGER:

Task 1: When at range 20000 km, observe limb using spectral analysis procedure.

Task 2: Measure spectral reflectance over 200 km X 500 km area centered on terminator at equator.

OBSERVATIONS – SPACECRAFT IMAGER (cont.).

5. Several dark areas on surface near limb at 40° north latitude. Perpendicular projection would show these to be roughly circular. Areas very dark in center but lightens (like an inverse conic function) toward edges.

CSCC TO SPACECRAFT IMAGE PROCESSOR:

Task 3: When in orbit, correlate contour map, spectral data, and geological analysis over this area to test for possible volcanic activity.

July 10, 2010

REPORT: CSCC

1:30 pm GMT – Disk size at 300 mm focal length indicates range 20100 km from satellite. Agrees ±100 km with microwave ranging system.

1:35 pm GMT – Terminator observed through 560 nm, 630 nm, and 1080 nm spectral bands. Relative average reflectance 0.2, 1.5, 0.1, respectively, indicating dense atmosphere refraction properties.

QUERY: CSCC – Since dense atmosphere hypothesis was reinforced, why sharp terminator?

CSCC TO SPACECRAFT IMAGER:

Task 4: Cancel Task 1.

Task 5: Measure geological activity in this area during orbit.

July 20, 2010

REPORT: CSCC

7:55 am GMT – Orbit insert completed at 6:00 am. Beginning Titan resource survey for Planet Model. Visible, IR, and microwave scanners operating. Data being processed and stored. Contour map being generated. Gravity anomaly experiment initialized.

8:00 am GMT – OBSERVATION – SPACECRAFT IMAGER

1. White haze in image slightly obscuring surface. Cloud-like rather than bright surface.

CSCC TO SPACECRAFT IMAGE PROCESSOR:

Task 6: Check Planet Model stereo imagery to verify haze above surface.

Response: Spacecraft Central Image Processor.

1. Verifies, white haze is above surface.

Task 7: Check stored spectral data to estimate spectral characteristics of haze.

Response: Spacecraft Central Image Processor.

1. Spectrum of samples areas matches methane.

OBSERVATION – SPACECRAFT IMAGER

1. Presently over black blotches noted on Titan approach.

CSCC TO SPACECRAFT CENTRAL IMAGE PROCESSOR:

Task 8: Initiate Task 3 – Examine topography data over black areas.

Response: Spacecraft Central Image Processor.

1. Contour map shows 3 circular black areas to be conic depressions 10 km, 5 km, and 6 km, respectively, in depth.
2. Spectral data indicates third IR band reflectance to be 1.6 times surrounding average in these areas.
3. Thermal IR indicates 100° K at centers of cones decreasing to 50° K at edges.

HYPOTHESIS: CSCC – Assume black areas to be craters.

CSCC TO SPACECRAFT CENTRAL IMAGE PROCESSOR:

Task 9: a. Analyze area surrounding black regions for height.

b. Examine geological data for lineaments in surrounding area.

Response: Spacecraft Central Image Processor.

1. In annuli ($R_1 < r < 1.1\ R_1; R_2 < R < 1.2\ R_2$; and $R_3 < r < 1.15\ R_3$) surrounding each black area (B_1, B_2, and B_3); depth decreases from inside edge to outside edge by 1 km, 0.5 km, and 2 km, respectively.
2. Radial lineaments exist inside annuli. None outside.

COMMENT: CSCC – Crater hypothesis reinforced, since very warm in center, assume to be volcanic in origin.

CSCC TO SPACECRAFT IMAGER AND CENTRAL IMAGE PROCESSOR:

Task 10: Initiate Task 5 – Analyze geological activity at terminator.

Response: Spacecraft Imager.

1. Texture in grey levels along terminator indicate mountain size shadowing.

Response: Spacecraft Central Image Processor.

1. Contour map indicates mountain range generally parallel to terminator from 70° south latitude.

REPORT: CSCC – Present knowledge indicates mountains can cause sharp cutoff of light. Unless contradictory evidence, assume mountain range at terminator causes observed sharp cutoff of light at terminator.

July 29, 2010

REPORT: CSCC

8:25 am GMT – Initial Planet Model completed. Atmospheric sounder indicates planned descent to surface procedure possible. Beginning procedure for finding highest priority landing target.

CSCC TO SPACECRAFT CENTRAL IMAGE PROCESSOR:

Task 51: Correlate parameters for most desirable landing site with surface model.

Response: Spacecraft Central Image Processor.

1. Highest priority areas – Those indicating possible carbon-based life or structures made by intelligent beings. – None indicated by Planet Model.
2. Next highest priority – Areas indicating possible H_2O. No indication of present H_2O by surface model. Possible ancient riverbed 25°-27°N. latitude and 34°-38° W. longitude. Topology map indicates possible target 25.1° N. latitude and 36.8° W. longitude. Target is within lander range of stratified river bank-like structures.

July 30, 2010

REPORT: CSCC

9:45 pm GMT – Have loaded site model into Lander Central Image Processor. Task sequences programmed to deploy parachute at 100 km, begin wind determination and parachute descent path modification at 45 km. Parachute to be ejected at 2 km. Landing rockets and obstacle avoidance imaging system to be initiated at 1500 m.

10:31 pm GMT – Initiating Lander Descent.

Response: Lander Central Image Processor.

1. Have locked onto target area.
2. Estimated ground drift and surface wind indicates parachute descent direction modification of 265° and 3.1 km necessary to hit target.
3. Obstacle avoidance system activated.

Response: Lander Guidance Image Processor.

1. Obstacle at site. Shift 30°, 0.16 km.
2. Site clean under lander, okay for vertical descent to surface.

July 31, 2010

REPORT: CENTRAL LANDER COMPUTER (CLC)

1:38 am GMT – Lander site assessment procedure initiated. Lander Guidance Imaging System turned on.

Response: Lander Central Image Processor.

1. Surface immediately surrounding lander mostly small rocks on relatively flat surface. Hill (slope < 30°) blocking view beginning 100 m away 248° to 0°. Surface of hill easily navigable. Stratified rock wall beyond 1 km 0° to 20°. Hill obscures wall beyond 0°.

CLC TO LANDER GUIDANCE IMAGE PROCESSOR.

Task 1: Initiate analysis to find safe path to climb to apex.

Response: Lander Guidance Image Processor.

1. Stereo depth and contour data added to site model.
2. Safe path calculated.
3. Initiating journey.
4. Apex of hill reached. Stereo depth and contour data being added to site model.
5. Safe path possible in forward direction.

OBSERVATION – LANDER CENTRAL IMAGE PROCESSOR.

1. Rock formation indicating upheaval at 240°.
2. No major obstacles indicated on Planet Model.
3. No major obstacles indicated in image pointed at 240°

CLC TO LANDER GUIDANCE IMAGE PROCESSOR.

Task 2: Initiate analysis to find safe path to formation.

Response: Lander Guidance Image Processor.

1. Safe path calculated for initial 100 m.
2. Beginning journey.
3. Dead reckoning from surface model and relative size indicates 0.5 of total distance covered.
4. Dead reckoning and size of upheaval in image indicates 300 m from upheaval.
5. Slowing down.
6. At base of upheaval. Rubble makes further progress in this path impossible.

CLC TO LANDER GUIDANCE IMAGE PROCESSOR.

Task 3: Initiate experiment No. 4379 – Rock specific density.

Response: Lander Guidance Image Processor.

1. Reconfiguring to manipulator vision configuration.
2. Located oval shaped rock 3 cm X 8 cm, not imbedded, within reach of manipulators.
3. Surface model for 0.5 of rock recorded.
4. Initiating manipulator to lift and weigh rock.
5. Rock weighs 15 N, Mass = 10 kg.
6. Initiating rotation of rock 180° with respect to initial position.
7. Surface model for remaining 0.5 of rock recorded.
8. Volume of rock is 0.010 m^3. Density is 1000 kg/m^3.

APPENDIX 3C

ILLUSTRATIVE HYPOTHESIS FORMATION SCENARIO

The scenario presented in table 3.6 suggests the great complexity of data processing and hypothesis generation involved in solving problems in a planetary investigation conducted by a fully-autonomous spacecraft. Table 3.6 shows a simulated report based on studies of the Martian oases following the Viking mission to the Red Planet in 1976 (Huguenin, 1978). Without delineating all logical functions required to arrive at the final suggested hypothesis, the series gives the reader a feel for the many steps involved in full-fledged scientific analysis of a new situation. It is presented in the format of a condensed message sent to Earth via statements which indicate what measurements were made and confirmed, and what reasoning was used to draw specific conclusions.

TABLE 3.6.– HYPOTHESIS FORMATION: AN ILLUSTRATIVE SCENARIO

"Condensates appear suddenly at dawn in two different locations: (-25°, 85°) and (-30°, 315°)."

"Condensates act to flatten the reflectance spectra and appear as highly reflecting at blue wavelengths."

"Blue cloud activity occurs during southern fall and winter, and mixed blue and yellow cloud activity occurs during spring and summer. Yellow clouds are indicative of dust. (This information will turn out not to be included in final hypothesis.)"

"In these two locations, condensates are brightest at dawn, indicating low-level hazes or frosts. If the brightening of the condensates occurred in the afternoon, convective cloud activity would be indicated."

"Both areas are also major centers of dust storms, the clouds typically appearing suddenly at dawn and fading from white to yellow by noon."

"Local winds are not sufficient to make airborne particles of the size observed at these sites."

"There appears to be a correlation between the time and the location of appearance of both the condensate hazes and the dust storms."

"Since winds cannot produce the dust storms, the hypothesis is that rapid evaporation just after sunrise of any water present in the soil is explosively ejecting dust particles into the surface atmosphere. The explosive action is the result of low ambient surface pressures and atmospheric densities."

3C.1 References

Huguenin, R. L.; *et al.*: Mars: Infrared Spectral Reflectance and Compositional Implications. J. Geophysical Research, vol. 83, no. B11, Nov. 1978, pp. 5433-5441.

CHAPTER 4

NONTERRESTRIAL UTILIZATION OF MATERIALS: AUTOMATED SPACE MANUFACTURING FACILITY

4.1 Introduction

The heavens have always been the subject of intense curiosity and longing, beckoning to our imaginations and, sometimes, to our desires for dominion over that which is not yet under human control. Recent American space exploration efforts represent only tentative steps toward increased human understanding of the Universe – indeed, lunar and planetary missions often have raised more questions than they have answered. Those in the forefront of space sciences believe that the ultimate horizons are as yet only dimly perceived. A substantial minority of the American public would like to see more effort devoted to deeper investigations of the planets, stars, and galaxies beyond Earth.

However, most people remain unconvinced that expanded activities in space gained apparently at the expense of other societal goals are worth the price (Overholt *et al.*, 1975). Clearly, future large-scale American space projects should embody a fundamentally new perspective – an overall shift from the existing policy of (primarily) exploration to one of integrated and direct utility for mankind. Such a pragmatic space utilization program may demand extensive use of nonterrestrial materials and an ever-increasing dependence on automation in all its dimensions. The Nonterrestrial Utilization of Materials Team has explored the need for such a program, and has laid the foundation for future NASA technology planning by examining in some detail a space project having the potential for physical growth with continually decreasing net materials import from Earth.

4.1.1 Objectives

A principal objective of the present study is to develop scenarios which show how, starting from current plans and capabilities, an extraterrestrial facility providing economic benefits for humanity can be established, maintained, and expanded in the near future. Ultimately, this permanent orbital factory will be dependent in large degree on nonterrestrial materials and autonomous robots programmed for advanced machine intelligence. The principal thrust of chapter 4 is to demonstrate the relationship between nonterrestrial utilization of materials and the growth of an orbital manufacturing facility beginning with a minimal "starting kit" of machines. The kit performs basic manufacturing processes necessary for facility expansion and the creation of a widening spectrum of the means of production.

This goal was chosen because only through the development of extraterrestrial resources can future space activities be pursued independently of terrestrial resource limitations and management constraints. The proposed scenarios for space facility maturation are essentially open-ended, so a variety of exploration and utilization options continuously become available once initial economic productivity is established.

The basic requirement explored in this approach to space industrialization is the establishment of two off-Earth facilities, one in space and one on the Moon. In this scenario, an Earth-orbital base will provide logistical support and production capability necessary for the creation of useful end products and its own expansion. A space platform should be established early, initially dependent on the Space Shuttle, as a demonstration of "starting kit" operation and advanced production methods. However, rapid factory growth necessitates the use of lunar or asteroidal resources. Therefore, a lunar processing and manufacturing facility (Dalton and Hohmann, 1972), possibly self-replicating (see chapter 5), is presupposed in the growth scenario.

The availability of nonterrestrial materials could make possible a decreasing dependence on Earth-based supplies. Growth of the Space Manufacturing Facility (SMF) subsequently would require no major additional Earth resource inputs. Given a supply of sufficiently inexpensive nonterrestrial materials, SMF output could be returned to Earth directly in the form of useful commercial products or indirectly in the form of solar power generation or satellite servicing.

The present analysis is explicitly guided by the goal of maximizing the use of automation and robotics during expansion of the processing and manufacturing facilities. Even in the early stages many operations can be conducted by remote teleoperators. As research in robotics continues, more and more system functions will be taken over by autonomous robots. While it is unlikely that the human presence soon can be completely eliminated, economic arguments favoring SMF deployment require the assumption of an increasing use of autonomous machine systems.

4.1.2 *Rationale for the Utilization of Space*

The American push into space, never fully backed by the public, appears in recent years to have slipped even lower on the list of national priorities (Lowman, 1975). The current unwillingness by the political leadership of the country to support space activities is reflected in the weakened budgetary position of NASA, the prime driving force behind the United States civilian space program. Major reasons for this lack of support include:

- A policy of piecemeal exploration
- An emphasis on limited-duration, "one-shot" projects
- Indirect rather than direct benefits achieved by space missions
- The view of human welfare as a byproduct rather than an explicit goal of space activity
- Too great an emphasis on purely scientific benefits
- "Selling" space to particular interest groups with insufficient regard for immediate public interests
- Too little public input in NASA planning.

The weak interest in NASA programs is, however, correctable. It must be established that major future NASA programs will be explicitly tied to the public welfare and that concrete, short-range benefits for individual members of society can and will be achieved. This may be accomplished either by demonstrating that an immediate threat to the American way of life can be averted through the implementation of a particular space program, or by showing that a mission will have a visible economic payback to the public.

Unquestionably the first method has the best chance of loosening the legislative purse strings. Indeed, the strongest public and legislative interest in space was expressed during the Sputnik crisis in 1958 (Overholt et al., 1975). A recent Woods Hole conference concluded that potential triggers for renewed efforts in space might be crisis-based. Among eighteen possibilities listed by panel participants were such events as impending asteroid collision with Earth, rapid changes in the polar icecaps, discovery of extraterrestrial life, some major accomplishment in space by another nation, or a credible military threat. Unfortunately, none of these possibilities suggest a positive planning process since by their very nature they occur unexpectedly (Sadin, private communication, 1980).

One externally generated crisis once thought to provide impetus for further space activities was the prediction by the "Club of Rome" of an impending shortage of critical terrestrial raw materials (Meadows *et al.*, 1972; Laszlo *et al.*, 1977). Subsequent researchers found significant flaws in the study, detracting from the immediacy of the threat (Kahn, 1976; Science Applications, Inc., 1978) and eliminating an impending world food crisis as a major space mission driver.

Still, it must be recognized that "need" is a relative term. For instance, a country (such as the United States) fundamentally committed to economic growth and vitality can find its horizons of economic "need" closing in much faster than, say, a global community committed only to survival. This public perception may inspire a recognition of the connection between the profitability of space ventures and the impending decline of a way of life. The issue of "need" thus reduces to the question of how best to utilize both terrestrial and nonterrestrial resources to avert a fundamental threat to the American standard of living.

Consonant with the above motivational framework, major future space missions must be clearly directed toward the utilization of space for the distributive benefit of the American public, and be designed to avert erosion of national living standards. In addition, the existing economic climate of the U.S. must be taken into consideration: Each project must show a near-term, growing productive capability; it must take appropriate measure of national priorities; it cannot rely too heavily on capital investment; and, finally, national leaders and the public must perceive it as directly beneficial to their own interests. The proposed Space Manufacturing Facility is designed to meet each of the criteria established above.

New resources. The SMF mission utilizes resources not presently available for the clear and direct benefit of the American public. This benefit may include (1) construction of solar power satellite stations to generate energy for Earth, (2) manufacture of useful products on the Moon for terrestrial use predominantly from lunar materials, (3) eventual production of consumer goods in the SMF for Earth, employing the unique qualities of the space environment plus lunar or asteroidal materials, (4) utilization of processes unsuitable, unsafe, or otherwise desirable for application on Earth, and (5) using the SMF as a springboard for further space resource exploration and industrialization.

Near-term growing productive capability. The Space Manufacturing Facility is intended to take full advantage of past, current, and future research in machine intelligence and robotics. Technological enablers now exist in automation, space transportation, and in the results from lunar research. Present competition in industrial robotics is intense, and rapid beneficial developments might be expected to occur even without NASA funding. Serious exploitation of robotics technology in an SMF scenario, however, will accelerate development and permit a growing productive capability from which immediate, near-term human benefits can be siphoned off. The proposed project is open-ended: Growth in productivity is expected with concurrent multiplication of the range of capabilities available, without infusing large amounts of additional capital.

Capital investment. The primary investment is for the establishment of two starting facilities, one on the Moon

and one in Earth orbit. It is anticipated that near-closure (see chapter 5) will be achievable and that minimal human presence will be necessary. Interaction between lunar and Earth-orbiting components allows growth materials required by the orbital module to be supplied from lunar sources, thus greatly reducing supply costs. A major gain with respect to capital investment and production costs is that fewer materials must be flown up from Earth and that almost all the required on-site labor can be performed by automata.

Distributive benefits. Certainly the SMF generates a number of indirect benefits for the public. It opens new horizons of knowledge, advantages American industry in international competition, provides new technologies, and reasserts the U.S. position of leadership in space. However, the public relates only vaguely to such interests, if at all. The establishment of a solar power satellite, on the other hand, is of more direct and tangible value. This kind of SMF product could have direct impact on energy costs now borne by the public and could lead to a visible decrease of dependence on foreign energy supplies.

Standards of living and public perceptions. If the capital investments required are accounted for, the proposed mission can help to stabilize the American standard of living and eventually permit it to continue to rise. Energy scarcity is widely perceived as the root cause of current economic difficulties, a viewpoint stressed repeatedly by the media. Rampant inflation and unemployment, justly or unjustly, are traced directly back to the cost of energy. Recently, however, it has become increasingly apparent that the issue is not simply energy supply but also energy cost. Given the education the public already has received, it should not require too much additional effort to make people aware that their own short- and long-term interests are well-served by the SMF. The poor economic climate actually may prove an added fiscal impetus for the mission rather than a restraint.

4.1.3 Summary of Chapter Contents

The study team focused its efforts on four areas related to the nonterrestrial utilization of materials:

- Material resources needed for feedstock in an orbital manufacturing facility (section 4.2)
- Required initial components of a nonterrestrial manufacturing facility (section 4.3)
- Growth and productive capability of such a facility (section 4.4)
- Automation and robotics requirements of the facility (section 4.5)

Section 4.2 presents an overview of energy and mass available in the Solar System, with special attention to those resources which may be available in the near future and to possible space materials processing techniques. A lunar-to-LEO shuttle system utilizing silane fuel and an Earth-based electromagnetic catapult are possible candidates for the transportation of raw materials and feedstock to low Earth orbit.

Scenarios for establishing an initial orbiting manufacturing facility are developed in section 4.3. To provide some basis for determining the minimum number and types of machines which might be available for space manufacturing and for constructing an automated shop capable of creating additional industrial equipment, a survey of basic manufacturing processes was performed by the team. "Starting kits" were conceptualized which might be useful in creating an ever-widening set of manufacturing devices requiring minimal initial inputs and using solar energy, vacuum, zero-gravity, and robotics to best advantage.

Section 4.4 demonstrates the growth and production potential of the Space Manufacturing Facility using the material resources and starting kits described earlier. Near-, mid-, and long-term examples of product manufacture are developed. These outputs, including Shuttle external tank conversion to simple structures (near-term), electronics components fabrication (mid-term), and the creation of space platforms, pure glasses, satellites, and robots (long-term), are presented as representative samples of SMF growth possibilities.

Section 4.5 concentrates on mission automation and machine intelligence requirements for an SMF. Limitations and functional demands of robotics in space are detailed, with recommendations for future machine intelligence developments. Mission technology drivers in major areas other than automation and machine intelligence are briefly summarized. Finally, section 4.6 provides a general discussion of the implications for society, potential consequences, and necessary sociocultural and political prerequisites for implementation of a space manufacturing mission.

4.2 Materials Background

A survey of Solar-System resources available to mankind in the near-, mid-, and distant-future is appropriate in evaluating the potential of the SMF mission concept. Such background is necessary to identify terrestrial and lunar resources, asteroidal materials, and various additional sources for space manufacturing feedstock. This section describes existing chemical extraction and materials processing alternatives including one new option identified during the course of the study (large-scale electrophoretic lunar soil processing) and expanded possibilities for the metallurgy of native lunar basalts, followed by a consideration of materials transport both from the Moon to low Earth orbit using silane-based propellants derived in part from lunar materials, and from the surface of the Earth to LEO using a ground-based electromagnetic catapult (Mongeau et al., 1981).

4.2.1 *Survey of Solar System Resources*

A survey of extraterrestrial resources reveals a number of major stores of energy and raw materials within the Solar System. Ultimately the most significant of these is the Sun itself. Total solar output is 4×10^{26} W, approximately 6×10^{13} times as much as mankind produced on Earth in 1980. An extremely power-intensive (15 kW/person) world society of 10 billion people drawing its materials resources solely from the common minerals of the Earth's crust would require only a trivial fraction (4×10^{-11}) of the available solar output (Goeller and Weinberg, 1976).

It is especially significant that the mass of capital equipment required to produce a unit of useful solar power in space is very low. It is anticipated that large-scale solar thermal power stations can be built for 0.1-1 metric ton equipment per megawatt (t/MW) and 1-10 t/MW for solar electric power. These estimates are calculated for 1 AU (i.e., Earth-orbital distances) from the Sun. Alternative terrestrial mass/power ratios are much larger – hydroelectric plants, 10^3 to 10^4 t/MW; projected nuclear fusion power stations, 10^3 t/MW; coal-fired plants, 2×10^2 t/MW (with 4000 tons of coal consumed per MW/yr); and terrestrial (ground-based) solar power, more than 10^3 t/MW. Thus, energy systems in space can grow at much faster rates using nonterrestrial materials than is possible on Earth. Energy payback times (time for recovery of initial energy investment) for construction of heliocentric orbital systems at 1 AU is on the order of 10 days. The intensity (I) of solar power varies inversely with the square of the radius (R) from the Sun ($I/I_0 = R_0^2/R^2$; $I_0 = 1.4$ kW/m^2, $R_0 = 1$ AU $= 1.54 \times 10^{11}$ m), so space energy systems may be operated at least out to the distance of Saturn (about 10 AU) before capital/energy efficiency ratios (measured in t/MW) approach values comparable to alternative terrestrial power systems. This is because very low mass optical reflectors can be used to concentrate the available sunlight.

Other power sources which eventually may become accessible to mankind include the kinetic energy of the solar wind (10^{14} MW); differences in the orbital and rotational energies of the Sun, planets, moons, and asteroids (perhaps allowing payloads to move between these bodies) (Sheffield, 1979); and the thermodynamic energies associated with the differentiation of chemical elements in planetoids across the Solar System. Tidal dams on Earth, terrestrial rocket launches, and space probe gravitational swing-bys have utilized trivial fractions of the potential and kinetic energies of the Earth and Moon, and Mercury, Jupiter, Saturn and their moons.

An appreciation of the magnitude of accessible matter resources in the Solar System is gained by noting that terrestrial industry processed about 20 billion tons (about 10 km^3) of nonrecoverable materials in 1972. (Annual tonnages of chemical elements used industrially are listed in table 4.1.) It can be estimated that humanity has processed slightly less than 10^{12} t (about 500 km^3) of nonrenewable materials since the start of the Industrial Revolution four centuries ago, assuming a 3% annual growth rate. For comparison, a 4.3 km-radius spherical asteroid (density 1.5 t/m^3) also contains 10^{12} t of matter. Thousands of asteroids with masses in excess of 10^{12} t already are known (Gehrels, 1979). Approximate total mass of the known minor planets is 2×10^{18} t, the moons 7×10^{20} t, and meteoritic and cometary matter roughly 10^{12} t. The planets have a total mass of 2.7×10^{24} t.

Mankind has launched about 5000 t into LEO since 1959. Most of this was propellant for Apollo lunar missions and for satellites hurled into geosynchronous orbits or into deep space. Approximately 1000 t was hardware. Averaged over the last 10 years, humanity has ejected mass from Earth at approximately 0.05 t/hr or 400 t/yr. Waldron *et al.* (1979) estimate that oxygen and possibly most of the fuel (silane based) for liquid-propelled rockets can be produced from lunar soil using chemical processing plants with intrinsic capital mass of 100 t/(t/hr) of output. Thus, a LEO propellant production plant weighing about 10 t could service all current major needs if provided with lunar materials. At some point in the future, major mass fractions of space facilities may be constructed of nonterrestrial matter. Space hardware should be produced in orbit at the rate of 10 kg/hr to match the 1970s and anticipated 1980s launch rates.

The United States Space Transportation System (STS), popularly known simply as the "Shuttle," is expected to establish approximately the same mass/year launching ratio during the 1980s at a cost of about \$1000/kg to LEO. Energy represents only a small fraction of this expenditure. Perfectly efficient conversion of \$0.05/kW-hr electricity into LEO orbital energy (about 10 kW-hr/kg) would cost roughly \$0.50/kg for materials transport to orbit, a factor of 2000 less than near-term STS lift prices. Projected bulk transport versions of the STS may lower Earth-to-LEO expenses to \$100/kg; still some 200 times greater than the equivalent cost of electrical energy at present-day rates. When launch charges reach \$1/kg a large Earth-to-LEO traffic becomes reasonable, since most terrestrial goods are valued at \$1-2/kg (Ayres *et al.*, 1979). However, if space industry someday is to approach cost distributions typical of terrestrial industries, then the supply of bulk or raw materials from the Moon and the asteroids must fall to a few ¢/kg (Criswell, 1977a, 1977b). On an energy basis alone, this goal appears achievable using high throughput lunar-mass drivers and relatively cheap solar energy (O'Neill, 1974).

Atmospheres of the various planets and moons are valuable as sources of materials and for nonpropellant braking of spacecraft (Cruz *et al.*, 1979). Deliveries of

TABLE 4.1.– COMPILATION OF AVERAGE COMPOSITION OF LUNAR SOILS FOR 80 ELEMENTS[a]

	Mare					Highland		Basin ejecta		
	High Ti		Low Ti							
	A-11	A-17	A-12	A-15	L-16	A-16	L-20	A-14	A-15	A-17
Al_2O_3, %	13.78	10.97	13.71	10.32	15.51	27.18	23.07	17.41	17.54	20.60
CaO, %	12.12	10.62	10.55	9.74	12.07	15.79	14.07	10.79	11.57	12.86
Cr_2O_3, %	.30	.46	.35	.53	.29	.107	.15	.22	.28	.26
FeO, %	15.76	17.53	15.41	19.75	16.41	5.18	7.35	10.36	11.58	8.59
K_2O, %	.15	.076	.27	.10	.10	.11	.08	.58	.17	.16
MgO, %	8.17	9.62	9.91	11.29	8.79	5.84	9.26	9.47	10.41	10.29
MnO, %	.21	.24	.22	.25	.21	.065	.11	.14	.16	.11
Na_2O, %	.44	.35	.48	.31	.36	.47	.35	.70	.42	.41
P_2O_5, %	.12	.07	.31	.11	.14	.12	.11	.50	.16	.14
SiO_2, %	42.47	39.87	46.17	46.20	43.96	45.09	44.95	48.08	46.59	45.08
TiO_2, %	7.67	9.42	3.07	2.16	3.53	.56	.49	1.70	1.32	1.62
Al, %	7.29	5.80	7.25	5.46	8.21	14.38	12.20	9.21	9.28	10.90
Ca, %	8.66	7.59	7.54	6.96	8.63	11.29	10.06	7.71	6.27	9.19
Cr, %	.21	.31	.24	.36	.20	.07	.10	.15	.19	.18
Fe, %	12.25	13.63	11.98	15.35	12.76	4.03	5.71	10.36	9.00	6.68
K, %	.12	.063	.22	.08	.08	.09	.066	.46	.14	.13
Mg, %	4.93	5.80	5.98	6.81	5.30	3.52	5.59	5.71	6.28	6.21
Mn, %	.16	.19	.17	.19	.16	.050	.085	.11	.12	.085
Na, %	.33	.26	.36	.23	.27	.35	.26	.52	.31	.30
O, %	41.6	39.7	42.3	41.3	41.6	44.6	44.6	43.8	43.8	42.2
P, %	.05	.03	.14	.05	.06	.05	.05	.22	.07	.06
S, %	.12	.13	.10	.063	.21	.064	.08	.088	.08	.06
Si, %	19.84	18.63	21.57	21.58	20.54	21.07	21.00	22.46	21.77	21.06
Ti, %	4.60	5.65	1.84	1.29	2.11	.34	.29	1.02	.79	.97
Ag, ppb	9.0	9.8	62.0	50.0	95.0	9.6	16.2	17.5	56.0	6.5
Ar, ppm	1.0	1.2	.3	.7	—	1.2	—	1.0	—	—
As, ppm	.32	—	.082	.010	.41	.14	.28	.066	—	—
Au, ppb	3.7	2.5	2.5	2.11	2.5	8.47	4.93	6.7	3.3	4.9
B, ppm	3.5	2.0	9.3	—	4.3	5.9	39.0	19.0	—	—
Ba, ppm	140	85.7	413	122	215	127.3	89.6	767.5	279	190
Be, ppm	2.0	—	5.0	1.31	2.2	1.2	—	5.5	2.8	—
Bi, ppb	1.5	7.7	1.5	.36	4.9	1.8	2.7	1.7	.17	—
Br, ppm	.239	.093	.165	—	.21	.217	.13	.41	.06	—
C, ppm	135	82	104	95	—	106.5	—	130	125	155
Cd, ppm	.045	.032	.046	.062	.80	.097	.048	.181	.042	.04
Ce, ppm	50.0	25.3	104.0	31.4	33.4	30.3	20.5	185.0	54.0	46.0
Cl, ppm	30.2	5.7	31.0	7.6	53.5	20.9	13.0	44.0	5.9	—
Co, ppm	32.0	35.0	43.0	54.4	37.0	25.3	40.5	35.8	42.0	33.0
Cs, ppm	.18	.30	.30	.23	.95	.11	.11	.63	.19	.18
Cu, ppm	11.5	11.0	10.3	8.2	31.0	8.26	19.0	11.1	7.9	6.4
Dy, ppm	20.2	12.2	24.6	8.6	10.9	6.8	5.0	39.0	13.6	11.0
Er, ppm	11.5	7.90	15.35	5.13	6.3	4.39	2.5	23.5	7.86	6.5
Eu, ppm	2.0	1.66	1.9	1.01	2.3	1.23	.98	2.64	1.30	1.35
F, ppm	278	—	132	45	242	72	37	219	60	—
Ga, ppm	4.3	7.5	4.3	4.43	4.4	4.5	3.7	6.8	3.6	4.7
Gd, ppm	16.3	11.4	25.7	8.1	9.8	6.7	3.06	34.8	11.74	10.07
Ge, ppm	1.0	.198	.32	.17	1.44	.76	.46	.70	.42	—
H, ppm	51.0	59.6	45.0	63.6	—	56.0	—	79.6	52.0	98.0
He, ppm	60.0	36.0	10.0	8.0	—	6.0	—	8.0	—	—
Hf, ppm	8.9	7.3	12.7	5.2	4.75	3.9	2.9	22.2	7.6	5.5
Hg, ppm	.015	—	.023	—	—	.004	—	—	—	—
Ho, ppm	5.4	—	5.3	1.7	2.5	1.50	.88	7.8	3.3	—
I, ppb	—	2.0	—	—	—	5.6	12.0	—	35.0	—
In, ppb	—	2.4	90.0	3.4	35.6	31.0	19.0	89.0	7.6	3.4
Ir, ppb	7.8	5.4	5.6	3.1	9.7	12.4	9.5	12.4	8.3	8.8
La, ppm	17.3	7.32	38.8	11.3	11.5	11.7	7.6	69.4	24.0	16.9
Li, ppm	16.5	9.77	19.5	9.09	9.7	7.4	5.7	29.8	10.8	11.7
Lu, ppm	1.6	1.03	1.93	.72	.84	.59	.40	3.10	.98	.88
Mo, ppm	.70	—	.34	—	—	.34	—	—	—	—
N, ppm	119	60	84	80	134	89	107	92	190	81
Nb, ppm	15.8	19.1	34.0	13.0	15.9	12.8	12.0	56.0	16.0	18.0
Nd, ppm	42.6	23.0	75.6	23.0	26.9	19.3	10.8	105.0	35.0	27.6

TABLE 4.1.– CONCLUDED

	Mare					Highland		Basin ejecta		
	High Ti		Low Ti							
	A-11	A-17	A-12	A-15	L-16	A-16	L-20	A-14	A-15	A-17
Ne, ppm	5.0	2.0	2.0	2.0	---	1.0	---	2.0	---	---
Ni, ppm	206	131	189	146	174	345	208	321	282	286
Os, ppb	14.0	---	6.0	1.79	30.0	22.0	---	---	---	---
Pb, ppm	2.9	.80	4.8	1.033	6.0	2.58	1.15	10.02	2.5	1.922
Pd, ppb	21.0	---	9.7	6.2	---	24.0	---	50.0	---	---
Pr, ppm	7.7	---	10.1	3.8	---	4.97	4.0	23.0	---	---
Rb, ppm	3.0	1.2	7.28	2.70	1.85	2.48	1.65	15.25	5.0	4.21
Re, ppb	5.26	.47	.34	.39	.36	.82	3.19	1.15	---	---
Rh, ppm	.1	---	.4	---	.077	---	---	---	---	---
Ru, ppm	.6	---	.047	---	.046	.010	---	---	---	---
Sb, ppb	4.1	25.4	47.0	30.0	3.8	9.7	5.7	3.4	---	26.0
Sc, ppm	62.8	65.0	39.2	37.1	39.9	8.9	17.0	21.9	22.0	18.0
Se, ppm	.39	.27	.30	.18	.39	.24	.30	.031	---	.23
Sm, ppm	11.7	8.0	20.3	5.85	8.8	5.38	3.39	30.9	9.6	8.1
Sn, ppm	.7	---	.3	---	1.7	.22	.8	---	---	---
Sr, ppm	193.0	166.0	138.9	104.2	234.0	168.0	140.8	183.8	152.0	150.0
Ta, ppm	1.5	---	1.58	.55	1.4	.50	.50	4.1	1.05	.87
Tb, ppm	3.3	2.63	4.07	1.4	1.5	1.07	.80	6.4	4.2	1.72
Te, ppm	.07	.01	.05	---	.088	.023	.051	.031	---	---
Th, ppm	2.24	.82	6.63	1.76	1.07	1.87	1.44	13.5	4.15	3.01
Tl, ppb	2.1	1.4	2.0	.94	1.6	7.7	6.2	22.0	---	2.4
Tm, ppm	1.5	---	2.02	---	.73	.67	.41	3.9	---	---
U, ppm	1.37	.26	1.61	.483	.300	.52	.45	3.48	.99	.90
V, ppm	66	128	110	191	73.5	25.5	38	49	84	52
W, ppm	.24	.14	.74	.31	---	.31	---	1.9	---	.52
Y, ppm	107	74	145	47	48	39.3	49	242	73	64
Yb, ppm	10.6	7.48	13.7	4.53	5.59	3.86	2.40	22.7	7.3	6.15
Zn, ppm	23.0	49.0	6.3	12.8	25.0	24.0	34.1	28.0	14.5	20.0
Zr, ppm	331	236	503	175	308	163.8	192	842	278	262

[a]Major elements (>0.1%) are reported first as both the usual oxide notation and elements. Data compiled from the Data Base Compilation of the Lunar Sample Curator, NASA Johnson Space Center, Houston, Texas.

propellants and fabricated parts to space from Earth may be sharply reduced by making full use of local (nonterrestrial) materials, energy, and linear and angular momentum.

Terrestrial materials. Progressive developments of more efficient Earth-to-LEO boosters are expected to reduce transport costs eventually to at least $10-20/kg, comparable to the price of transoceanic air travel (Akin, 1979). The major tradeoff is between development costs of new launch systems and rates of transport in t/yr. Thus, Earth-to-LEO shipment of higher-value products (above $10/kg) needed in low annual tonnages is acceptable and should not seriously restrict the growth of space industries (Criswell, 1977a, 1977b). Space manufacturing directly leverages the effectiveness of any system for transporting goods and materials off-Earth if the value added to the space products is less than the value added by launch of functionally similar goods from Earth (Goldberg, 1981).

STS components such as exhausted hydrogen/oxygen propellant tanks can be used for raw materials. Shuttle external tanks could provide approximately 140 kg/hr of aluminum and 10 kg/hr of other elements (e.g., plastics, residual propellants) for early development of manufacturing procedures and products, assuming 30 Shuttle flights per year. (See sec. 4.4.2.)

Earth's upper atmosphere also may prove a valuable source of nitrogen and oxygen for use at LEO and beyond. At 200 km altitude a scoop 1 km in radius oriented perpendicular to the orbital motion intersects approximately 4 t/hr of molecular nitrogen and 3 t/hr of atomic oxygen. Physical convergent nozzles might be used to collect either N_2 or O^+, and a convergent magnetic field might be employed to recover O^+. Power must be supplied to liquefy the gases and to accelerate a portion of the gathered material to maintain orbital velocity.

Lunar resources. Table 4.1 lists major oxides and elements found in samples of the mare and highland areas of the Moon and returned to Earth during the Apollo and Soviet programs. Table 4.2 summarizes the major lunar minerals and the general uses to which each could be put (Arnold, 1977). The Moon is extremely rich in refractories, metals (Fe, Mg, Ti, Al), oxygen and silicon. Extensive

TABLE 4.2.– TYPICAL LUNAR RESOURCE AVAILABILITY

Material	Representative uses	Source	Source material concentration	Beneficiation and processing considerations	Abundance and occurrence
Regolith, not chemically or mechanically separated	Reaction mass, radiation shielding, thermal shielding, spun glass, sintered building material	Regolith	100% of surface material	Handling of dust, excavating	Ubiquitous
Basalt, not chemically separated	Cast basalt for construction	Basaltic flows into maria	100% of subregolith and scattered fragments	Hard rock	Abundant in maria
Nonmetallics	Construction materials, special uses	Plagioclase and processing by-products	70 to 95% of highlands anorthositic rocks; 10 to 40% in mare basalts	Use anorthositic regolith or crush friable anorthosite; basalt is generally tough	Abundant in highlands
Al, Al_2O_3, Ca, CaO, Na, Na_2O, Si, SiO_2, O_2	Metals for construction, ceramics, solar cells, reactants for chemical processing, life support	Plagioclase	70 to 95% of highlands anorthositic rocks; 10 to 40% in mare basalts	Use anorthositic regolith or crush friable anorthosite; basalt is generally tough	Abundant in highlands
Fe, FeO, Ti, TiO_2, O_2	Metals, pigments, life support, special uses	Ilmenite	2 to 20% in mare basalt and mare regolith	Size separation of regolith to concentrate ilmenite	Abundant in maria
Mg, MgO, Fe, FeO, Si, SiO_2, O_2	Metals, ceramics, solar cells	Olivine	0 to 20% in mare basalt; 95% in dunite	Difficult to separate from basalt	Dunite is rare in sample collection, as breccia clasts
H_2, H_2O	Life support, fuels	Cold-trapped volatiles at lunar poles	Unknown	Significant technological development required	Occurrence has not been demonstrated
H_2, C, N	Life support, organics	Solar wind trapped in regolith and soil breccia and buried possibly in polar cold traps	100 ppm in mature regolith and soil breccia	Direct thermal extraction; concentration of ilmenite or <60-μm fraction enhances yield	Ubiquitous, but low grade
Zn, Pb, Cl, S, F, other volatile elements	Industrial materials	Surface deposits on volcanic spherules and regolith fines	5 to 100 ppm concentrated at surfaces; may be higher locally	Requires technique development for low-grade extraction	Two known sources; others possible
P, Zr, F, Cl, Y, Cr		Major components in accessory minerals in KREEP, basalts, etc.	Minerals present in abundance <1% of rock; elements are substantially lower in abundance; local concentrations are conceivable	Exceedingly difficult to concentrate from dispersed source	No known concentrations

knowledge of lunar resources permits the immediate investigation and development of processing techniques to be employed at an early time in space or on the Moon (Criswell, 1978, 1979; Green, 1978; Inculet and Criswell, 1979; Pomeroy and Hubbard, 1977). Further lunar exploration from orbit (European Space Agency, 1979; Minear *et al.*, 1976) and on the surface using machine intelligence techniques (Duda *et al.*, 1979) almost certainly will reveal additional resources. Discovery of volatiles, such as icy-dirt in permanently shadowed craters at the poles (Arnold, 1978; Watson *et al.*, 1963), certainly would expedite the growth of space industries.

The major components of the dark mari surfaces are basalt in the form of lithified or basalt-derived lunar soil and anorthositic plutonic rocks. "Granitic" glass is present in the light highlands.

On the basis of data and samples gathered by the Apollo and Luna missions it has been established that lunar surface basalts can be divided into two classes – high Al/Si (highland basalts) and low Al/Si (mare basalts). The major difference is in feldspar content, which is high in highland basalts and low in mare samples. Major minerals, and others found as minor constituents or traces in lunar basalts, are tabulated in table 4.3.

Pyroxenes occur as enstatite ($MgSiO_3$), wollastonite ($CaSiO_3$), ferrosilite ($FeSiO_3$), and mixtures of all three. Olivines are found as solid solutions of forsterite (Mg_2SiO_4) and fayalite (Fe_2SiO_4), with most falling in the range of 50-75 mole-percent forsterite. Plagioclase feldspars occur as solid solutions of anorthite ($CaAl_2Si_2O_8$) and albite ($NaAlSi_3O_8$), with most in the range of 80-100 mole-% anorthite.

A normative chemical analysis of "typical" lunar basalts is shown in table 4.4. It must be remembered that these values are for only two samples of basalt, and therefore, may not represent all lunar basalts. The composition of lunar soil is essentially the same as for lunar basalt, with grain constituents including agglutinates, basalt clasts, anorthite clasts, plagioclase, olivine, ilmenite, and glass. The average grain size is approximately 40 μm, but lunar soils often display bimodal size distributions.

Plutonic anorthosites are present in the highland areas. The mineral distributions in three anorthosite samples collected during Apollo missions are given in table 4.5.

TABLE 4.3.– MINERAL DISTRIBUTION IN LUNAR BASALTS

Mineral and composition	Highland composition, vol %	Mare composition, vol %
Feldspar (plagioclase)		
$(Ca,Na)Al_2Si_2O_8$	40-98	15-35
Pyroxene		
$(Ca,Mg,Fe)SiO_3$	0-40	40-65
Olivine		
$(Mg,Fe)_2SiO_4$	0-40	0-35
Ilmenite $FeTiO_3$	<2	0-25
Spinels		
$(Fe,Mg,Al,Cr,Ti)O_4$	<2	<2
Troilite FeS	<2	<2
Apatites		
$Ca_5(PO_4)_3(F,Cl)_3$	<0.2	<0.2
Zircon $ZrSiO_4$	<0.2	<0.2
Amphibole		
$(Ca,Mg,Fe)(Si,Al)_8O_{22}F$	<0.2	<0.2
Rutile TiO_2	<0.2	<0.2
Magnetite Fe_3O_4	<0.2	<0.2

TABLE 4.4.– A NORMATIVE ANALYSIS OF TYPICAL LUNAR BASALTS

Compound	Weight, %
SiO_2	37–48
TiO_2	4-13
Al_2O_3	7-11
FeO	16-22
MnO	0.2-0.3
MgO	6-16
CaO	8-13
Na_2O	0.2-0.5
K_2O	0.1-0.5
Cr_2O_3	0.3-0.6
P_2O_5	0.1-0.4
S	0.1-0.8

TABLE 4.5.– MODAL MINERALOGY OF LUNAR ANORTHOSITE

Mineral	Weight, %		
	Sample 15415	Sample 60025	Sample 76535
Pyroxene	3	1	4-5
Plagioclase	97	98-99	37-60
Olivine	---	---	35-58

Lunar glasses occur in two forms, basaltic and "granitic." Basaltic glass has roughly the same normative chemical distribution as lithified basalt. "Granitic" glass is somewhat anomalous, and may represent the quenched product of magma fractionation. The normative chemical composition of lunar glasses is shown in table 4.6.

TABLE 4.6.– NORMATIVE CHEMISTRY OF LUNAR GLASS

Compound	Basaltic, weight, %	Granitic, weight, %
SiO_2	44.6	73.2
TiO_2	2.1	.5
Al_2O_3	8.7	12.4
Cr_2O_3	.6	.4
FeO	22.5	3.5
MgO	11.4	.1
CaO	9.4	1.3
Na_2O	.3	.6
K_2O	.4	5.9

Asteroidal materials. Asteroids, especially those with near-terrestrial orbits, are expected to offer a wider range of useful minerals and elements than is available on the Moon (Gehrels, 1979). These bodies may be able to supply many minerals rare or absent on the Moon. For instance, spectroscopic analysis of outgassed volatiles suggests that some asteroids may have abundant water-ice (Degewij, 1980). Those bodies with carbonaceous chondritic composition should contain abundant carbon (up to a few percent by weight), an element which is comparatively rare in lunar soil. The water and carbon expected to be obtainable from asteroids could allow use of water-based and organic chemistry in space factories, techniques otherwise infeasible on a dry, carbonless Moon (though careful recycling will still be necessary). Asteroidal iron-nickel fractions should contain metals in the reduced state and may be rich in platinum-group elements. These resources complement those already found on the lunar surface.

It is conceivable that small quantities of meteoritic material have been trapped in the "gravity wells" (Lagrangian points L4 and L5) of the Earth-Moon (Freitas and Valdes, 1980) and Earth-Sun (Dunbar, 1979) systems. Should such materials exist, very little energy would be required to retrieve them to LEO.

As of 1978, 40 asteroids were known to have trajectories passing close to or inside of the Earth's heliocentric orbit. It has been estimated that 500-1000 Apollo and Amor objects have diameters in excess of 700 m (mass about 1-5$\times10^6$ t), together with more than 100,000 objects greater than 100 m diam with a mass of about 10^6 t each (Arnold and Duke, 1978; Gehrels, 1979). Although most of these asteroids have high velocities and inclinations with respect to Earth's motion around the Sun, a few percent have low inclinations and perihelion near Earth orbit. One, Anteros, can be reached from LEO with less delta-V than is required for transfer to the Moon (Hulkower, Jet Propulsion Lab, private communication, 1980). Several detailed studies have been conducted to examine the possibility of returning one or more of these objects to the vicinity of Earth for use in space manufacturing (Bender et al., 1979; Gaffey et al., 1979; O'Leary et al., 1979). Methods considered for retrieval have included mass drivers, pellet launchers, solar sails, or detonation propulsion, perhaps expedited by gravitational swing-bys of Mars, Venus, Earth, or the Moon, as required. Extensive increases in ground-based searches and exploration missions to favorable objects should be initiated to fully characterize these resources in preparation for utilization. Table 4.7 summarizes the compositional information now available on Apollo/Amor asteroids, some of which are expected to be a far richer source of volatile materials than low-latitude lunar soils.

Between the orbits of Mars and Jupiter lie thousands of asteroids. These range in diameter from 1000 km down to the limits of telescopic visibility – a few kilometers

TABLE 4.7.– CHARACTERIZATION OF APOLLO/AMOR OBJECTS
(Adapted from Billingham *et al.*, 1979)

Number	Name	Aphelion, AU	Perihelion, AU	Surface type[a]	Albedo[b]	Diameter,[b] km
433	Eros	1.458	1.13	Olivine, pyroxene, metal (~H chondrite)	0.17	23
887	Alinda	2.516	1.15	Olivine, carbon (~C3 carbonaceous chondrite)	.17	4
1036	Ganymede	2.658	1.22	"S" – probably silicate or metal-rich assemblage	---	(~35)
1566	Icarus	1.078	.19	Pyroxene (olivine, metal?)	.17	1
1580	Betulia	2.196	1.12	"C" – opaque-rich assemblage, possibly carbonaceous	.05	6
1620	Geographos	1.244	.83	"S" – probably silicate or metal-rich assemblage	.18	3
1627	Ivar	1.864	1.12	"S" – probably silicate or metal-rich assemblage	---	(~7)
1685	Toro	1.368	.77	Pyroxene, olivine	.12	3
1864	Daedalus	1.461	.56	"O" – probably silicate or metal-rich assemblage	---	(~2)
1960 UA	---	2.26	1.05	"U" ?	---	---
1976 AA	Arnold	.97	.79	"S" – probably silicate or metal-rich assemblage	.17	1

[a]Where adequate spectral data are available, mineralogical characterizations and meteorite equivalents are given (from work by Gaffey and McCord, 1977). Where only UBV colors (i.e., C, S, O, U) are available, the Chapman-Morrison-Zellner classification of the object as summarized by Zellner and Bowell (1977) is given. Underlined classification symbols indicate those based on a single classification criterion. Probable mineral assemblages are indicated.

[b]Albedos and diameters as summarized by Morrison (1977). The diameters in parentheses were derived assuming an average albedo for the "O-S" class of the object and should be considered as indicative only.

(Gehrels, 1979; Morrison and Wells, 1978). Certainly still smaller bodies exist but cannot be seen from Earth. Table 4.8 summarizes available information on the widely variable surface compositions of asteroids (Lunar, 1978). The predicted large quantities of rare elements, such as chromium and vanadium, and common metals such as iron and nickel might ultimately have great importance to terrestrial markets and space industries (Gaffey and McCord, 1977; Kuck, 1979). Industrial facilities and habitats constructed from asteroidal materials would make possible the rapid spread of humanity throughout the Solar System.

Investigation and development of asteroidal resources will require at least a three-phase approach. First, it is important to find and catalogue the populations and spectral classes of near-Earth asteroids. This could begin at once with a modest investment in a dedicated automated telescope and television camera system which, it is estimated, should be able to find approximately one new Earth-crossing asteroid every night (Gehrels, 1979).

Second is the necessity for direct exploration and sample-return missions. Although there is evidence suggesting that asteroids are equivalent to terrestrial meteorites in composition, the precise physical structures of these bodies are unknown. They may be solid, "fluffy," or more like "raisin bread" with rocks and metals distributed in some matrix. Refining and processing system designs would be significantly affected by the structural configurations of asteroids.

The third and final phase involves large-scale utilization of asteroidal materials either on-site or following transport into near-Earth space. There is a need to develop systems for despinning asteroids, emplacing powerful thrusters, then returning the body to near-Earth space. Ultimately, whole factories might be delivered to or evolved upon individual asteroids. One unusual possibility is that automated factories sent to asteroids could "blow" local materials (metals, glasses, composites) into large, thin, glass-like bubbles many kilometers across, or into metal-coated film

TABLE 4.8.– ASTEROID DATA

[Compiled by Clark Chapman and Ben Zellner from the TRIAD[a] data file]

Asteroid	Semimajor axis, AU	Eccentricity	Inclination, deg	Absolute magnitude B(1.0), mag	Color U-B, mag	Color B-V, mag	Albedo[b]	Diameter,[c] km	Rotation period, hr	Type,[d]	Inferred mineralogy,[e]
Asteroids larger than 200 km in diameter (listed in order of size)											
1 Ceres	2.767	0.0784	10.61	4.48	0.42	0.72	0.053	1020	9.078	U	Silicate (olivine?) + opaque (magnetite?)
4 Vesta	2.362	.0890	7.13	4.31	.48	.78	.235	549	5.34213	U	Clinopyroxene (+ plagioclase?)
2 Pallas	2.769	.2353	34.83	5.18	.26	.65	.079	538[f]	7.88106	U	Silicate (olivine?) + opaque (magnetite?)
10 Hygiea	3.151	.0996	3.81	6.50	.31	.69	.041	450	18	C	Phyllo-silicate + opaque (carbonaceous?)
511 Davida	3.187	.1662	15.81	7.36	.35	.71	.033	341	5.12	C	Phyllo-silicate + opaque (carbonaceous?)
704 Interamnia	3.057	.1553	17.31	7.24	.25	.64	.035	339	8.723	C	Silicate (olivine?) + opaque (magnetite?)
31 Euphrosyne	3.154	.2244	26.30	7.28	---	---	---	(333)	---	CM	---
451 Patientia	3.061	.0772	15.23	8.05	.31	.67	.026	327	7.11	C	---
65 Cybele	3.434	.1154	3.55	7.99	.28	.68	.022	308	---	C	---
52 Europa	3.092	.1138	7.47	7.62	.35	.68	.035	290	11.2582	C	Phyllo-silicate + opaque (carbonaceous?)
16 Psyche	2.920	.1390	3.09	6.88	.25	.70	.093	252	4.303	M	Nickel-iron (+ enstatite?)
324 Bamberga	2.686	.3360	11.16	8.07	.29	.69	.031	251	8	C	Phyllo-silicate + opaque (carbonaceous?)
3 Juno	2.670	.2557	12.99	6.51	.42	.82	.151	248	7.213	S	Nickel-iron + olivine + pyroxene
15 Eunomia	2.642	.1883	11.73	6.29	.44	.82	.167	246	6.0806	S	Nickel-iron + silicate (olivine ≫ pyroxene)
13 Egeria	2.576	.0889	16.50	8.15	.45	.75	.033′	241	7.045	C	---
45 Eugenia	2.721	.0806	6.60	8.31	.27	.68	.030	228	5.700	C	---
87 Sylvia	3.481	.0985	10.85	8.12	.24	.69	---	(225)	---	CMEU	---
19 Fortuna	2.442	.1576	1.56	8.45	.38	.75	.030	221	7.46	C	Phyllo-silicate + opaque (carbonaceous?)
216 Kleopatra	2.793	.2520	13.09	8.10	.24	.72	---	(219)	5.394	CMEU	---
532 Herculina	2.771	.1789	16.35	8.05	.43	.86	.120	217[g]	9.406	S	Pyroxene + olivine + nickel-iron? + opaque?
624 Hektor	5.150	.0248	18.26	8.65	.26	.76	.038	216[h]	6.9225	U	---
107 Camilla	3.489	.0699	9.92	8.28	.29	.70	.037	210	4.56	C	---
7 Iris	2.386	.2303	5.50	6.84	.47	.83	.160	210	7.135	S	Nickel-iron + olivine + minor pyroxene
24 Themis	3.138	.1208	.77	8.27	.34	.69	---	210	8.375	C	---
409 Aspasia	2.575	.0733	11.26	8.31	.31	.71	---	208	---	C	---
88 Thisbe	2.768	.1619	5.22	8.07	.28	.67	.045	207	6.0422	C	Phyllo-silicate + opaque (carbonaceous?)
747 Winchester	2.994	.3438	18.15	8.84	.32	.71	.024	205	8	C	---
702 Alauda	3.194	.0347	20.54	8.29	.31	.66	---	205	---	C	---
165 Loreley	3.128	.0802	11.24	8.81	.31	.74	---	203	---	C	---
Other interesting asteroids											
8 Flora	2.202	0.1561	5.89	7.73	0.46	0.88	0.125	153	13.6	S	Nickel-iron + clinopyroxene
25 Phocaea	2.401	.2531	21.61	9.07	.51	.93	.184	65	9.945	S	Nickel-iron + pyroxene + clinopyroxene
44 Nysa	2.422	.1517	3.71	7.85	.26	.71	.467	72	6.418	E	Enstatite?
80 Sappho	2.295	.2008	8.66	9.22	.50	.92	.113	86	>20	U	Silicate (olivine?) + opaque (carbonaceous?)
158 Koronis	2.868	.0559	1.00	10.95	.38	.84	---	36	---	S	---
221 Eos	3.014	.0958	10.85	8.94	.41	.77	---	(97)	---	U	Silicate (olivine?) + opaque (carbonaceous?)
279 Thule	4.258	.0327	2.34	9.79	.22	.77	---	(60)	---	MEU	---
349 Dembowska	2.925	.0862	8.26	7.24	.55	.97	.260	145	4.7012	R	Olivine ≫ pyroxene (+ nickel-iron?)
433 Eros	1.458	.2220	10.83	12.40	.50	.88	.180	16	5.2703	S	Silicate (olivine = pyroxene) + minor nickel-iron
434 Hungaria	1.944	.0736	22.51	12.45	.24	.70	.300	11	---	E	---
785 Zwetana	2.576	.2029	12.72	10.73	.17	.64	.078	45	---	U	---
944 Hidalgo	5.820	.6565	42.49	12.05	.23	.74	---	(39)	10.0644	CMEU	---
1566 Icarus	1.078	.8267	22.99	17.32	.54	.80	---	(1.7)	2.2730	U	---
1580 Betulia	2.196	.4905	52.04	15.66	.27	.66	---	6.5	6.130	C	---
1620 Geographos	1.244	.3351	13.33	15.97	.50	.87	---	2.4	5.2233	S	---
1685 Toro	1.368	.4360	9.37	16.20	.47	.88	---	(7.6)	10.1956	U	Pyroxene + olivine?

[a]TRIAD ≡ Tucson Revised Index of Asteroid Data is the source of all data, except as noted in subsequent footnotes. Contributors to this computerized file are: D. Bender (osculating orbital elements), E. Bowell (UBV colors), C. Chapman (spectral parameters), M. Gaffey (spectrophotometry), T. Gehrels (magnitudes), D. Morrison (radiometry), E. Tedesco (rotations), and B. Zellner (polarimetry). TRIAD is described in *Icarus* 33, 630-631 (1978). To use TRIAD, contact: B. Zellner, Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721.

[b]Albedos are geometric albedos from radiometry. They are not always consistent with tabulated diameters.

[c]Except as noted, diameters are from Bowell *et al.* (*Icarus*, Sept. 1978). Values are less reliable for asteroids for which no albedo is listed in previous column. Especially unreliable diameters are listed in parenthesis.

[d]Taxonomic type, related to surface composition, is from Bowell *et al.* (*Icarus*, Sept. 1978) wherein the types are defined.

[e]From Gaffey and McCord (*Proc. Lunar Sci. Conf. 8th*, p. 113-143, 1977), here augmented in several cases by C. Chapman. Refers only to optically important phases.

[f]Stellar occultation diameter (Wasserman *et al.* and Elliot *et al.*, *Bull. Amer. Astron. Soc.*, Oct. 1978).

[g]Stellar occultation diameter (Bowell *et al.*, *Bull. Amer. Astron. Soc.*, Oct. 1978).

[h]Hartman and Cruikshank (*Icarus*, in press, 1979).

(*Lunar*, 1978)

for use as solar sails (Drexler, 1980; Nichols, 1979, unpublished report, CIT, Pasadena, Calif.), or as mirrors.

Each of the three asteroid resource development phases is an excellent driver for machine intelligence, robotics, and teleoperation technologies. Long mission times to asteroids favor automation over manned missions. However, it appears that emplacement of thrusters, large-scale bubble-blowing and processing are beyond state-of-the-art, especially in the absence of teleoperation.

Other Solar System resources. Eventually, the resources of the planets (Greeley and Carr, 1975) and their major satellites may become accessible to mankind. Initial attempts at utilization probably will focus on the moons of Mars to support permanent exploration of that planet as well as travel between Mars and the Earth, and beyond. Much of the technology needed for maintaining permanent occupancy of LEO and the Moon should help make extensive exploration of Mars economical. The atmospheres of Venus, Jupiter, Saturn, and their moons and rings are likely early-target resources within these planetary systems (Table 4.9, Lunar, 1978). Later the surface materials of many of the moons may be accessed (Burns, 1977). The radiation belt of Jupiter constitutes a major impediment to utilization of that diverse system. (Access to inexpensive mass for shielding would permit both manned and unmanned penetration of the Jovian magnetosphere for extended periods of time.) Methods have been suggested for extracting energy directly from the particle radiation of the belts by means of secondary emission of charged particles.

Comets, the solar wind, and the Sun are the last major material resources within the Solar System. Most comets pass through the inner Solar System at very high velocities and inclinations (table 4.10, Lunar, 1978). To dependably retrieve large quantities of cometary material it may be necessary to locate and intercept these bodies in the outer Solar System or beyond and provoke repeated gravitational encounters with various planets to effect capture for near-Sun use. These bodies should be exceptionally rich sources of C, N, H, Na, and other volatile elements. Wetherill (1979) estimates that comets, on a 100,000-year timescale, become new Apollo/Amor objects at the rate of 10^{11} t/yr. Deliberate capture probably could increase this rate by several orders of magnitude.

The solar wind is the outward flow of fully ionized gases (at least 10^{13} t/yr) from the Sun. Presumably, all elements present in the Sun are represented in the solar wind. Table 4.11 gives estimates of the annual output tonnages of the elements, assuming each is present with the same distribution as the cosmic abundance (Allen, 1976). (It is assumed that the hydrogen flux (2×10^{18} ions/cm^2-sec) is omnidirectional from the Sun.) Some type of magnetodynamic systems would clearly be required to collect the solar wind. Perhaps flux-braking by the solar gravitational force is possible, via many convergent magnetic nozzles. Significant fractions of the solar wind might be condensed into grains in convergent regions of such magnetic loops. Though the team can offer no conceptual designs for such systems, it is intriguing that the solar wind output of most elements rivals or far exceeds their corresponding current annual terrestrial production rates. In particular, this source could provide enormous masses of hydrogen throughout the Solar System. Given a means of collecting large fractions of the solar wind, eventually it may be possible to tap tiny portions of the 2×10^{27} ton mass of the Sun itself. Such a "star-centered" resource technology capability could decouple the extrasolar spread of humanity and its artifacts from the need for detailed knowledge of the star system of destination.

4.2.2 Extraction and Materials Processing Alternatives

The development of material processing techniques suited to nonterrestrial conditions is absolutely essential if the proposed SMF growth scenario is ever to take place. Studies have been conducted on the gathering of lunar materials for use in situ and elsewhere (Criswell (see Carrier), 1980; Fields and Weathers, 1967). Ultimately, SMF output must be fabricated from feedstock derived from lunar, asteroidal, or other space materials. The production of such diverse components as lubricants, coils, semiconductor chips and structural components requires a versatile and efficient raw material processing capability. Furthermore, this processing system must be fully automation-compatible. Mass multiplication is one key consideration in a growing space-processing facility. Every effort should be made to minimize both the quantity of processing materials brought from Earth per unit of nonterrestrial products, and the mass of the capital equipment (both terrestrial and nonterrestrial) per unit of output per unit of time. It is desirable for the fraction of all such terrestrial material supplied per unit of output product, called the "Tukey Ratio" (Heer, unpublished draft notes of the Proceedings of the Pajaro Dunes Goal-Setting Workshop, June 1980), to approach zero as deployment and growth proceed – or, alternatively, for the mass multiplication (referenced to Earth-originating materials) to approach infinity.

Another important aspect of SMF design is the ability of the primary processing equipment to accept a wide range of input materials, thus minimizing the need for intensive and extended exploration and characterization of source materials. It appears that this approach already may be possible for the explored regions of the Moon due in part to the limited variety of lunar materials and glasses (Waldron et al., 1979). Additionally, mass multiplication factors in excess of 100 can be anticipated for one or more proposed lunar materials processing schemes (Criswell, 1978, 1979).

TABLE 4.9.– PLANETARY ATMOSPHERES

[Compiled by Glenn Orton]

	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune
Surface pressure	93 ± 2 bars	1.00137 bars (mean sea level)	5-10 mb, variable	Undefined	Undefined	Undefined	Undefined
Temperature at surface (or at one bar of pressure where surface not defined)	741 ± 7 K	~290 K	210-240 K, variable	170 ± 10 K	140 ± 20 K	80 ± 30 K	80 ± 30 K
Number mixing ratio of principal constituents	0.95 ± 0.05 CO_2	N_2 0.79 O_2 0.21	0.953 CO_2	H_2 0.90 ± 0.05 He 0.10 ± 0.05	*a* H_2 (most abund.) He	*a* H_2 He?	*a* H_2 He?
Number mixing ratio of minor constituents	0.05 ± 0.05 Ar/N_2 vaiable H_2O	CO_2 3.3×10^{-4} $H_2O \lesssim 5 \times 10^{-3}$ (variable)	0.027 N_2 0.010 Ar 0.0013 O_2	CH_4 $0.5\text{-}2.0 \times 10^{-3}$ $NH_3 \leqslant 2 \times 10^{-4}$	CH_4[a] NH_3	CH_4[a] NH_3?	CH_4[a] NH_3?
Known trace constituents	CO HCl He HF H_2SO_4	N_2O CO CH_4 O_3 and many more	CO H_2O Ne Kr Xe O_3	H_2O CO GeH_4 PH_3 C_2H_2 C_2H_6	PH_3 C_2H_2? C_2H_6		C_2H_6?
Principal sources of aerosols	Concentrated H_2SO_4 droplets	H_2O ice surface dust	H_2O ice CO_2 ice wind blown surface "dust"	NH_3 ice hydrocarbons? NH_4SH?	NH_3 ice? hydrocarbons?	CH_4 ice? NH_3 ice? hydrocarbons?	CH_4 ice? NH_3 ice hydrocarbons?
Mean horizontal wind at surface[b]	0-2 m sec^{-1}	0-10 m sec^{-1}	2-9 m sec^{-1}	Differential zonal velocities with max value ~200 m sec^{-1}	Differential zonal velocities with max value ~450 m sec^{-1}	*a*	*a*
Maximum wind in atmosphere[c]	~100 m sec^{-1}	~45-65 m sec^{-1}	~60-80 m sec^{-1}	Differential zonal velocities with max value ~200 m sec^{-1}	Differential zonal velocities with max value ~450 m sec^{-1}	*a*	*a*

Error bars are one standard deviation.
[a]Too uncertain for quantitative entry.
[b]For midlatitudes in spring or autumn where global and seasonal effects are present.
[c]Where distinguishable from surface.

(*Lunar*, 1977)

TABLE 4.10.– COMETS

[Compiled by Ray Newburn]

Comet[a]	Period,[b] yr	Perihelion distance,[b] AU	Inclination,[b] deg	Absolute magnitude,[c] H_{10}	Remarks
Selected short-period comets					
Encke (1974 V) (1977 –)	3.30	0.34	12.0	9.7 (pre-perihelion)	Shortest period comet known. Smallest perihelion distance of any comet with P < 100 years. Fan-shaped coma. Sometimes shows weak ion tail. No continuum in spectra. Often mentioned as space-probe target.
Tempel 2 (1977d)	5.26	1.36	12.5	8.4	Third shortest period. Fan-shaped coma. No tails. Strong continuum in spectra. Possible space-probe target. Brightest about 3 weeks after perihelion.
Tuttle-Giacobini-Kresak (1973 VI)	5.56	1.15	13.6	11.5	A few days before perihelion and 8-35 days after perihelion in 1973, this comet flared up in brightness by nine magnitudes (a factor of 4000).
d'Arrest (1976 IX)	6.23	1.17	16.7	9.5	Visible to naked eye briefly in 1976, close approach to Jupiter in 1979 will increase perihelion to 1.29 AU. Large, fan-shaped coma. Brightness usually increases 3 magnitudes *after* perihelion.
Giacobini-Zinner (1978h)	6.52	.99	31.7	10.0	Striking tail for so faint a comet. Meteoroids apparently very low density, "fluffy" objects.
Schwassmann-Wachmann 1 (1974 II)	15.03	5.45	9.7	~7.5	Smallest eccentricity (0.105) and largest perihelion of any short period comet. Shows periodic flares in brightness of ~5 magnitude.
Halley (1910 II)	76.09	.59	162.2	5.0	Brightest comet with P < 100 years. Well developed ion and dust tails. Strong continuum in spectra. Poor apparition as seen from Earth in 1986. Possible target for space probe flyby.
Selected long-period comets					
Ikeye-Seki (1965 VIII)	800	0.008	141.9	6.5	One of a family of comets with similar orbital elements that passes through the solar corona (Kreutz family).
Tago-Sato-Kosaka (1969 IX)	~110,000	.47	75.8	6.6	First comet observed at Lα wavelengths (1216A). Vast hydrogen corona found (~10^7 km diameter).
Bennet (1970 II)	~1,700	.54	90.0	4.3	Spectacular naked eye object. Most thoroughly studied comet until Kohoutek.
Kohoutek (1973 XII)	~79,000	.14	14.3	5.2 (pre-perihelion)	The public was disappointed, but science gained tremendously from a well-planned, coordinated program of research.
West (1976 VI)	~6,400,000	.20	43.1	4.4 (post-perihelion)	Nucleus split into four pieces near perihelion. Very spectacular naked eye object. Unusual banded dust tail.

[a]Comets are normally named after their discoverers, but no more than the first three independent observers are so recognized. A few comets have been named after those who made extensive studies of their motion (e.g., Halley and Encke). Comets with periods less than 200 years are arbitrarily called "short-period" and are often written with a preceeding P/ for periodic (e.g., P/Tempel 2). Long-period comets are sometimes preceeded by C/ just to designate they are comets (e.g., C/West). When one observer or combination of observers discovers more than one *short-period* comet, the names are followed by an Arabic numerical in order of their discovery to tell them apart (e.g., P/Tempel 1 and P/Tempel 2).

In addition to their discoverers' names, comets are given a temporary designation of the year followed by a lower case letter indicating the order of discovery or recovery. Thus new comet C/Meier is 1978f, while 1978h is P/Giacobini-Zinner making its tenth observed appearance. A few comets such as P/Encke can be observed completely around their orbits. These are designated "annual" comets and are not given temporary letter designations. After about two years, when it is reasonably certain that all of the comets with perihelion passage in a given year have been discovered, each comet is given a permanent designation of the year and Roman numeral indicating the order in which it passed perihelion. Thus P/Encke was 1971 II and 1974 V, and sometime in 1979 it will receive a 1977 permanent designation. These permanent designations are often used together with the name for long period comets [e.g., C/West (1976 VI)] since many discoverers have more than one long-period comet to their credit. The comet 1978a was also discovered by West, for example, while 1973e (1973 VII) and 1973f (1973 XII) were both discovered by Kohoutek, as for that matter were 1970 III, P/Kohoutek (1975 III), and P/West-Kohoutek-Ikemura (1975 IV).

Recent review papers on comets can be found in *Comets Asteroids Meteorites Interrelations, Evolution and Origins,* A. H. Delsemme (ed.), University of Toledo, 1977 and in *The Study of Comets,* Proceedings of IAU Colloquium No. 25, B. Donn, M. Mumma, W. Jackson, M. A. Hern and R. Harrington (eds.), NASA SP-393, 1976.

[b]From Marsden "Catalog of Cometary Orbits, 2nd Edition." This is the primary source of orbital data for all comets.

[c]Magnitude at 1 AU from Earth and from Sun (extrapolated to that distance, if the comet doesn't actually achieve it, using a 4th-power law with heliocentric distance).

(*Lunar,* 1978)

TABLE 4.11.– COMPOSITION AND MASS OF ANNUAL SOLAR WIND OUTFLOW

Element	Cosmic abundance	Estimated solar wind abundance, ton/yr	Earth output (1977), tons
Actinium	0	0	---
Aluminum	9.5E 4	1.9E 9	1.5E 8
Americium	0	0	---
Antimony	.2	1.8E 4	7.9E 4
Argon	1.5E 5	4.4E 9	5.5E 5
Arsenic	4	2.2E 5	3.8E 4
Astatine	0	0	---
Barium	3.7	3.7E 5	2E 7[a]
Berkelium	0	0	---
Beryllium	20	1.3E 5	107
Bismuth	.1	1.5E 4	8.8E 6
Boron	24	1.9E 5	3.1E 6
Bromine	13	7.6E 5	338
Cadmium	.9	7.4E 4	1.6E 4
Calcium	4.8E 4	1.4E 9	1.6E 4
Californium	0	0	---
Carbon	3.5E 6	3E 10	7E 7[a]
Cerium	2.3	2.4E 5	3E 5[a]
Cesium	.5	4.8E 4	80[a]
Chlorine	9000	2.4E 8	5E 8[a]
Chromium	7800	3E 8	1E 7
Cobalt	1800	7.8E 7	3E 4
Copper	210	9.8E 6	8E 6
Curium	0	0	---
Dysprosium	.6	7.1E 4	---
Einsteinium	0	0	---
Erbium	.3	3.6E 4	200[a]
Europium	.2	2.2E 4	---
Fermium	0	0	---
Fluorine	1500	2.2E 7	3E 7[a]
Francium	0	0	---
Gadolinium	.7	6.1E 4	5E 3[a]
Gallium	11	5.6E 5	400[a]
Germanium	51	2.7E 6	8.2
Gold	.1	1.4E 4	1.6E 3
Hafnium	.4	5.2E 4	8E 4
Helium	3.1E 9	9.1E 12	1E 5[a]
Holmium	.1	1.2E 4	200[a]
Hydrogen	4E 10	3E 13	6E 7[a]
Indium	.11	9.2E 3	44
Iodine	.8	7.5E 4	1.1E 4
Iridium	.8	1.1E 5	3[a]
Iron	6E 5	2.5E 10	7.5E 9
Krypton	51	3.1E 6	1E 3[a]
Lanthanum	2	2E 5	1E 5[a]
Lawrencium	0	0	---
Lead	.5	7.6E 4	3.1E 6
Lithium	100	5.1E 5	4.9E 3
Lutetium	.05	6.4E 3	200[a]
Magnesium	5.1E 5	1.6E 10	1.5E 5
Manganese	6900	2.8E 8	2.4E 5
Mendelevium	0	0	---
Mercury	.3	4.4E	7.6E 3
Molybdenum	2	1.4E 5	1E 5
Neodymium	1.4	1.5E 5	1E 5[a]
Neon	8.6E 6	1.2E 11	2E 4[a]
Neptunium	0	0	---
Nickel	2.7E 4	1.1E 9	8.5E 5
Niobium	1	6.8E 4	6E 4[a]
Nitrogen	.66E 7	6.8E 10	6E 7
Nobelium	0	0	---
Osmium	1	1.3E 5	2[a]
Oxygen	2.2E 7	2.5E 11	5.6E 7
Palladium	.7	5.4E 4	1E 3[a]
Phosphorus	1E 4	2.2E 8	2E 8[a]
Platinum	1.6	2.3E 5	218
Plutonium	0	0	---
Polonium	0	0	---
Potassium	3200	9.1E 7	2E 8[a]
Praseodymium	.4	4.1E 4	3E 4[a]
Promethium	0	0	---
Protactinium	0	0	---
Radium	0	0	---
Radon	0	0	---
Rhenium	.135	1.8E 4	5
Rhodium	.2	1.5E 4	100[a]
Rubidium	7	4.4E 5	40[a]
Ruthenium	1.5	1.1E 5	---
Samarium	.7	7.7E 4	---
Scandium	28	9.2E 5	---
Selenium	68	8.9E 6	1.3E 3
Silicon	1E 6	2.1E 10	2.5E 7
Silver	.3	2.3E 4	1.1E 4
Sodium	4.4E 4	7.4E 8	8E 6
Strontium	19	1.2E 7	3.4E 4
Sulfur	3.75E 5	6.8E 9	5.9E 7
Tantalum	.07	9.2E 3	430
Technetium	0	0	---

[a]Freitas, R., 1980a.

TABLE 4.11.– CONCLUDED

Element	Cosmic abundance	Estimated solar wind abundance, ton/yr	Earth output (1977), tons
Tellurium	4.7	4.4E 5	8E 3[a]
Terbium	.1	1.2E 4	---
Thallium	.1	1.5E 4	40[a]
Thorium	0	0	1.1E 4
Thulium	.03	3.7E 3	2E 3[a]
Tin	1.3	1.1E 4	2.6E 5
Titanium	2400	8.4E 7	4.8E 4
Tungsten	.5	6.7E 4	4.7E 4
Uranium	.1	1.7E 4	3.2E 4
Vanadium	220	8.2E 6	3.2E 4
Xenon	4	3.8E 5	50[a]
Ytterbium	.2	2.5E 4	200[a]
Yttrium	9	5.7E 5	312
Zinc	490	2.3E 7	6.8E 6
Zirconium	55	3.7E 6	4.7E 5

As on Earth, a continuing tradeoff between availability of primary materials, processing options, and substitution of materials can be expected. Systems designed for the Moon might not be appropriate for Mars, an iron asteroid, or Titan. Still, most of this section describes silicate minerals processing as these are the dominant components of lunar soil and seem likely to be representative of the composition of many asteroids, Mercury, and the moons of Mars. Since the Solar System offers a much wider range of compositions and conditions, many alternative types of manufacturing facilities may be expected to evolve, many of which may eventually prove useful on Earth.

Chemical extraction techniques. The first most important component of the SMF is the chemical processing facility. The ultimate success of the space manufacturing venture hinges upon the ability to process nonterrestrial materials without importation of terrestrial reagents. This task is further complicated by the additional requirement that the processing capability grow at a rate equal to or greater than the overall growth rate of the SMF. The applicability of a number of established chemical engineering technologies to the processing of low-latitude lunar materials, including (1) carbothermic reduction, (2) carbochlorination, (3) electrolysis, (4) NaOH treatment, and (5) HF acid leaching, has been suggested (Waldron et al., 1979).

In carbothermic reduction anorthite is broken down and refined. The aluminum oxide reacts with carbon to produce useful metallic aluminum and carbon monoxide (Phinney et al., 1977). The thermodynamics of this process requires that the processing vessel be maintained at 2400 K. High-temperature condensates such as SiC, Al_4C_3, and Al_4O_4C are present, along with the gases Al_2O, SiO, Al, and Si. These are likely to prevent the key reactions from achieving equilibrium (Waldron et al., 1979).

In the carbochlorination process, titanium, iron, and aluminum are refined from anorthite and ilmenite by reaction with carbon and chlorine (Rao et al., 1979). This process does not require high reaction temperatures. However, chlorine recycling involves very massive equipment (Waldron et al., 1979).

Electrowinning of aluminum from anorthite powder dissolved in a mixture of alkaline earth chlorides at 75 K has been considered (Criswell (Das et al.), 1980). This approach requires only a moderate amount of energy.

Iron and titanium can be refined from ilmenite by treatment in molten NaOH (Rao et al., 1979). TiO_2 is soluble in NaOH, unlike Fe_2O_3, and thus the two compounds can be separated and refined. High temperatures (1000–1300 K) are necessary for this process.

Lunar soil may be broken down into its elemental constituents by the HF leaching technique (Waldron et al., 1979). This process begins with the dissolution of lunar soil in a heated HF solution, followed by a series of steps including ammonium salts fusion, silicon hydrolysis, metal oxide production, acid recovery, fluoride hydrolysis, ion exchange and platable-metals separations, precipitation and crystallization, and metal oxide reduction.

Most of the reagents used in the above processes are rare on the Moon compared to the known average lunar composition. Thus, recycling and leakage must be regarded as critical problems. Thermal dissipation is another major problem because many techniques involve exothermic reactions which generate heat that is difficult to dispose of due to the unavailability of direct conductive cooling in space. HF acid leaching appears to be the most promising for interim processing and short-term growth of the SMF. More (valuable) elements can be extracted in this way than any other process studied to date. However, while the HF process appears quite efficient there are several potential pitfalls associated with the deployment of an acid leach system. HF usually is stored in polymer containers because it dissolves most metals and all silicates. Such polymers cannot easily be derived from lunar soil. Containerless reaction technology cannot be employed because of the sublimation problem. Possibly etch-resistant solid silane containers could be developed, but these would have to be maintained at 75 K or colder, resulting in prohibitively sluggish reaction rates. Yet another potential problem is leakage. The numerous steps involved in the HF acid technique significantly increase the likelihood of accidental loss of vital process fluids.

It is important that the reagents, plumbing, and containment vessels for the chemical processing plant eventually be produced from nonterrestrial materials – importation of these commodities is not feasible if the long-term growth rate is to be exponential. As to the first of these necessities, calculations by Freitas (1980b), based on an HF leach factory module capable of processing roughly 4000 t/yr of lunar soil, indicate that sufficient hydrogen and fluorine can be produced to allow replication of the required reagents. The calculations assumed 95% recovery of hydrogen and 50% recovery of fluorine due to leakage, which may be too optimistic. On the other hand, these limitations may be offset by discoveries of richer sources of hydrogen (e.g., Arnold, 1980) and fluorine on the Moon or by changes in physical-to-chemical processing ratios. It appears that at least short-term growth of SMF capability is possible with the use of HF acid leach extraction. The remaining problems of producing plumbing and containment vessels from nonterrestrial materials appear insoluble at present; however, importation of polymeric plumbing and make-up reagents is feasible for short-term growth.

The methods discussed above are well-suited to short-term nonexponential SMF growth. Table 4.12 summarizes the recommendations of a recent workshop on silicate and other lunar-like minerals processing (Criswell, personal communication, 1980). New processing methods which do not require aqueous solutions or reagents composed of rare nonterrestrial elements might help to achieve a long-term self-sufficient, exponentially growing SMF (Grodzka, 1977). Possible new avenues of research may include silicon- and oxygen-based processes, advanced zone refining or fractionation techniques, induced immiscibility in melts, and rapid controlled-crystal-nucleation methods.

Electrophoretic processing. An important initial step in the generation of new processing options for dry, granular materials found on the Moon is the development of an effective mineral separation or primary beneficiation process. If the primary materials of interest for a particular refined product (such as lunar anorthite plagioclase for aluminum and silica) can be isolated, then the problem of developing a self-sufficient chemical beneficiation process is far less difficult (Rao et al., 1979).

Every chemical processing option for beneficiating lunar soil suggested to date requires chemicals that are relatively scarce on the Moon. Some of these options may demand high levels of automation not presently available. It is therefore desirable to develop new processing options that can be expanded with little or no importation of terrestrial materials and that are either self-automated or automation-compatible. A promising new primary beneficiation technology opportunity appears to be electrophoretic separation, a one-step, self-automatable technique (Dunning and Snyder, 1981).

Electrophoresis is defined as the transport of electrically charged particles in a direct current electric field (Bier, 1973). The movement occurs as a result of the electrostatic potential between the layer of ions adsorbed from the suspension medium onto the surface of particles and the bulk suspension medium. The layer of adsorbed ions is called the "Helmholtz double layer" or the "electrical double layer." It consists of the potential determining layer (the surface of the particulate material), the Stern layer (the layer of adsorbed ions from the atmosphere), and the Guoy layer (the bulk fluid) (Bier, 1973; Jungerman, 1970). The electrophoretic potential is defined as the electrostatic potential between the Stern layer and the bulk fluid. If the electrophoretic potential is positive or negative, a particle moves towards one of the electrodes in the system. The direction of movement depends on the relative charge signs of the particle and the electrode, and the velocity is a function of the magnitude of the electrophoretic potential. If the potential of a particle is zero (the isoelectric point), particles remain stationary and suspended. Electrophoretic separation depends on differential migration rates for particles in the bulk suspension medium (although electrode-reaction electrophoresis is employed for electroplating). The major requirement for successful beneficiation is that the particulate matter be sufficiently fine-grained to remain suspended in the bulk medium. The ideal grain size for geologic materials is 25–60 μm (Westwood, 1974).

Electrophoresis has been used by physiologists and biologists since the 1930's as a tool for separation and identification of enzymes, proteins, lipids and blood cells. Tests were performed on blood cells during the Skylab and Apollo-Soyuz experiments with good success (Henderson and Vickery, 1976; Schoen et al., 1977), and the electrophoretic phenomenon has been utilized as a terrestrial separation technique for clays and limestones.

Of the numerous electrophoresis technologies only a few are suitable for geologic materials. One technique – high-voltage zone electrophoresis – is particularly well-suited to lunar soil separation because it is a one-step, self-automated separation method. Typically, a tank is filled with suspension medium into which two electrodes are inserted. Filter paper is mounted on both electrodes. When an electric field is applied, mineral particles move toward the filter paper and are trapped in various positions along its length. Each mineral phase migrates to a discrete area depending on the magnitude and sign of the electrophoretic mobility. These phases then may be removed in a single, simple automated step.

Lunar soil is ideally suited to electrophoretic separation. Average grain size is 40 μm (Williams and Jadwick, 1980), well within the optimal range cited earlier for geologic materials. This grain size distribution is also very poorly

TABLE 4.12.– RESEARCH DIRECTIONS FOR THE DEVELOPMENT OF NEW PROCESSING TECHNOLOGIES FOR UTILIZATION OF LUNAR AND SILICATE MINERALS
(Criswell, 1979)

1. Physical separations:
   - Verify degrees and rates of physical separabilities of distinctive components (major minerals, free-iron grains, amorphous combinations) by direct and combined means (magnetic, electrostatic, sieving, crushing, vibrations, electrophoretic, etc.). Use analog materials and very limited quantities of lunar samples.
2. Glass and ceramics:
   - Apply the extensively developed technologies and basic materials knowledge of terrestrial glasses and ceramics to determine the products and production characteristics for the direct and early use on the Moon and in space of bulk lunar soils, physical separates (mineral, vitreous and metallic), and chemical separates of the soils.
   - Verify the indicated degree and rate of recovery of gases from lunar soils which will be released by heating in melting operations and by means of low-energy desorption processes (extreme oxidizing and reducing conditions at low gas pressures).
3. Chemical processing:
   - Demonstrate the electrorefining and alloying of metallic "free" iron.
   - Demonstrate with simulated lunar soils on the bench-scale level the HF acid leach, ammonium salt fusion, and mixed acid leaching based on adaptations of well-known terrestrial industrial and laboratory procedures for extracting major oxides and elements (O, Si, Al, Mg, Ti, Ca, Fe) from a wide range of bulk lunar soils. Rates of throughput, recycle efficiencies, and separability data should be determined in these demonstration experiments. Implications of reagent composition from native lunar materials should be determined.
   - Recycle chemistry: Investigation of alternative methods of salt splitting or recycling acids and fluorides.
     Topics: Pyrolysis of NH_4F. Conversion of metal fluorides to compounds more readily pyrolyzed – sulfites, formates, oxalates, etc. Conversion to hydroxides with NH_3. Conversion of NaF (from sodium reduction) to Na, HF, and O_2 via NaOH and Castner cell, or from fused fluorides using consumable anodes.
   - Literature studies of methods to recover minor and trace element fractions obtainable from immiscible liquid extraction of magmas (molten fluids) such as would occur in glass production.
4. Electrochemical processing: Investigation of direct high-temperature electrolysis of silicates or other semi-refractory source materials, either as molten systems or dissolved in high-temperature fused salt systems such as fluorides or carbonates.
   Topics: A. Science – Solubility and miscibility limits in specific systems. Distribution coefficients between magmatic and fused salt phases where liquid immiscibility exists. Potentiometric studies of specific elements in molten, fluoride and carbonate systems.
   B. Engineering – Preliminary economic and engineering feasibility studies. Cell materials compatibility studies for magmatic, fluoride and carbonate systems, including container and anode materials. Special emphasis to be directed to finding nonconsumable electrodes.
   C. Establishment of kg-scale electrochemical feasibility tests for molten silicate and fused salt (fluoride and carbonate) systems.

   Systems analyses and operations tests:
   - Examine economic attractiveness of the manufacturing of machines of production (including materials processing devices) and products in a minimum-mass facility using native lunar iron, glass, ceramics, and derived products. Facility should be based on current state-of-the-art semi-automatic numerical production and remote monitoring.
   - Theoretically examine the use of silane-based fuels for use in Moon-Earth liquid-fueled transfer rockets. Determine whether lunar hydrogen can be obtained in sufficient quantities to transport Moon materials back to Low-Earth Orbit, significantly reduce Earth-lift requirements of propellants, and provide feedstock in LEO for materials industries.
   - Examine construction of large-volume sublunar living and manufacturing chambers by melting of the lunar soil into self-sealed lava tubes.

suited to conventional mineral separation techniques involving electrostatic or electromagnetic (cf. Inculet and Criswell, 1979), flotation, or density characteristics. The low gravity of the Moon and the absence of gravity in space should be extremely beneficial to the electrophoresis process because settling is either minimal or nonexistent (Henderson and Vickery, 1976; McCreight, 1977; Saville and Ostrach, 1978; Vanherhoff and Micale, 1976; Weiss et al., 1979). Electrophoretic separation of minerals is only moderately temperature-dependent, thus eliminating another source of potential difficulty (Bier, 1978). Finally, the isoelectric points of lunar minerals have enough variation to ensure extremely efficient separation. A few of these values are tabulated in table 4.13.

Suspension media options are a major research area in the development of lunar electrophoretic separation technology. Aqueous solutions commonly are used for bulk suspension due to the availability and ionization potential of water. For this reason, isoelectric points customarily are defined in terms of aqueous pH. Carbon tetrachloride also has been used as a high-voltage zone electrophoresis medium. Aqueous and carbon tetrachloride suspensions may be impractical for lunar separation facilities because of the relative scarcity of carbon, hydrogen, and chlorine on the Moon. Further, leaks in the system would be devastating if all major reagents must be imported. Some means must be found to thoroughly dry the output stream and to return these fluids to the bath. Alternative bulk media derived wholly from lunar materials might possibly be devised; for instance, silane or low-temperature basalt slag suspension fluids. The problem is hardly trivial, though it appears to present no fundamental insurmountable technological barriers.

Using high-voltage zone electrophoresis, only one medium is needed for a wide range of minerals. Other techniques require the ionic concentration of the operating fluid to be varied to match the isoelectric point (expressed in activity or concentration of a particular ion analogous to aqueous pH) of the desired mineral for each electrophoresis cell. This seems an unnecessary complication.

Other problem areas include fused mineral grains and iron coatings. Fused mineral grains, which are relatively common in lunar soil (10-20% by volume, Criswell, personal communication, 1980), are not amenable to electrophoretic separation because the isoelectric points are ill-defined. This may actually be beneficial since only pure mineral grains will be separated, thus eliminating the need for additional more complicated separation techniques. Iron coatings on mineral grains caused by "sputtering" also may be present in lunar soil. If coatings are thicker than about 30 nm, efficiency of the electrophoretic process decreases. Fortunately, the very existence of these coatings is open to some question, and there is no evidence at present that they are thicker than 10 nm. Also, if the coatings do not entirely cover the grain surfaces the problem of lessened electrophoretic activity is significantly reduced.

Electrophoretic separation appears highly adaptable to automation. The process itself is largely self-regulating and the collection of separated minerals appears to be a trivial robotics task. An automated biological electrophoresis system already has been designed and is under construction (Bartels and Bier, 1977).

An automated high-voltage zone electrophoretic separation system for lunar materials might require a large tank with two electrodes and filter paper (perhaps comprised of spun basalt fibers) suspended between them. The tank would be filled with some liquid medium closely matching the isoelectric point of a particular mineral of interest. After insertion of lunar soil a direct current electric field is applied to initiate separation. Grains of the mineral whose isoelectric point has been selected plate out near the center of the paper, the other minerals in discrete bands nearby.

TABLE 4.13.– AQUEOUS ISOELECTRIC POINTS OF LUNAR MINERALS

Mineral	Aqueous isoelectric point (pH)	Source
Spinel ($MgAl_2O_4$)	9.1	Bloom and Gutmann, 1977
Hydroxyapatite ($Ca_5(PO_4)_3OH$)	7	Bloom and Gutmann, 1977
Fluorapatite ($Ca_5(PO_4)_3(FOH)$)	6	Bloom and Gutmann, 1977
Rutile (TiO_2)	5.8	Bloom and Gutmann, 1977
Fayalite (Fe_2SiO_4)	5.7	Feurstenav and Raghaven, 1978
Olivine ($(Mg\ FE)_2SiO_4$)	5.7	Feurstenav and Raghaven, 1978
Hematite (Fe_2O_3)	4.8	Bloom and Gutmann, 1977
Forsterite (Mg_2SiO_4)	4.1	Feurstenav and Raghaven, 1978
Clinopyroxene (Ca, Na, Mg, Fe_2, Mn, Fe_3, Al, $Ti(Si\ Al)_2O_6$)	2.7	Feurstenav and Raghaven, 1978
Anorthite ($Ca(Al_2Si_2O_8)$)	2.4	Feurstenav and Raghaven, 1978
Albite ($Na(Al\ Si_3O_8)$)	1.9	Feurstenav and Raghaven, 1978

Individual mineral species are then extracted by robot scoops as the filter paper rolls continuously through the tank.

The proposed automated mineral separator consists of an input port, a suspension tank, two electrode cells, a bond of basalt fiber filter paper, a spectral scanner calibration unit, robot extraction scoops, and repository bins. These components are illustrated in figure 4.1. The sequence of automated operations, as suggested by figure 4.2, is roughly as follows:

(1) Lunar soil is introduced via the input port into the suspension tank.

(2) Lunar soil goes into suspension and begins to separate and move towards the electrodes.

(3) Individual mineral species move towards the electrodes along paths and with velocities which are a function of their electrophoretic potential.

(4) Various mineral species are trapped and plated onto a bond of filter paper continuously rolled through the suspension tank. The paper is connected to both electrode cells. Each mineral phase plates out in a unique area which is a function of the electrophoretic potential of that phase, resulting in discrete bands of pure minerals arranged across the filter paper.

(5) The paper is rolled through the extraction module where the width and composition of each band of trapped grains are measured and verified by a spectral scanner and vision module. Robot scrapers remove individual mineral phases and deposit them in receptacles.

The suspension, mobility, separation, and plating or entrapment steps in this process are self-regulating. The only steps requiring new automation are input, calibration, and extraction. The separator most probably can be scaled up to the requisite size for any given throughput rate, as present-day electrophoresis cells vary a great deal in capacity. The ratio of the volume of suspension medium to the volume of suspended soil can be as high as 1:1 (Micromoretics, Inc., personal communication, 1980).

Metallurgy of basalt. The occurrence of large quantities of tholeiitic (olivine-poor) basalt on the Moon has focused attention on its "metallurgy" (Kopecky and Voldan, 1965; Kopecky, 1971) and on its possible uses as a material for SMF construction. Early work in France involved substituting melted basalt for glass and was not directed toward improving the product over the raw material. German researchers advanced another step by evolving a technology

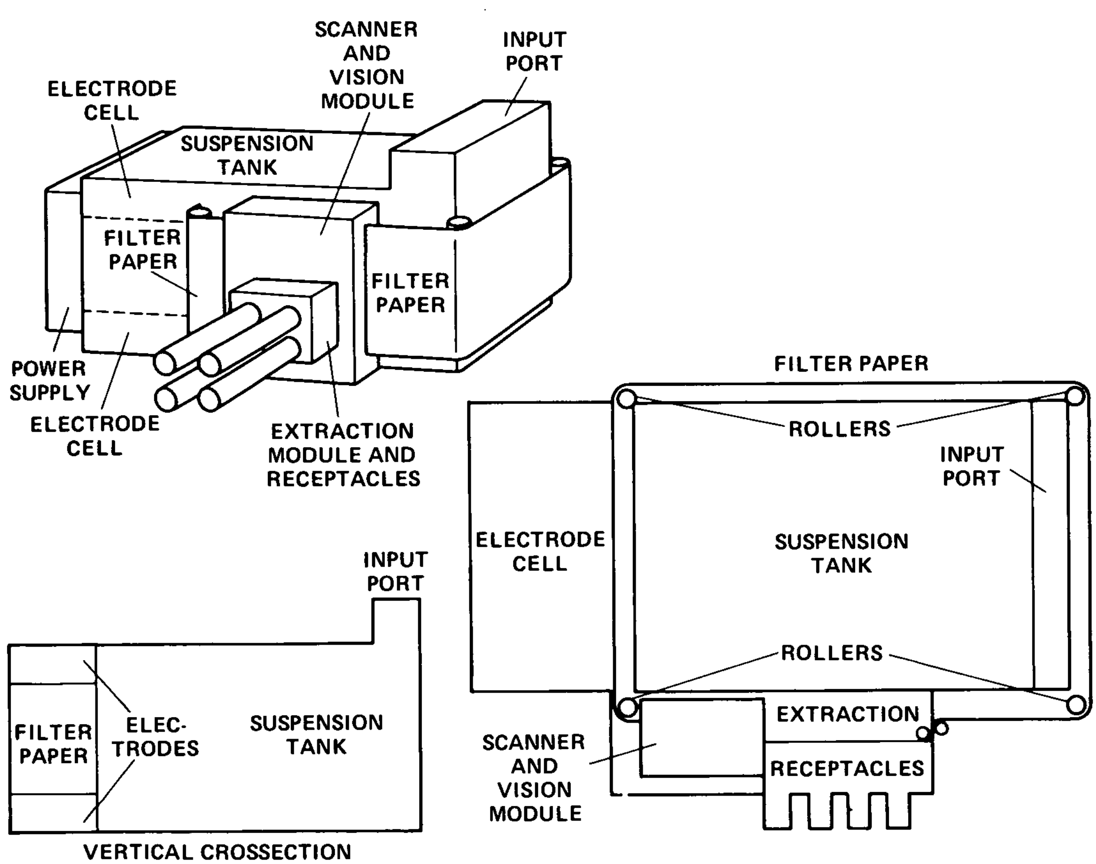
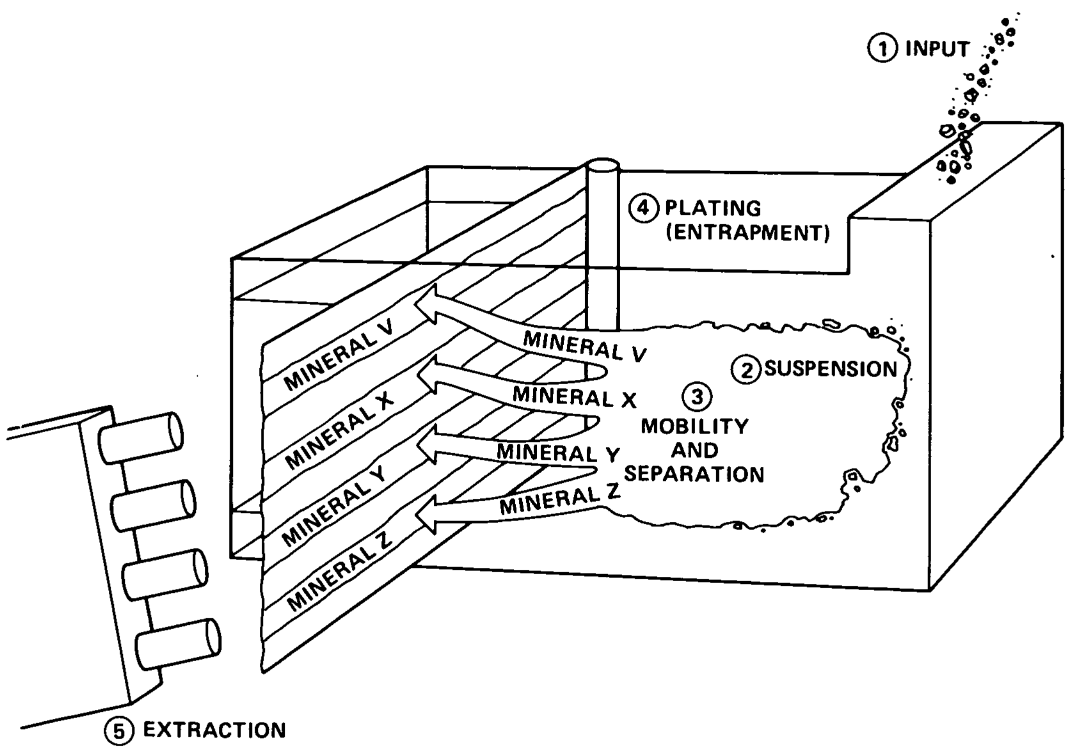


Figure 4.1.– Components of the proposed automated electrophoretic lunar materials separator.



Fgiure 4.2.– Sequence of operations in the proposed electrophoretic lunar materials separator.

for recrystallizing the melt and casting it into simple shapes. The Soviet Union began experimentation with basalt in the 1930s at the Moscow Rock Foundry Works. Processed basalt currently is being manufactured either on a pilot or factory scale in Czechoslovakia, Poland, Sweden, Italy, and the United States. Czechoslovak Ceramics distributes its products mostly to Sweden and England (see fig. 4.3). Many basic patents are held by Mr. H. L. Watson of the now-dissolved Compagnie Generale du Basalte in France.

From laboratory studies and operational experience, raw feedstock basalt should contain pyroxene ($(Ca,Mg,Fe)SiO_3$) in excess of 60%, as it imparts desirable qualities (such as resistance to abrasion, mechanical strength, and chemical resistivity) to the recrystallized mass. Magnetite (Fe_3O_4) and olivine ($(Mg,Fe)_2SiO_4$) also are important because they induce crystallization, but their concentration should not exceed 10%. Higher fractions would reduce the SiO_2 content, leading to the formation of larger crystals that promotes bursting on annealing. (Also olivine, which has a high melting point and thus is difficult to melt, would not dissolve in the short time available for fusion, especially if present as large crystals.) Feldspars ($(Ca,Na)Al_2Si_2O_8$) influence the viscosity and regulate the rate of crystallization. Nepheline ($NaAlSiO_4$) and plagioclase feldspars should be present within the ratios 1:1 to 1:3, with a total content of about 20%. Other rock types such as melaphyres (alkali feldspars) and tephroites (Mn_2SiO_4) have been investigated (Kopecky and Voldan, 1965), but technological difficulties prevent their exploitation at present.

In addition, the material must be fine-grained, homogeneous, unweathered, nonporphyritic, and uncontaminated. A melting temperature range of 1500-1600 K must be associated with a relatively low viscosity (100-1000 poises) in order to cast well. The casts should recrystallize easily in a fine-grained state and not crack after cooling. Favorable factors for lunar basalt include the uncontaminated, unweathered nature of the material as well as an extraordinarily low viscosity.

However, little work has been done to assess certain other factors which might affect lunar basalt casting. For instance, in the manufacture of cast and sintered basalt different successions of minerals crystallize out depending upon the rate of cooling of the melt. By slow cooling and annealing of the casts the following succession is observed: magnetite, olivine, monoclinic pyroxene, plagioclase, then monoclinic amphibole. With rapid chilling, involved in the sintering process, the succession is: magnetite, pyroxenes, amphibole, olivine, and finally plagioclase. Inasmuch as crystallization of the castings depends on melt viscosity, control of that viscosity determines the quality of the final product. Turbulent flow arising from very low viscosity enhances the production of crystals of unequal size and creates swirls in the finished coating, so silica may have to be added to increase the viscosity of thin lunar basaltic melts. On the other hand, excessively high viscosities produce an undesirable laminar structure. The optimum is defined by a Reynolds number of about 1000. On the Moon, reduced gravity should slightly improve the casting

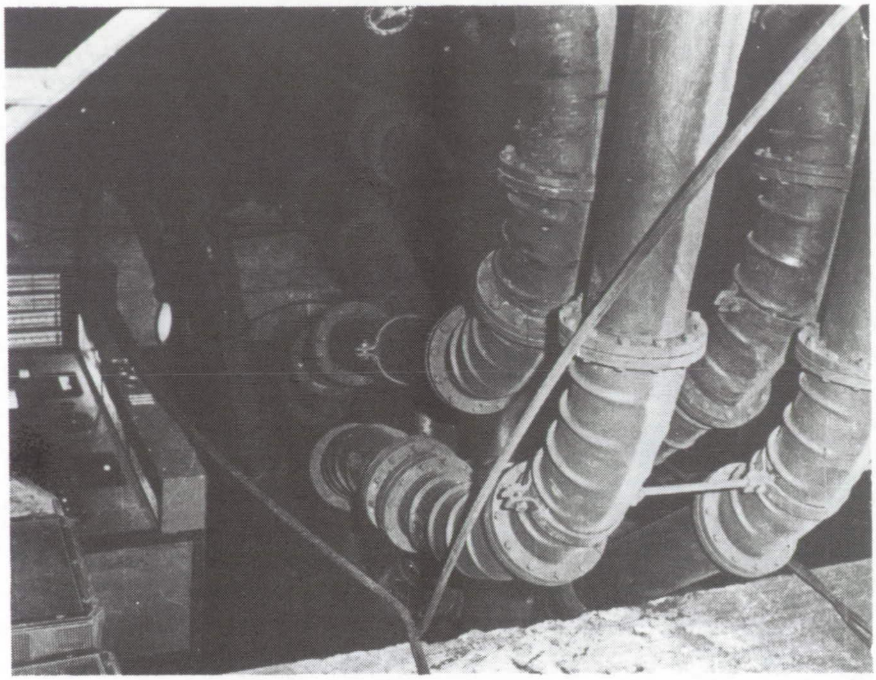


Figure 4.3.– Cast basalt pipe used in coke transfer. (Courtesy of Czechoslovakia Ceramics, Inc., Prague.)

process by reducing the onset of turbulence for a given crystal size. Stokes' equation would apply to a higher value for the terminal velocity of particles, therefore, laminar flow on the Moon would persist at higher terminal velocities than on Earth. Perhaps the effect of gravitational separation of mineral phases often seen during melting, and the inhomogeneities produced in casting, would also be less apparent in lunar cast basalt.

The results of laboratory gradient melting studies by Kopecky and Voldan were applied to the manufacture of cast basalt. The low crystallization speed of plagioclase (3-10 min) prohibits the crystallization of this mineral and it persists as a residual glass phase. Other newly formed crystalline phases of the pilot plant closely resemble the gradient furnace products, except that the cast basalt minerals are more skeletal and dendritic. The most apparent feature in cast basalt is the zonality of the product, which is a function of the cooling rate.

In commercial manufacturing operations in Czechoslovakia, the raw material (8-15 mesh basalt) is melted at 1575-1625 K in vertical gas-fired Lehr furnaces, a process similar to open-hearth steel production. The molten material then is conducted into a homogenizer drum where, at carefully controlled and slightly reduced temperatures, the melt begins to crystallize. The subsequent casting is similar to conventional metallurgical techniques except for differences imposed by the greater viscosity and cooling rates. Static casting in the sand molds originally employed produced a product having rough surfaces and poor tolerances. Metal molds (fig. 4.4) have now replaced sand molds and currently are used in the production of tiles, plates, and fittings. Recently, centrifugal casting methods (fig. 4.5) have resulted in a product of superior quality. Annealing furnaces (fig. 4.6) are used to cool the castings from 1100 K to room temperature over a 24-hour period. Careful control of temperature reduction is required to prevent bursting and other imperfections on annealing.

Most of the castings weigh 3-80 kg. The largest, representing the limits of present-day equipment, weighs 300 kg; the smallest is 0.34 kg, a 60-mm diameter ball. Tiles usually are made in thicknesses of 25-40 mm; pipe walls typically are 15-20 mm thick, with a maximum of 50 mm. The lower limit of thickness is determined by the rate of heat loss and the danger of vitreous solidification. Research is needed on the effects of reduced gravity and on the maximum mass of various castings.

The sintering process is similar to that employed in powder metallurgy (see sec. 4.3.1). The basalt frit made from molten metal is finely ground (1600 mesh), impregnated with a plasticizer, shaped under a pressure of 1000 kg/cm^2, then sintered in electric furnaces at 1395-1415 K. Sintered basalt is valuable in the manufacture of small articles such as nozzles, wire-drawing dies, spheres, and other special fabrications.

Basalt fibers for industrial and commercial applications also currently are produced overseas. Basalt fiber research

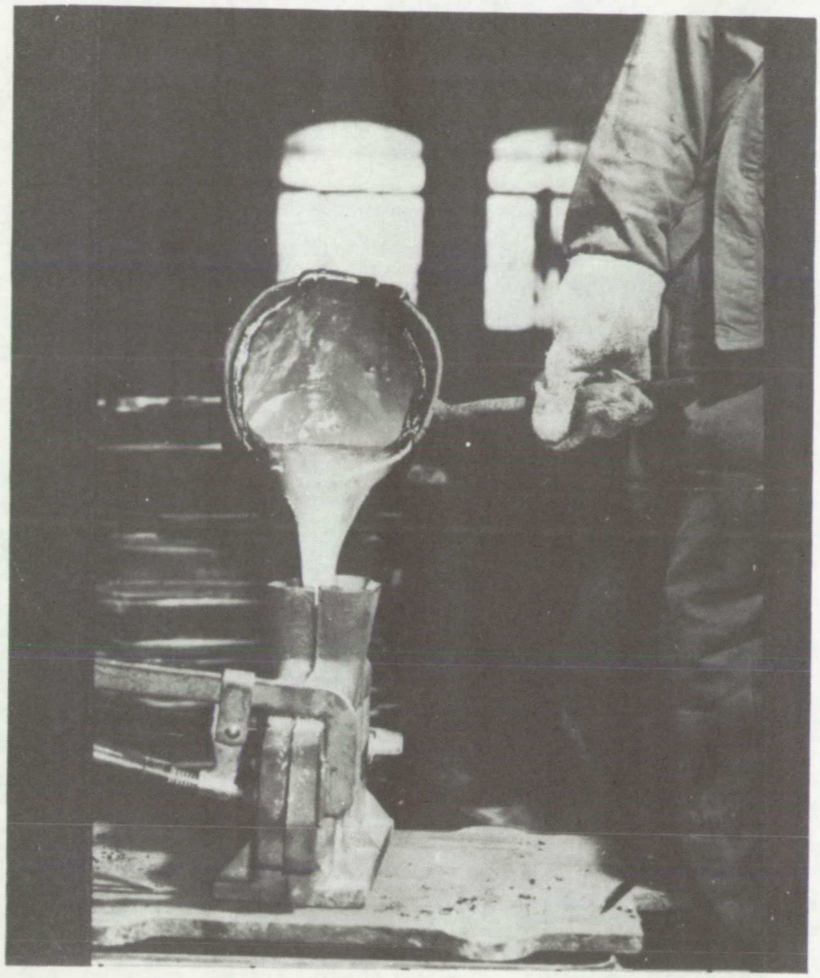


Figure 4.4.– Ladling of molten basalt into metal molds. (Courtesy of Czechoslovakia Ceramics, Inc., Prague.)

Figure 4.5.– Centrifugal casting of basalt. (Courtesy of Czechoslovakia Ceramics, Inc., Prague.)

Figure 4.6.– Basalt casting removed from centrifugal casting drum and positioned for placement into annealing oven. (Courtesy of Czechoslovakia Ceramics, Inc., Prague.)

TABLE 4.14.– CHEMICAL COMPOSITION AND STRENGTHS OF FIBERS FROM BASALTS OBTAINED FROM VARIOUS LOCATIONS

Sample number, rock flow, source	X-6, Lolo Flow, Whitman County, WA	RC-11. Imnaha, Rocky Canyon, ID	K-9068, Elephant Mt., Saddle Mt., WA	K-9064, Middle Yakima, Saddle Mt., WA	O-2, Sweet Home, OR	K9048, Pomona, Saddle Mt., WA	RC-3, Imnaha, Rocky Canyon, ID	K-9017, Lower Yakima, Saddle Mt., WA	BCR-P, Yakima Flow	E-glass Owens-Corning
Analysis, %										
SiO_2	49.10	49.41	50.02	50.49	50.50	50.86	51.41	53.61	54.50	52.20
Al_2O_3	13.80	17.92	13.29	13.62	16.00	15.18	15.14	15.14	13.60	14.80
TiO_2	3.16	2.41	3.48	3.06	2.17	1.69	2.22	1.84	2.20	0.00
Fe_2O_3	2.00	2.00	2.00	2.00	---	2.00	2.00	2.00	2.00	.30
FeO	11.98	9.66	13.27	12.62	13.80	9.21	11.32	9.60	10.50	.00
MnO	.21	.17	.21	.22	---	.19	.21	.18	.18	.00
CaO	9.43	9.06	8.59	8.48	10.00	10.62	9.33	8.43	6.92	18.70
MgO	5.25	5.64	4.28	4.45	4.30	6.49	5.05	4.98	3.46	3.30
K_2O	1.26	.81	1.35	1.54	.35	.80	.68	1.14	1.70	.00
Na_2O	3.09	2.57	2.95	2.93	3.20	2.62	2.28	2.73	3.27	.30
P_2O_5	.68	.35	.55	.58	---	.33	.36	.35	.36	.20
B_2O_3	.00	.00	.00	.00	.00	.00	.00	.00	.00	10.20
Temperature of drawing, °C	1325	1300[a]	1250	1250	1250	1250	1250	1250	1360[a]	1250[a]
Speed of drum, rpm	515	250	370	515	370	250	370	515	250	250
Number of samples	40	29	30	18	25	20	20	20	31	23
Average diameter of fiber, μ	13.0	12.2	13.5	9.0	11.4	11.8	11.3	10.2	12.1	12.2
Average tensile strength,										
GPa	1.97	1.99	2.13	2.23	2.08	2.08	2.25	2.45	1.97	2.52
psi	286,000	288,000	309,000	323,000	302,000	301,000	326,000	355,000	285,000	365,000
Young's Modulus,										
GPa	82.76	77.93	77.93	87.59	90.34	82.76	87.59	87.59	71.03	81.38
Millions psi	12.0	11.3	11.3	12.7	13.1	12.0	12.7	12.7	10.3	11.8

[a] Air jet used.

(Subramanian et al., 1976)

programs and demonstration units have been implemented at Washington State University (Subramanian et al., 1975, 1976, 1977, 1978, 1979) and at the University of California at Los Angeles (Mackenzie and Claridge, 1979). Production methods for spinning basalt include: (1) continuous fiber simple extrusion and reeling, similar to standard glass fiber production (Andreevskaya and Plisko, 1963), and (2) staple fiber extrusion augmented by air or steam jets including centrifugal spinning methods (Dubovkaya and Kosmina, 1968). Both methods warrant further research for robotics applications and automated manufacturing (Kato et al., 1978) in lunar environments. The typical composition of spun basalt (in wt %) is represented by SiO_2 (50%), Al_2O_3 (15%), TiO_2 (3%), FeO (11%), Fe_2O_3 (2%), MnO (0.2%), CaO (9%), MgO (5%), K_2O (1%), Na_2O (3%), and P_2O_5 (1%). The fibers are brown in color because of their iron content. Table 4.14 provides a list of compositions of raw feedstock and other fiber characteristics. Tensile strengths are comparable to those of E-glass.

Both continuous and staple fibers can be made from basalt. Continuous fibers are produced using standard glass fiber production equipment. After the feedstock is fused in an electric furnace, the melt is fed to electrically heated platinum-rhodium bushings containing 200-300 perforations. As shown in figure 4.7 (Subramanian et al., 1975), a drum winding pulls the fibers from the platinum-rhodium bushing perforations. Fiber diameter is a function of melt temperature and drum or centrifugal nozzle speed. Temperatures range from 1525-1675 K; thread diameters usually are in the 10-15 μm range, although superfine fibers 0.2-4.0 μm thick reportedly have been manufactured in Russia.

Staple fibers are produced using melting tank furnaces that feed electrically heated centrifugally spun platinum-rhodium bushings. Jets of air or steam moving parallel to a fiber extruded from the centrifugally spun nozzles tear it into short lengths (about 30 mm) which fall onto a porous drum under vacuum. Either continuous or centrifugal spinning staple methods may be applicable for lunar fiber production.

Silanes (organosilicon compounds) have been evaluated as coating materials on basalt fibers to permit adhesion of the fibers to epoxy composites (Subramanian et al., 1976, 1979). The results showed that silane coupling agents are effective in improving interfacial bond strength in basalt fiber-polymer systems and that basalt fiber has excellent potential as a reinforcing fiber for polymer composites. The tensile strength and tensile elastic moduli of epoxy composites of silane-treated basalt fibers are presented in figures 4.8 and 4.9, respectively, as a function of volume fraction V_f.

Processed or machined basalt has been suggested as a logical construction material with which to produce the component parts of large space and lunar structures. The strength of this basalt and of other construction materials must be compared. In table 4.15 the proportional limit, ultimate strength, and modulus of elasticity of sintered basalt are measured against those of carbon steel, cast iron, malleable cast iron, wrought iron, cast aluminum, aluminum alloy 17ST, rolled brass, cast bronze, and drawn copper.

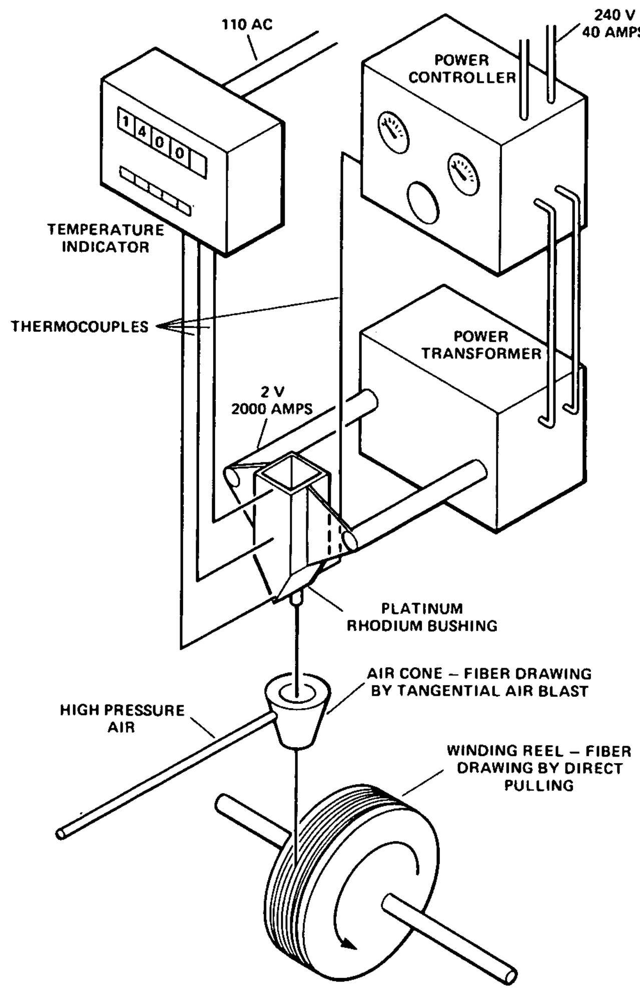


Figure 4.7.– Single fiber drawing equipment for basalt fiber production.

The physical properties of basalt compare quite favorably with those of conventional construction materials. The compressional strength and elastic modulus are quite high; that is, basalt as a construction material is far more rigid than other substances listed, a quality of some importance in large space structures. One drawback is tensile strength, roughly an order of magnitude lower for basalt than other typical construction materials. This problem can be overcome either by designing structures such that basalt components are not exposed to high tensile or extensional stress states or by producing a compound basalt reinforced with fibers. The first alternative is impractical, as large structures contrived to reduce tensile stresses on basalt components

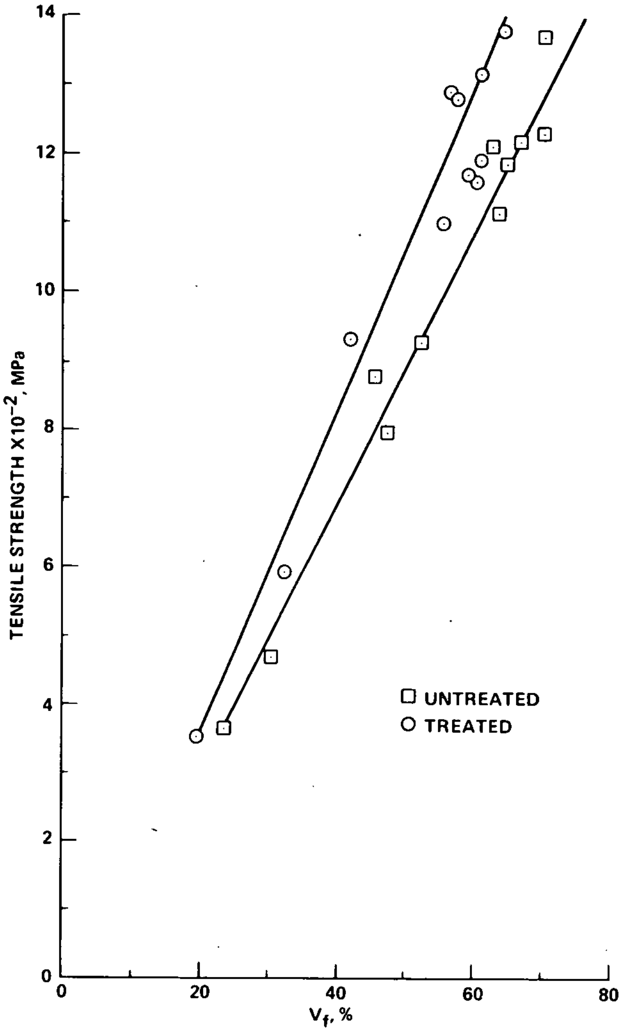


Figure 4.8.– Tensile strength of epoxy resin composite DGEBA reinforced by untreated and silane A-1100 treated basalt fibers.

would be difficult to design and of limited utility. Compound basalts could be prepared by sintering basalt-sodium flux materials and imbedding the melt with a cross-hatched pattern of basalt filaments to increase tensile and shear strength without sacrificing rigidity. (The sodium flux reduces the fusion point of the mixture so that the basalt filaments do not themselves melt.) Finally, the low thermal expansion coefficient (7.7×10^{-6}/K around room temperature) and thermal conductivity of sintered basalt (8×10^{-4} J/m^{2}sK) are very suitable for lunar application, enhancing the structural rigidity of sintered basalt.

One last potential problem is machinability. Cast basalt has a rather irregular surface, a property inappropriate for some construction components, and needs some surface and internal grinding. Also, the hardness of cast and sintered basalt is high, 8.5 on the Moh's scale. A grinding compound of higher hardness is therefore needed, preferably some substances found on the lunar surface. A logical choice is spinel (Moh's value 9.0), which probably can be extracted from lunar soil by an electrophoretic technique.

A summary of possible methods and applications of processed lunar basalt is presented in table 4.16.

4.2.3 Transport to Low Earth Orbit

In the near term two sources of raw materials may be tapped to supply a space processing center in LEO – the Earth itself and the Moon (see fig. 4.10). Earth may provide material, primarily feedstock, by way of the Shuttle and derived vehicles. The possibility of using a land-basea electromagnetic accelerator for ground-to-LEO transport offers the tantalizing promise of greatly reduced supply costs for feedstock payloads able to withstand the 10^4-10^5 m/sec^2 accelerative loads required for direct launch from Earth (Mongeau *et al.*, 1981).

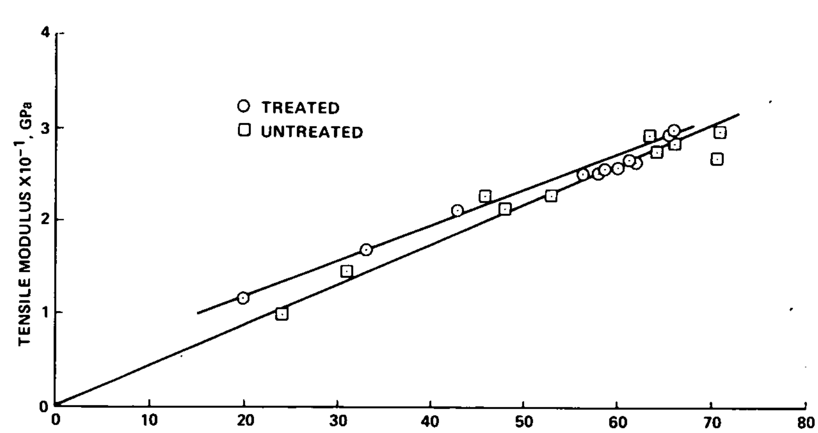


Figure 4.9.– Tensile modulus of epoxy resin composite (DGEBA) reinforced by untreated and silane A-1100 treated basalt fibers.

TABLE 4.15.– COMPARISON OF PHYSICAL PROPERTIES OF BASALT WITH OTHER BULK MATERIALS

Material	Proportional limit, MN/m^2			Ultimate Strength, MN/m^2			Modulus of elasticity, MN/m^2	Source
	Tension	Compression	Shear	Tension	Compression	Shear		
0.2% Carbon steel (hot rolled)	238	241	142	408	248	306	2.0×10^5	1961 AISC Manual
0.2 Carbon steel (cold rolled)	408	414	244	544	416	408	2.0×10^5	
0.8 Carbon steel (hot rolled)	476	491	286	816	503	714	2.0×10^5	
Cast iron	40.8	43.1	---	367	510	---	1.0×10^5	
Malleable cast iron	414	428	156	367	261	326	1.6×10^5	
Wrought iron	204	231	122	340	242	373	1.8×10^5	
Cast aluminum	61.2	74.0	---	88.4	76.1	---	6.7×10^4	
Aluminum alloy 17ST	217	223	143	381	242	218	4.1×10^4	
Rolled brass	170	179	102	374	186	327	5.0×10^4	
Cast bronze	136	140	---	224	381	---	3.3×10^4	
Drawn copper	258	272	156	374	284	---	5.1×10^4	
Sintered basalt	36	550	---	36	550	---	1.1×10^5	Kopecky and Voldan, 1965

Bock et al. (1979) have studied the retrieval of lunar materials to various points in space, using chemical rockets burning lunar LOX and aluminum powder or terrestrial H_2. The objective is to transport from the Moon to cislunar orbital space many times more mass than could be supplied from Earth at equal cost. A particularly appealing proposal for near-term acquisition of lunar resources using chemical propulsion has been suggested by Waldron *et al.* (1979). The potential fuel is lunar silicon and terrestrial hydrogen combined to form silanes, which then are burned as rocket fuel with lunar oxygen. Even if mass drivers supplant this use of lunar-derived propellants for bulk transport, the silane/LOX system, if feasible, would still be useful in trajectory correction (RCS), stationkeeping, and related specialized applications.

The costs and mechanics of STS launch and operations are treated extensively in the literature and will not be reviewed here. Two relatively new proposals – the lunar silane/LOX propellant scenario and the Earth-based electromagnetic catapult – are treated in more detail below. Calculations are presented for the total and net lunar mass that could be delivered to LEO in terms of multiples of the hydrogen needed from Earth.

Lunar supply of a LEO station. To demonstrate early net growth in space the team considered the problem of supplying a LEO station with bulk materials from the Moon. There will be only moderate initial supply from Earth and very limited resupply thereafter. A LEO facility able to accept raw lunar stock and a very small factory able to extract oxygen from and load lunar soil into arriving spacecraft for Moon-to-LEO transport are assumed already to exist. The initial supply base will likely be located at a previously visited Apollo site. A more sophisticated version of the lunar base produces both oxygen and silane (from lunar silicon and Earth-supplied hydrogen). The overall plan requires an Orbital Transfer Vehicle (OTV), a Lander, and a supply of hydrogen from Earth. OTV and Lander are fueled by terrestrial-supplied hydrogen and lunar-derived oxygen or by silane and lunar-derived oxygen. Lander is loaded with lunar soil to be processed and delivers it to the OTV. The OTV returns to the manufacturing facility in low Earth orbit. There, at the SMF, part of the soil is used to produce oxygen (or oxygen and silane) to refuel the OTV and Lander. The remainder is available as raw material for the manufacture of useful output. Either the H_2-O_2 or the SiH_4-O_2 combination allows significant multiplication of resource mass beyond that supplied from Earth.

This scenario could be accomplished according to the following sequence:

(1) The OTV carrying Lander and the required hydrogen leaves LEO with impulse ΔV_1 m/sec.

TABLE 4.16.– LUNAR FACTORY APPLICATIONS OF PROCESSED BASALT

Cast basalt	Sintered basalt	Spun basalt (fibers)
Machine base supports (lathes, milling machines)	Nozzles	Cloth and bedding
Furnace lining for resources extraction	Tubing	Resilient shock absorbing pads
Operations	Wire-drawing dies	Acoustic insulation
Large tool beds	Ball bearings	Thermal insulation
Crusher jaws	Wheels	Insulator for prevention of cold welding of metals
Pipes and conduits	Low torque fasteners	Filler in sintered "soil" cement
Conveyor material (pneumatic, hydraulic, sliding)	Studs	Fine springs
Linings for ball, tube or pug mills, flue ducts, ventilators, cyclers, drains, mixers, tanks, electrolyzers, and mineral dressing equipment	Furniture and utensils	Packing material
Tiles and bricks	Low load axles	Strainers or filters for industrial or agricultural use
Sidings	Scientific equipment, frames and yokes	Electrical insulation
Expendable ablative hull material (possibly composited with spun basalt)	Light tools	Ropes for cables (with coatings)
Track rails	Light duty containers and flasks for laboratory use	
"Railroad" ties	Pump housings	
Pylons	Filters/partial plugs	
Heavy duty containers for "agricultural" use		
Radar dish or mirror frames		
Thermal rods or heat pipes housings		
Supports and backing for solar collectors		

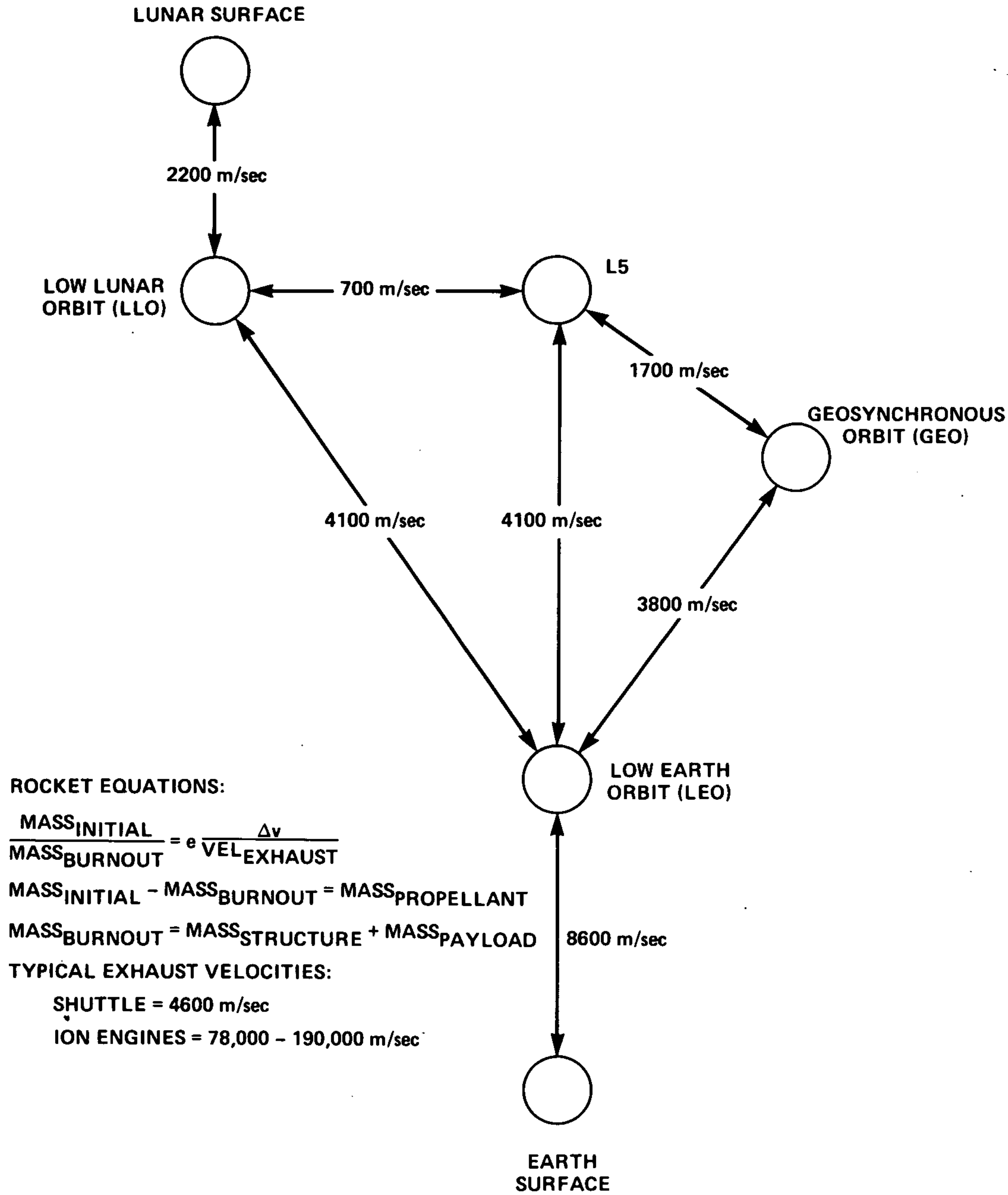


Figure 4.10.– Delta-V's for various orbital transfers.

(2) OTV passes low over the lunar surface (50 km altitude) and releases Lander, then returns to LEO on a free-return trajectory using aerobraking. No propulsion is assumed for any of these maneuvers.

(3) Lander burns fuel (ΔV_2 m/sec) to enter an elliptical lunar orbit with apolune at the point of separation and perilune at the surface of the Moon.

(4) Lander burns fuel (ΔV_3 m/sec) to land and rendezvous with the already emplaced lunar soil processor. Lander arrives carrying only the hydrogen required for a return to LEO.

(5) The lunar processor supplies Lander with native oxygen. If the silane alternative is used, the processor also takes Lander's hydrogen and converts it to silanes (predominantly SiH_4) using lunar silicon.

(6) Lander is loaded with a cargo of lunar soil destined for the LEO manufacturing facility.

(7) Lander lifts off from the Moon (ΔV_4 m/sec) and returns via aerobraking to LEO to rendezvous with the orbiting manufacturing facility.

(8) Lander and OTV are refueled for a return trip to the Moon.

The above procedure has been worked out parametrically without specifying the masses of OTV and Lander. The same fuel and oxidizer are used at each burn. It is desired to determine the incremental cost of one kilogram of lunar payload delivered to LEO which is not needed for fuel in terms of incremental mass lifted to LEO from the Earth. The full mathematical analysis is presented in appendix 4A – only the results are given here.

Let a be the tankage fraction needed to carry the payload from the Moon, B the propellant tankage fraction, and B_H the fraction of the total fuel-plus-oxidizer combination that is hydrogen. If X is as given in equation (2) of appendix 4A, and P is the mass of the payload not needed for propellant replenishment, then the mass of extra hydrogen that must be lifted from Earth to deliver 1 kg of extra lunar payload to LEO (dM_{Hlift}/dP) is given by equation (3) of appendix 4A. The following values are given for H_2-O_2 propellants:

- $c = 4.414$ km/sec ($I_{sp} = 450$ sec)
- $B_H = 1/9$
- $a = B = 0.038$
- $\Delta V_1 = 3.2244$ km/sec
- $\Delta V_2 = 0.84303$ km/sec
- $\Delta V_3 = 1.69147$ km/sec
- $\Delta V_4 = 2.51872$ km/sec
- $X = 0.39718$
- $dM_{Hlift}/dP = 0.2075$,

so the multiplier is $(0.2075)^{-1} = 4.82$ kg of extra payload gained for every kilogram lifted to LEO from Earth. For SiH_4-O_2 propellants:

- $c = 3.463$ km/sec ($I_{sp} = 353$ sec) $B_H = 1/24$
- $X = 0.49420$
- $dM_{Hlift}/dP = 0.12921$,

so the multiplier is 7.739 kg/kg.

If the OTV is eliminated and Lander alone leaves LEO and returns, then for H_2-O_2:

- $X = 0.39799$
- $dM_{Hlift}/dP = 0.20335$, so the multiplier is 4.92;

and for SiH_4-O_2:

- $X = 0.47696$
- $dM_{Hlift}/dP = 0.12395$, so the multiplier is 8.067.

The team concludes that significant multiplication of resources at LEO is attainable if part of the propellant required to run the system is drawn from the Moon. Lunar oxygen production allows 4.82 kg of raw material to be brought to LEO from the Moon for every kilogram of hydrogen lifted from Earth. If the OTV is removed, this multiplier factor rises to 4.92. Production of silanes as well as oxygen may allow 7.74 kg of raw material to be brought to LEO from the Moon for every kilogram of Earth-supplied hydrogen. If no OTV is used, this figure rises to 8.07. (Allowing Lander to complete the round trip without an orbital transfer vehicle increases performance slightly if the fuel for the first propulsive burn is stored in the space allotted to the payload on the return trip.) The foregoing parametric analysis indicates the advisability of continuing with this line of research. A very small initial plant on the Moon could permit the utilization of lunar materials in LEO early in space manufacturing experimentation.

Earth impulse launch supply of a LEO station. The use of launchers to propel material from the lunar surface has been a key element in space manufacturing and colony-building scenarios for many years (Grey, 1977). Even more revolutionary is the concept of an impulse launcher to lift cargo off the surface of the Earth (Mongeau et al., 1981). If payloads are of sufficient size and are projected almost vertically, atmospheric resistance reduces velocity by only about 15% (see Kolm in Grey, 1977). Since the launch must be nearly perpendicular to minimize atmospheric drag, it is not feasible to supply a LEO station directly. (About 7 km/sec of horizontal velocity would have to be added after launch, so there would be no advantage in using an impulse launcher.) But if payloads are lofted to geostationary altitude (GEO), a burn there of only 1.5 km/sec puts the cargo in an orbit tangential to the Earth's atmosphere. Aerobraking then lowers the apogee until a final burn circularizes the orbit and allows rendezvous with the LEO facility.

Although modern rockets are very thermally efficient, only about 0.5-1.0% of the energy originally available in the propellant tanks is finally delivered to the payload; the rest is expended accelerating propellants and vehicle mass. The impulse launcher is vastly more efficient, allowing all but about 3% of the energy required to reach LEO to be imparted to the payload while it is on the ground. The 3% expenditure is made by a booster fired at apogee to raise perigee to the upper levels of Earth's atmosphere.

Two methods of impulse launch have been proposed. The first is a simple version of the rail gun as shown in figure 4.11. It suffers from major inefficiencies (I^2R losses) but illustrates the principle. In this system, current flow through a plasma causes magnetic pressure to be exerted by the arc on the projected base. The second type of impulse launcher uses superconducting coils as suggested by von Tiesenhausen (personal communication, 1980) and Kolm (in Grey, 1979). For a given acceleration and final velocity, the second (induction motor) launcher is 2-3 times longer than the first, since payloads are hurled forward in a bucket and the bucket eventually must be decelerated. The projectile is a 1000 kg mass in the form of an ogive 1.1 m diam and 6.3 m long. The launcher operates at 300 kW average impact power and launches the payload at 11.05 km/sec.

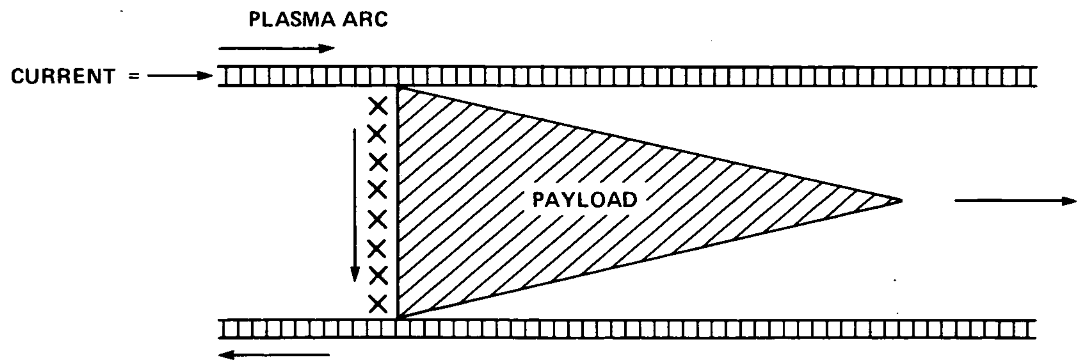


Figure 4.11.– Schematic of a rail-gun impulse launcher.

If 80% efficiency and power storage in homopolar generators between launches is assumed, one shot can take place every 3.5 min. The firing tube is about 1.5 km long for a 5000-g launch, or 2.5 km including bucket-slowing if the linear induction motor impulse device is employed. At 80% efficiency a launch requires 7.63×10^{10} J or 21,200 kW-hr of energy. Electricity costs about \$0.05/kW-hr, therefore the equivalent cost of an impulse launch in terms of power requirements is \$1060.

The projectile slows to 10.22 km/sec by 100 km altitude, the limit of the sensible atmosphere. Ten percent of the launch mass and 16.9% of the launch energy have been lost by this point due to ablation. When the projectile reaches GEO altitude it orients itself horizontally and the solid booster fires, providing a delta-V of about 1500 m/sec. This places the payload on an atmosphere-grazing trajectory allowing aerobraking and orbital circularization. If the solid booster (I_{sp} = 300) has an inert mass of 100 kg and the aerobraking shield is 25 kg, then the net mass delivered to LEO is:

$$(1000\text{-}100)\exp(-1500/9.8 \times 300) - (100 + 25) = 415 \text{ kg}$$

This represents a power cost of just \$2.55/kg. Even if the upper stage motor costs as much as \$100,000, the total expense to LEO is \$304/kg. If the cargo is launched around the Moon to obtain the requisite horizontal velocity by a gravitational assistance maneuver, the mass to LEO is (1000-100) - 25 = 875 kg and the cost of launch energy rises to about \$1100, or \$1.25/kg. Even if guidance and personnel requirements raise this figure by an order of magnitude it still is only 2% of the most optimistic estimate of expected Shuttle costs. The major savings for impulse launching occur because the usual need of accelerating large masses of propellants in addition to cargo is avoided.

4.3 Initial LEO "Starting Kit" Facilities

It seems clear that a wide range of industrially useful feedstocks can be economically provided for LEO and lunar utilization, using materials delivered first from low Earth orbit, later from the Moon, and ultimately from asteroidal and other resources. Sufficient knowledge of lunar materials exists to permit development and implementation of a variety of processing options; similar technology definition for asteroidal materials awaits more detailed information on specific bodies or the development of more generalized processing schemes appropriate to the space environment.

Approximately 10 man-years of research effort already have been devoted to lunar materials processing alternatives (Billingham *et al.*, 1979; Criswell, 1978, 1979; Waldron *et al.*, 1979) on the Moon and in space. The assembly of large structures in space from pre-formed parts has also received much study. Most of this work is reviewed in the MIT (Miller and Smith, 1979) and General Dynamics (Bock, 1979) studies on the manufacture of components for satellite solar power stations using lunar and terrestrial materials processed in factories deployed wholly from Earth.

Options available for manufacturing a wide range of machines or systems of production in space or on the Moon from locally available industrial feedstocks have received far less study. Virtually no effort has been directed toward answering the following questions: (1) What mass fraction of available and foreseeable machines of production can be produced in space from available materials, and (2) how might a hierarchy of production technologies be "grown" in space to create an ever-increasing variety of product and production options? Thus, the growth of industrial capacity can be partially or totally decoupled from terrestrial export of key processing resources.

A broad survey and analysis of a number of basic terrestrial manufacturing processes for their potential nonterrestrial applicability suggests several alternative starting kit scenarios, as described in section 4.3.1. Special attention is then given to "starting kits" in section 4.3.2. A "starting kit" is an initial space manufacturing unit of minimal mass and complexity which, given a supply of feedstock material, can produce second-generation tools (and some products) with which production capability may be gradually expanded further.

4.3.1 *Survey of Terrestrial Manufacturing Processes*

A survey of basic terrestrial manufacturing processes was accomplished by examining a representative sample of reviews of the field (Amstead *et al.*, 1979; Bolz, 1974; Campbell, 1961; DeGarmo, 1979; Lindberg, 1977; Moore and Kibbey, 1965; Schey, 1977; Yankee, 1979) and then generating from this "review of reviews" the taxonomy of approximately 220 manufacturing processes in table 4.17. A listing created in this manner is reasonably comprehensive, though probably not complete. Four major categories emerged: (1) casting and molding (powder metallurgy), (2) deformation (forming and shearing), (3) machining (milling, drilling, and lathing), and (4) joining.

The remainder of this Section consists of reviews and analyses of the processes in each of the four major categories that are potentially useful in space. All methods have been closely scrutinized with respect to a substantial fraction of the criteria listed in table 4.18. Many conventional techniques are rejected because they do not meet these unique requirements for space manufacturing. For instance, most standard machining operations are unsuitable due to the cold weld effect which occurs in a vacuum environment. Many joining techniques require prohibitively large quantities of imported consumables, and thus are inappropriate for a self-sustaining space industrial complex. Some casting and molding practices must be rejected since they require gravitational forces. Many deformation techniques are eliminated because of their tendency to produce inconvenient waste debris.

Casting, powder metallurgy, and plastics. Casting is a process in which melted fluid is introduced into a mold, allowed to cool to produce a solid product, and then this product is ejected. The primary limitation in terms of potential space utilization is the gravity required for all casting processes except permanent mold, centrifugal, die, and continuous casting. However, terrestrial gravity and atmosphere also create most of the major difficulties associated with these techniques on Earth. For example, liquid metals have a lower kinematic viscosity than water, and develop significant velocity by falling only a few centimeters. This condition creates turbulence, erosion of mold materials, and entrapment of air and mold gases. Manipulation of molten materials under controlled, low-gravity conditions and in vacuum may provide significant advantages (Adams, 1977).

There are two basic approaches to casting. The first, expendable mold casting, is the simplest process and the least likely to go wrong. However, gravity is necessary to feed fluid into the mold. It is not easy to replace gravity feed because expendable mold castings tend to be fragile; any type of pressure feed will likely damage the mold and ruin the final product. Another problem is that expendable molds draw heavily on inputs comparatively difficult to supply nonterrestrially. Some materials for temporary molds, such as sand in sand casting, can be recycled, but processes such as investment casting may require significant Earth inputs to remain viable space manufacturing alternatives.

Nonexpendable mold casting, on the other hand, relies less on the conditions of gravity and pressurized atmosphere. The molds tend to last for a greater number of runs. The main disadvantages are that (1) production devices tend to be large, on the order of tons, and (2) the processes are more complicated than for expendable mold casting. A more complete review of both methods from the standpoint of space applications may be found in appendix 4B.

The key problem appears to be mold/pattern preparation, the heart of the casting process. This problem provides an excellent focus for future artificial intelligence and robotics technology development efforts: A robot which can produce a mold/pattern to close tolerances is required (appendix 5F). Such manipulation might be initially performed via teleoperation, followed by a gradual evolution toward complete automation. Mold/pattern design is a fine art for which some type of expert system may be required for near-autonomous operation. The development of more precise robots with enhanced feedback and access to an expert system for casting technology should alleviate the mold production problem.

Casting processes have some definite advantages with respect to space applications. For instance, expendable mold casting is simple and nonexpendable mold casting requires no gravity. A potential solution to the gravity problem for expendable molds might be the generation of artificial gravity via centrifuge. Centrifuges are capable of applying great pressures, although force gradients inevitably will be present even in large rotating systems. Research is needed to identify and circumvent the difficulties of mold/pattern production in space.

Another casting/molding manufacturing technique is powder metallurgy. In this process, primary material is powdered and then placed in a suitable mold or extruded through a die to produce a weakly cohesive part. High pressures and temperatures then are applied to fuse powder particle contact points until a sufficient flow of material closes all pore spaces. Powder metallurgy can be conducted in a minimum facility able to produce an ever-widening range of increasingly complex parts and tools (Jones, 1960). A considerable theoretical and applications knowledge base already exists to help extend powder technologies into space (Bradbury, 1979).

Any material which can be melted can be powdered. Reformation does not necessarily require complete liquefaction, so the usual "phase rules" of melting may be ignored. The formation process thus has much greater flexibility than casting, extrusion forming, or forging. Controllable characteristics of products include mechanical, magnetic, porosity, aggregation, and alloying properties of metals and nonmetals. Many useful production options are

TABLE 4.17.– TAXONOMY OF MANUFACTURING PROCESSES

I. Casting and molding
  - A. Casting
    - 1. Sand
    - 2. Plastic mold
    - 3. Shell mold
    - 4. Investment (lost wax, precision)
    - 5. Permanent mold
    - 6. Centrifugal
    - 7. Die
    - 8. Slush or slurry
    - 9. Full mold
    - 10. Low pressure
    - 11. Continuous
  - B. Molding
    - 1. Powered metal
      - a. Compaction plus sintering
    - 2. Plastics
      - a. Injection
      - b. Compression
      - c. Transfer
      - d. Extrusion
      - e. Blow
      - f. Rotational
      - g. Thermoforming
      - h. Laminating
      - i. Expandable bead
      - j. Foam
      - k. Rotomolding
      - l. Thermoforming
      - m. Vacuum plug assist
      - n. Pressure plug assist
      - o. Matched mold

II. Deformation (forming and shearing)
  - A. Forming
    - 1. Forging
      - a. Smith
      - b. Hammer
      - c. Drop
      - d. Press
      - e. Impact (see also extrusion)
      - f. Upset
      - g. No draft
      - h. High-energy-rate
      - i. Cored
      - j. Incremental
      - k. Powder
    - 2. Rolling
      - a. Shape
      - b. Ring
      - c. Transverse
      - d. Orbital
      - e. Cross-rolling
      - f. Thread
    - 3. Stretching (expanding)
    - 4. Drawing (shrinking) of wire bar or tube
      - a. Embossing
      - b. Coining
      - c. Stamping
      - d. Sizing
      - e. Redrawing
      - f. Bulging
      - g. Necking
      - h. Nosing
      - i. Ironing
    - 5. Deep drawing
    - 6. Swaging
    - 7. Extrusion
    - 8. Spinning
    - 9. Bending
    - 10. Miscellaneous other
      - a. Peening
      - b. Guerin process
      - c. Wheelon process
      - d. Magnetic pulse
      - e. Explosive
      - f. Electroforming
      - g. Staking
      - h. Seaming
      - i. Flanging
      - j. Straightening
  - B. Shearing
    - 1. Line shearing (slitting)
    - 2. Blanking
    - 3. Piercing or punching
    - 4. Follow-up on #2 and #3
      - a. Trimming
      - b. Shaving
      - c. Notching
      - d. Perforating
      - e. Nibling
      - f. Dinking
      - g. Lancing
      - h. Cutoff

III. Machining (material removal)
  - A. Milling
    - 1. Peripheral (slab)
    - 2. Face
    - 3. Chemical
  - B. Turning
    - 1. Facing
    - 2. Boring
    - 3. Spinning (flow turning)
    - 4. Knurling
    - 5. Cutoff (parting)
  - C. Drilling
    - 1. Reaming
    - 2. Countersinking
    - 3. Tapping
  - D. Sawing
    - 1. Filing
  - E. Broaching
  - F. Shaping
    - 1. Horizontal
    - 2. Vertical
    - 3. Special purpose
  - G. Planning
    - 1. Double housing
    - 2. Open-side
    - 3. Edge or plate
    - 4. Pit-type
  - H. Grinding (abrasive machining)
    - 1. Abrasive jet machining
    - 2. Honing
    - 3. Lapping
    - 4. Superfinishing
    - 5. Barrel finishing
    - 6. Vibratory finishing
    - 7. Spindle finishing
    - 8. Abrasive belt
    - 9. Polishing
    - 10. Buffing
    - 11. Burnishing
    - 12. Grit- or shot-blasting
    - 13. Tumbling
    - 14. Wire brushing
    - 15. Electropolishing
    - 16. Electro-chemical grinding
  - I. Routing
  - J. Hobbing (hubbing)
  - K. Ultrasonic
  - L. Electrical discharge
  - M. Electron beam
  - N. Electrochemical

TABLE 4.17.– CONCLUDED

O. Chemical
P. Photochemical
Q. Laser beam

IV. Joining
- A. Welding
  - 1. Arc
    - a. Shielded metal
    - b. Gas metal
      - 1. Pulsed
      - 2. Short circuit
      - 3. Electrogas
      - 4. Spray transfer
    - c. Gas tungsten
    - d. Flux-cored
    - e. Submerged
    - f. Plasma arc
    - g. Carbon arc
    - h. Stud
    - i. Electroslag
    - j. Atomic hydrogen
    - k. Plasma-MIG (metal inert gas)
    - l. Impregnated tape
  - 2. Oxyfuel gas
    - a. Oxyacetylene gas
    - b. Methylacetylene propadiene (MAPP)
    - c. Air-acetylene
    - d. Oxyhydrogen
    - e. Pressure gas
      - 1. CO_2
  - 3. Resistance
    - a. Spot
    - b. Projection
    - c. Seam
    - d. Flash butt (flash)
    - e. Upset (butt)
    - f. Percussion
  - 4. Solid state
    - a. Ultrasonic
    - b. Explosive
    - c. Diffusion
      - 1. Hot press
      - 2. Isostatic hot gas
      - 3. Vacuum furnace
    - d. Friction
    - e. Inertia
    - f. Forge
    - g. Cold
    - h. Roll
  - 5. Electron beam
  - 6. Laser beam
    - a. Solid-state
    - b. Axial-flow gas
    - c. Cross-flow gas
  - 7. Thermit
  - 8. Induction
    - a. Low frequency (50–450 Hz)
    - b. High frequency (induction resistance; 200–450 kHz)
  - 9. High frequency resistance
  - 10. Electromagnetic
  - 11. Flow
- B. Brazing
  - 1. Torch
  - 2. Induction
  - 3. Furnace
  - 4. Dip
  - 5. Resistance
  - 6. Infrared
  - 7. Vacuum
- C. Soldering
  - 1. Iron
  - 2. Resistance
  - 3. Hot plate
  - 4. Oven
  - 5. Induction
  - 6. Dip
  - 7. Wave
  - 8. Ultrasonic
- D. Sintering (of powdered metals)
- E. Adhesive bonding (incomplete)
  - 1. Thermo-setting and thermoplastic
    - a. Epoxy
    - b. Modified epoxy
    - c. Phenolics
    - d. Polyurethane
  - 2. Adhesive alloys
  - 3. Miscellaneous other powders, liquids, solids, and tapes
- F. Metal fasteners
  - 1. Screws
  - 2. Nuts and bolts
  - 3. Rivets
  - 4. Pins
    - a. Cotter
    - b. Groove
    - c. Tapered
    - d. Roll
  - 5. Retaining rings
  - 6. Quick-release
- G. Stitching
- H. Stapling
- I. Shrink fitting
- J. Press fitting
- K. Plastic
  - 1. Hot-air-welding
  - 2. Friction
  - 3. Heated metal plate
  - 4. Solvent
  - 5. Dielectric
  - 6. Magnetic
  - 7. Ultrasonic
  - 8. Radio frequency welding

TABLE 4.18.– SELECTION CRITERIA FOR SPACE MANUFACTURING OPTIONS

- Make other options: Can this process be used to manufacture other basic process equipment?
- Productivity: Is the production rate adequate for the intended purpose? Production rate should be high relative to machine mass.
- Required consumables: What materials are consumed by the process (e.g., gasoline and oil for internal combustion engines)? Note that electrical power is not considered a "consumable" in this analysis.
- Production energy: How much electrical power, fuels, and other energy resources are required to operate the process? (Some figures in these analyses may be underestimates by a factor of 2–4, as they indicate power input to or output from a final stage rather than the total power required by the system.)
- Preparation steps: What is involved in making the process machine(s) and in preparing materials for processing by such machines?
- Production environment: What special environmental characteristics are necessary in order to allow the process to operate effectively? Of particular concern are atmospheric pressure (can the process operate in a vacuum, or is some form of atmosphere required?) and gravity (can the process operate in zero-g, or low lunar gravity, or is terrestrial gravity necessary or desirable?).
- Automation/teleoperation potential: Is it feasible to consider automating the process, or at least operating it manually from a remote location?
- People roles: What roles must people play, if any, either on Earth, the Moon, or in space?
- R&D required: Does the process appear to have a good potential for nonterrestrial use, and what research and development (R&D) steps may be necessary to enhance the viability of the process in such a setting? (Techniques to be used for production in the early phases of space manufacturing should be testable on Earth or in early LEO systems.)
- Tukey ratio: What fraction of the amount of materials required to utilize a process can be obtained from nonterrestrial sources as opposed to terrestrial sources? (Inverse of mass multiplication ratio.)

possible through powder metallurgy. For instance, cold welding and porosity control are two aspects which can more easily be manipulated in space than on Earth.

Cold welding first was recognized in the 1940s as a widespread effect between like metals. If two flat, clean surfaces of metal are brought into contact, they join at the molecular level and the interface disappears. Cold welding is strongly inhibited by surface flaws such as oxide layers, especially in those which are softer than the parent metal. Such films do not form quickly on fresh metallic surfaces of grains manufactured in the hard vacuum of space, as they do on Earth. Thus, metal powders will naturally form very cohesive structures upon contact or slight compression.

On Earth it is difficult to achieve porosities of less than 10% in uncompressed or lightly compressed powder forms. Significant changes in dimensions of parts may occur following a sintering or pressing operation. Theoretically, it should be possible to achieve arbitrarily low porosities by combining grains of many different sizes. However, this is not practical on Earth due to gravitational separation effects. In space, and to a lesser extent on the Moon, gravity effects can be so drastically reduced that uncompacted porosities of less than 1-3% may be possible. As an added benefit, in space individual parts can be gently transported to heating or pressure modules without the danger of fragmentation by gravity or rough handling.

Sintering, an increased adhesion between particles resulting from moderate heating, is widely used in the finishing of powder parts. In most cases the density of a collection of particles increases as materials flow into grain voids, and cause an overall size decrease in the final product. Mass movements permit porosity reduction first by repacking, then by evaporation, condensation, and diffusion. There are also shift movements along crystal boundaries to the walls of internal pores, which redistribute internal mass and smoothen pore walls.

Most, if not all, metals can be sintered. Many nonmetallic materials also sinter, including glass, alumina, silica, magnesia, lime, beryllia, ferric oxide, and various organic polymers. A great range of materials properties can be obtained by sintering and subsequent reworking. It is even possible to combine metals and nonmetals in one process. Solar energy may be used extensively for sintering operations in space.

Several techniques have been developed for the powdering of metals. Streams of metal can be atomized with or without gases; thrown against rotating surfaces and sprayed out; thrown off high-speed rotating wheels (especially those being melted as source material); projected against other

streams of metal, liquids such as water, or gases; or electrified. Solar thermal energy may be used in any of these processes, which represent the major energy-intensive step in powder metallurgical manufacturing.

A very large range of products is possible. Virtually any item which can be manufactured by forging, extruding or casting can be duplicated either directly or with appropriate reworking. In addition, special articles such as high-strength or highly refractory composites, filaments, linings for friction brakes, metal glasses, heat shields, electrical contacts, magnets, ferrites, filters, and many other specialized products can be made. Very complicated parts composed of metal and refractory components are directly producible.

The "flow" nature of powder metallurgical techniques is amenable to automation and remote control at all stages from design through production and inspection. The virtually complete separation of the major energy input stages from the design embodiment stage permits the early use of precise but low-force-level devices for near-final shaping. Powder metallurgy can use lunar iron and aluminum, is appropriate for vacuum manufacturing, is insensitive to particle or photon radiation, and can take advantage of zero- and reduced-gravity conditions. It is worth noting that vapor deposition of materials can also be considered as an alternative or supplemental process to powder metallurgy in some applications – such as the production of sheets or large areas of metals. An extended discussion of powder metallurgy appears in appendix 4C.

Plastics are mostly hydrocarbon-based. Raw materials necessary for their preparation are relatively rare in lunar soil. Hence, they must be extracted from bulk materials of carbonaceous chondritic asteroids or eventually from the atmospheres of other planets, their moons, or the solar wind, or else be brought up from Earth. Except for special uses in critical cases, it does not make sense to plan the extensive utilization of plastics in the early phases of space industrialization. These substances may be replaced by sintered or pressure-formed metals or by ceramic parts in many applications. A critical new research area is the possibility of replacing plastics in resin and composite applications with materials derived primarily from inorganic elements found in lunar soil in greater abundance (Lee, 1979).

There exists a great commonality between forming techniques in powder processes and in plastics. In addition, powder techniques are capable of making most, if not all, of the equipment necessary for plastics forming. Thus, if supplies of hydrocarbons ever should become more easily available (see section 4.4.2), the machinery and automation support already would be in place or readily adaptable to this purpose.

Deformation. Deformation includes ten major operations in forming and four in shearing, each of which may be further subdivided as indicated in table 4.17. Major aspects of these processes related to current industrial robot applications and possible automated space manufacturing options are provided in appendix 4D. Highlights of forming processes especially suitable for extraterrestrial utilization are given below. All shearing processes may involve cold welding, and can be performed best by laser beam or other techniques. The team noted that many space structures (such as photovoltaic cells) will be very thin, and thus are more appropriate for laser or E-beam cutting than the comparatively thicker members of typical terrestrial structures.

Regarding forming processes in space, low-weight electromagnetically driven forges may be optimal in view of the special technology created for the electromagnetic mass launcher (Kolm, 1977). At present, "mass-driver" forges are not used on Earth, although magnetic impact welding is being explored industrially at Maxwell Laboratories in San Diego, California.

Powder forging, inasmuch as it would apply to metal- and basalt-sintering options, deserves special consideration for research and nonterrestrial deployment. Powder forging is a relatively new technique able to produce more accurate parts at a lower cost than alternative methods. Unlike other processes, 1600-mesh basalt or lunar "soil" (plus plasticizer) pre-forms could possibly be forged in one operation by a single blow from a set of preheated closed dies. (For terrestrial basalts the temperature would be in the range of 1495-1515 K.) The terrestrial coining process to increase part density by reducing voids may be unnecessary in space, since vibratory or electrostatic quenching techniques may serve the same purpose to optimize forces in powders. Prior to forging, pre-forms are usually coated with graphite to prevent oxidation and provide lubrication. It is not presently known if graphite is required in the vacuum of space, since oxidation versus lubrication tradeoffs have not yet been quantified.

Rolling processes are well-suited to lunar operations, particularly when combined with the ribbon aluminum production line detailed by Miller and Smith (1979; see appendix 4D). In particular, thread rolling is an adaptation of the rolling process that may be ideally suited to high-vacuum manufacturing environments. Conventional die-cutting methods for threaded fasteners produce cutting chips. In space, these chips could contact-weld and foul other equipment if released as isolated fragments. Thread rolling overcomes both problems. Because threads are impressed, no fragments are produced, thus obviating chip vacuum welding. This cold-forming process has long been used in the fastener industry to produce precision threads at high production rates. Other applications have been recently devised, including forming small gear teeth, splines, and knurl patterns. It is possible that backing pieces for the moving and stationary dies needed for thread rolling could be made of cast basalt.

Extrusion has high potential for space manufacturing, as suggested previously in connection with powder metallurgy.

Conventional fabrication methods may be modified to produce lunar spun basalt using advanced automation techniques. An argument for pressurized lunar/space factories can be made if basaltic fiber manufacture is planned, since micron-diameter fibers exhibit vaporization losses under high vacuum (Mackenzie and Claridge, 1979).

A considerable amount of research and development is needed in all phases of vacuum metal extrusion operations. Little is known of dissimilar feedstock/die material cold welding effects, or of enhanced ductility. For basalt melt extrusion, studies are required to determine whether a spun product can be made from low-viscosity lunar basalt either by mechanical drawing or centrifugal spinning (see appendix 4D). Research on the following engineering variables would be useful: (1) Viscosity control; (2) speed of the winding drum; (3) duration of preload remelt; (4) chemistry of raw feedstock; (5) surface tension of melt; (6) temperature coefficient of viscosity; and (7) alternate cooling techniques (other than water). Favorability criteria driving this research include availability of basalt, availability and suitability of electrical energy on the Moon or in space for basalt processing, amenability of robots to high temperature components handling, and usefulness of the product in lunar and cis-lunar systems.

Four of the ten miscellaneous forming methods listed in table 4.17 deserve particular attention because they may be applicable to lunar or asteroid surface operations: shot-peen forming, vapor deposition, magnetic pulse forming, and electroforming. Although electroforming is well-suited to the production of thin-walled vessels it also requires an electrolytic working fluid, which downgrades it to a lower priority than magnetic pulse forming for space manufacturing. (Vapor deposition and electroforming accomplish similar functions.)

Vapor deposition of both polycrystalline and amorphous silicon has been chosen by Miller and Smith (1979) as part of their design for a space manufacturing facility. Their study found deposition rates of 0.5–4.0 μm/min to be a reasonable output for an energy input of 6 kW. Scaling up such procedures could result in the production of single crystal parts such as rivets or other more complex items; hence, vapor deposition provides a possible alternative to powder metallurgy. Hybrid structures, in which thin layers of vapor-deposited structures (such as mirrors) are later stiffened with basalt or basalt composites, are yet another possibility. Vapor deposition also is ideal for gossamer structures. Among the most significant products of this type which could be constructed might be solar sails (Drexler, 1980), devices in the shape of 10-ton spheres 100 nm thick and 3 km diam (see section 4.4.4).

Shot-peen forming is the method of choice for manufacturing airfoil sections with compound curves, where it is desired to form the metal leaving little residual stress. A computer-controlled shot-peen former is currently in use by Wheelabrator-Frye, Inc. of Gardena, California.

Magnetic-pulse forming could draw upon the magnetic accelerator technology now under development for lunar ore transport, as reported in the 1979 Princeton Conference on Space Manufacturing (Grey and Krop, 1979). Forming is accomplished using very intense pulsating magnetic field forces lasting only a few microseconds. Electrical energy stored in capacitors is discharged rapidly through a forming coil. (The capacitor bank currently used in the Princeton mass accelerator research program can supply 4×10^6 W.) In magnetic pulse forming, high-intensity magnetic fields behave much like compressed gases. The metallic workpiece can be uniformly impressed with pressures of up to 340 MN. Three basic methods of magnetic pulse forming are shown in figure 4.12.

Combined with a magnetic driving foil, magnetic pulse forming may be particularly amenable to shaping nonmagnetic superplastic metals (Mock, 1980). A new ternary eutectic of aluminum, zinc, and calcium (Alloy 08050) has been developed by the Alcan Aluminum Corporation which could possibly be pulse-formed into complex shapes. Products currently manufactured using magnetic-pulse forming technology include steering gears, drive shafts, ball joints, shock absorbers, and the assembly of vial caps, potentiometers, instrument bellows, coaxial cables and electric meters.

Electroforming is a modification of electroplating in which metal parts are deposited onto an accurately machined mandrel having the inverse contour, dimensions, and surface finish required of the finished part (fig. 4.13). Thin-walled structures (less than 16 mm) can be fabricated using this technique, with dimensional tolerances to 2.5 μm and 0.5 μm surface finishes (DeGarmo, 1979). Metals most commonly deposited by electroforming include nickel, iron, copper, and silver. Mandrels may be made of aluminum, glasses, ceramics, plastics, or other materials, although if nonmetals are used the form must be rendered electrically conductive. Plating temperatures and current densities must be carefully controlled to minimize internal stresses in the formed product. The final part must be carefully removed from the mandrel if the latter is to be reused. The electroforming process is suitable for automated techniques because few moving parts are involved and the operations are relatively simple.

Electroforming is considered a promising option for lunar and other nonterrestrial applications. Extremely thin-walled products can be manufactured, and mandrels may be prepared from aluminum and sintered/cast basalt. The need for an electrolyte-plating solution requires the electroforming unit to be pressurized and, possibly, operated only in an accelerated frame. The anode plate is consumed during the forming process, but iron and titanium are widely available for this purpose. The electrolyte is recycled (except when leakages occur), and energy constraints appear minimal.

Research on aluminum-coated cast basalt and shell reinforcement by spun basalt is of critical importance in

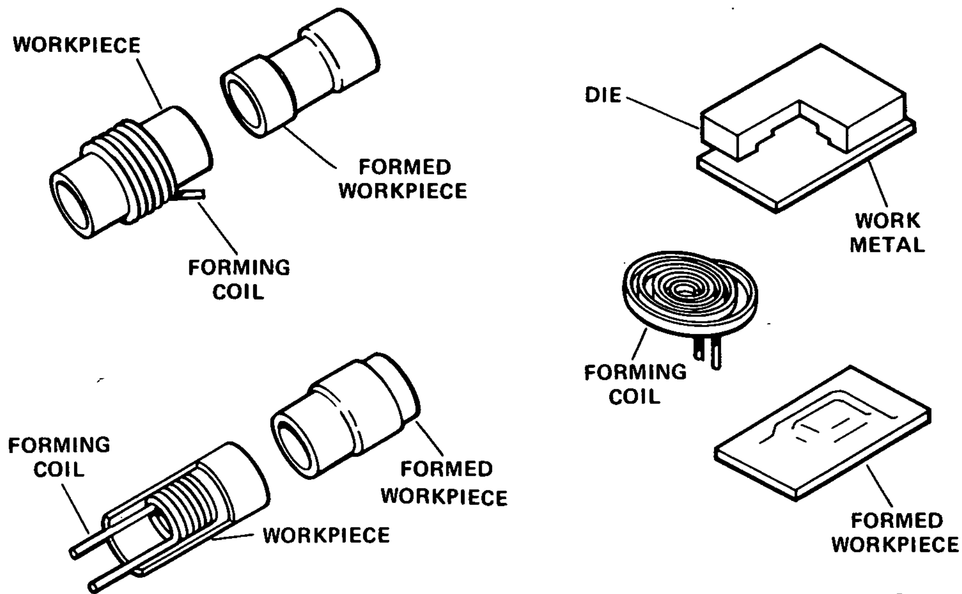


Figure 4.12.– Three basic methods of electromagnetic forming: (a) compression forming, (b) expansion forming, and (c) contour forming.

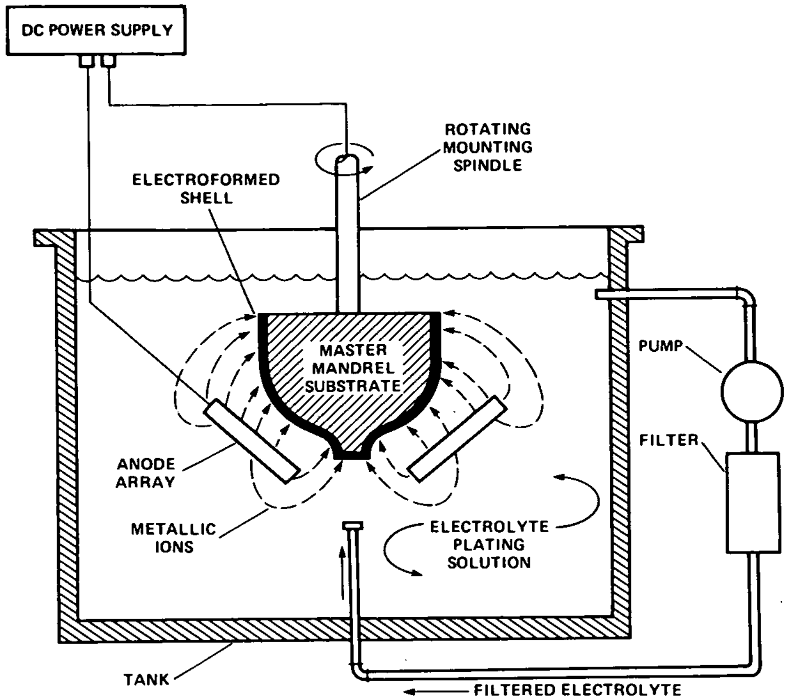
determining the feasibility of the electroforming manufacturing option. Automated processing also should be investigated, particularly with regard to monitoring electrical current densities as a function of metal deposition rate and techniques of mandrel-shell separation (while keeping electrolyte losses to a minimum).

Machining. Machining processes, for the most part, suffer several limitations as manufacturing methods in automated lunar, asteroidal, or orbital factories. The major limitation is the sensitivity of these techniques to the atmospheric configuration. Production efficiency, consumable requirements, and the ratio of machine mass to machine productivity further limit the utility of machining methods (table 4.19). The most promising options currently available are grinding and laser beam machining, techniques which appear to be both useful and adaptable to the space environment.

Milling can be divided into three basic categories – mechanical, chemical, and ion. Mechanical milling of metals in a high vacuum environment is exceedingly difficult with current technology because of the cold-welding effect. The machine mass/production ratio, required consumables, production energy requirements, and mass-multiplication or Tukey ratio are not favorable. Chemical milling is feasible only if reagents are produced from nonterrestrial materials; if not, the mass-multiplication ratio is prohibitive. Also, the efficiency and adaptability of chemical milling in high vacuum are low. Ion milling is also energetically inefficient.

Cold welding also is an inherent problem in turning operations under hard vacuum. In conventional lathing a metal tool is used to fabricate metal stock; hence, cold welding of the tool and stock becomes a serious potential problem. Basalt stock possibly could be turned, or basalt tools designed, to help alleviate this difficulty. Cutting fluids of the conventional type are unsuitable for space and lunar applications due to vacuum sublimation and the need for fluid reconstitution. The production energy, required consumables, and machine productivity ratio for turning are equivalent to those for mechanical milling, as are the required transportation costs.

Cold welding should not occur during grinding unless very fine abrasive grit is employed. However, tool life (e.g., of abrasive wheels) is likely to be short if grinding techniques are used exclusively to shape and mill in the same manner as mechanical milling and turning. Production energy, consumables, and mass/production ratio again are about the same as for mechanical milling. Grinding equipment transportation costs are relatively high, partly because of the massive machines involved that are often larger than milling equipment. Offsetting this disadvantage is the widespread availability of abrasives such as spinel (Al_2O_3) in lunar soil.



Herbert W. Yankee, MANUFACTURING PROCESSES, ©1979, pp. 349, 528, 533. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.

Figure 4.13.– A typical electroforming setup.

TABLE 4.19.– COMPARISON OF BASIC MACHINING PROCESSES

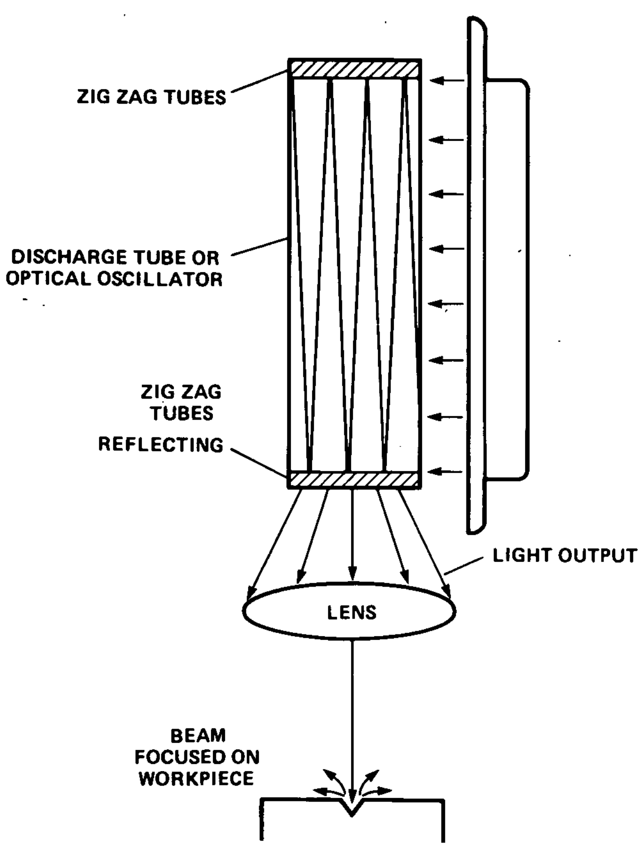
Technique	Production energy,[a] J/kg	Consumables required,[b] kg used/kg produced	Machine mass/productivity,[c] kg/(kg/hr)
Mechanical milling	$2\text{-}21 \times 10^6$	1.1-3.0	10-1000
Chemical milling	(3.1×10^5)[d]	1.01-1.5	0.5-10
Ion milling	$1\text{-}10 \times 10^7$	1.0-1.1	1000
Turning (lathing)	31×10^6	1-2	100-1000
Drilling	$10^4\text{-}10^5$	1.01-1.1	10-100
Grinding	$10^6\text{-}10^7$	1-3	100-10,000

[a]Production energy = energy required/mass of product.
[b]Consumables required = mass of starting materials/mass of product.
[c]Machine mass/productivity = machine mass/(mass of product/hr).
[d]HF milling solution (concentrate) calculated from heat of formation.

Laser beam machining (LBM), first demonstrated in 1960, may prove an extremely useful machining technique in future space manufacturing applications. On Earth, LBM already has attained "production machine" status. There are four types of laser processes theoretically available (solid-state, gas, liquid, and semiconductor), but only solid-state and gas systems are currently used in industrial machining.

Solid-state lasers employ a ruby, yttrium-aluminum-garnet (YAG), or neodymium-doped glass (Nd-glass) crystal rod that converts incoherent light from a krypton arc or tungsten-aluminum flash lamp to coherent optical radiation at a discrete wavelength. Solid-state devices are somewhat wavelength-limited (0.69-1.06 μm; Yankee, 1979) at the present time, and hence are of limited utility as generalized machining tools because the material to be worked must be wavelength-compatible with the laser. Solid-state systems can be employed effectively in some metal processing applications, although efficiency is lower than for gas lasers (Way, 1975) and only pulsating-mode operation is possible.

Gas lasers (fig. 4.14) have discharge and zig-zag tubes filled with argon or carbon dioxide (CO_2) which convert incoherent optical flash lamp radiation to coherent light with a wavelength of about 10.6 μm. Gas lasers are employed in continuous mode for nonmetal machining and in pulsed mode for metal machining. Since metallic substances are highly reflective at the CO_2 wavelength a pulsed beam (10^{-9}-10^{-6} sec bursts; Gross, personal communication, 1980) is needed to penetrate the surface and vaporize the metal (which causes a drop in reflectivity and enhanced energy absorption). The efficiency of metal machining with gas lasers also is not high.



Herbert W. Yankee, MANUFACTURING PROCESSES, ©1979, pp. 349, 528, 533. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.

Figure 4.14.– Typical CO_2 gas laser system.

Laser beam machining has a wide variety of applications in manufacturing. Indeed, some tasks can only or best be accomplished by utilization of laser techniques, such as internal welding, high-accuracy dynamic balancing, case hardening, photoetching, flash trimming, insulation and coating stripping, drilling, measurement and testing to accuracies of ±0.2 μm (Yankee, 1979), flaw detection, and impurity removal (e.g., black carbon inclusion removal in diamonds). Still, LBM remains a micromachining technique and cannot reasonably be expected to replace bulk machining tools such as surface grinders or mills. Lasers are inherently inefficient; LBM requires a great deal of energy to machine comparatively minute amounts of material (Product Engineering, 1970; Way, 1975; Yankee, 1979). The energy of production, required consumables, and machine productivity ratios are unfavorable for bulk mass-fabrication at the present state of the art. Laser research projects funded by DOD and various military agencies have developed tunable helium-neon and xenon-fluoride lasers with relatively high (30%) conversion efficiency. The predicted peak efficiency with minor redesign, according to the developers, should approach 50% (Robinson and Klass, 1980). This is far in advance of contemporary machine shop LBM technology, which offers only 0.1-5% efficiency for solid-state lasers and 10% efficiency for CO_2 gas devices (Belforte, 1979). The advantage of tunable lasers is their ability to match lasing wavelength to the optimal absorption wavelength of the workpiece material.

LBM is very well suited to automated operation. Automatic laser beam machining of plastic flash already has been accomplished (Belforte, 1979; Product Engineering, 1970; Yankee, 1979), and a certain degree of automation is employed in laser welding. Robotics and teleoperated processes could be implemented using current automation technology in laser cutting, measuring, and flaw detection because sophisticated computer vision is not required. Laser operations such as case hardening, shaping, and impurity detection require more sophisticated machine intelligence technology than is presently available. Most LBM techniques today involve a certain degree of teleoperation, which suggests a potential compatibility with broader automation.

The lack of atmosphere and gravity in space are not serious impediments to the use of LBM; in fact, the absence of air may make lasers slightly more efficient in orbit or on the Moon. The only difficulty arising from the lack of atmosphere is plasma removal. In terrestrial LBM a gas jet removes vaporized material (plasma) from the workpiece. The gas jet technique is less feasible in space because it is difficult to generate gases without a great deal of energy. Fortunately, an electrostatic field probably could be utilized to carry away the highly ionized plasma, perhaps using a coil as a kind of "plasma vacuum cleaner."

The major limitation of LBM involves the production of its component parts. A solid-state laser requires a garnet, ruby, or Nd-glass crystal and a halogen, krypton, or xenon flash lamp; a gas laser requires CO_2 or neon gas. These materials are not easily produced in a near-term SMF. For example, 10–100 tons of lunar soil must be processed to produce enough carbon (by sublimation upon heating) for the CO_2 in one laser tube (Criswell, 1980; Williams and Jadwick, 1980; see also appendix 5F). Halogens, xenon, and krypton are not present in sufficient abundance on the Moon to easily produce the flash lamps (Williams and Jadwick, 1980) – at the pulse rates normally employed in solid-state lasers, flash lamp life is between 10 hr and 1 week under continuous operation. Garnet, ruby, and neodymium are not known to be present on the Moon or in space, although spinel (available on the lunar surface) might possibly be used instead of garnet. All these components must be produced in space if the SMF ultimately is to expand in a self-sufficient manner.

Joining techniques. Joining processes of some sort are universally required for manufacturing. Materials joining techniques include welding, brazing, soldering, adhesive bonding, metal fastening, stitching, shrink fitting, and press fitting. Sintering, the joining process associated with powder metallurgy, has already been discussed. Methods for joining plastics are not covered because these materials are inappropriate in the context of early space manufacturing; besides exhibiting poor mass-multiplication ratios due to their hydrocarbon composition, most plastics are volatile and degrade quickly when irradiated by strong ultraviolet light. Many joining techniques used on Earth, and all which appear feasible in space, are readily automatable. A detailed analysis of welding, brazing, and soldering techniques may be found in appendix 4E. A review of adhesives, fasteners and fitting technologies and their possible applicability in SMF operations appears in appendix 4F.

Welding leads to the permanent joining of materials, usually metals, through the application of some suitable combination of temperatures and pressures (DeGarmo, 1979). Approximately 40 different welding techniques have been utilized on Earth (Lindberg, 1977), the majority of which fall into one of five major categories: electric arc welding, oxyfuel gas welding, resistance welding, solid-state welding, and "electronic welding."

Contact welding occurs almost too easily in the vacuum environment of space. Prevention of undesired cold welding is probably a more challenging problem than weld creation during manufacturing. Friction welding may be combined with vacuum welding to facilitate removal of protective coatings from workpieces as well as to enhance bonding.

Electronic welding techniques (electron beam, laser beam, and induction/high-frequency resistance welding) all appear feasible for space applications. NASA has already made considerable effort to investigate these processes, including successful experiments with E-beam and laser beam welding in space (Schwartz, 1979). E-beams and laser beams are extremely versatile technologies. For example, lasers can drill, cut, vapor deposit, heat treat, and alloy, as well as weld an incredible variety of materials. High-frequency resistance and induction methods can also weld many materials with greater efficiency (60% vs 10%; Schwartz, 1979) than lasers can, though lasers and E-beam welders are capable of more precise work.

E-beam devices probably are the easiest of the electronic welders to construct in space. Major requirements include a vacuum, an electron-emitting filament or filament-plus-cathode, deflection plates, and a high-voltage power supply. Filament consumption rates range from 2–1000 hr/filament. Lasers, on the other hand, require precision-ground mirrors, flash lamp and rod (or gas and heat exchanger), etc. These parts are more numerous, more complex, and demand far greater precision of manufacture than those of an E-beam welder. As indicated in the previous section, gases needed for flash lamps in solid-state and gas lasers appear to be in short supply on the Moon, suggesting a poorer mass-multiplication or Tukey ratio. Likewise, neodymium-doped yttrium-aluminum-garnet (Nd:YAG) rods for solid-state lasers are difficult to produce from lunar resources. Both E-beam and laser-beam welders may draw tens of kilowatts of electrical energy in normal operation.

Brazing and soldering differ from welding in that a molten filler metal joins the workpieces at a lower temperature than is required to melt the workpieces themselves. Of the 15 brazing and soldering techniques identified in table 4.17, only vacuum (fluxless) brazing displays exceptional compatibility with the space environment. Compared with vacuum welding, vacuum brazing requires some heat to melt filler material but can bond a greater variety of materials – refractory and reactive bare metals, ceramics, graphites, and composites (Schwartz, 1979).

Under the general classification of "adhesives" are glues, epoxies, and various plastic agents that bond either by solvent evaporation or by bonding agent curing under heat, pressure, or with time. The recent introduction of powerful agents such as "super-glues" that self-cure permits adhesive bonds with strengths approaching those of the bonded materials. Epoxies are combined with metallic and nonmetallic fibers to form composites. Use of such materials, whose strength-to-weight ratios equal or exceed those of many metals, will perhaps constitute the primary application of adhesives in space.

Most glues are carbon-based. The relative scarcity of this element in space suggests that carbon-based glues should be used only where they cannot be replaced by other materials. Boron and carbon, the two most common substances used in composites on Earth, are both rare in space; aluminum and iron fibers may replace them in nonterrestrial fabrication of composites. Energy for fabrication

and glue curing is quite small compared with requirements for welding, and production of iron and aluminum fibers for epoxies should consume less energy than forming solid metal pieces. The major energy expenditure for glues is transportation from Earth. Careful studies are needed to determine tradeoffs between using glues as bonding materials or in composites, and welding or metal-forming requirements.

Space utilization of glues and composites imposes several restrictions yet also offers several advantages. Zero-gravity has little impact – the absence of atmosphere is much more significant. Many resins and glues used on Earth are fairly volatile and deteriorate under vacuum; however, some of them, once cured, are vacuum compatible. The planned early use of composite beams for space construction requires that such compatible bonding agents be available. (Actual use of these agents may need to be under atmosphere.) Many hydrocarbon-based glues weaken under the influence of radiation, and more research is required to develop radiation-resistant adhesives and bonding agents. The unsatisfactory Tukey ratio for current carbon-based adhesives is one of the major hindrances to their use in the long run. Manufacture of composite structural parts from nonterrestrial materials and the possibility of silicon-based bonding agents offer the promise of dramatic increases in mass-multiplication for nonmetallic bonding agents.

Metal fasteners may be grouped into two categories – those producing a semipermanent bond and those requiring either a releasable bond or a sliding bond. Screws, nuts, bolts, rivets, brads, retaining rings, staples and clamps are used for semipermanent fastening of objects when stress bonds or environmental conditions preclude gluing, do not require welding, or where the bond is intended for an indefinite service life. They are semipermanent in that they may be undone for some purpose such as repair. Nonpermanent fasteners include quick-release clips and clamps meant to come off at a specified time, and pins which allow relative movement of fastened parts. Pins are used where movements are not as rigidly constrained, as with bearings.

Metal fasteners are "consumed" during the process of fastening, but since they can be fashioned primarily from abundant lunar iron and aluminum the need for consumables and energy is about the same as that required to fabricate parts from these metals. The machines to manufacture and apply metal fasteners on Earth are serviceable in space applications if modified for zero-g and vacuum-compatibility.

Iron, aluminum, and titanium are abundant on the Moon; such nonterrestrial resource candidates will likely receive early attention. This suggests a favorable Tukey ratio for fasteners. The manufacture of iron and titanium units from lunar or simulated lunar material is a worthwhile early materials-processing experiment. The space environment enables metal fasteners to replace welds in many applications because the loads are generally lower in zero-g. Vacuum welding may strengthen bonds meant to be permanent. Surface poisoning or the use of incompatible metals would be required for breakable bonds.

Stitching is the process of joining parts by interweaving a piece of material through holes in the items to be coupled. The bond is frictional if the linked pieces are not rigid or tension-produced if they are. Interlace fasteners on Earth are made of organic threads of various sizes and compositions and are used mostly for joining fabrics. A major space-related use of interlace fasteners is in the manufacture of fabrics, primarily for space suits. Threads, strings, and ropes have been fabricated from nonvolatile inorganic materials having superior tensile strength and flexibility. There is little need for consumables except for bonding agents in the making of ropes. Ultrafine threads can be produced in space because the zero-g conditions enhance controllability of the extrusion pull rate.

The possibilities offered by metal and basalt threads (see section 4.2.2) and the comparatively unsophisticated character of fabric-stitching, rope-, and cable-making equipment promise exceedingly low Tukey ratios for these processes. The high-radiation and vacuum environment of space precludes the use of many terrestrial thread materials because of volatility and susceptibility to radiation deterioration. Basalts and metals appear capable of filling this applications gap. Lunar iron can be used to manufacture threads, strings, ropes and cables; Moon-like basalts already have been spun into 0.2–4.0 μm fibers (an established commercial process). Thread- and wire-production machines can be used in space with no specific modifications, and stitching-, rope-, and cable-making devices require only simple alterations to take best advantage of zero-g conditions. Even in applications where the fabric must hold pressure, metal and basalt fibers should prove adequate with minor design changes. The Space Activity Suit (Annis and Webb, 1971), for instance, maintains pressure by tension rather than by retaining a cushion of air.

Shrink fitting is accomplished by heating one piece so that a hole in it expands to accept (usually under pressure) another piece within that hole. Contraction with cooling then locks the two together. Press fitting is a related process requiring higher pressures but no heat. These two techniques are prime candidates for space assembly operations. Because no additional materials are employed, only power is consumed. Both processes are far more energy- and material-efficient than welding, and produce strong bonds. Beams made from rigid materials and many parts can be joined this way. (For example, gears are routinely attached to shafts by shrink fitting.) No bonding agents are required, and the parts materials (metals) are abundant in space. Zero-g permits lower-energy/lower-strength bonds. Shrink or press fitting is preferable to welding for light bonding; however, vacuum welding may provide added strength. Metals and other conductors may be heated by induction

techniques, making possible an extremely high mass multiplication.

4.3.2 *Summary of Analysis of Production Options for Space*

The survey in section 4.3.1 provided necessary background information for selection of processes which are especially appropriate for nonterrestrial materials utilization, summarized in table 4.20. All major manufacturing categories (casting, molding, deformation, and joining) are represented by at least five techniques. Containerless processing, with many potential applications for space, is an entirely new category possible only under zero-g conditions.

As previously noted, these techniques were chosen because of their advantages with respect to the selection criteria given in table 4.18. It is anticipated that the R&D necessary to adapt the techniques to useful productive tasks in space will be significantly less than that associated with processes where development must await investigations of a fundamental nature or more extensive space operations (either unmanned or manned). It should be possible to incorporate the consequences of the earliest possible applications of these techniques in space to the planning of space operations in the mid-1980s and beyond.

Table 4.21 summarizes 12 generic functional components required for space production of devices or products which could be manufactured by the techniques listed in table 4.13 using lunar-derived materials. (A brief discussion of these components appears in section 4.4). All functional elements except #9 (glasses) and #12 (lasing media) can be made directly by adaptations of powder metallurgy-based "starting kits." These two items would require the creation of derivative or second-generation production systems.

The team did not reject the use of the nearly 200 manufacturing procedures listed in table 4.10 for eventual use in space. However, most of these options require special support (e.g., supplies from Earth, special atmospheric conditions) or generally are low-ranked by the criteria in table 4.18. Flexible techniques such as provided by a terrestrial machine shop may be feasible and even necessary during future development of growing space industrial operations, but appear less fruitful to implement in the near-term.

In any event, a number of manufacturing options apparently exist that are sufficiently adaptable to the SMF mission, and a growing hierarchy of materials processing and manufacturing systems, in principle, is possible. Section 4.3.3 considers a subset of the general hierarchy in table 4.20 which appears to offer virtually a one-step method for manufacturing most of the devices of production (and other products) from both native-lunar and refined-terrestrial feedstocks. Section 4.4.1 examines near- and mid-term development of an expanding manufacturing complex in LEO.

4.3.3 *Starting Kits*

More than 40 manufacturing techniques were found appropriate for a near-term evolutionary SMF. The logical limit of this analysis is to determine whether or not there are technological subsets which could be embodied in compact systems to produce most of the mass of subsequent generations of machines of production. These bootstrapping systems or "starting kits" should take advantage of local available materials and be compatible with the use of automation and robotics. Most likely many such kits can be created, their designs strongly influenced by the materials available locally for manipulation.

The present effort focused on the handling of metals and ceramics known to be available from lunar or asteroidal materials, or potentially importable from Earth at low unit cost. No attempt was made to produce conceptual systems able to operate in the hydrocarbon-helium atmospheres of the outer planets and their moons, or in the sulfur-rich atmosphere of Venus or surface of Io. One major approach to starting kits suitable for near-term space manufacturing useful on the Moon involves powder metallurgy. This case was examined in some detail to help clarify the concept. Another approach using large blocks of metal was also briefly considered.

General comments on powder metallurgy and space. An extensive discussion of the development of powder metallurgy appears in appendix 4C. Powder metallurgy appears to offer several basic advantages for space manufacturing. Virtually all the energy for powdering metals, glasses, and possibly ceramics, can be provided by direct solar thermal power. Thus, primary energy systems (e.g., solar mirrors) can be very low in mass per unit of output and reasonably simple to fabricate. Grains of powder created, stored, and manipulated in a very hard vacuum should have minimal surface contamination and therefore will be susceptible to useful contact welding. Good internal bonding of powders thus may occur through grain contact, sintering, and melting. Lack of gas bubbles in a vacuum-manufacturing environment will also aid the production of well characterized parts.

It should be possible to achieve 90% or better of the ultimate powder density in "green" compact parts prior to final forming, if made under low-g conditions. This is because, in the zero-g operating environment of the SMF, very fine grains of the appropriate size and shape distributions could be placed in the void spaces between larger grains. On Earth this cannot be done reliably, since gravity causes smaller grains to settle toward the bottom of the green compact, producing parts of irregular density, composition, and strength (proportional to final density).

On Earth, large presses, sometimes also operating at high temperatures, are required to squeeze the parts to 99% or

TABLE 4.20.– MANUFACTURING PROCESSES APPLICABLE TO SPACE

Based on terrestrial experience	
Preferable	**Usable with recycling or adaptation**
Casting	
a. Permanent b. Centrifugal c. Die d. Full-mold e. Low-pressure f. Continuous	g. Sand h. Shell i. Investment
Molding	
a. Powder metals and ceramics	
Deformation	
a. Thread rolling b. Magnetic pulse forming c. Electroforming (basalt electrolyte) d. Rolling – reversing mill	e. Forging (with electrical drives) f. Lead-in mill g. Extrusion (basalts) h. Spinning (glass and basalt)
Machining[a]	
a. Laser b. Electron beam	c. Turning (basalts) d. Drilling (basalts) e. Grinding (recycle binder, using Al_2O_3-grit)
Joining	
a. Cold/friction welding (metals) b. Laser-beam welding c. Electron-beam welding d. Induction/HF-resistance welding e. Fluxless/vacuum brazing f. Focused solar energy g. Metal fasteners (permanent) h. Stitching (metal or inorganic threads) i. Staples j. Shrink and press fitting	k. Metal fasteners (need fusion preventers) l. Glues (need carbon)
Containerless	
a. Surface tension b. Fields – 1. E & M 2. Centrifugal 3. Gravity gradients c. Direct solar heating (differential) d. Vapor deposition	e. Metal and/or ceramic clays (binder loss)
Containments	
a. Powder/slab – cold welding b. Foaming (metals/ceramics)	c. Metal and/or ceramic clays (binder recycling and loss)

[a]In a vacuum environment most machine techniques will require a pressurized container to prevent cold-welding effects.

TABLE 4.21.– FUNCTIONAL COMPONENTS REQUIRED IN NONTERRESTRIAL MANUFACTURING AND AVAILABLE MATERIALS

Functional components	Materials
1. Structures	Metals (Fe, Al, Ti, Mg) Ceramics/glasses/basalts Reinforced materials
2. Refractories Molds, orifices	Major lunar minerals Chromia, titania, titanium silicide, glasses
3. Dies:	Steels (C, Si, Ni, Co) Silica carbides
4. Heaters: Direct solar Electric[a]	Mirrors (Al and/or inorganic shaped materials) Si (and others) solar cells
5. Insulation (electric and thermal) (glass fiber mattes)	Basalts, ceramics, inorganic fibers, glasses Soil, wools, foams – inorganic
6. Magnetic material (motors, separators)	Iron and alloys Magnetic ceramics
7. Electrical conductors: (motors, electromagnets, control circuits[a])	Al, Fe, Ca – low temperature
8. Grinders	Spinel in glass matte/Ca wheels
9. Glasses (optics)[a]	Si, SiO_2 (+ mixes of major and minor elements)
10. Adhesives and coatings	Metals, ceramics
11. Lubricants and fluids[b]	Sulfides, SO_2 (trace CO_2, H_2O, and compounds of K, O, N, Na, H)
12. Lasing media[a,b]	CO_2

[a]These specific products require second-generation or higher-generation production hierarchies.
[b]This component is a major problem because it requires chemical elements which are rare on the Moon.

more of final density from original densities of 70-90%. Major changes in physical dimensions may occur. It is conceivable that the need for such pressing operations can be eliminated almost entirely for many products and the changes in physical dimensions between green compacts and final product largely avoided by using either direct sunlight or electric heating in space for forming final parts. If very dense green compacts of near net-shape can be prepared then final parts should require minimal cutting or trimming which makes the use of laser or electron-beam devices in final shaping conceivable. Such devices are presently relatively inefficient for materials removal but are capable of very fine-tolerance operations.

Much terrestrial experience is available on powder technologies applicable to both metallic and nonmetallurgical materials. Many of the experiments necessary to adapt this technology to space could be performed in early Spacelab missions. In addition, there can be strong interaction among designers in the planning of parts derived from powders (e.g., overdesign size of parts for additional strength) and the evolution of in-space production techniques.

Impact molder system for production from powders. Figure 4.15 illustrates the impact molder powder process starting kit which consists of a powder/liquid injector (7) and a two-dimensional die (2) enclosed in a scatter shield (3). The shield prevents grains which are misaimed or which do not stick to the working face from drifting out of the production area. Wasted grains can be removed and eventually recycled. The injector directs particles (8) sequentially across that portion of the working face (1) of a part which needs building up, continuously adding thickness as desired at any particular point. Insertable shields can be used to create voids and produce internal patterns (not shown). Metal grains are cold-welded at the

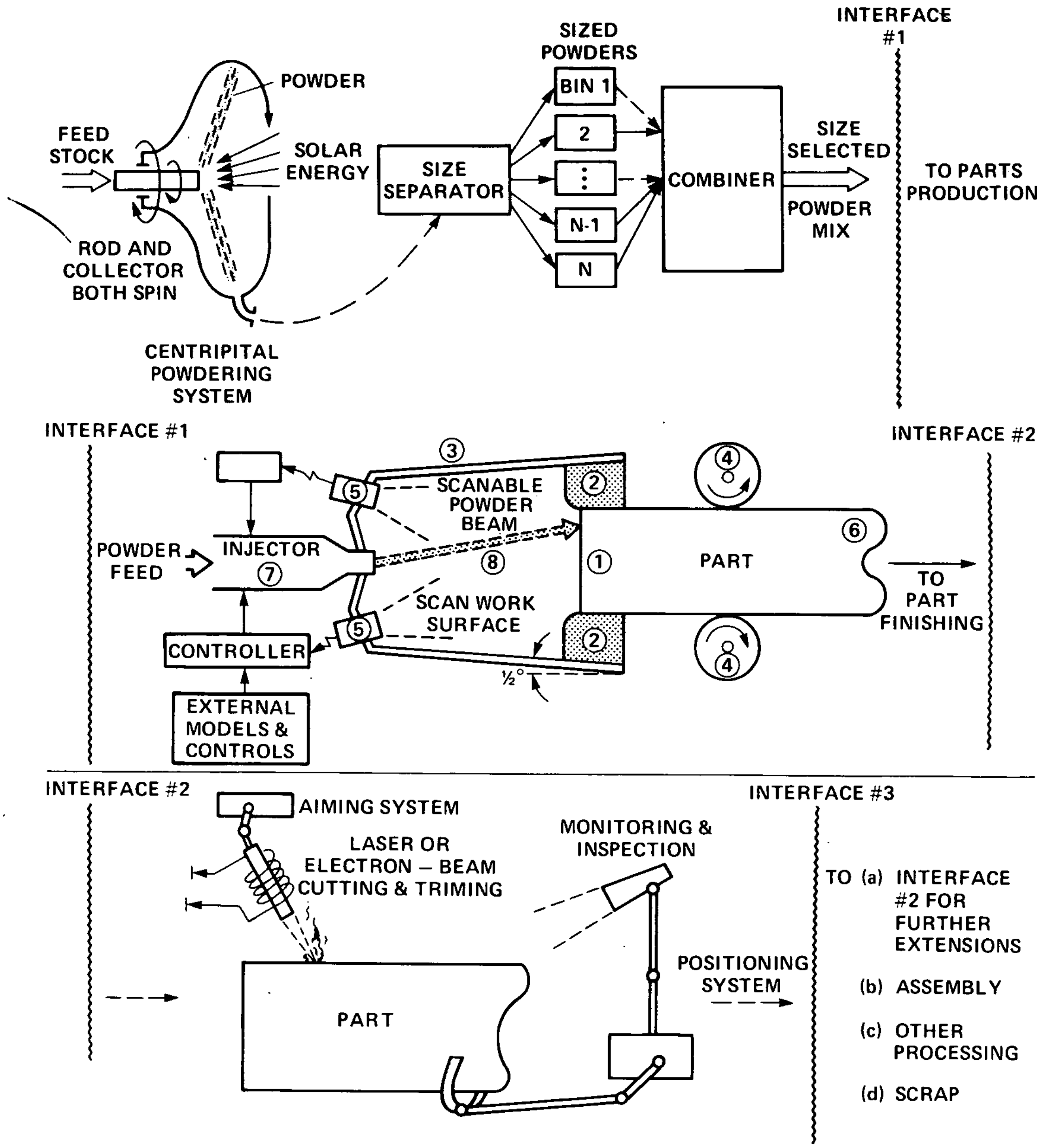


Figure 4.15.– Impact molder powder process starting kit.

instant of impact and coalesce by cooling. Size-distribution management of injected metal powder particles should make possible parts of minimum porosity (i.e., no greater than 3-5%). Vapor-deposition techniques might be useful in decreasing the porosity still further.

The developing workpiece is actively inspected by scanning electron microscopes or optical sensors ⑤ which guide the beam to areas where the surface is rough, appears too porous, or has not adequately been filled. Beam cross-section is fixed by the interior shape of the ceramic die. This die can be made by a casting process or by cutting out blank disks. Rollers or other grippers ④ slowly extract the workpiece from the die as it is formed. A starting surface ⑥ must be provided upon which powder forming can begin and to which extraction devices may be attached.

After formation, parts move to an inspection station for final trimming by a high-energy laser (which exerts no force on the workpiece) or other cutting device. If necessary, pieces are sliced perpendicular to the formation plane to produce more complex parts than can be manufactured directly from the die. It should be possible for a precision, low-mass robot to hold pieces for final trimming. Final choice of finishing tool depends on the tolerances achievable in parts formation as well as tool efficiency.

The impact-molder system produces rodlike components in the first operation of the procedure. It should be possible to build more complex parts by repositioning rod components perpendicular to the die (2) and using the side of the finished part as the starting point for appendages. The process can be repeated as often as necessary so long as access to the die mouth is possible.

Throughput varies depending on the velocity of scanning beam material, number density of particles, mass of individual particles, and cooling rates obtained at the casting die when powders are used. Parts which can tolerate large porosity prior to sintering possibly may be produced at the rate of 1-10 kg (of machinery)/kg-hr. Parts demanding low initial porosity (less than 5%) and very high tolerances must be composed of a wide range of grain sizes, and smaller grains must be placed most precisely by the ejector. The anticipated production rate of these parts is 0.01 kg/kg-hr or less.

Several different injection systems may be used depending on the velocity and mass of the grains to be accelerated. More massive particles must be emplaced by mechanical ejectors, perhaps to be operated by electric motors. Smaller particles (less than or about 1 μm) may be propelled by precision electrostatic systems. Deposition rate M (kg/hr) is of the order $M = fpvA$, where f = filling factor of the beam, p = density of input metal (taken as 5000 kg/m^3), v = injection velocity, and A = injection nozzle area (assumed 1 mm^2). If the reasonable values $f = 0.1$ and $v = 100$ m/sec can be obtained, then $M = 180$ kg/hr. Specific input power P (W/kg) is given by $P = 1/2\, pfAv^3 = Mv^2$, hence $P = 500$ kW/(ton/hr) in the above example. Equipment mass is dominated by the ejector electrical supply (at $v = 100$ m/sec), suggesting a total system productivity of about 5 ton machinery/(t/hr product) and assuming a solar array with specific power rating 10 ton/MW. Note that M scales with v whereas P scales with v^3 – at early stages of production it may be advantageous to operate at low ejection velocities and accept the implied lower throughputs. These estimates are significantly lower than those for mechanical milling – about 2 MW/(ton/hr) and more than 10^4 ton/(ton/hr) – given in table 4.19.

Most of the energy required for the powder-making process can be supplied as direct focused sunlight by systems with intrinsic power of 300 MW/ton. Thus, the solar input subsystem represents a small contribution to the total mass of the powder processor. Little material should be consumed in the production process, with die wear dominating losses.

One major disadvantage of this approach is its primary applicability to production of metal parts or metal-coated ceramic parts. Most other materials must be passively restrained during the sintering process. Parts appropriate to the preparation of ceramics or fused basalts or other nonmetallic materials require the creation of a subsequent set of tools for the construction of ceramics and basalt manufacturing facilities.

There are several areas for applications of robotics and advanced automation techniques in production, process monitoring and parts handling. Process monitoring is required in powder preparation, sorting, storage, and recombination. Very high speed monitoring is necessary at the impact surface of the part under production, especially if a wide range of grain sizes is needed to reduce porosity. Many options for such monitoring that will include active means (e.g., scanning electron beams, sonar interior scanning, radiation transmission measurements) and passive means (e.g., optical examination, temperature) must be examined. In effect, machine intelligence is applied at the microscopic level of the materials handling process. Very detailed analysis of macro-handling of parts is necessary, including such operations as extraction, moving parts in physical space without impacting adjacent objects, parts repositioning for trimming, cutting, or sintering, and monitoring the effects of these operations. Finally, parts are passed to assembly robots or automated lines. Many of the procedures are extensions of present technologies of automatic transfer in terrestrial practice. However, there will be far more emphasis on reliability, scheduling, flexibility, and repairability.

Metal- and ceramic-clay-based starting kit. According to Jones (1960), the concept of manufacturing metal objects from powders formed into clays using spinning or sculpting techniques is a very attractive one. This is true especially if it is possible to avoid drying out periods and obtain high densities with relatively brief sintering times. Binders are feasible for Earth applications – polystyrene and polythene in particular, each of which is recoverable and nonreactive with the more common metals, and both are suitable for the production of clay-like metal masses. While such recyclable organic binders may be useful in space and on the Moon, certainly it would be more advantageous to obtain binders from local sources. Desired characteristics include the following:

- The binder should impart a stiff clay-like quality to the metal or ceramic mass and permit easy manipulation, have a sufficiently low volatility under the desired working conditions to allow a reasonable working period, and leave no residue following the completion of sintering.
- The binder should not require removal prior to placing formed clay into the sintering oven, but should not disrupt the molding during volatilization.
- The rigidity of the molding should be maintained during the early phase of sintering.
- The binder and its solvent (if needed) should not react chemically with the powder either at working or elevated temperatures, nor should they attack furnace components or elements of the recovery system.

- Binder and solvent should be nontoxic under the working conditions in which they are used.

Table 4.22 identifies several binders appropriate for use on Earth. The last compound listed is preferred on the basis of slow evaporation rate, high boiling point, and high flash point. Thermoplastic binders such as polybutene dissolved in xylene with a hydrocarbon wax, or ethyl silicate, are other possibilities. These are introduced into molding furnaces at moderate (430 K) temperatures and have permitted the successful molding and sintering of small objects. Unfortunately, workpiece rigidity is insufficient for terrestrial manufactures bigger than 5 cm; larger items tend to slowly collapse at room temperatures. Clearly, bigger parts could be made on the Moon, and there is no serious limit on the size of objects which could be sculpted in space.

TABLE 4.22.– METAL/CLAY BINDERS APPROPRIATE FOR TERRESTRIAL USE

Binders	Boiling range, °C	Flash point, °C	Evaporation rate[a]
Methyl amyl acetate	143-150	110	47
Ethylene glycol diacetate	186-195	205	2
2-ethylhexyl acetate	195-205	190	3
2-methoxyethyl acetate	137-152	140	31
Ethyl benzene	134-137	85	91
Carbitol acetate	213-223	230	<1
Decahydronaphthalene	190-200	160	10
Tetrahydronaphthalene	203-220	185	1

[a]H-butyl acetate = 100.

Binders in space may be able to function in two additional ways. First, the compounds may be selected to inhibit contact welding between grains to facilitate the greatest packing of voids by filler grains. Second, initial binder evaporation could expose surfaces to permit preliminary contact welding prior to full sintering of the part. An extensive literature search should be conducted to determine whether or not such compounds can be derived from lunar and asteroidal materials. Lee (1979) has suggested several liquid silicon-based and Ca-O-Al compounds that could be derived predominantly from lunar materials. Perhaps such fluids (for which recovery is not as critical) could be adopted for vacuum forming.

The powder metallurgy approach to manufacturing has considerable potential in nonterrestrial low- or zero-g applications. There is virtually a complete separation of the three basic stages of production: (1) creation of working materials (high energy), (2) embodiment of a design into a mass of clay to form a part, and (3) hardening of the part by contact welding and sintering. Very complicated designs can be produced by machines able only to apply relatively small forces, allowing considerable quantities of mass to be formed for very little energy but potentially with high precision.

Figure 4.16 illustrates three techniques for pattern impression. One possibility is to inject the clay into a mold. This mold may be very intricate provided it is sacrificed after sintering, a modest penalty because of the low initial temperatures. Second, clay could be packed around "melt forms" (recoverable from the vapor) to make pipes, conduits, and other structures with internal passages. Third, parts could be sculpted directly from masses of clay. These masses could be initially amorphous or might be preshaped to some extent by molds or spinning techniques as in the manufacture of pottery on Earth.

Advanced automated pottery techniques are not limited to the production of metal parts because sintering is used in the final stage. For instance, metal and ceramic parts could be interleaved in the clay stage to produce, say, electrical machinery. In such applications the porosity of the different ceramic and metal powders in the various portions of the respective clays is carefully controlled so that differential expansions and contractions during the formation process do not ruin the part. In addition, hollow metal grains would permit local metal volumes to decrease under planned stresses as necessary during the sintering process. Conceivably, this could allow very complicated metal paths to be melted directly into the body of a ceramic material having a much higher melting point and also to produce exceedingly complex composites.

It is interesting to speculate on the ultimate limits of the above techniques with respect to the size and complexity of the final object. Rates of expansion, heating and cooling of the workpiece (which presumably can be well controlled over long periods of time in space using solar energy), gravity gradients, rotation and handling limitations during the formation phase must all be considered. It may be that the largest objects must be formed in very high orbits so that continuous sunlight is available during critical periods and gravitational tidal effects remain small. Perhaps, in the ultimate limit, major mass fractions of spacecraft, space stations or habitations could be manufactured in monolithic units by this process.

Clay metal and ceramic technologies suggest a number of theoretical and experimental projects or demonstrations related to both near- and long-term terrestrial and nonterrestrial operations. Experiments on grain size distribution, dimensional changes, compositions of metals and ceramics, and choices of binders with regard to porosity, new molding and forming techniques which might be employed in space, and the general area of automatic production, inspection, and robot handling are all appropriate research topics. Indeed, one of the most important characteristics of starting kits is the easy automatability of the tools involved.

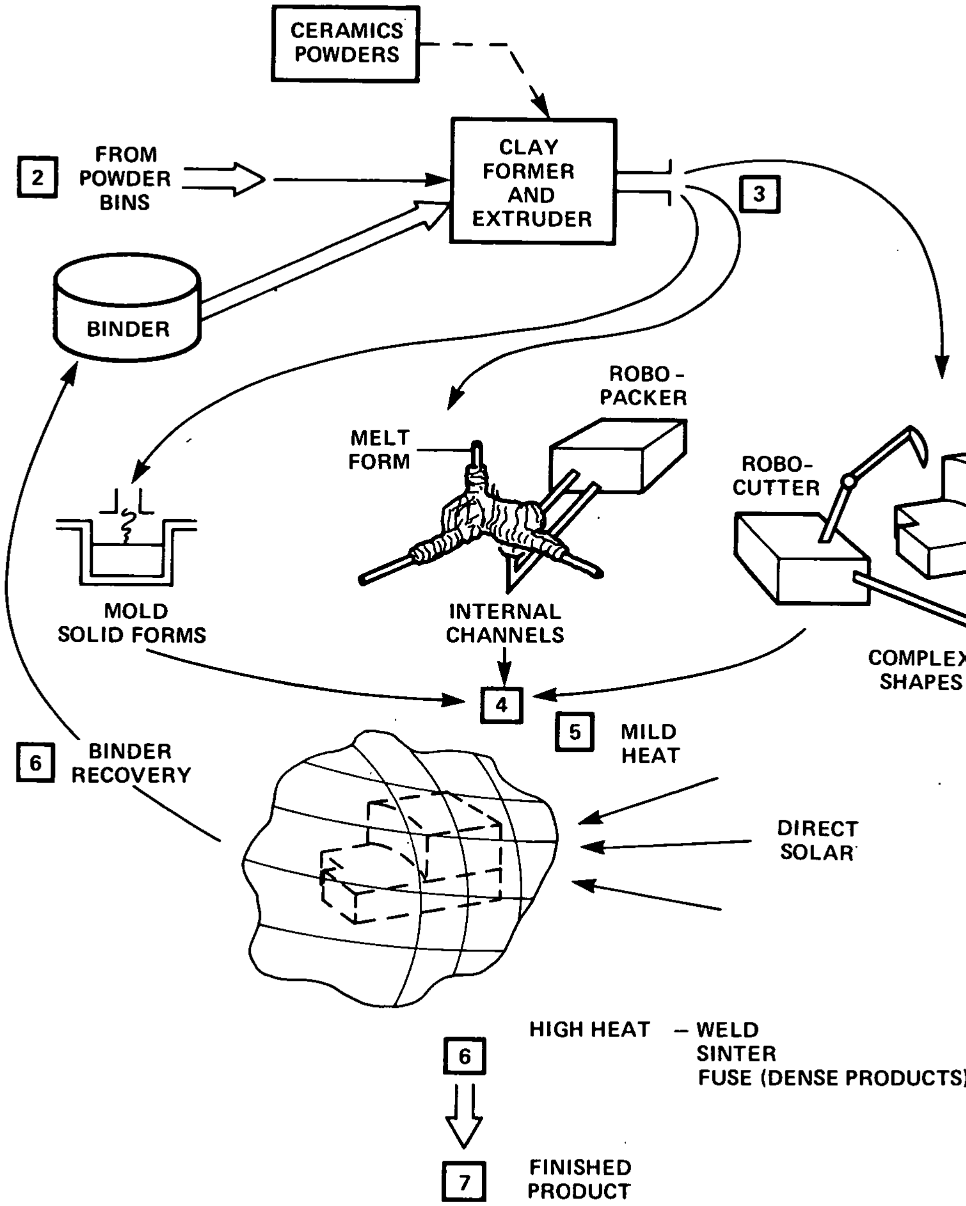


Figure 4.16.– Metal clays and pottery manufacturing.

In the basic kit, forming and shaping functions of the fabrication robot are farthest from deployable state of the art. But tools and techniques have been chosen that can generate a wide variety of products of differing complexity using relatively few simple modes of operation. These starting kits could be deployed in the near-term as part of a fault-tolerant, easily reprogrammable prototype SMF.

Macro-blocks and contact welding. It is conceivable that many useful tools and products, especially very large parts, could be quickly manufactured from metal blocks of various sizes. The same or similar metal blocks with clean surfaces will cold-weld when pressed together with sufficient force. One problem with this approach is that pressures in excess of 10^7 Pa may be required even for blocks with extremely smooth surfaces, making large powerful presses impractical in the early phases of an incremental space industrialization program. One possible solution is to manufacture a very fine "dust" of hollow particles of the same metal as the pieces to be joined. Dust particles should have approximately the same radius as the asperities of the large blocks. This "dust" is then evenly distributed over the contact surface of one of the pieces to which it would adhere by cold welding and the second piece is pressed upon it. Joining pressure need only be sufficient to flatten the hollow spheres, permitting them to flow into and fill voids between the two macrosurfaces. Electrical current

passing across the gap between the blocks could heat the dust and further promote joining.

This approach to construction would allow the use of a small number of furnaces and molds to produce standard sets of blocks from appropriate sources of metals. The blocks could then be contact-welded to manufacture a wide range of structures. While such blocks would not allow detailed flexibility of design as might be permitted by the two powder metallurgy systems described earlier, the throughput of the system for the construction of large repetitive objects would likely be significantly higher. A major potential difficulty requiring far more study is the degree of smoothness necessary prior to joining and the precise size distributions of hollow powders used to fill the gaps between the blocks. This may limit the maximum size of blocks which can be joined with minimal preworking.

Starting kit technology development. Sufficient knowledge exists with respect to powder metallurgy, space operations in LEO and on the lunar surface, and about lunar materials near the Apollo landing sites for development of starting kits to begin. Naturally, the relevant concepts should be fully reviewed by experts in the respective fields. These reviewers might also define key experiments and tests necessary for convincing near-term demonstrations (see section 5.6 for a useful relevant methodology). For instance, it would be useful to demonstrate (perhaps in low-g aircraft or sounding-rocket flights) the sintering of multisized powders which are well-mixed prior to sintering. Detailed consideration should also be given to the design of subsequent components by conceivable starting kits.

Demonstration of the full capabilities of contact welding may not be possible from Shuttle-supported facilities in LEO without incorporating a molecular shield into the mission and performing the key tests beyond the immediate vicinity of the Shuttle. Even at LEO there is sufficient ambient gas (e.g., highly reactive atomic oxygen) that surface contamination may be significant. However, LEO experiments should be able to show the full potential of powder techniques with respect to powder forming using solar energy, zero-g, and green mold densification, final product sintering or fusing using solar energy, and working with metallic/ceramic clays in space including binder recovery techniques.

The powder approach possibly may be useful on the lunar surface. Fine-grained (1-10 μm) metallic iron is present in lunar soils to 0.1% by weight. This metal can be extracted magnetically and separated from adhering glass and minerals by direct heating. Such iron may be used as a structural, electrical, or magnetic engineering material. Various other lunar soil components can be used for structural and insulating purposes. Hence, it appears possible to effectively utilize native iron using little more than a thermal processing technology capability. If so, then the "starting kit" approach can be employed to create much larger iron-processing facilities on the Moon over a period of time by "bootstrapping" what is essentially a very simple system.

Chapter 5 of this report explores the initial deployment of "starting-kit-like" devices capable of self-replication as well as growth.

4.4 SMF Growth and Evolution

Following its deployment, the starting kit begins to manufacture second-generation tools, as well as replacement parts for itself. These tools can be used to produce additional types of equipment and early product lines. Eventually, space-compatible equivalents of all major terrestrial manufacturing processes and new systems evolved in space must be available to the evolving SMF.

Further growth and increased complexity are required if the SMF is to evolve from the starting kit into a sophisticated manufacturing center which depends less and less on Earth for raw materials resupply. One key growth area especially significant in view of the heavy requirements for computers and robotics in space is the automated fabrication of integrated circuitry and other electronics components. Certain unique characteristics of the space environment, combined with anticipated advances in laser-, electron-, and ion-beam technologies, may make possible automated machinery capable of manufacturing highly sophisticated integrated circuits as well as resistors, capacitors, printed-circuit boards, wire, and transformers in space, using raw materials supplied entirely from the Moon, and ultimately a wide variety of additional complex products.

4.4.1 Starting Kits for SMF Growth

Having considered a range of possible starting kits, the Team next explored the possibility of an ever-widening collection of production machinery using kits described in section 4.3.3. This aspect of the analysis is crucial to growth and evolution, since the taxonomy of manufacturing processes is distinct from the list of functional components comprising the implements of manufacturing. Table 4.21 showed the major functional machine components which must be available in a growing SMF. Nonterrestrial, especially lunar, materials can be used in most cases. The most serious deficiencies are the lubricants and fluids needed for pressure transfer or solution-processing (electrolytes, wetting agents), though silanes may be serviceable in lunar applications. High-powered lasers are convenient for cutting and finishing in space. The Moon is somewhat deficient in the most common gases used in tunable power lasers, He, Ar, Xe, but fortunately each gas is readily recyclable.

Manufacturing components listed in table 4.21 were reviewed specifically for derivability from starting kits, with

the assumption that appropriate processed materials would be supplied as feedstock to the SMF:

- Structures – A wide variety may be produced directly from any starting kit as described in section 4.3.3. These range from very small solid pieces such as shafts or dies to much larger components including rigid members for heavy presses. Metals, ceramics, and ceramic/metal combinations can also be prepared.
- Refractories and dies – can be manufactured using the powder metallurgical components of the starting kit. Laser trimming can be performed as required after solidification and inspection of the part. These components then become available for casting complex shapes and for extruding both long-dimension components and parts designed to sustain very high temperatures and pressures.
- Heating – by direct solar energy may initially be accomplished using aluminum deposited on spherical surfaces. These surfaces may be shaped by rotation of unitary structures of appropriate radii of curvature extruded using the starting kit. Alternatively, metal vapor deposition on interior subsections of bubbles grown in zero-g may be used. The existence of solar-electric devices is assumed.
- Insulation – for both thermal and electrical needs can be derived from fiberglass mattes produced by a spinning process involving the extrusion of molten glass through small orifices. Electrical insulation exhibiting mechanical softness or compliance is achieved by pressing fiber mattes into long thin ribbons and then wrapping these tightly around the wires, followed by partial sintering. Basalt fibers may be useful in this application (see section 4.2.2).
- Magnetic materials – can be manufactured directly from the starting kits or by powder metallurgical technologies. Dies and heating equipment produced in earlier steps are probably required for maximum versatility.
- Electrical conductors – particularly wires for motors, busbars and other purposes, may be extruded (original starting kit equipment) or fabricated using rollers and dies derived from structure and refractory manufacturing components produced earlier.
- Grinders – are needed for precision finishing of surfaces. These tools should be producible by pressing and casting operations available with the starting kits. Grinders may be composed of spinel grains (a lunar-abundant grinding agent) embedded in glass fiber mattes perfused with calcium for mechanical softness and binding.
- Glasses and fibers – can be manufactured by using casting, grinding, and die-extrusion operations. Grinding is required for optical-quality glass shapes. Electron-beam and laser techniques are useful for final finishing of optical surfaces.
- Adhesives and coatings – of metals and ceramics can be applied by the starting kits or a specialized kit suited to the particular geometries of certain parts.
- Lubricants and fluids – present special problems because of deficiencies in presently known lunar raw materials resources. It may be that self-lubricating powder metallurgy bearings containing brass and lead in very small quantities are feasible. Also, silicon-based compounds requiring a minimum of relatively rare lunar carbon and hydrogen should be extensively investigated.
- Lasing media – It is also important to determine to what extent lasing media for high-power lasers can be derived primarily from lunar materials. Undoubtedly a considerable literature applicable to such devices already exists, but is classified for military reasons.

Control systems and electronics (see section 4.4.3) are also necessary, especially for automated manufacturing facilities in space.

Several technologies with limited terrestrial applications may prove extremely useful in space. One example is containerless production, in which objects are formed directly from melts. Overall shape is controlled by surface tension, external forces, and directed solar heating. Vapor deposition is another potentially favorable technique which should be given high research priority. Also, as the human presence in space expands, special production environments that allow the use of gases and liquids will become more commonplace. Thus chip-producing machinery, foaming and other processes requiring the recovery of production fluids may eventually become feasible in space.

It is easy to see how a starting kit might generate production equipment required for other space-compatible manufacturing techniques. (Shearing operations are assumed to be within the capabilities of starting kit laser beam units). For example, laser techniques for scribing reverse threads onto hardened steel rolling dies is a foreseeable technology (fig. 4.17). The availability of chromium on the Moon (0.6% by weight and higher in beneficiated iron grains) and lunar basalt for base plates makes thread rolling a valuable adjunct to the starting kit extrusion system.

A second example is magnetic-pulse-forming equipment. The two main components of the magnetic-pulse former are the forming coil and the capacitor. Robots with appropriate wrist actions should be capable of conventional winding operations to manufacture forming coils from extruded wire. The capacitor may consist of a basalt/aluminum or

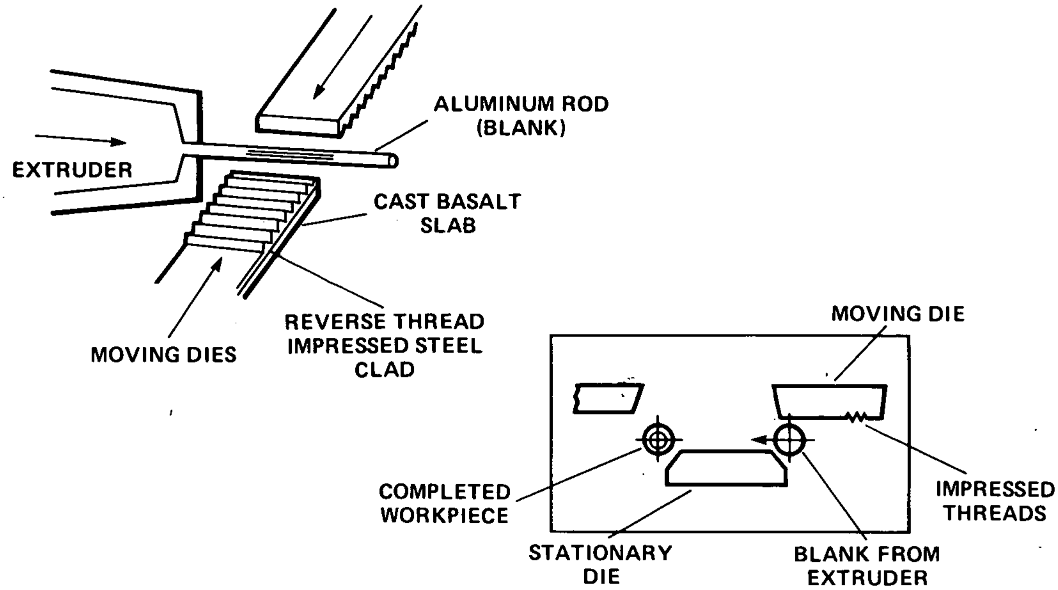


Figure 4.17.– Schematic of the principle of thread rolling.

alumina/aluminum sandwich based on the standard formula $C = kEA/d$, where C is capacitance, k is the dielectric constant of basalt or alumina (4.5-8.4 at 10^6 Hz), E is the permittivity of free space, A is capacitor plate area, and d is plate spacing.

A third example is electroforming technology. As discussed in section 4.3.1, the components of an electroforming unit are somewhat more complex than those of magnetic-pulse formers because of the need for an electrolytic plating solution. The tank containing the solution may be fabricated using the extruder, then welded together by a laser beam unit. The mandrel (fig. 4.13) may be formed of cast or sintered basalt over which aluminum is vapor-deposited. Iron or titanium anode plates are no problem for the starting kit extruder, and centrifugally spun basalt may be used in the electrolyte filter. Cast basalt pipes, an off-the-shelf terrestrial casting technology, provide necessary plumbing for the entire electroforming system.

4.4.2 Near-Term Manufacturing Demonstration: Shuttle Tank Utilization

The Space Shuttle external tank (Martin Marietta Corporation, 1974) carries liquid fuel for the Shuttle main engines and separates from the spacecraft just prior to orbital insertion at an altitude of about 128 km. The cylinder then follows a ballistic re-entry path, crashing into the ocean far from inhabited areas. The cylinder is not recovered or reused. But the tank, when dropped, has already achieved roughly 99.7% of orbital velocity. The added delta-V needed for tank orbital insertion is only 46 m/sec, about 10% of available Shuttle Orbiter thrust. Alternatively, the tank could be orbited by burning the main engines for a slightly longer time, or with the aid of a jet-assisted takeoff (JATO) booster. The cylinder itself measures 8.4 m diam, 47 m long (a volume roughly equivalent to that of a 10-story condominium), and 33,503 kg in inert weight. Most of this mass is pure structural aluminum, though about 100 kg of outer skin insulation contains organic materials which could serve as the basis for early organic chemistry at the SMF (carbon, plastics, biological products, and so forth). A few tons of unused propellants (LOX and LH_2) may also be present, and surplus materials from Shuttle operations (hydrazine, helium, food, etc.) could be stored in orbit for later use.

Any Shuttle flight carrying a volume-limited cargo can bring the external tank to orbit with near-zero propulsion costs. Valued as payload at about $1000/kg, an empty tank is worth about $33.5 million, less additional propulsion costs but plus added value derived from conversion of tank mass to useful products by the SMF. If Shuttle flies every 2 weeks, the payload value of the tank masses inserted into orbit would be the equivalent of roughly $1 billion per year. To an orbital space manufacturing economy this represents new additional income, in this case the equivalent of about 20% of the current annual NASA budget.

For such a cost-effective program to be implemented, the means for orbital insertion of the tank must first be perfected. Next, a system (teleoperated or robotic) should be designed which is capable of scraping off valuable external insulation. Cutoff valves must be added to prevent excess propellant from venting (permitting it to be stored in orbit rather than lost to space).

The starting kit provides a means of reducing the tank to powder or liquid form. The kits described earlier can

accomplish this directly without the necessity of manufacturing additional process equipment. Another possibility is a solar-powered milling device (with portable atmosphere) which clamps onto the external tank and carves it into small pieces, most likely under teleoperator control. Tank fragments are then melted by a solar furnace consisting of a spherical mirror constructed by aluminizing a thermoplastic bubble hemisphere (Moore, 1980). The plastic allows sunlight to enter but retains infrared radiation by internal reflection, keeping the work materials hot. A hatch is cut in the mirror to permit insertion of metal shards, which join the growing droplet of molten aluminum at the focus. The melt volume of an entire tank would be about 12 m^3, easily maneuverable through a small opening if processing proceeds in a dozen or so smaller batches.

Once tank material is molten a variety of manufacturing options become available. Ingots or simple bulk castings could be prepared as feedstock for other SMF processing operations. Liquid or vapor metal streams could be directed into molds or sprayed onto lighter structures for stiffening. For instance, thin thermoplastic bubbles may be aluminized to make pressure vessels, mirrors, or heavy solar sails; then plastic is stripped off and recycled. A more elegant method is to blow uniform metal bubbles directly, an ideal zero-g application. Aluminum is a good thermal conductor and reflector, and hence radiates heat slowly while retaining an even temperature distribution. Small tin bubbles have recently been blown experimentally in drop towers (Wang and Kendall, 1980), but far more research remains to be done.

Quite large volumes can be enclosed by structures manufactured using metal derived from a single Shuttle external tank. Aluminum pressure vessels 50 mils thick can retain one-third normal Earth atmosphere (O'Neill, 1977). Average tank thickness is about 250 mils, so a pressure vessel of roughly 13,000 m^3 can be made from just one tank. This is more than fifty times the volume of the Space Shuttle cargo bay (240 m^3).

4.4.3 Middle-Term SMF Expansion: Manufacture of Electronics Components

The present study urges a dramatic increase in the utilization of computerization and automation in nearly every conceivable future NASA mission. It is likely that a nonterrestrial source of computers and robots eventually will prove both useful and cost-effective in space. The team analyzed currently available and anticipated electronics components manufacturing technologies to determine which will satisfy two major criteria: (1) compatibility with a low- or zero-g factory environment, and (2) possibility of deriving required consumables from lunar resources.

Key components in computer systems include integrated circuits (ICs), capacitors, resistors, printed circuit (PC) boards, and wire. Fabrication capability in these five critical areas will permit most other necessary components to be produced as well. For instance, an IC fabrication facility could manufacture at least some varieties of transistors, diodes (rectifiers, small-signal, and zener), varactors, thyristors, silicon-controlled rectifiers (SCRs), and others. It would, however, have difficulty producing light-emitting diodes (LEDs) due to the scarcity of gallium and arsenic on the Moon. Thus, the intent of the following analysis is to present feasibility arguments concerning how lunar materials near-closure might generally be achieved. Substitution and comprehensive manufacture of electronics components are beyond the scope of the present study. Even with this limited review, it is encouraging to note the number of instances in which space equals or is superior to terrestrial factory environments for the manufacture of electronic components.

Integrated circuits. Conventional wafer fabrication techniques (Oldham, 1977) are, for the most part, not feasible in a lunar-supplied SMF. On the other hand, the vacuum of space greatly enhances the applicability of several techniques which are at or beyond the current terrestrial state-of-the-art.

Silicon (chemical refining required) is plentiful on the lunar surface, about 20% by weight (Phinney *et al.*, 1977). While it is not clear precisely how lunar silicon will be transformed into boules of the pure element, it is reasonable to assume that this can be accomplished. Hard vacuum should facilitate the processes of crystal-pulling and zone-refining purification of elemental silicon (Grossman, 1976). Conventional zone refining requires induction heating (Grossman, 1976; Manasse, 1977), a space-compatible technique.

High-speed ICs using silicon-on-sapphire (SOS) technology are currently being fabricated by Hewlett-Packard (Pighini, personal communication, 1980) and others for custom applications. Should it appear desirable to produce such high-speed devices in the SMF, it is worth noting that spinel is plentiful on the Moon. Spinels are closely related to sapphire and actually provide a better crystallographic match to silicon, leading to higher mobility and less aluminum autodoping than in conventional SOS processing (Glaser and Subak-Sharpe, 1977). (The only major problem with spinel is the difficulty of finding high-quality crystals of correct composition.) Epitaxial growth of silicon on spinel substrates may be accomplished by the pyrolysis of silane (Glaser and Subak-Sharpe, 1977) according to:

$$SiH_4 \text{ (gas)} \xrightarrow[(1300\text{ K})]{} Si \text{ (solid)} + 2H_2 \text{ (gas)}.$$

Hydrogen is in short supply on the Moon, roughly 0.01% by weight (Phinney *et al.*, 1977), but fortunately only small amounts of it are required in this procedure. Silane is also an intermediate product in the chemical refining scenario described by Waldron *et al.* (1979).

Conventional photolithography and diffusion techniques are not feasible for space electronics fabrication. Many of the required chemical elements are present in lunar soil only at the ppm or ppb level. Photoresists consist largely of hydrocarbons, substances whose atoms are rare and which deteriorate rapidly in the space environment. The best alternatives may be laser, electron beam, and ion beam technologies. It is anticipated that these methods could lead to greater reliability on an increasingly miniaturized scale, particularly under the high-quality vacuum conditions characteristic of space (Carter and Grant, 1976).

Ion implantation already has begun to supplant diffusion techniques in the practices of many semiconductor firms. This technology allows greater control over quantities of impurities introduced, depths and widths of doped volumes, concentration gradients, etc. Of particular interest for a future wafer fabrication plant in space is the potential for computer-controlled, maskless, multilayer implantation of multiple device types with submicron geometries (Namba, 1975; Wilson and Brewer, 1973). While further research and development must be conducted to translate this tremendous potential into practical reality, other features of ion implantation make it a highly desirable interim choice. Masking may be accomplished by aluminum or other metals, passivation layers, resists, etc. Doping also is possible using passivation layers, an approach which could lead to reduced leakage and better yields (Wilson and Brewer, 1973).

One drawback to ion implantation is crystal lattice damage. A recently developed technique permits extremely localized annealing by laser beam (Tebo, 1979). This process, unlike its thermal annealing predecessor, completely restores damaged crystalline structures through epitaxial regrowth. The net result is a lower resistivity material more suitable for semiconductor use, with fewer defects and higher yields. If this laser technique can be computer-controlled like the multilayer ion process described earlier, automated production of three-dimensional integrated circuitry in space is entirely conceivable.

Pre-3D wafer technologies adaptable to more conventional production sequences also are available. Chemical and plasma etching processes require chemicals (e.g., HF, H_2SO_4, CF_4-O_2) which cannot conveniently be supplied in sizable quantities from lunar soil. A feasible substitute may be ion beam etching. While the closely related process of sputter-etching requires high-pressure argon gas, ion-beam etching at the rate of 10–300 nm/min can be achieved in a 10^{-4} torr argon atmosphere (Glaser and Subak-Sharpe, 1977). Titanium oxide is a suitable etch mask for this process. Argon and titanium are available from lunar sources (1 ppm and 1-5%, respectively) in the necessary quantities.

One chemical vapor deposition technique is perfectly space-compatible. An electron beam easily evaporates materials such as aluminum *in vacuo,* so metal masking and metallization pose no unusual problems. Oxidation of silicon for masking or passivation purposes probably is most easily achieved thermally using anhydrous oxygen gas, rather than chemical vapor deposition methods which require hydrogen compounds. An alternative oxidation process might involve the use of a laser to create extremely localized heating (Tebo, 1979). Aluminum and oxygen are plentiful in lunar soil (5-14% and 40–45% by weight, respectively).

One final critical issue is cleanliness. Particulates should pose fewer problems in space than on Earth because of the absence of atmosphere for convective transfer. An aperture in the fabrication facility enclosure opposite the SMF velocity vector, suitably baffled, should provide a clean vacuum source. Some versions of such orbital devices are called molecular shields, and can provide less than 10^{-14} torr environments at LEO. Internally, moving parts and outgassing are probable sources of particulates which must be minimized (Naumann, personal communication, 1980). Condensibles may prove a bigger cleanliness problem than particles. Techniques for coping with them include avoiding line-of-sight exposure to sources, use of materials with high vapor pressures, and installation of cold traps.

Capacitors. Basic elements of discrete fixed capacitors include metal plates or foil, dielectric material, and wire leads. The plates or foil and leads can be contrived from readily available aluminum. Alumina, silica, and a variety of glass and ceramic materials provide suitable dielectrics. All of these substances are readily available from lunar sources.

Two capacitor fabrication techniques – thin- and thick-film – are compatible with silicon integrated circuit technology, though discrete capacitors generally are preferred over thick-film versions (Glaser and Subak-Sharpe, 1977). Thin-film capacitors usually are made with tantalum (Ankrum, 1971; Grossman, 1976; Khambata, 1963). However, thin-film capacitors with higher working voltages but lower capacitance and slightly poorer temperature stability can be constructed of alternating aluminum and alumina (or silica) layers over silicon dioxide and the silicon substrate (Ankrum, 1971; Glaser and Subak-Sharpe, 1977; Khambata, 1963). Titanium dioxide is another possible dielectric – its dielectric constant is four times that of alumina (Glaser and Subak-Sharpe, 1977). Maximum capacitance values obtainable using thin-film technology are on the order of thousands of picofarads, and automated laser trimming can produce a high-accuracy (±0.05%) product (Grossman, 1976).

Resistors. Since carbon is a relatively scarce lunar resource, only wire-wound, metal or metal-oxide-film, and semiconductor resistors (Dummer, 1970; Glaser and Subak-Sharpe, 1977) will be seriously considered for use in space applications.

Wire-wound devices are appropriate in applications requiring relatively high power dissipation, such as bleeder resistors in power supplies. Nichrome wire (80% nickel, 20% chromium) can probably be supplied in limited quantities from lunar materials (abundances 0.01-0.03% and 0.1-0.4%, respectively). Titanium, another possibility, is abundant on the Moon, and has a resistivity (42 M ohm-cm) which is approximately half that of nichrome.

However, most resistors used in computer circuitry need not dissipate much power. Thin-film and semiconductor devices appear most promising in this regard. Thin-film resistors are fabricated by evaporation or by sputtering 0.025-2.5 μm of metal or metal alloy onto a substrate of alumina or silica (Glaser and Subak-Sharpe, 1977; Grossman, 1976; Khambata, 1963; Manasse, 1977). While some metallic materials commonly used in resistor manufacture are too rare in lunar soil for serious consideration (e.g., tantalum, nichrome, tin oxide, chromium), titanium offers a sheet resistance of 2 k-ohms/cm^2 and a temperature coefficient of resistance (TCR) of -100 ppm/°C (Ankrum, 1971; Dummer, 1970; Grossman, 1976; Khambata, 1963). Thus, the electron-beam evaporation and laser-beam trimming technologies discussed above may be utilized to prepare fine-tolerance, thin-film titanium resistors (Glaser and Subak-Sharpe, 1977; Grossman, 1976; Khambata, 1963; Manasse, 1977). At present it is unknown how closely these technologies can approach contemporary terrestrial tolerance and manufacturing standards (better than ±0.005%, TCR = 1 ppm/°C; Rothschild *et al.*, 1980).

Semiconductor resistors can be made with a technology already discussed. Ion implantation of boron into silicon produces sheet resistances of up to 12 k-ohms/cm^2, suggesting that high discrete values are readily achievable. While less precise than their thin-film counterparts, ion-implanted semiconductor resistors have been shown to offer yields on the order of 90% after packaging (Wilson and Brewer, 1973).

Printed circuit boards. Printed circuit (PC) boards are made of phenolic resin reinforced with paper, or an epoxide resin reinforced with paper or fiberglass cloth, which is then clad with copper (Coombs, 1979; Scarlett, 1970). Unfortunately, resins deteriorate in space and are difficult to prepare from lunar resources; also, copper is rare on the Moon (8 to 31 ppm by weight; Phinney *et al.*, 1977). A new approach to PC board manufacture is necessary. Two possibilities are basalt rock slabs and silane-coated basalt fibers (Green, personal communication, 1980). Basalt is an excellent insulator and can be drilled and aluminized to form an etchable conductive surface (Green, personal communication, 1980; Naumann, personal communication, 1980). Boards made of silane-coated basalt fibers would be lighter and easier to drill, but it is unknown whether aluminum can be vapor deposited onto such a surface. If evaporation problems should arise, a thin layer of titanium could serve as an excellent deposition primer (Glaser and Subak-Sharpe, 1977). Ion beam etching might be used selectively to remove aluminum to form any desired circuit pattern. This process is likely to be amenable to precision computer control.

Wiring. The lunar availability of aluminum will permit its widespread use as a conductor for PC board claddings and for all space wiring in general. Its low resistivity (2.8 μohm-cm) compares favorably with that of copper (1.8 μohm-cm), and it readily forms a protective anodic oxide upon exposure to air (Glaser and Subak-Sharpe, 1977). The major terrestrial drawback to aluminum conductors is their incompatibility with conventional soldering and welding methods (Glaser and Subak-Sharpe, 1977). Fortunately, the preferred welding techniques for use in space (see section 4.3.1) should bond this metal nicely. Basalt or glass fibers are possible materials for sheathing aluminum wire (Green, personal communication, 1980), and Miller and Smith (1979) have devised a space-qualified wire insulation wrapping machine.

Before leaving the topic of aluminum wire, it should be noted that high-quality inductors also may be made of this material. One class of inductors – transformers – represents a particularly important component of many computer systems. Iron is plentiful on the Moon (4-15% by weight; Phinney *et al.*, 1977) so transformer cores present no serious problems for the proposed electronics components fabrication facility.

4.4.4 Complex products

The ultimate goals of the SMF are independence from terrestrial resupply, *in situ* production of all components needed to maintain and expand existing space facilities, and the manufacture of high-value products for consumption on Earth (fig. 4.18). Following deployment of the initial starting kit and manufacture of second-generation tools, development of a product line of ever-greater complexity must occur if the ultimate goals are to be attained. The evolution of complex product manufacturing is outlined below with a focus on just a few important potential products typical of each stage of increasing production sophistication.

Platforms. Expansion of the SMF requires a concomitant enlargement of the facility platform. Such construction represents an early evolutionary threshold, a step requiring little materials processing innovation with some advancement in robotics capability. Component parts may be manufactured from cast or sintered basalt or from aluminum beams, any of which could be produced by the initial starting kit and second-generation tools embodying a synthesis of advancements which already have occurred in

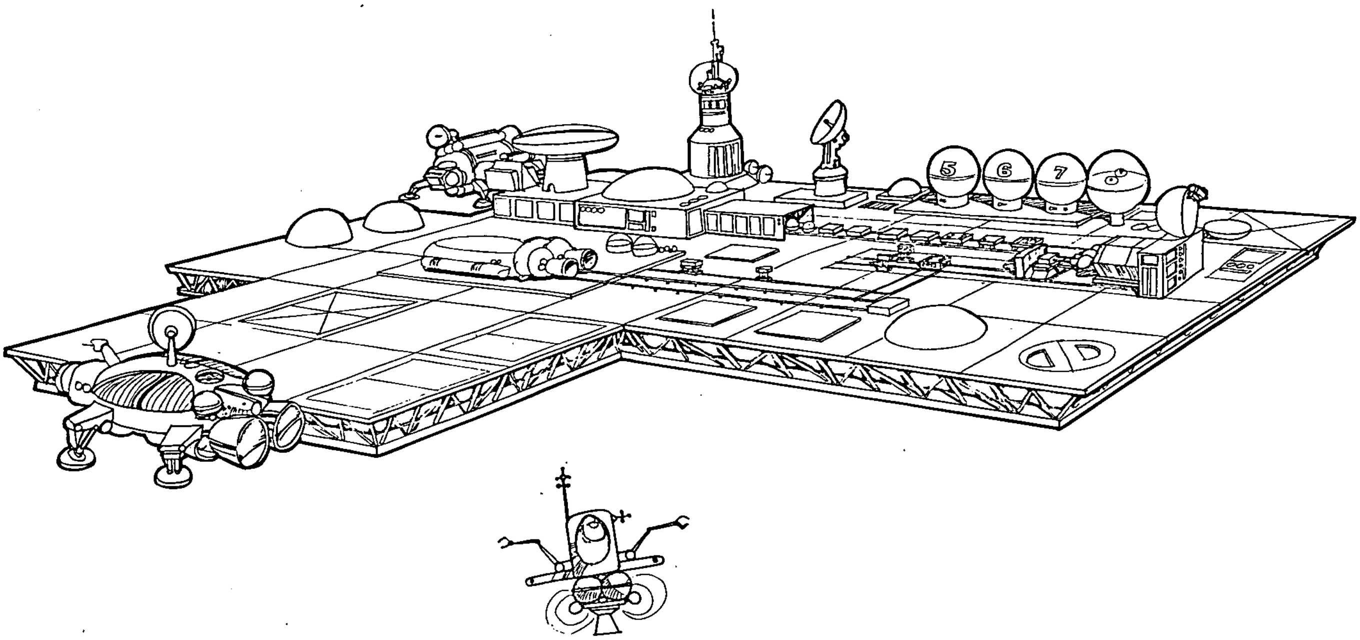


Figure 4.18.– An advanced Space Manufacturing Facility (SMF).

industrial automation and mobile autonomous robotics (Leonard, 1980; Lovelace, personal communication, 1980). Robot mobility studies by the Vought Corporation for Marshall Space Flight Center indicate that construction of space platforms is within the grasp of state-of-the-art automation technology. For instance, robot-compatible fasteners have been developed (Borrego, 1977) and deployed in simulation studies at Langley Research Center (Lovelace, 1980).

Pure glasses and synthetic crystals. The manufacture of complex products containing sophisticated electronic specialized materials may require the preparation of pure glasses and synthetic crystals. Production steps that need to be developed include material separation and sophisticated materials processing.

Consider, for example, the manufacture of synthetic quartz semiconductor materials. Plagioclase first is separated from lunar soil by electrophoresis or other techniques. The refined mineral is then fused and its chemical composition altered to induce quartz to crystallize from the cooling solution. Successful fractionation of quartz from an altered plagioclase melt requires significant advances in the techniques of controlled nucleation, crystallization, and zone refining. Development of a special materials-production capability will permit the manufacture of space-made solar panels, solid-state lasing crystals, fiber optics, and perhaps solar sails. New terrestrial materials techniques such as quick-freezing of molten metals to make "glassy metals" (Giuse and Guida, 1980) may find extensive use in space or on Earth.

Satellites. In-space production of satellites will require the manufacture of special components for control, observation, and communication, and a significant evolutionary advance in automation technology. Satellites may represent the first highly complicated, coordinated construction challenge to be undertaken entirely by teleoperators or robots in space. The construction of solar-power panels, antennas, and sophisticated computer control and communications modules demands a versatile new manipulator system. This system should be equally adaptible to the high-resolution construction tasks necessary in computer assembly and the lower-resolution, high-spatial-range construction jobs required for the assembly of hulls, antennas, and solar panels. Current capabilities of automated assembly are not yet sufficiently well-developed to enable construction of a complete satellite from its constituent parts (Holland *et al.*, 1979; Leonard, 1980; OAST, 1980; Vought Corporation, 1980).

Robots and teleoperators. Two of the most important advanced products to be manufactured in space are robots and teleoperator mechanisms. The ultimate goals for SMF cannot be attained without a significant expansion of the automation equipment initially deployed from Earth. Space robots and teleoperators eventually must be designed from working experience following initial deployment of the starting kit, and then manufactured in space. These second- and third-generation devices must be far more versatile and fault-tolerant than present-day machines. Logistics requirements for production of equipment of this complexity are

staggering. The design must incorporate new features based on earlier experiences with robots and teleoperators in space facilities, and should include either a high degree of self-preservation "instinct" or else a highly adaptive servo-feedback system using extensive space computer facilities as decisionmakers.

The manufacture of robots and teleoperators in space necessitates the automated production of intricate component parts, a task of far greater complexity than current automated assembly systems can handle (Hart, personal communication, 1980). Automated assembly of advanced devices is perhaps no more difficult than the automated assembly of satellites, which already will have been accomplished during an earlier phase of SMF evolution. The most crucial technologies to be developed for the manufacture of second- and third-generation robots and teleoperators are space-adaptive sensors and computer vision. The current state of machine tactile and vision sensor research is insufficient for sophisticated space robots and automated assembly operations (Holland *et al.,* 1979). The best computer-vision package currently available, CONSIGHT-1, can determine the position and orientation of a wide variety of parts with preprogrammed specifications (Holland *et al.,* 1979). Enhanced decisionmaking and self-preservation features must be added to computer-vision systems such as CONSIGHT-1 for use in space robots and teleoperators. A dedicated computer for teleoperator control, programmed to make decisions based on previous experience and insight, would be an instrumental achievement requiring levels of heurism and hypothesis formation unavailable in present-day software (Sacerdoti, 1979).

Solar sails. The solar sails briefly mentioned in section 4.3.1 constitute an unusual but provocative complex product which might be manufactured at the SMF. Sails with a design capability of delivering about two 200-ton payloads per year to the heliocentric distance of Mars have been proposed (Drexler, 1980). Assuming that the viability of self-replicating factories has been demonstrated on the Moon by this point in time (see chapter 5), an interesting scenario would involve the transport of 100-ton self-reproducing "seed" machines (Freitas, 1980c; Freitas and Zachary, 1981) from a lunar-source facility to other moons and planets in the Solar System.

Other complex products. A number of complex products representing various evolutionary steps not yet mentioned or discussed might include impulse landers, biological products, storage tanks, mobile rovers, nuclear-power stations, agricultural products, and many others integral to the evolution of a complex products manufacturing capability. The time sequence of these steps is a function of the desired technologies which must be developed at one stage and integrated at a later stage to make products of ever-increasing complexity.

SMF establishment and growth requires a vigorous parallel development of the three basic materials/energy functions – raw materials and materials processing, manufacturing and technology, and energy production. As the SMF increases in output and creates new net resources, unit output costs should fall and an ever-increasing array of commercially interesting products and services will come into existence. Figure 4.19 and table 4.23 illustrate some of the higher-order systems and services which might be expected ultimately to develop.

4.5 Automation and Manufacturing Technology Requirements

To realize the full potential of space manufacturing, a variety of technological development programs should be initiated in the near future. It is strongly recommended that NASA focus research attention on improvements in teleoperation and robotics, automated manufacturing techniques, and advanced materials processing.

Space manufacturing efforts will draw heavily on teleoperation at first, gradually evolving over many decades towards the extensive use of autonomous robots. Additional research in teleoperation is needed immediately on sensors – tactile, force, and visual, and on sensor and master-slave range scaling. Robotics requirements include improvements in decisionmaking and modeling capabilities, sensors and sensor scaling, mobility, adaptability to hazardous conditions and teleoperator safety (Schraft *et al.,* 1980), natural language comprehension, and pattern recognition. Many of these needs are presently under review by the Engineering Services Division of Goddard Space Flight Center as part of their ongoing CAD/CAM program.

Better automated control systems for space-manufacturing processes are imperative. Machine intelligence controlled laser-, electron-, and ion-beam technologies will make possible the highly sophisticated cutting and trimming operations, integrated circuit fabrication, and other related functions necessary for an efficient SMF operation. Further work should be aimed at devising new fabrication techniques specifically designed for space, such as automated beam builders.

In the materials processing area, effective use of undifferentiated materials such as cast basalt should be stressed. Beneficiation systems better suited to nonterrestrial conditions must be developed to achieve production of differentiated materials with maximum process closure.

4.5.1 Teleoperation and Robotics

Teleoperator development is especially important in the early stages of the space manufacturing effort because the sophistication of current robots in sensory scaling, adaptive control, learning, and pattern recognition is inadequate to establish an autonomous space manufacturing capability. These skills are embodied as subconscious processes in the

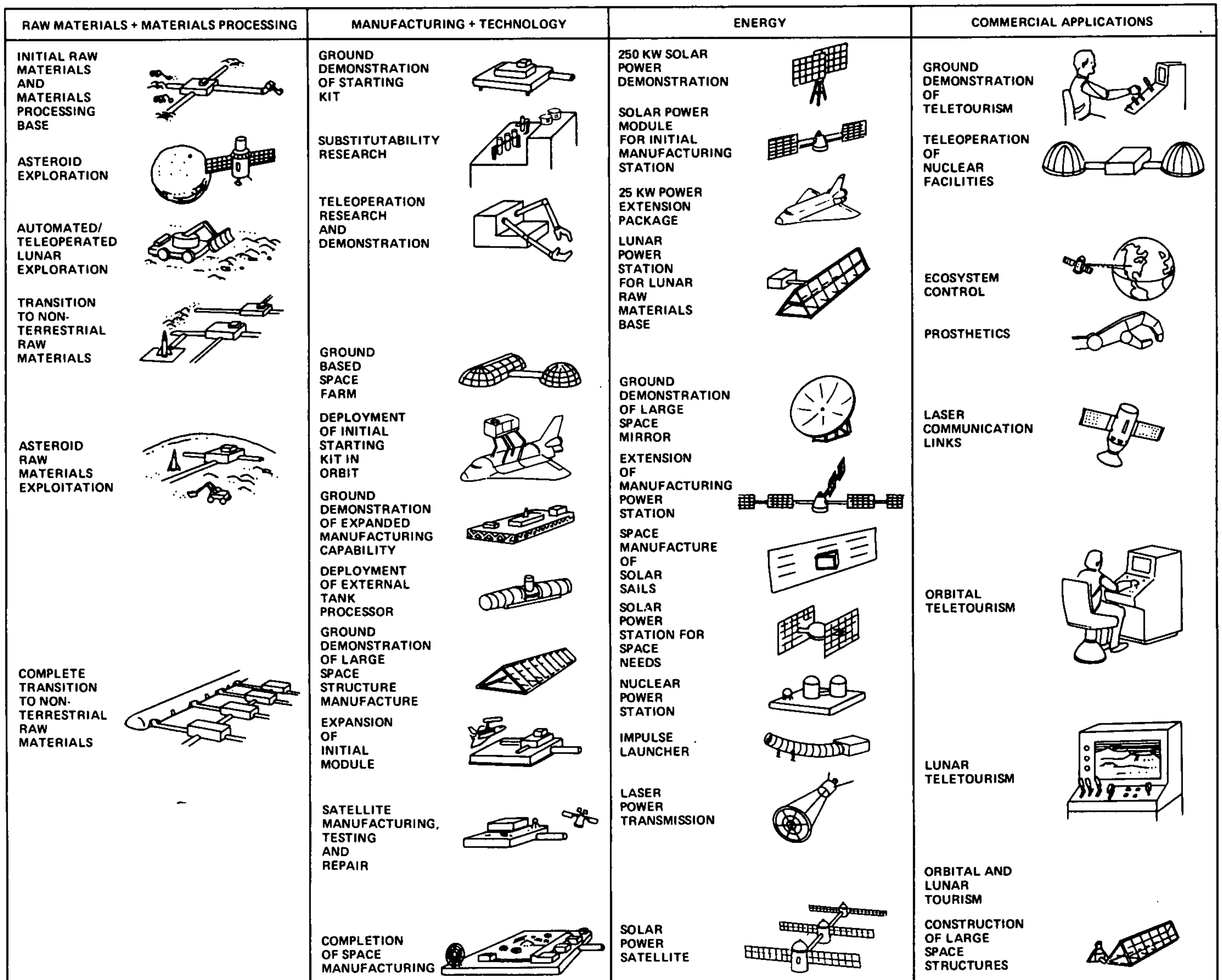


Figure 4.19.– Space manufacturing milestones.

TABLE 4.23.– INTERMEDIATE GOALS IN THE EVOLUTION OF SPACE MANUFACTURING

Raw materials and materials processing
Initial lunar raw materials and processing base– Small processors, soft-landed on the Moon, will extract iron, begin electrophoretic separation of desired mineral phases, and produce silane propellants and oxygen. *Asteroid exploration–* A dedicated telescopic asteroid search will be directly succeeded by exploration of Anteros and other asteroids using rovers and orbiters. *Automated/teleoperated lunar exploration–* Orbital and highly mobile teleoperated or automated rovers will explore the lunar surface (particularly the poles) for possible alkalic basalts and volatiles. The discovery of significant volatiles, especially water, would reduce the complexity of achieving growth and independence. *Transition to nonterrestrial raw materials–* Production of aluminum, titanium, processing chemicals and many other materials will be initiated. This will require a transport system and/or mass-driver facility. This effort will begin with modest goals, later culminating in complete raw materials independence for the SMF. *Asteroid raw material utilization–* Water, carbon, platinum group metals and other materials will be returned to LEO and utilized by SMF. *Transition to nonterrestrial materials completed–* Eventually lunar and asteroidal resources will make a completely independent space economy possible.
Manufacturing and technology
Ground demonstration of starting kit– A ground demonstration of the initial starting kit will be carried out and the development of second-generation tools by the starting kit will be examined. *Substitutability research–* Ground-based research employing simulated lunar and asteroidal materials will be carried out to develop substitute materials for commonly used terrestrial materials which are scarce in lunar or asteroidal soils. *Teleoperator research/demonstration–* Teleoperator research will be directed toward the most sophisticated dextrous operations. *Ground-based space farm–* A small agricultural facility with a closed-controlled atmosphere will be built to examine the feasibility of space agriculture. *Deployment of initial starting kit in orbit–* The deployment, by the Shuttle, of the initial starting kit will be carried out, and second-generation tools will be constructed by the starting kit from Shuttle external tanks. *Ground demonstration of large space structure manufacture–* Large space structures will be manufactured and assembled by teleoperators and robots in a water tank simulator. *Ground demonstration of expanded manufacturing capability–* Second-generation tools will be employed to manufacture and assemble products. The feasibility of third-generation tools for greater manufacturing versatility will be examined. *Expansion of initial module–* With additional feedstock derived from additional external tanks, the manufacturing facility can be expanded from the initial module. *Satellite manufacturing, testing, and repair–* Satellites will be constructed by the SMF mobile units that will refuel existing satellites and modify and test experimental satellites. *Completion of space manufacturing–* At this stage, all products required in space are manufactured from nonterrestrial materials. High-unit-value products may be transported back to Earth.
Energy
250 kW solar power demonstration– A proof of concept demonstration for conversion of solar energy into microwaves and transmission of microwaves to a distant station as usable energy will be implemented. *25 kW power extension package–* The PEP will be deployed in order to enable the Shuttle to remain at the station longer and perform more complex missions. *Solar power module for initial manufacturing station–* This module enables autonomous operation of the initial manufacturing module. This unit is a descendant of the power extension package for the STS. *Lunar power station for raw materials base–* A solar or nuclear power plant will be deployed on the Moon to supply power for exploration, acquisition, and processing of raw materials. This facility will be large enough to allow for transportation of materials to LEO.

TABLE 4.23.– CONCLUDED

Energy – Concluded
Ground demonstration of large space mirror– A proof-of-concept demonstration of manufacture and construction of large space mirrors will be a necessary precursor for the Solaris mission.
Extension of manufacturing facility power station– The manufacturing power station will be expanded to accommodate the expanded manufacturing capacity. Additional power is required for expanding the acquisition and utilization of nonterrestrial materials.
Space manufacture of solar sails– Thin-film solar sails, which are difficult to construct on Earth and very difficult to deploy, will be manufactured in space. The solar sails will be employed to transport payloads within the inner Solar System.
Solar power station for space power needs– An SPS will be constructed to supply electrical energy to stations in space. The power may either be used where it is developed or transmitted over distances to remote stations.
Nuclear power station– Fission or fusion energy will be employed in those situations where solar energy is impractical. A nuclear power station will be constructed for outer Solar System missions and lunar night power.
Impulse launcher– A mass-driver reaction engine will be developed and deployed as a part of the materials and products transport system.
Laser-power transmission– A laser-power transmission system will be developed and deployed. The precise frequency of laser light will enable tuned photocells to be used to convert the laser beam into useful power.
Solar power satellite– The development of the power station for space use and the laser-transmission system culminates in the development of the SMF solar power station which will be capable of delivering multi-gigawatts of energy for transmission to the Earth.

Commercial Applications
Ground demonstration of teletourism– Development of teleoperators for space might lead to "teletourism." People could "travel" to exotic places via teleoperation.
Teleoperation of nuclear facilities– Advanced teleoperator technology could eliminate radiation exposures in nuclear facilities by eliminating human operation in dangerous areas.
Ecosystem control– Enhanced remote-sensing technologies developed in the manufacturing facility could provide monitoring and "fine tuning" of terrestrial ecosystems.
Prosthetics– Research on advanced teleoperators and robots would greatly enhance the field of prosthetics. Sensory display devices, for instance, might be adapted as aids for blind and deaf persons.
Laser communication links– High bandwidth laser data links to space and Earth stations will be developed. This will greatly ameliorate the radio band allocation situation.
Orbital teletourism– High bandwidth communications satellites, manufactured by the SMF, could be employed for orbital teletourism.
Lunar teletourism– Manipulators and viewers on the lunar surface could provide the ability to develop lunar teletourism.
Orbital and lunar tourism– Fully reusable Shuttle-derived lift vehicles will permit orbital plus lunar tourism packages.
Construction of large space structures– Such large space structures as medical centers, space and worldwide communications centers, and hotels would provide a survival capability in the event of a terrestrial catastrophe.

human nervous system. The development of teleoperators with sufficient interface dynamics would provide "telepresence" (Minsky, 1979, 1980) in the early stages of SMF development while significant new robotics research is undertaken.

The team surmises that within the next 50 years robot systems will be capable of handling a large fraction of the needs of a general-purpose SMF. The feasibility of robot systems making sophisticated judgments is less certain. Controls likely will evolve from teleoperated to semiautomated, then to fully automated (Bejczy, 1980). Cost requirements in orbit or on the Moon or asteroids may encourage development of adaptive robots with flexible control systems (Asada and Hanafusa, 1980). According to research currently underway at the School of Electrical Engineering at Purdue University, a limiting requirement may be manipulator motion (Paul *et al.*, 1980). Manipulators in an SMF must be capable of working on a moving assembly line – the maximum "reach" of current Cyro robots is 3 m – and of accepting visual position information. It is also important to determine the degree to which real time computational constraints can be relaxed in controlling robot motions in Cartesian coordinates. In extraterrestrial environments, the dynamic behavior of each link in a manipulator arm must be considered. Centrifugal and coriolis accelerations (in spinning systems) and gravity loading are significant factors governing the relationship between forces and moments of successive links.

Limits on control requirements also have been considered by Yushchenko (1980), who has written algorithms for semiautomatic robot operations. Since semiautomatic robots undoubtedly will precede fully automatic robots into space, the three major techniques of direct human master control – velocity, force, or position – must be considered. Velocity methods are rapid but manipulator motions are imprecise. Force methods control manipulators through human feedback in Yushchenko's study, but these techniques provide little regulation of acceleration during object motion. Limitations in force-sensing controls for mating of parts have been reviewed by Korolev *et al.* (1980) and by the Draper Laboratories, the latter quantifying clearance and friction factors. The positional method ensures proportionality of linear and angular displacements of manipulator grip through the handle of a master control device.

Manipulators need to be greatly improved. Current master-slave devices require 2-3 times longer to accomplish a given task than do human hands (Bradley, personal communication, 1980). The mass of teleoperator appendages is high compared to the weight they can lift. With better visual and tactile feedback, the heavy, rigid manipulator arms could be replaced by lightweight, compliant, yet strong arms. To accomplish this, the low-resolution, low-stability, low-dynamic-range force reflection tactile systems must be replaced with servofeedback systems including suitable touch display modules. Viewing systems will require additional research and development – the most advanced system currently available is a monocular head-aimed television. This system should be redesigned as a binocular system with auto-focus, variable resolution, and color. Sensory scaling to compensate for differences in size between slave and master manipulators is necessary for fault-tolerant teleoperation. This may be accomplished by adjusting the scale of the master visual image or by incorporating error signals into the visual display.

Limitations also arise by virtue of the space environment itself, whether in LEO, on the lunar surface, or on asteroids. Hard vacuum demands redesign of robot joints and manipulator end-effectors to minimize undesired cold welding if de-poisoning of metal surfaces occurs. Radiation bursts during solar flares could possibly induce embrittlement of metal components of automata. Likewise, electronic components could be degraded or altered by temperature extremes.

4.5.2 Functional Requirements for Automation

The functional requirements for an automated SMF, taken in part from Freitas (1980d), are listed below roughly in order of increasingly sophisticated capability: robot language systems, product assembly, product inspection and quality control, product modification, product repair, product adjustment, product improvement, remedial action by reason of emergency or subtle hazard, robot self-replication. It is assumed in each case that the impediments to meeting these requirements (e.g., control techniques, "packaging" to withstand hostile ambient environments, etc.) will somehow be overcome. The first three functional requirements are described briefly below, followed by a general discussion of the more advanced requirements.

Robot control languages. Numerous machine languages exist for the control of semiautomated machine tools (Lindberg, 1977). These include APT (automatic programming tool) and ICAM (integrated computer aided manufacturing). McDonnell Douglas Aircraft Company has recently extended APT to MCL (manufacturing control language) in order to program a Cincinnati Milacron T3 robot to rivet sheet metal. Higher-level robot control languages, obvious requirements for advanced automated space systems, include VAL (versatile assembly language) for the Puma robot and "HELP" for the Pragmac robot (Donata and Camera, 1980). The problem of extending high-level languages from comparatively simple machine tools to more sophisticated multiaxis integrated robot systems which may be found in future automated space factories must be viewed as a top priority research item.

Product assembly. At SRI International, requirements for the five basic operations in factory assembly have been evaluated by Rosen *et al.* (1976). These include (1) bin picking, (2) servoing with visual feedback, (3) sensor-controlled manipulation, (4) training aids, and (5) manipulator path control.

The team has recognized the need for improved performance in bin picking of, say, assorted cast basalt and metal objects. Multiple electromagnetic end-effectors certainly could pick out just the metal casings. Variably energized end-effectors might be used to separate and select metal parts of varying magnetic susceptability randomly arranged in a bin (i.e., aluminum vs iron vs titanium parts). But general bin picking from random parts assortments is not yet possible, though it might be essential in a fully automated SMF operation.

SRI has applied visual servoing by combining a General Electric television (100 X 100 element solid-state) camera with an air-powered bolt driver incorporated into an end-effector. Three-dimensional cameras may be required for highly contoured objects fabricated in space (Agin, 1980; Yachida and Tsuji, 1980). Such cameras have already been applied to automated bin selection tasks by the Solid Photography Company in Melville, New York.

Computer-vision technology needs to be merged with discoveries from biological studies. Automatic gain control, gray-scale imaging, and feature detection must be included in computer-vision technology if robot autonomy is the goal. Parallel computer-control systems will ensure the speed of reaction and self-preservation "instincts" required for truly autonomous robots, but will require a decrease in existing computer memories both in size and access time by several orders of magnitude. Consideration should be given to associate and parallel memories to couple perceptions to the knowledge base in real time.

To achieve sensor-controlled manipulation, somewhat greater precision is required of robot arms than can be obtained now. Present-day Unimates (control arm precision of 2.5 mm) have been used in a one-sided riveting operation using strain-gauge sensing of the rivet gun mandrel, but there is still a need for more rapid finding, insertion, and fastening by passive accommodation, servo adjustment, and search algorithms. A novel "eye-in-the-hand" adaptation for rapid assembly in space may utilize acoustic sensors. The Polaroid Corporation in 1980 applied its camera ranger to end-effectors for tool proximity sensing. The unit emits a millisecond pulse consisting of four ultrasonic frequencies (50, 53, 57, and 60 kHz). Ultrasonic techniques are potentially quite useful in air or other fluid-filled bays in nonterrestrial manufacturing facilities, especially in view of the acoustic positioning systems developed by the Jet Propulsion Laboratory for containerless melt manipulation. Under vacuum conditions when precise positioning is necessary, laser interferometry may provide the answer (Barkmann, 1980).

Regarding training aids, more sophisticated coordinate transformation programs are required to operate manipulators for diverse tasks. A possibility for the future is "show and tell," a new technique for robot training (see chapter 6). Ultimately, a robot itself could train future-generation machines through some means of "training-by-doing." A related issue – the problem of robot obsolescence – will not be trivial.

Finally, manipulator path control should be fully automated in SMF where, for example, rock melts must be transported along smoothly controlled paths (see the discussion of basalt fiber spinning in section 4.2.2). In the manufacture of bearings or fibers where high-speed trajectories are involved, manipulator halts at corners must be avoided by developing better path control strategies. In the near-term, it may be possible to extend the capabilities of the Unimate:PDP-11/40 couple. For every machine proposed for the SMF, including the starting kit extruder, it is simplest to use a coordinate system based on that machine to interact with robot manipulators continuously to redefine forbidden regions and motions. Thus, a major requirement in robot factory assembly is to specify the coordinate systems of the component machines.

Product inspection and quality control. The need for visual methods of inspection and quality control by automata must be defined for each class of SMF product envisioned. For instance, the application of electroforming on the Moon to produce thin-walled fragile shapes, aluminum ribbon extrusion, or internal milling of Shuttle tanks, definitely demands inspection and quality control. Terrestrial automated inspection systems currently are in use at General Motors, Western Electric, General Electric, Lockheed Recognition Systems, Hitachi Corporation, SRI International, and Auto-Place Corporation. A detailed synthesis of the vision requirements for each is given by Van der Brug and Naget (1979). Off-the-shelf television systems with potential for robotics applications already provide measurements to 1 part in 1000 of the height of the TV image, e.g., the EyeCom Automated Parts Measurement System manufactured by Special Data Systems, Inc. in Goleta, California. Finally, the use of fiber optics in quality control, as demonstrated by systems now in use by Galileo Electronics, Inc., warrants further development.

Advanced functions and recommendations. The needs of space manufacturing for automated product modification, repair, adjustment and improvement, as well as robot adaptation to emergencies and self-replication, depend in large part on the capabilities of future automata control systems and the environment in which they are applied. The hazards of space to human beings are well known, whereas their impact on robot systems is less well understood. Potential dangers include rapid pressure changes, spillage of corrosive fluids or hot melts due to vessel rupture, radiation effects

from solar flares (e.g., embrittlement), anomalous orbital accelerative perturbations producing force-sensor errors, and illumination-intensity variations caused by space platform tumbling or nutation (producing visual observation problems such as shadow effects in fiber optics sensors).

Robotic intelligence must be vastly increased if these devices are largely to supplant human workers in space. This may be accomplished by deploying a versatile intelligent multipurpose robot or by developing a number of specialized, fixed-action-pattern machines. Multipurpose intelligent robots lie well beyond state-of-the-art robotics technology, yet they still are an important ultimate goal. In the interim, sophisticated fixed-action-pattern robots suitable for restricted task scenarios should be developed. The behavior of such robots would be not entirely different from that of many plants and animals endowed with very sophisticated fixed action patterns or instincts.

Before true machine intelligence can be applied to factories in space, the requirements for automated nonterrestrial manufacturing systems must be determined by an evaluation of the state-of-the-art in this field. A complete and updated computerized library containing abstracts of all available robotics research and applications publications, accessible through ARPANET, should be implemented to enhance automation technology transfer. Among the subject categories which should be emphasized are controls, arm/work envelopes, robot adaptability, applications, and costs. Knowledgeability in the field requires contact with firms listed below to better understand how solutions of the practical problems of today can be extrapolated to help solve those of tomorrow: Unimation, Inc.; Cincinnati Milacron; ASEA, Inc.; Prab Conveyors, Inc.; Planet Corporation; Devilbiss/Trallfa; Nordson Corporation; Binks, Inc.; Thermwood Machinery Corporation; Production Automation Corporation; AutoPlace Company; Modular Machine Company; Seiko Instruments, Inc.; Jones Oglaend Corporation; Fujitsu Fanuc Corporation; Okuma Machinery Corporation; Advanced Robotics Corporation; Hitachi Corporation; and Benson-Varian Corporation.

4.5.3 Space Manufacturing Technology Drivers

The successful deployment of a large, growing, independent SMF requires technologies not presently available. Three technical areas in particular will require major developmental efforts: manufacturing technologies, materials processing, and space deployment. Many of the technology drivers and required advancements discussed previously are currently the subject of some R&D activity at various industrial and government research facilities. The first and perhaps most crucial step in any technology drive to make the SMF a reality is a thorough synthesis and coordination of current and previous research. A determined effort must then be made to augment technical competence as required to sustain a successful space manufacturing venture.

Manufacturing technologies. The control system for an automated manufacturing facility must be sophisticated, fault tolerant, and adaptive. Technological advances required for a factory control system are primarily software developments. A "world model" for the facility must comprehend variable throughput rates, breakdowns, and unexpected commands from Earth-based supervisors. The control system also must be able to formulate and execute repair plans, retooling exercises, and scheduling options. Such a system needs flexible hypothesis formation and testing capabilities, which in turn demands heuristic programming employing some measure of abductive reasoning without requiring unreasonably large memory capacities (see sec. 3.3).

Advances in ion-, electron-, and laser-beam technologies are necessary for welding, cutting, sintering, and the fabrication of electronic components. The efficiency and power of weapons-grade tunable lasers now under development by Department of Defense contractors (Robinson and Klass, 1980) already are high enough to fulfill most cutting and sintering needs of the SMF. Heat dissipation is a substantial problem inherent in laser utilization for space manufacturing. Space-qualified heat exchangers must be developed for laser-beam machining to achieve its full potential as a viable macromachining space technology. In addition, industrial lasers must be designed to re-use the working gases.

In the manufacture of electronics components, ion-beam devices are required for implantation and etching in space. Lasers are helpful in facilitating annealing and oxidation processes and in trimming fine-tolerance capacitors and resistors. Electron beams have applications in silicon crystal purification and deposition of metals, though lasers also may be employed. Other uses for each beam type are readily imaginable. High-resolution automated control technologies must be developed for implantation, annealing, etching, and trimming processes in particular.

Contact welding is a highly useful feature of the vacuum space environment. Of course, in some instances cold welding must be avoided so surface poisoning methods must be developed. Terrestrial poisoning agents such as hydrogen, hydroxyl, and various surfactants are not readily produced from nonterrestrial materials. Highly adsorptive oxygen-based surface active agents appear to be the most feasible solution to the cold welding problem.

Materials processing. Extensive research is needed in the field of processing of raw materials if a self-sufficient space manufacturing presence is to be established. Several possible avenues include fractionation, zone refining, and oxygen-based chemical processing. Fractionation of a wide variety of elements including fluorine, hydrogen, silicon, boron, phosphorus, and many others is a prerequisite to

independent manufacturing in space. Raw material separation prior to processing (primary beneficiation) is a logical step in the total beneficiation process. The preliminary isolation of particular compounds or mineral species could significantly reduce the problems inherent in developing suitable chemical-processing options.

Space deployment. There are a number of mission tasks associated with space manufacturing for which technological developments must be made. Sophisticated rendezvous techniques are needed for SMF resupply, in-orbit assembly, and satellite tending. Deployment of repair rovers is required for satellite maintenance and troubleshooting. Long-term satellite autonomy is not possible without repair and refueling capabilities which are not currently available. Large-mass deployment and retrieval procedures must likewise be developed if feedstock, raw materials, and products are to be delivered to or from the SMF. Multimission compatibility must be designed into satellites, shuttles, and transport vehicles if self-sufficiency is to be achieved within a reasonable time.

4.5.4 Generalized Space Processing and Manufacturing

A generalized paradigm for space industrialization is presented in figure 4.20. Solar energy powers the systems which gather nonterrestrial materials for conversion into refined materials products. These "products" can be additional power systems, materials gathering/processing/manufacturing systems, or simply support for other human and machine systems in space. Earlier chapters examined observational satellites for Earth and exploration systems for Titan having many necessary features of a generalized autonomous robotic system designed to explore the solid and fluid resources of the Solar System (item (1) in fig. 4.20) using machine intelligence. However, in the materials and manufacturing sectors a qualitatively new interface must be recognized because "observations" explicitly are intended to precede a change of objects of inquiry into new forms or arrangements. These machine intelligence systems continuously embody new variety into matter in such a way that preconceived human and machine needs are satisfied. This "intelligently dynamic interface" may be explored as two separate notions: (1) a generalized scheme for materials extraction, and (2) the (fundamentally different) generalized process of manufacturing (see also chap. 5).

Generalized materials processing system. Figures 4.21 and 4.22, developed by R. D. Waldron (Criswell, 1979), offer a very generalized overview of the options and logic involved in the selection of a processing system for an arbitrary raw material input. By way of illustration, note that the extraction (in either reduced or oxide form) of the seven most common elements found in lunar soils requires

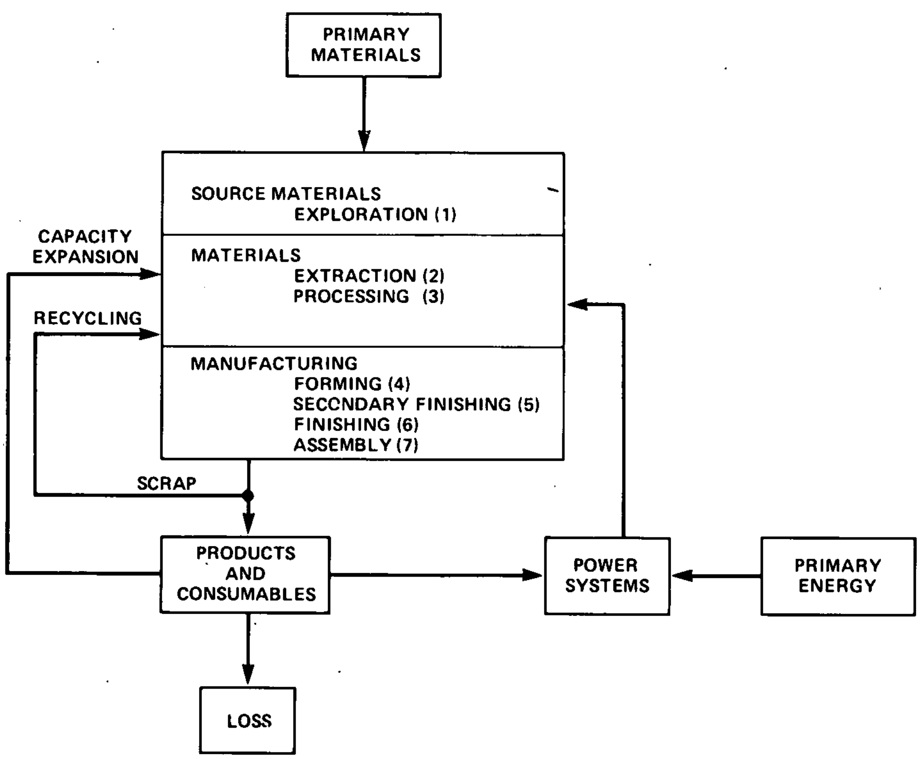


Figure 4.20.– A generalized paradigm for space industrialization.

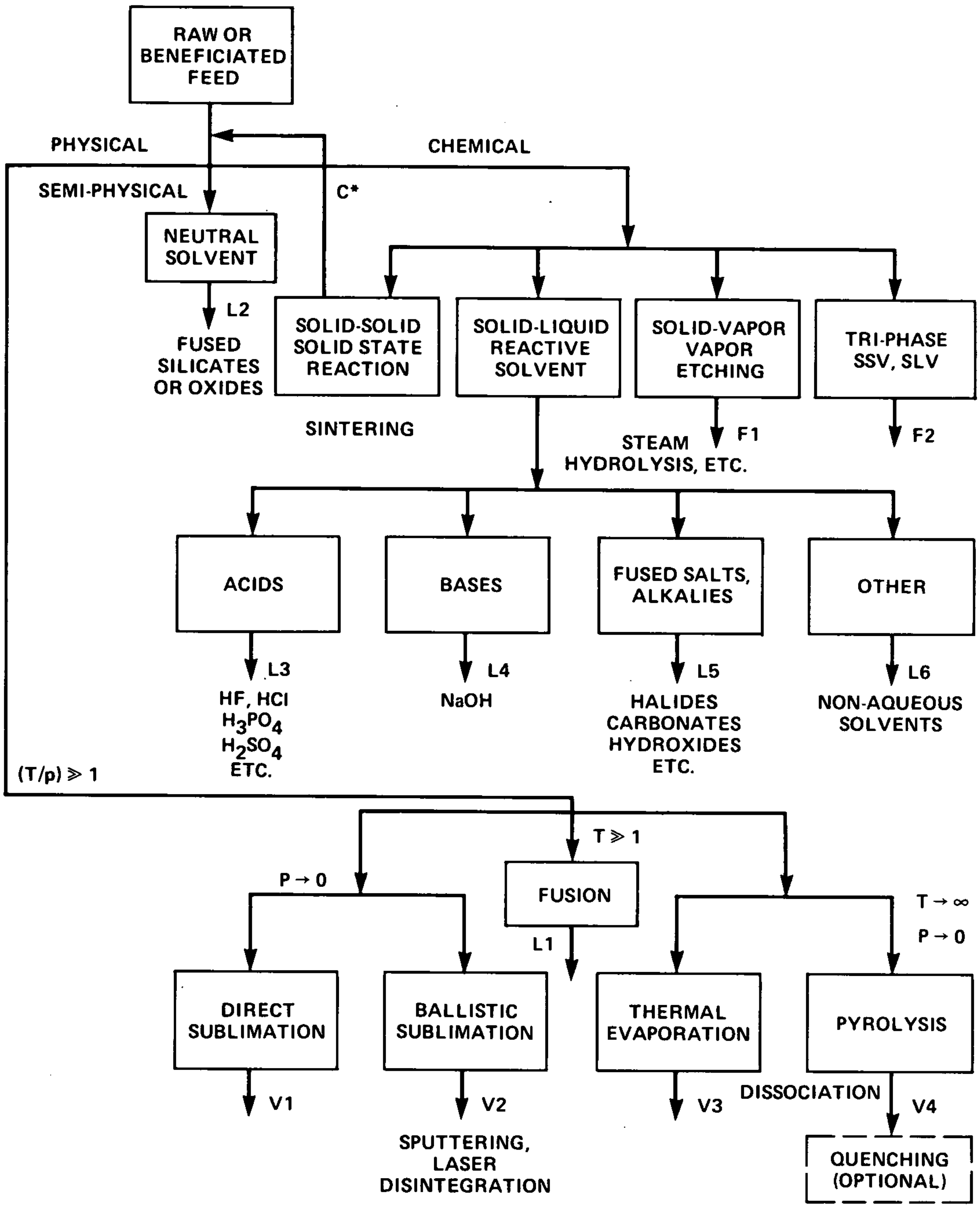


Figure 4.21.– Mobility/diffusibility materials processing options.

at least six separation steps, with yet additional steps for reagent recycling. Even if a single separation technique from each of the 22 categories shown in figure 4.21 is considered for each of the six lunar elements, more than 113,000,000 combinations (22^6) of separation segments would be possible. The 13 categories of mobility/diffusibility options further increase the total number of process variations available.

Clearly, an enormous range of materials-processing alternatives can be indexed by a finite number of decision nodes. One might imagine a very large, complex, but finite extraction machine comprised of 35-40 process categories, each capable of performing an operation described in figures 4.21 or 4.22 (e.g., ballistic sublimation, liquid-solid absorption/ion exchange). In addition, each category subsystem is capable of fully monitoring its own input, internal, and output materials streams, and environmental or operating conditions, and must have access to detailed knowledge of relevant data and procedures in chemical engineering, physics, and the mathematics necessary to

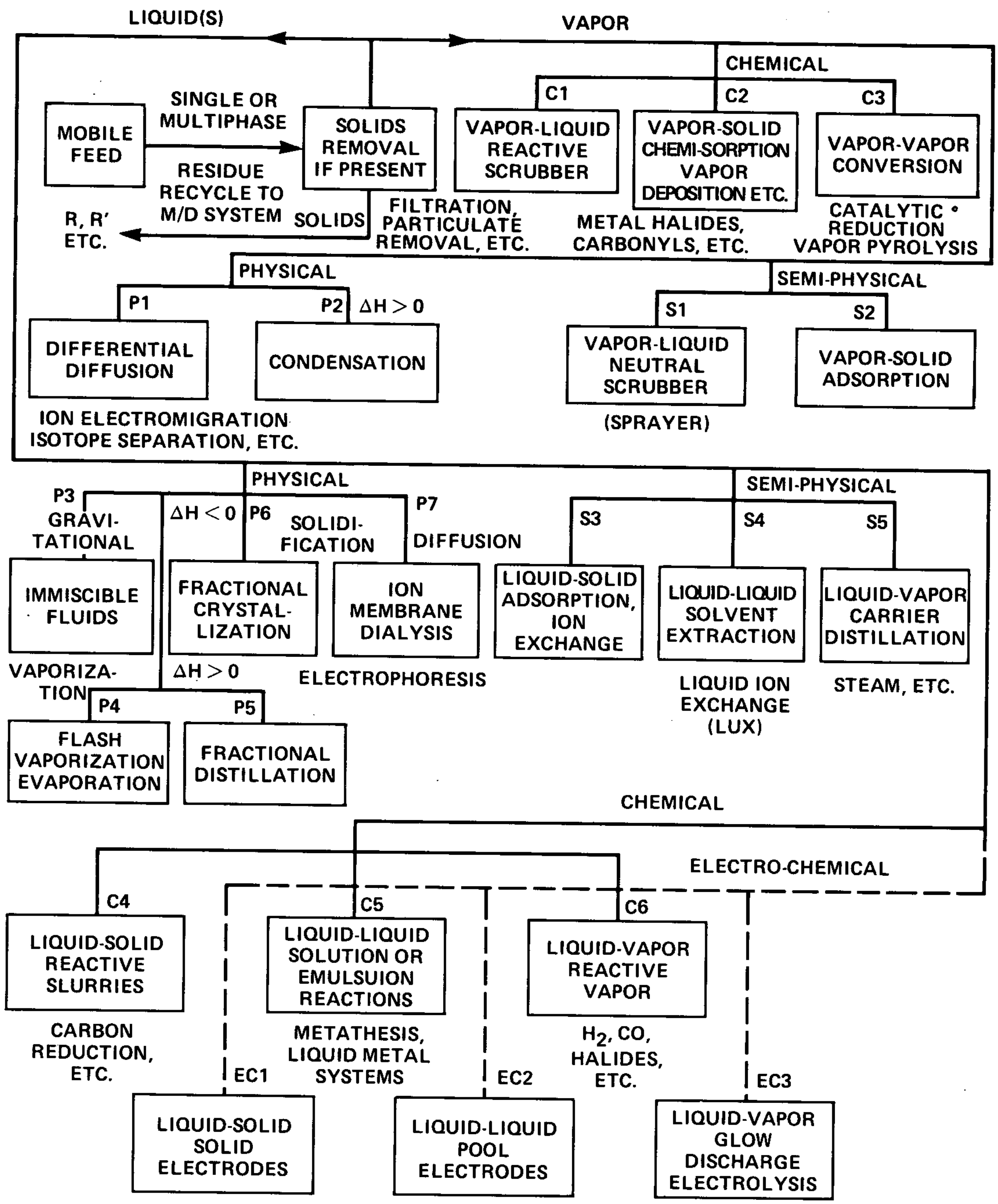


Figure 4.22.– Separation materials processing options.

maintain stable operation or to call for help from an overview monitor system. Each processing subsystem communicates extensively with an executive system to select process flows consistent with external factors such as available energy, excess materials, local manufacturability of process components, necessary growth rates and the general environment.

During deployment, the complete package is delivered to a materials source. Representative local raw materials are sampled to select appropriate overall processing options. After selection is made, throughput rates in the process stream are upgraded to full production levels. Output materials are delivered to a generalized manufacturing system which builds larger specialized production units and support systems such as power supplies, mining, and other materials-gathering equipment, transporters, and related items.

In the most general terms, the Materials Processing System reduces variety in the local environment by absorbing unknown or chaotic resources and producing numerous

output streams of well characterized industrial materials. Variety reduction is accomplished by definite and finite sequences of analytic operations. The analysis task, though large, is finite. The next step, manufacturing, involves the production of possibly an infinite number of forms, hence will likely require different mathematical and computational approaches.

The concept of a self-contained regenerative processing unit affords an interesting didactic tool. What tasks would be required for the unit to manufacture a collection of locally appropriate processing subsystems? What "cognitive structures" are necessary to organize and to direct the activities of the manufacturing units and the 35–45 analytic cells? Further questions regarding possible tasks include:

- What physical operations and observations must be conducted in each process category?
- What equipment types are common to various categories of materials processing, materials transfer, and storage needs?
- What chemicals are essential for the materials processing capabilities desired?
- Have any process categories been omitted?
- What physical knowledge of processing operations must be embedded in directly associated machine intelligence (MI) units?
- What are the necessary relations between extent of exploration observations, initial test processing, and build-up to large-scale processing?
- How many process paths should the overall system physically explore? To what extent, and how, should theoretical understanding and limited observations be used to rule out the vast majority of processing alternatives to permit early focus on adequate production sequences?
- How can new knowledge acquired in operations in new environments and with new compounds be incorporated into the MI system?
- What principles of overall management must the system obey to ensure survival and growth?
- What are the fundamental ultimate limits to the ability of self-regenerative systems to convert "as found" resources into industrial feedstock? Are there any essential elements which limit growth by virtue of their limited natural abundance?
- How can an understanding of physical principles be incorporated into the overall management system to direct operations?

Generalized manufacturing. Figure 4.23 illustrates the generalized manufacturing process. Units 2-8 suggest the flow of formal decisions (along a number of "information transfer loops") and material items which finally result in products. The management unit directs the entire enterprise in response to internal and external opportunities and restrictions. Development of new products requires participation of the entire system, whereas manufacture of repetitive output focuses on providing smooth production flows through units 4-8 guided by management. This schema explicitly refers to the manufacture of "hard products" such as telephones, automobiles, and structural beams, but a generally similar methodology also applies in the preparation of made-to-order chemical compounds. Thus, the reduced chemical feedstock discussed earlier may supply material to logistics (8) for input to manufacturing processing.

Considerable progress in automation and computer assistance have been made in the functional areas of design (2:computer aided design), parts fabrication (4:computer aided manufacturing), logistics (7:computer aided testing), and management support (1). If extension of state-of-the-art practices is focused on space operations, further advancements readily may be visualized in parts fabrication (4: e.g., flexible machining systems), materials handling (5: e.g., automated storage systems and transfer lines, retrieval, parts presentation), assembly (6: e.g., robots with vision and human-like coordination), and inspection and system testing (7: e.g., physical examination using vision, sonics, X-rays, or configuration as when checking computer microchip integrity).

Major additional research is necessary in process planning (3), handling (5), assembly (6), and inspection and system testing (7) in order to fully develop autonomous SMF. Although machine intelligence systems are appropriate in all phases of manufacturing, the most advanced applications will be in management, design, and process planning.

There is a fundamental difference between generalized materials processing and manufacturing. In the former (production of "standardized" industrial materials) the system is designed to reduce variety of originally random or unstructured resources. There are a finite number of chemical elements and a finite but extremely large collection of processes and process flows by which chemical elements may be derived from primary native materials. On the other hand, manufacturing processes presumably can impress virtually an infinite range of patterns upon the matter and energy of the Universe. Substitutions of materials and alternate solutions to various engineering challenges are manifestations of the diversity possible. Parts fabrication is the "materials" focus of manufacturing; as shown in figure 4.23, there are four major steps – parts formation, secondary finishing, finishing, and assembling – with matter flowing generally from one stage sequentially to the next.

Table 4.24 by Waldron (Criswell, 1979) presents a noninclusive functional taxonomy of manufacturing processes

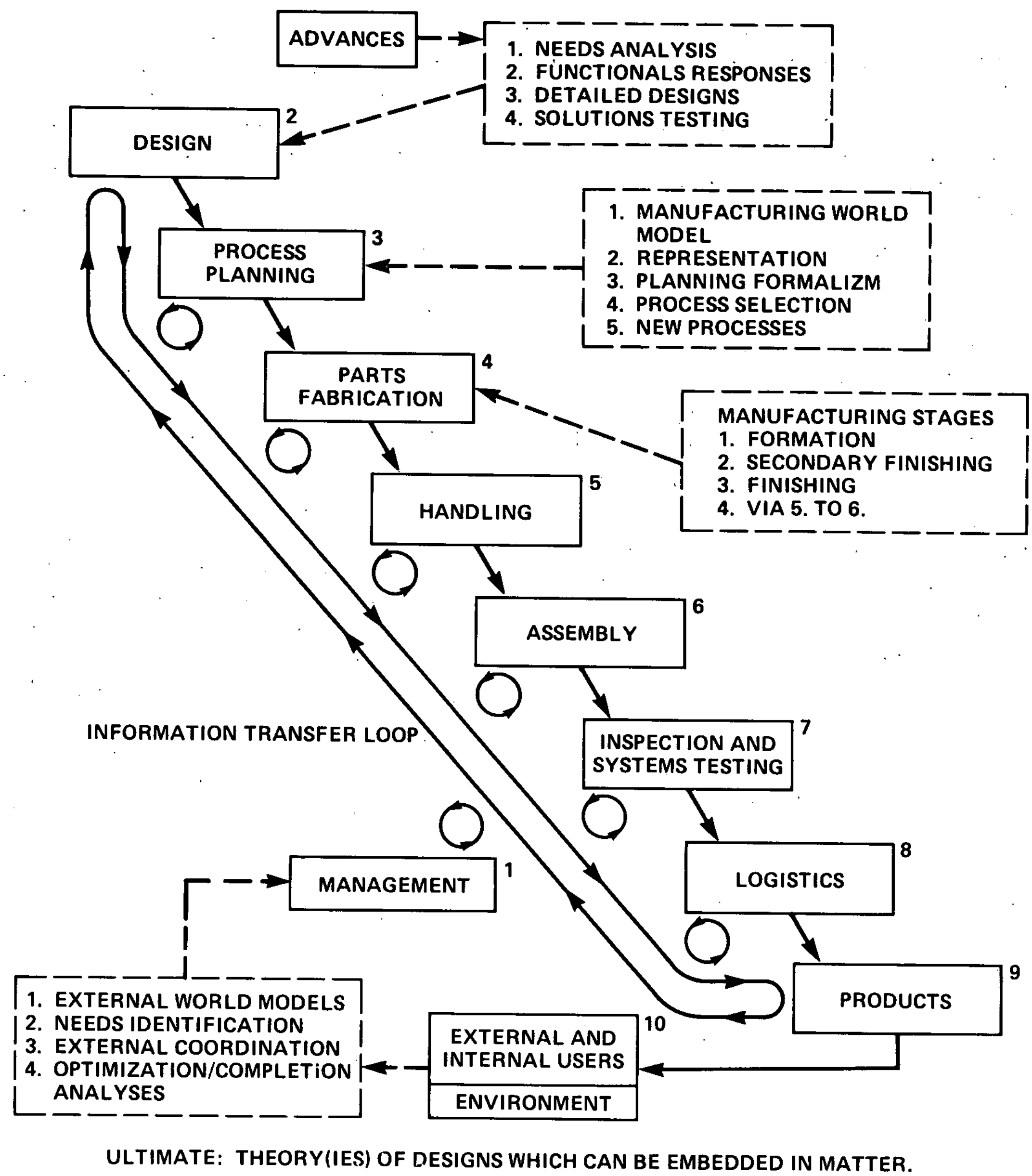


Figure 4.23.– Comprehensive manufacturing schema.

which is organized differently from table 4.17. With few exceptions all may be applied to advantage in one or all of the four stages of manufacturing. Each can be used to produce parts of arbitrary size, form, dimensional accuracy, composition, and other collective properties (e.g., magnetic susceptibility, tensional strength, thermal conductivity, switching speeds), so it is clear that a continuously growing diversity of products is possible. Thus, manufacturing intrinsically requires machine intelligence systems to create novel forms embedded in nonterrestrial materials. In turn, these "matter patterns" might be used to control nonmaterial flows of electric and magnetic patterns, momentum, photons and infomation – the key to further propagation of new pattern production.

The following is a list of research challenges extending from the broadest issues of "matter patterns" to the present state-of-the-art of machine intelligence as applied to design, process planning, and management units depicted in figure 4.23:

- Creation of world models and methods of identifying "needs" for materials, energy sources, products, etc., which the system must provide for further growth.

TABLE 4.24.– GENERAL CLASSIFICATION OF MANUFACTURING PROCESSES

1. Kinematic
   - K1 Fiber operations (felting, paper, textiles)
   - K2 Particulate operations (mixing, conveying, clustering, dispersing)
   - K3 Fluid operations (mixing, pumping, heating, cooling, phase separations)
   - K4 Discrete operations (small scale assembly)
2. Forming
   - F1 Hot forming (casting, forging, powder metallurgy, sintering, hot rolling, bending)
   - F2 Cold forming (extruding, bending, punching, drawing, machining, grinding, cold molding)
   - F3 Unconventional forming (electromagnetic)
3. Surface treatment
   - S1 Removal (washing, etching, electropolishing)
   - S2 Addition (coating, plating, anodizing)
   - S3 Modification (bonding, hardening, shot peening)
4. Internal treatment
   - I1 Internal heating (resistance, induction, dielectric)
   - I2 Irradiation (X-ray, gamma ray, electron beam, UV)
   - I3 Miscellaneous (magnetic poling, ultrasonic)
5. Bonding
   - B1 Hot processes (welding, brazing, soldering)
   - B2 Cold processes (adhesives, chemical bonding, ultrasonic welding)
   - B3 Unconventional processes (explosive, electromagnetic)
6. Large-scale assembly
   - L1 Reduced gravity (lunar surface)
   - L2 Microgravity (orbital assembly)

- Observational and communications means and strategies by which world models can be extended, compared to external realities, and then needs recognized and fulfillments confirmed.
- Computational strategies for optimal uses of the means of production and the resources for creating new products.
- A method of creating, analyzing, and testing new designs derived from validated theoretical concepts or empirically justified knowledge (i.e., that something works). A similar need exists in the task area of assembly in which knowledge of the desired functions of a device or system can be referred to in the assembly procedure rather than referencing only configurational information or combinatorial blocks in a sequence of assembly steps.
- Some means of representing the resources of a production system and a formalism for process planning tasks.

The scientific and engineering communities continually strive, in a somewhat uncoordinated manner, to develop new comprehensive physical theories and then apply them to the creation of new material systems. A new scientific/engineering discipline is needed which explicitly and systematically pursues the following related tasks:

- Document the historically evolving capability of humanity to impress patterns onto matter, the quality of life as patterning ability becomes more sophisticated, the physical dimensions of pattern impressment, the interaction of new patterns by which even more comprehensive orderings may evolve, and the relationship between physical control over matter-energy and the socially based field of economics.
- Investigate on very fundamental levels the interrelations among information, entropy, negative entropy, self-organizing systems, and self-reproducing systems. This study should incorporate the latest thinking from the fields of physics, mathematics, and the life sciences in an attempt to create a model or theory of the extent to which regenerative and possibly self-aware designs may be impressed onto local and wider regions of the Universe – a "general theory of matter patterns."

- Seek the transforms which can be employed at any stage of development to create higher orders of matter patterns.

Human thoughts and conversations typically are conducted using "object"- and "action"-based words learned during childhood. Deeper and more widely applicable symbolic manipulations may be derivable from the mathematical fields of group/set theory, topology, and from the physical and social sciences. A long-term research program should seek to construct a "relationally deep" natural language for human beings and to develop systems for teaching the language both to adults and children. In effect this program would strive to understand intelligence as an entity unto itself and would attempt to explore, identify, and implement more capable "intelligence software" into both life-based and machine-based systems.

4.6 Conclusions, Implications, and Recommendations for Implementation

The Nonterrestrial Utilization of Materials Team developed scenarios for a permanent, growing, highly automated space manufacturing capability based on the utilization of ever-increasing fractions of nonterrestrial materials. The primary focus was the initiation and evolutionary growth of a general-purpose SMF in low Earth orbit. The second major focus was the use of nonterrestrial materials to supply the SMF. A third major focus was on SMF automation technology requirements, particularly teleoperation, robotics, and automated manufacturing and materials processing techniques.

The team adopted a systems approach, beginning with a review of inputs to the SMF system (including sources of raw materials in the Solar System), processes for converting nonterrestrial materials into feedstock, and costs of transporting raw materials and feedstock to LEO. Initiation and growth of the SMF then were considered. A taxonomy of terrestrial manufacturing techniques was developed and analyzed to determine space-compatibility, automatability, and cost-, mass-, and energy-efficiency. From this selection process emerged several "starting kits" of first-generation equipment and techniques. One such "kit," for example, was based on powder metallurgy, extrusion/spray forming, laser machining, robotic forming (by cold welding), and process control via central computer or a distributed network. These tools and techniques would provide an initial space manufacturing presence for the production of second-generation machines and more sophisticated outputs.

As the SMF grows it evolves in several dimensions beyond mere expansion of manufacturing capability. First, the original factory is highly dependent on Earth for its raw material inputs. This dependency lessens as nonterrestrial sources of raw materials – especially the Moon and the asteroids – are developed. Second, the initial facility is run almost entirely by teleoperation (equipment operated by people located at sites remote from the SMF, such as on Earth), but later these teleoperators may largely be replaced by autonomous robots. Finally, the SMF system originally manufactures solar power stations, communications satellites, and a number of other products difficult or impossible to make anywhere but in space (e.g., certain biomedical substances, and foamy metals), but should eventually also begin to produce some outputs for use in other NASA missions in space or back on Earth. Examples include hulls and pressure vessels, integrated circuits and other electronics components for robots and computers, laser communication links, gigantic antennas, lunar teletourism equipment, and solar sails.

The establishment and growth of such a facility would have far-ranging and significant effects on human social and economic institutions. Stine (1975) has called space manufacturing the "third industrial revolution" to highlight the tremendous potential for transforming civilization, much as did the introduction of powered machinery in the 19th century and computers in the 20th century. It is impossible to predict the exact nature of the implications of an SMF for Earth since many would be second- and third-order perturbations. However, several areas of maximum impact were outlined by the team to aid in developmental planning and to minimize potential negative effects.

From an economic standpoint, the SMF scenario is expressly designed to reduce its demands on Earth resources – both material and monetary – as it is developed. Thus, initial costs are the major issue, and proposals have been made for reducing these. Other studies suggest that an SMF can provide a very reasonable return on investment. Certainly, the government will be highly involved in both the approval of the project and its implementation. The establishment of an SMF has definite legal implications, and close cooperation among several nations may be necessary in order to create a mutually satisfactory system. Finally, the public stands to benefit from the establishment of space solar power systems, the creation of new wonder drugs, superpure materials and other products unique to space, and the potential for unusual and fascinating vacations via teletourism.

Besides reducing environmental pollution hazards and increasing world interdependence, the advanced SMF in the long term will undoubtedly have major impacts on private enterprise, labor, industrial capacity, and social conditions in general. While expanded capacity and increased product variety seem likely to be a positive contribution, competition for markets and jobs must certainly be a concern. Careful planning plus a very gradual evolution will help to minimize disruption. A system for equitable involvement of private enterprise in space manufacturing must be devised. The gradual retraining of labor to carry out supervisory and high-adaptability roles for which humans are uniquely suited is already necessary because of advancing automation

on Earth. But it is important to note that this retraining, though initially potentially painful, casts human beings in the fundamentally most appropriate role: telling machines what to do for the benefit of all mankind.

4.6.1 Long-Term Implications for Humankind on Earth

The implications of a growing SMF are unquestionably complex and to some degree unforeseeable. The following discussion is limited to just a few major impact areas, conceptually isolated to convey the enormous potential consequences of the undertaking. A large-scale research effort in the area of societal consequences is required to provide an adequate assessment of the possible scope of the effects.

Environment. The direct environmental impact of the SMF will be significant and positive, mostly because of the relief it will provide from the twin pressures of resource exhaustion and industrial pollution. Many processes now conducted in Earth-based laboratories and factories which pose health hazards could be transferred to the SMF. Biological investigations of recombinant DNA and physics experiments with nuclear or other dangerous materials could be carried out on space platforms using teleoperators.

Indirectly, SMF could serve as construction bases for space solar power systems (SPS). Easily accessible sources of nonrenewable fuels are being consumed at an alarming rate, and the increased use of capital intensive nuclear energy is meeting stiff public opposition in this country and elsewhere. The sane alternative is to use the Sun as a source of "free" energy. Even using terrestrial resources, space solar power stations appear economically attractive (Grey *et al.*, 1977; Johnson and Holbrow, 1977). About 100 5-GW stations would suffice to supply a majority of current Earth-based electric power requirements.

Environmental benefits of placing the energy plant in space are manifold: There would be no danger from natural disasters such as earthquakes, no thermal or particulate pollution, and no risk of explosions or other failures which might conceivably cause harm to human populations (Grey *et al.*, 1977; Mayur, 1979). On the negative side must be weighed the possibility of leakage of microwave transmissions (Barr, 1979; Glaser, 1979; Johnson and Holbrow, 1977) and the security of installations which become the main U.S. energy source. Still, it is clear that SPS technology has the potential to relieve much of the current global energy shortage. Some global cooperation would be inevitable, suggesting major impacts in the sphere of world interdependence.

World interdependence. Somewhat paradoxically, establishment of an SMF may contribute to global interdependence. The fully productive SMF can be compared, in scope, to current multinational corporations, with one important difference: if the economic investments required are so great that governments rather than private sources must be partners in the venture, then nations will share in the wealth generated by the SMF instead of individual investors. Active cooperation would then be required to find some equitable means to ensure that under developed countries have an opportunity to share the fruits of state-of-the-art manufacturing technology (Mayur, 1979). Thence, all nations will come to regard industrialization in a more homogeneous manner, enabling less-developed countries to concentrate greater effort on improving the social and economic conditions within their own borders.

Private enterprise. Space industrialization has the potential for enormous impact on the economic system of the United States. Some of the potentially negative effects can be avoided through proper planning. SMF may be national, perhaps even multinational, enterprises, with the potential of transferring a great deal of the productive capacity of the U.S. into government hands in part because of the anticipated long lead time for economic return. Since SMF output would continually increase, it could eventually dominate the U.S. GNP, in which case Earth-based American manufacturing industries may no longer be competitive. While net national productivity would expand because of the input of additional nonterrestrial materials and energy, the contribution of private enterprise might diminish.

Some means must be found to transfer some of the investment potential back into private hands as the time-frame for economic return on the SMF grows shorter. This opportunity must, however, be equitably distributed, or a few large corporations could gain oligopolistic or monopolistic control over nonterrestrial resources. One suggestion for avoiding the problems associated with the economics of space industrialization is to encourage individuals to become investors (Albus, 1976). While this would avoid the problem of monopolistic control, some means should be devised to ensure that the scheme would not further widen the gap between rich and poor in this country.

Space industrialization will hardly leave the present private enterprise system unmodified. SMF planners must consider to what extent modification of the economic system is acceptable in future generations.

Automation and labor. The scenarios developed by the study team presuppose a high degree of SMF automation. In the long term this means that much of the productive capability of the world will be in the hands of robots, a trend already abundantly apparent in Earth-based manufacturing. Thus, the SMF merely accelerates an existing trend in industrial robotics deployment, with a multiplier effect throughout commercial manufacturing At issue, then, is to what extent the SMF threatens existing jobs while eliminating the possibility of alternative employment – not simply whether machines can replace humans in some roles.

Many Americans define self-worth through their work. A potentially grim scenario resulting from rapid automated space manufacturing development is that many people might be left suddenly "worthless," shut off from productive activity. The best antidote to such an unwholesome situation is early recognition of the problem. Alternative employment possibilities must be created, perhaps by returning to a strong craftsman or handicraft tradition. Some means must be found to permit participation in automated activities, perhaps through teleoperation or higher-order supervisory control (Chafer, 1979). Finally, more creative leisure time activities must be developed. The educational system must be re-oriented to support the notion that human beings need not derive their worth solely through work. Personal relationships, expanded hobbies, and private research are just a few of the many possible alternatives.

A more subtle result of increased automation is greater human dependency on "the machine." Many people may begin to sense a lack of autonomy in relation to their robot creations. This feeling will be exacerbated by the seeming remoteness of the the SMF, far from the immediate control of people on Earth. This may be a real psychological problem for the general public, so great care should be taken to ensure that the move toward complete automation is sufficiently gradual to allow people the opportunity to adjust to a new man-machine relationship.

Industrial capacity. An expanding SMF must eventually greatly augment the industrial output of the U.S. and the world as a whole. New materials and energy resources will become available at an ever-decreasing cost. Care must be taken to ensure that this capacity is used in a socially responsible manner. Extensive planning may be required to determine what products will have the most beneficial impact for the least cost in terrestrial resources. Several long-term "complex" products have been suggested in section 4.4.4.

It must be recognized that one important function of the SMF is to provide an industrial capability not currently available. The unique space environment makes possible the production of substances not easily duplicated under atmospheric and high-gravity conditions. These materials include serums and vaccines now produced only in very limited quantities, new composite substances, porous metallic structures, and high-purity metals and semiconductors (Grey *et al.*, 1977). Thus directed, the increased industrial capacity derived from the SMF would supplement rather than supplant existing terrestrial industry, and therefore alleviate potential problems of unemployment.

Society. The spirit of the American people has taken an introspective turn. Many are no longer convinced that unexplored horizons still exist. Predictions of global calamity are commonplace, and the philosophy of "small is beautiful" has become popular (Salmon, 1979). Given only limited terrestrial resources, such predictions and prescriptions might indeed be appropriate.

However, establishing an SMF opens new horizons with the recognition that planet Earth is just one potential source of matter and energy. Recognition of the availability of lunar and asteroidal materials and the abundant energy of the Sun can revitalize the traditional American belief in growth as a positive good and can generate a new spirit of adventure and optimism. It is unnecessary to speculate on the directions of growth in its various dimensions because it is clear that American society would continue its historic tradition of exploring new horizons and avoiding stagnation in an ever-changing Universe (Dyson, 1979).

On a more fundamental level, the proposed mission is species-survival oriented. Earth might at any time become suddenly uninhabitable through global war, disease, pollution, or other man-made or natural catastrophes. A recent study has shown that an asteroid collision with Earth could virtually turn off photosynthesis for up to 5 years which, together with massive kilometer-high tsunamis, would virtually extinguish all higher life on this planet (Alvarez *et al.*, 1980). The proposed mission assures the continued survival of the human species by providing an extraterrestrial refuge for mankind. An SMF would stand as constant proof that the fate of all humanity is not inextricably tied to the ultimate fate of Earth.

4.6.2 Near-Term Requirements for SMF Implementation

The foregoing analysis suggests that no single consequence of building an SMF is inevitable. Societal impacts may be channeled by proper planning, i.e., by taking a look at the proposed technological development within the entire global cultural framework. Recognition of the consequences of building an SMF is the first step in determining the requirements for making space industrialization a reality.

Some additional short-term planning requirements are reviewed below. The set of preconditions for the mission also may serve as a set of recommendations for action in NASA planning. A number of technological and societal factors are instrumental in determining whether the scenario proposed in this chapter can ever be actualized. The present discussion makes explicit many requirements tacitly assumed in previous sections, and provides both a general review of the broader significance of the mission and a set of recommendations for future NASA planning.

New technologies. SMF research and development should proceed concurrently with research in materials processing and the design of human space-transportation systems. While lunar and orbital starting kits could be deployed using current techniques, resupply of the SMF,

conveyance of raw materials from Moon to LEO, and delivery of SMF products all demand additional technological development to be economically feasible. Likewise, independence of terrestrial resupply (nonterrestrial materials closure) and economic feasibility go hand in hand.

Technical requirements for limited human habitats in space and on the Moon also must be considered. The proposed mission attempts to minimize the human presence in space through automation. Nevertheless, some supervisory functions must be performed by people, at least in the initial and midterm stages of space industrial development. In addition, as the machine systems evolve they will be able to create increasingly economical and secure nonterrestrial habitats for people.

Economics. Implementation of the proposed scenario, even given its strong emphasis on the utilization of space resources and automation, will require large-scale investment. The mission is designed, however, to build on existing and planned space programs. Some venture capital from private individuals and corporations may be expected as the project draws closer to the point of economic payback.

The space-manufacturing mission is designed to draw less and less on terrestrial economic resources as it develops. The primary initial investment will be for the emplacement of the orbital and lunar starting kit facilities, the concurrent technical development required to create the machines contained within these packages, and people involved in the maintenance of operations.

Actual economic calculations are beyond the scope of the present study. However, previous studies have shown that space facilities not based on the principles of growth and automation can provide an economic return on investment (Johnson and Holbrow, 1977; Science Applications, 1978; Rockwell, 1978), so it is expected that the economics of the present proposal should show an even greater potential for return on investment.

Government. Studies have shown that the Executive branch of the federal government must be a driving force behind the implementation of a large-scale space program (Overholt *et al.,* 1975). The Chief Executive must be convinced that the project will have real value for the nation's citizens, preferably during his own Administration. The present mission emphasizes quick, highly visible results; the SMF is a constantly growing accomplishment with clearly visible benefits for all.

Planners of the project should strive for interagency cooperation and financial support. DOD and DOE are obvious candidates, but other agencies such as NIH might also be interested if properly approached. The SMF could become a truly national facility, particularly if it is recognized as transcending the interests of any single government agency. The grand potential for space utilization may require some revision in NASA's charter, which presently is directed primarily towards exploration (Logsdon, 1979).

Space law. The legal difficulties associated with an SMF and a lunar mining facility have not yet been resolved or exhaustively examined. The latest draft Moon Treaty emphasizes that the use of the Moon should be "for the benefit of all mankind." Interpretations of this phrase vary (Jankowitsch, 1977), but an advisable approach would be to allow other nations to participate in the benefits achieved from the SMF. A second possibility is to ensure that the space and lunar activities proposed for space manufacturing will be explicitly declared legally permissible in the final version of the Treaty.

Global requirements. Since the SMF will eventually have impacts ranging far beyond the borders of the United States, active cooperation with other nations should be sought during project implementation. In this way the problems inherent in a narrow territorial perspective, as well as possible legal objections by other nations, can be avoided. Plans should be drawn for the distribution of some, if not most, SMF products on a global basis. Capital investment in the project by other nations should be encouraged. Less-developed countries should be given an opportunity to participate in any way they can, even if they are unable to invest money in the project (Glaser, 1979). These attempts at open-handedness will do much to alleviate international apprehensions concerning a large-scale project of this sort initiated and conducted solely under U.S. auspices.

Public sector. Americans must be convinced they will derive some immediate and visible advantages from the SMF if it is to become a politically viable concept. People tend to view past space efforts primarily as prestige-oriented events (Overholt *et al.,* 1975). An extensive public education program should be undertaken to demonstrate that automated space manufacturing can produce real economic benefits to the nation as a whole (Barmby, 1979). It should be emphasized that the project can help combat the problem of inflation now facing the country, and that space solar power systems will offset energy shortages that aggravate worldwide economic problems (Science Applications, Inc., 1978). Thus, the project can be shown to have the necessary links to problems of immediate and long-range concern to the ordinary citizen (Overholt *et al.,* 1975). Special applications in the areas of health and tourism should also be emphasized.

It is essential to reassure the nation that fully autonomous robots will be only gradually introduced, that most

existing jobs will not be replaced but rather enhanced, and that automata employed on the SMF will not be completely beyond human control. (Quite the contrary; for example, terrestrial construction crews could teleoperate bulldozers, cranes, or machine tool equipment in space or on the surface of the Moon.) People tend to fear machines which they feel they cannot control, even when this apprehension is unjustified (Taviss, 1972). Great care should be taken to discredit demagogues who may try to create false images and fears in the minds of the public.

Useful production. Planning for an SMF must include detailed consideration of the outputs expected to be produced. The most reasonable approach to production is to view it as an incremental process. Primary initial output of the SMF should be that which allows the facility to expand its own productive capability – an expanded set of machinery. These new devices may then construct hulls or pressurized vessels to provide a larger working environment. At this point some small-scale preparation of biological materials could be carried out (Grey *et al.,* 1977).

Large-scale expansion of the facility requires large-scale teleoperation and robotics. Second-generation products should consist of parts essential to the construction of robots such as integrated circuits, capacitors, resistors, printed circuit boards, and wire. Some of this may be shipped back to Earth as useful production, together with increasing quantities of rare biomedical substances and other materials unique to space.

An expanded SMF makes possible full-scale production of space products and permits utilization of the facility for other purposes. Space platforms, pure glasses and synthetic crystals, satellites, and robots are ideal outputs. In addition, the SMF could undertake the major construction of solar power stations and provide a variety of other commercial applications.

4.7 Final Remarks

The analysis presented in this chapter shows that utilization of the space environment is a viable possibility for future manufacturing strategies. A shift in emphasis from a significant human-supported presence of people in space to a substantially automated and expanding orbital facility which can expand and support both machine systems and people should offer significant economic advantages over previous space industrialization scenarios. The team has elucidated some of the major technological requirements for the actualization of the project. Additional technology feasibility and social impact studies are, of course, required. The main purpose of the present work was to provide a realistic framework for further research and development which may culminate in the construction of an automated Space Manufacturing Facility sometime in the next several decades.

4.8 References

Adams, C. Mel: Production, Assembly, and High Vacuum Fabrication, in J. Grey, ed., Space Manufacturing Facilities (Space Colonies), Proceedings of the Princeton/AIAA/NASA Conference, 7-9 May 1975, AIAA, New York, March 1977, pp. 99-103.

Agin, G. V.: Computer Vision Systems for Industrial Inspection and Assembly, Computer (IEEE), vol. 13, May 1980, pp. 11-20.

AISC (American Institute of Steel Construction) Manual, 1961.

Akin, D. L.: Optimization of Space Manufacturing. In Space Manufacturing III, G. Grey and C. Krop, eds., Proc. 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979, pp. 257-266.

Albus, James S.: People's Capitalism: The Economics of the Robot Revolution, New World Books, College Park, Maryland, 1976.

Allen, C. W.: Astrophysical Quantities. The Athlone Press, London, 1973. Third ed.

Alvarez, L. W.; Alvarez, W.; Asaro, F.; and Michel, H. V.: Extraterrestrial Cause for the Cretaceous-Tertiary Extinction. Science, vol. 208, 6 June 1980, pp. 1095-1108.

Amstead, B. H.; Ostwald, P. F.; and Begeman, M. L.: Manufacturing Processes. John Wiley and Sons, New York, 1979. Seventh ed.

Andreevskaya, G. D.; and Plisko, T. A.: Some Physical Properties of Continuous Basalt Fibers. Steklo i Keramika, no. 8, August 1963, pp. 15-18; CA 60, 1964, p. 1470.

Ankrum, P. D.: Semiconductor Electronics. Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

Annis, James F.; and Webb, P.: Development of a Space Activity Suit. NASA CR-1892, 1971.

Arnold, J. R.; and Duke, M. B.; eds.: Summer Workshop on Near-Earth Resources. NASA CP-2031, 1978.

Asada, H.; and Hanafusa, H.: An Adaptive Tracing Control of Robots and its Application to Automatic Welding. Joint Automatic Control Conf., 12-15 August 1980, San Francisco, Section FA7-D.

Ayres, R. U.; Ayres, L. W.; and Criswell, D. R.: Economic Considerations in Space Industrialization. In Space Manufacturing III, J. Grey and C. Krop, eds., Proc. of the 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979. Paper 79-1390. pp. 209-221.

Barkmann, W. E.: Laser-Interferometer Position Feedback for Precise Machine Tools. Joint Automatic Control Conference, 12-15 August 1980, San Francisco, vol. 1, sec. WA7-E.

Barmby, John G.: Improving the Justification for Space Industrialization. In Space Manufacturing III, J. Grey and C. Krop, eds. Proc. 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979, pp. 163-166.

Barr, Nathaniel F.: An Environmental Assessment of the Satellite Rescuer System Reference Design. In Space Manufacturing III, J. Grey and C. Krop, eds. Proc. 4th Princeton/AIAA Conf., 14-17 May 1979, AIAA, New York, 1979, pp. 441-446.

Bartels, Peter A.; and Bier, Milan: Design, Development, Test, and Evaluation of an Automated Analytical Electrophoresis Apparatus. NASA CR-150443, 1977.

Bejczy, A. K.: Sensors, Controls, and Man-Machine Interface for Advanced Teleoperation. Science, vol. 208, June 20, 1980, pp. 1327-1335.

Belforte, D.: Seeking Answers Overseas. Electro-Optical Systems Design, August 1979.

Bender, David F.; Dunbar, R. Scott; and Ross, David J.: Round-Trip Missions to Low-Delta-V Asteroids and Implications for Material Retrieval. In Space Resources and Space Settlements, J. Billingham, W. Gilbreath, B. Gossett, and B. O'Leary, eds. NASA SP-428, 1979, pp. 161-172.

Bier, M.: Electrophoresis. In An Introduction to Separation Sciences, Barry Karger, Lloyd Snyder, and Csaba Horvath, eds., John Wiley and Sons, New York, 1973, pp. 497-526.

Billingham, J.; Gilbreath, W.; Gossett, B.; and O'Leary, B., eds. Space Resources and Space Settlements, NASA SP-428, 1979.

Bloom, H.; and Gutmann, F.: Electrochemistry. Plenum Press, New York, 1977.

Bock, Edward: Lunar Resources Utilization for Space Construction. General Dynamics Convair Division, GDC-ASP79-001, Contract NAS9-15560, 1979.

Bolz, R. W.: Production Processes: The Productivity Handbook. Conquest Publications, Novelty, Ohio, 1974.

Borrego, J.: Space Grid Structures. MIT Press, 1968.

Bradbury, S., ed.: Source Book on Powder Metallurgy. Am. Soc. for Metals, Metals Park, Ohio, 1979.

Burns, Joseph A., ed.: Planetary Satellites. U. of Arizona Press, Tucson, AZ, 1977.

Campbell, J. S.: Principles of Manufacturing Materials and Processes. McGraw-Hill, New York, 1961.

Carter, G.; and Grant, W. A.: Ion Implantation of Semiconductors. John Wiley and Sons, New York, 1976.

Chafer, Charles M.: The Role of Public Interest Groups in Space Policy. In Space Manufacturing III, J. Grey and C. Krop, eds. Proc. 4th Princeton/AIAA Conf., 14-17 May 1979, AIAA, New York, 1979, pp. 183-192.

Coombs, Clyde F., ed.: Printed Circuits Handbook (Second ed.), McGraw-Hill, San Francisco, California, 1979.

Criswell, David R.: Demandite, Lunar Materials, and Space Industry. In Space Manufacturing Facilities II – Space Colonies, J. Grey, ed. Proceedings of the 3rd Princeton/AIAA Conference, 9-12 May 1977, AIAA, New York, 1977, pp. 131-137.

Criswell, David R.: Lunar Materials for Construction of Space Manufacturing Facilities. In Space Manufacturing Facilities II – Space Colonies, J. Grey, ed. Proceedings of the 3rd Princeton/AIAA Conference, 9-12 May 1977, AIAA, New York, 1977.

Criswell, David R.: Extraterrestrial Materials Processing and Construction. (Final Report, Lunar and Planetary Institute, Houston, Texas), NASA CR-158870, 1978 and 1980.

Cruz, M. I.; Armento, R. F.; and Giles, W. H.: Aerocapture – A Systems Design for Planetary Exploration. Paper No. 379-160, 30th International Astronautical Federation Congress, Munich, Fed. Rep. Germany, 16-22 September 1979, 16 pp.

Dalton, C.; and Hohmann, E.: Conceptual Design of a Lunar Colony, 1972 NASA/ASEE Systems Design Institute. (University of Houston, NASA MSC, Rice University, NASA Grant NGT-44-005-114), NASA CR-129164.

DeGarmo, E. P.: Materials and Processes in Manufacturing, Fifth ed., Macmillan, New York, 1979.

Degewij, J.: Spectroscopy of Faint Asteroids, Satellites, and Comets. J. Astron., vol. 85, 1980, pp. 1403-1412.

Donata, G.; and Camera, A.: A High Level Programming Language for a New Multi-Arm Assembly Robot. First International Conference on Assembly Automation. Brighton, England, D.E.A. SPA, CSO, Torino, 21-27 March 1980.

Drexler, K. E.: Lightsails – An Overview. In Discussion Meeting on Gossamer Spacecraft (Ultralightweight Spacecraft): Final Report, Roy G. Brereton, ed., JPL Publication 80-26, Jet Propulsion Laboratory, Pasadena, California, 15 May 1980, pp. 8-12.

Dubovkaya, T. S.; and Kosmina, N. E.: Chemical Stability of Staple Basaltic Fibers. Strukt. Sostav. Svoistva Formovanie Steklyanykj Volokon, 1968, pp. 1, 140; CA 73, 112426 (1970).

Duda, R. O.; Hart, P. E.; Konolige, K.; and Reboh, R.: A Computer-Based Consultant for Mineral Exploration. Prepared for the National Science Foundation by SRI International, Menlo Park, California, SRI Project 6415, 1979. 185 pp.

Dummer, G. W. A.: Materials for Conductive and Resistive Functions. Hayden Books, New York, 1970.

Dunbar, R. S.: The Search for Asteroids in the L4 and L5 Libration Points in the Earth-Sun System. In Space Manufacturing III, J. Grey and C. Krop, eds. Proceedings of the 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979, pp. 560-568.

Dunning, Jeremy D.; and Snyder, Robert S.: Electrophoretic Separation of Lunar Soils in a Space Manufacturing Facility. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, New Jersey.

Dyson, Freeman: Disturbing the Universe. Harper and Row Publishers, New York, 1979.

European Space Agency: Polar Orbiting Lunar Observatory (POLO) Assessment Study, ESA-SCI (71), National Technical Information Service, 1979. 46 pp.

Feurstenav, P.; and Raghaven, S.: In Rektor der Bergakedie Freiberg, A 593, VEB Deutsch er Verlag Fur Grundstoffindustrie, 1978.

Fields, S. A.; and Weathers, H. M.: Problems and Techniques of Lunar Surface Mining. NASA TM X-53560, 1967.

Freitas, Robert A., Jr.: A Self-Reproducing Interstellar Probe. J. British Interplanetary Soc., vol. 33, July 1980, pp. 251-264.

Freitas, Robert A., Jr.: LMF Chemistry Sector: A Preliminary Design. Summer Study Document, 27 July 1980a, 38 pp.

Freitas, Robert A., Jr.: Baseline Reference Design for a Growing, Self-Replicable Lunar Manufacturing Facility. Summer Study Document, 8 July 1980b.

Freitas, Robert A., Jr.: Automated Assembly System for a Growing, Replicating Lunar Manufacturing Facility. Summer Study Document, 8 August 1980c.

Freitas, Robert A., Jr.; and Valdes, Francisco: A Search for Natural or Artificial Objects Located at the Earth-Moon Libration Points. Icarus, vol. 42, June 1980d, pp. 442-447.

Freitas, Robert A., Jr.; and Zachary, William B.: A Self-Replicating, Growing Lunar Factory. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, New Jersey.

Gaffey, M. J.; Helin, E. F.; and O'Leary, B.: An Assessment of Near-Earth Asteroid Resources. In Space Resources and Space Settlements, J. Billingham, W. Gilbreath, B. Gossett, and B. O'Leary, eds., NASA SP-428, 1979, pp. 191-204.

Gaffey, M. J.; and McCord, T. B.: Mineralogical Characterizations of Asteroid Surface Materials from Reflectance Spectroscopy: A Review. AIAA Paper 77-527. In Space Manufacturing Facilities II, J. Grey, ed., Proceedings of the 3rd Princeton/AIAA Conference, 9-12 May 1977, AIAA, New York, 1977, pp. 143-154.

Gehrels, T., ed.: Asteroids. The Univ. of Arizona Press, Tucson, 1979.

Giuse, D.; and Guida, G.: User Defined Control Structures in Programming Languages for Robots. 10th International Conference on Industrial Robots, 5-7 March 1980, Milan, Italy, pp. 73-88.

Glaser, A. B.; and Subak-Sharpe, G. E.: Integrated Circuit Engineering. Addison-Wesley, Reading, Massachusetts, 1977.

Glaser, Peter E.: The Benefits of Solar Power Satellites. In Space Manufacturing III, J. Grey and C. Krop, eds., Proc. 4th Princeton/AIAA Conf., 14-17 May 1979, AIAA, New York, 1979, pp. 235-243.

Goeller, H. E.; and Weinberg, Alvin M.: The Age of Substitutability. Science, vol. 191, 20 February 1976, pp. 683-689.

Goldberg, Aaron: The Economics of Bootstrapping Space Industries – Development of an Analytic Computer Model. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, New Jersey.

Greeley, R.; and Carr, M. H.; eds.: A Geological Basis for the Exploration of the Planets. NASA SP-417, 1975.

Green, J.: The Polar Lunar Base – A Viable Alternative to L-5 in the Future of the U.S. Space Program. Proceedings of the 25th Anniversary Conference of the American Astronautical Society, 30 October-2 November 1978. Paper No. AAS78-191, pp. 385-425.

Grey, J., ed.: Space Manufacturing Facilities (Space Colonies). Proc. Princeton/AIAA/NASA Conference, 7-9 May 1975, AIAA, New York, March 1977.

Grey, J., ed.: Space Manufacturing Facilities II. Proc. 3rd Princeton/AIAA Conference, 9-12 May 1977, AIAA, New York, 1977.

Grey, J.; Downey, P.; and Davis, B.; eds.: Space: A Resource for Earth. AIAA, New York, 1977.

Grey, J.; and Krop, C.; eds.: Space Manufacturing III. Proc. 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979.

Grodzka, P. G.: Process Chemistry of L5 in Space Manufacturing Facilities. In Space Manufacturing Facilities (Space Colonies), J. Grey, ed. Proceedings of the Princeton/AIAA/NASA Conference, 7-9 May 1975, AIAA, New York, March 1977.

Grossman, M.: Technician's Guide to Solid-State Electronics. Parker Publishing Company, West Nyack, New York, 1976.

Heer, Ewald, ed.: Proceedings of the Pajaro Dunes Goal-Setting Workshop, unpublished draft notes, June 1980.

Henderson, R. A.; and Vickery, A.: Floating Zone Electrophoresis Facility for Spacelab, Phase A Study. Final Report ESA-CR(P)-807, (N76-27313), March 1976.

Holland, S.; Rossol, L.; and Ward, Mitchel R.: CONSIGHT 1. In Computer Vision and Sensor Based Robots, Richard O. Duda, L. Rossol, eds. Plenum Press, New York, 1979.

Inculet, I. I.; and Criswell, D. R.: Electrostatic Beneficiation of Ores on the Moon Surface. Inst. Physics Conference, Ser. No. XX, Chapter Z, Un. Oxford Press, England, 1979. LPI Contribution #352.

Jankowitsch, Peter: International and Legal Considerations. In Space Manufacturing Facilities II, Space Colonies, Jerry Grey, ed. Proc. of the 3rd Princeton/AIAA Conference, 9-12 May 1977, AIAA, New York, 1977, pp. 337-342.

Johnson, Richard D.; and Holbrow, Charles, eds.: Space Settlements: A Design Study. NASA SP-413, 1977.

Jones, W. D.: Fundamental Principles of Powder Metallurgy. Edward Arnold Ltd., London, 1960.

Jungerman, E., ed: Cationic Surfactants. Marcel Dekker, Inc., New York, 1970.

Kahn, Herman; Brown, William; and Martel, Lem: The Next 200 Years. William Morrow & Co., New York, 1976.

Kato, H.; Minezaki, M.; and Ozaki, Y.: Automatic Forging Press Line – 5000 tf. Metallurgia, vol. 45, no. 10, Oct. 1978.

Khambata, Adi J.: Introduction to Integrated Semiconductor Circuits. John Wiley and Sons, New York, 1963.

Kolm, Henry: An Electromagnetic 'Slingshot' for Space Propulsion. Technology Review, vol. 79, June 1977, pp. 61-66.

Kopecky, L.: Crystallization Temperatures of Rocks from Tranquility Base Compared with Experimental Crystallization of Dry Silicate Melts. In Geological Problems in Lunar and Planetary Research, J. Green, ed. AAS Science and Technology Series, vol. 25, 1971, pp. 685-698.

Kopecky, L.; and Voldan, J.: The Cast Basalt Industry. In Geological Problems in Lunar Research, Annals of the New York Academy of Science, Harold E. Whipple, ed., vol. 123, 1965, pp. 1086-1105.

Korolev, V. A.; Sergeev, S. M.; and Jurevich, E. I.: Application of Pneumatic Industrial Robots for Assembly. First International Conf. on Assembly Automation, 25-27 March 1980, Brighton, England.

Kuck, D. L.: Near-Earth Extraterrestrial Resources. In Space Manufacturing III, J. Grey and C. Krop, eds., Proceedings of the 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979, pp. 97-108.

Laszlo, Ervin *et al.*: Goals for Mankind. Dutton, New York, 1977.

Lee, Stuart M.: Lunar Building Materials – Some Considerations on the Use of Inorganic Polymers. In Space Resources and Space Settlements, J. Billingham, W. Gilbreath, B. Gossett, and B. O'Leary, eds. NASA SP-428, 1979, pp. 233-241.

Leonard, Raymond S.: Automated Construction of Photovoltaic Power Plants. Bechtel National, Inc., San Francisco, California, 1980.

Lindberg, R. A.: Processes and Materials of Manufacture (Second Ed.). Allyn and Bacon, Boston, 1977.

Logsdon, J. M.: Space Manufacturing and Public Policy Formulation. In Space Manufacturing III, J. Grey and C. Krop, eds. Proc. 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979, pp. 151-155.

Lowman, Paul D., Jr.: The Apollo Program: Was It Worth It? World Resources: The Forensic Quarterly, vol. 49, 1975, pp. 291-302.

Lunar and Planetary Institute: Proceedings of the Ninth Lunar and Planetary Science Conference, Houston, 1978, Pergamon Press.

Mackenzie, J. D.; and Claridge, R. C.: Glass and Ceramics from Lunar Materials. In Space Manufacturing III, J. Grey and C. Krop, eds., Proceedings of the 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979, Paper No. 79-1381, pp. 135-140.

Manasse, F. K.: Semiconductor Electronics Design. Prentice-Hall, Englewood Cliffs, New Jersey, 1977.

Martin Marietta Corporation: Space Shuttle External Tank System Definition Handbook: Vol. 1 – Configuration and Operational Data. Martin Marietta, Michoud Assembly Facility, New Orleans, Louisiana, 1974.

Mayur, Bashmi: Space Industrialization: An International Perspective. In Space Manufacturing III, J. Grey and Christine Krop, eds., Proc. 4th Princeton/AIAA Conf., 14-17 May 1979, AIAA, New York, 1979, pp. 63-70.

McCreight, L.: Industrial Development in Zero-G. In Space Manufacturing Facilities (Space Colonies), J. Grey, ed. Proc. of the Princeton/AIAA/NASA Conference, 7-9 May 1975, AIAA, New York, 1977, pp. 87-97.

Meadows, Donella H.; Meadows, Dennis L.; Randers, Jorgen; and Behrens, William W., III: The Limits to Growth. Universe Books, New York, 1972.

Miller, Rene H.; and Smith, David B. S.: Extraterrestrial Processing and Manufacturing of Large Space Systems. NASA CR-161293, 1979.

Minear, J. W.; Hubbard, N.; Johnson, T. V.; Clarke, V. C.; and Lunar Polar Orbiter Sciences Working Team: Mission Summary for Lunar Polar Orbiter, Doc. #660-41, Jet Propulsion Laboratory, Pasadena, California, September 1976, 38 pp.

Minsky, M.: Toward a Remotely Manned Energy and Production Economy. MIT Artificial Intelligence Laboratory, A.I. Memo No. 544, Boston, MA, September 1979, 19 pp.

Minsky, M.: Telepresence. Omni, vol. 2, June 1980, pp. 44-52.

Mock, J. A.: Metals That Form Like Plastic. Materials Engineering, vol. 92, August 1980, pp. 36-39.

Mongeau, P.; Fitch, O.; Kolm, H.; Williams, F.; Greneau, P.; and McKinney, K.: The Case for Electromagnetic Earth Launching. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, New Jersey.

Moore, H. D.; and Kibbey, D. R.: Manufacturing: Materials and Processes. Irwin, Homewood, Illinois, 1965.

Moore, R. G.: Bubble Structures for Gossamer Spacecraft. In Discussion Meeting on Gossamer Spacecraft (Ultralightweight Spacecraft): Final Report, Roy G. Brereton, ed. NASA CR-163275, 1980.

Morrison, David: Asteroid Sizes and Albedos. Icarus, vol. 31, no. 2, 1977, pp. 185-220.

Morrison, D.; and Wells, W. C.; eds.: Asteroids: An Exploration Assessment, NASA CP-2053, 1978.

Namba, S., ed.: Ion Implantation in Semiconductors. Plenum Press, New York, 1975.

New Development in High Power Lasers, Product Engineering; December 1970.

Office of Aeronautics and Space Technology (OAST), NASA Space Systems Technology Model, 1-3 May 1980. NASA TM-81041 through NASA TM-81043. Washington, D.C., National Aeronautics and Space Administration, 1980.

Oldham, William G.: The Fabrication of Microelectronic Circuits. Scientific American, vol. 237, September 1977, pp. 110-128.

O'Leary, B.; Gaffey, M. J.; Ross, D. J.; and Salkeld, R.: Retrieval of Asteroidal Materials. In Space Resources and Space Settlements, J. Billingham, W. Gilbreath, B. Gossett, and B. O'Leary, eds., NASA SP-428, 1979, pp. 173-189.

O'Neill, Gerard K.: The Colonization of Space. Physics Today, vol. 27, September 1974, pp. 32-40.

O'Neill, G. K.: Maximum Strength, Minimum Mass Structures. In Space-Based Manufacturing from Extraterrestrial Materials, Gerard K. O'Neill, ed. AIAA Progress in Astronautics and Aeronautics Series, vol. 57, AIAA, New York, 1977, pp. 161-171.

Ostwald, Phillip F.: Cost Estimating for Engineering and Management, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

Overholt, W.; Wiener, A. J.; and Yokelson, D.: Implications of Public Opinion for Space Program Planning, 1980-2000. (Hudson Institute Inc., Croton-on-Hudson, N.Y., N76-17984), NASA CR-144730, 1975.

Paul, R.; *et al.*: Advanced Industrial Robot Control Systems, School of Electrical Engineering, Purdue U., Lafayette, Ind., NSF Contract no. APR-77-14533, 1980.

Phinney, W. C.; Criswell, D. R.; Drexler, E.; and Garmirian, J.: Lunar Resources and their Utilization. In Space-Based Manufacturing from Extraterrestrial Materials, Gerard K. O'Neill, ed. AIAA Progress in Astronautics and Aeronautics Series, vol. 57, AIAA, New York, 1977, pp. 97-123.

Pomeroy, John H.; and Hubbard, Norman V.: The Soviet-American Conference on Cosmochemistry of the Moon and Planets, NASA SP-370, (Part 2). 1977.

Rao, D. Bhogeswara; Choudary, U. V.; Erstfeld, T. E.; Williams, R. J.; and Chang, Y. A.: Extraction Processes for the Production of Aluminum, Titanium, Iron, Magnesium, and Oxygen from Nonterrestrial Sources. In Space Resources and Space Settlements, J. Billingham, W. Gilbreath, B. Gossett, and B. O'Leary, eds., NASA SP-428, 1979, pp. 257-274.

Robinson, C. A., Jr.; and Klass, P. J.: Technical Survey: Particle Beams, Laser Weapons. Aviation Week and Space Technology, vol. 113, 28 July 1980, pp. 32-66 (Part 1) and vol. 113, 4 August 1980, pp. 44-68 (Part 2).

Rockwell International: Space Industrialization. Vol. 1: Executive Summary. NASA CR-150720, 1978.

Rosen, Charles A.; Nitzan, D.; Agin, G.; Andeen, G.; Berger, J.; Blean, R.; Gleason, G.; Hill, J.; Park, W.; and Paul, R.: Machine Intelligence Research Applied to Industrial Automation. Stanford Research Institute, Grant APR 75-13074, November 1976.

Rothschild, K. J.; Jeziorski, P.; Cunningham, A.; and Silverman, G.: Spotlight on Resistors. Electronic News, Supplement, 21 July 1980.

Sacerdoti, Earl D.: Problem Solving Tactics, SRI Project 7910, Technical Note 189, SRI International, Menlo Park, California, July 1979.

Salmon, J. D.: Space Manufacturing and Technocratic Centralism. In Space Manufacturing III, J. Grey and C. Krop, eds. Proc. 4th Princeton/AIAA Conference, 14-17 May 1979, AIAA, New York, 1979, pp. 167-172.

Saville, D. A.; and Ostrach, S.: Fluid Mechanics of Continuous Flow Electrophoresis. NASA CR-150689, 1978.

Scarlett, J. A.: Printed Circuit Boards for Microelectronics. Van Nostrand Reinhold Company, London, England, New York, 1970.

Schey, J. A.: Introduction to Manufacturing Processes. McGraw-Hill, New York, 1977.

Schoen, E.; Dorl, G.; and Lemke, H. J.: Electrophoresis Experiment MA-014 in the Apollo-Soyuz Test Project. Part 3: Future Applications. Final Report, ESA-TT-473 (N78-22131), September 1977. In German.

Schraft, R. D. *et al.*: Possibilities and Limits for the Application of Industrial Robots in New Fields. 10th Int. Symp. on Industrial Robots, 5-7 March 1980, Milan, Italy.

Schwartz, M. M.: Metals Joining Manual. McGraw-Hill Book Co., 1979.

Science Applications, Inc.: Space Industrialization, Vols. 1-4 (NASA Contract NAS-8-32197, SAI-79-(602-605)-HU, Science Applications, Inc., Huntsville, Alabama, 15 April 1978). NASA CR-150752 through NASA CR-150755.

Sheffield, Charles: How to Build a Beanstalk. Destinies, vol. 1, Aug./Sept. 1979, pp. 41-68.

Stine, G. H.: The Third Industrial Revolution. G. P. Putnam and Sons, New York, 1975.

Subramanian, R. V.; and Austin, H. F.: Basalt Fibers. In Handbook of Fillers and Reinforcements for Plastics, Harry S. Katz and John V. Milewski, eds. Van Nostrand Reinhold Co., New York, 1978, pp. 504-510. (Chapter 26A).

Subramanian, R. V.; Austin, H. F.; Raff, R. A. V.; Sheldon, G.; Dailey, R. T.; and Wullenwaber, D.: Use of Basalt Rock for the Production of Mineral Fiber. Pacific Northwest Commission Annual Report, Contract No. NR-3001, College of Engineering, Washington State Univ., Pullman, Washington, June 1975.

Subramanian, R. V.; Austin, H. F.; and Wang, T. J. Y.: Basalt Fibers. Pacific Northwest Regional Commission Annual Report, Contract No. NR-3017, College of Engineering, Washington State Univ., Pullman, Washington, September 1976.

Subramanian, R. V.; and Kuang-Huah, Shu: Interfacial Bonding in Basalt Fiber-Polymer Composites. 34th Annual Technical Conference, Reinforced Plastics/Composites Institute, Section 17C, 1979, pp. 1-10.

Subramanian, R. V.; Wang, T. J. Y.; and Austin, H. F.: Reinforcement of Polymers by Basalt Fibers. Sampe Quarterly, vol. 8, no. 4, July 1977, pp. 1-10.

Subramanian, R. V.; and Austin, H. F.: Method for Forming Basalt Fibers with Improved Tensile Strength. U.S. Patent 4,149,866, 1979.

Tebo, A. R.: Laser Annealing – and Beyond. Electro-Optical Systems Design, December 1979.

Taviss, Irene: A Survey of Popular Attitudes Toward Technology. Technology and Culture, vol. 13, October 1972, pp. 606-621.

Vanderbrug, G. J.; and Nagel, R. N.: Image Pattern Recognition in Industrial Inspection, NBSIR-79-1764, PB 80-108871 (N80-21771), U.S. Bureau of Standards, avail. September 1980.

Vanderhoff, John W.; and Micale, F. J.: Electrophoresis Experiment for Space. NASA CR-149925, 1976.

Vought Corporation: A Review of Vought's Previous Work on the Teleoperator Maneuvering System. Orientation Meeting, A Study Contract for Marshall Space Flight Center, 29-30 May 1980.

Waldron, R. D.; Erstfeld, T. E.; and Criswell, D. R.: The Role of Chemical Engineering in Space Manufacturing. Chemical Engineering, vol. 86, no. 4, 12 February 1979, pp. 80-94.

Wang, T. G.; and Kendall, J. M.: Fluid Dynamic Studies Relating to the Production of Large Fusion Targets. Inertial Confinement Fusion Conference, Conference on Laser Electronics and Optics (CLEO), San Diego, California, February 1980.

Way, F.: The Pulsed Nd:YAG Holographic Laser: Present Status and Applications. Proc. Technical Program, Electro-Optical Systems Design Conference and International Laser Exposition, Anaheim, California, 11-13 November 1975.

Weiss, R. A.: Lanham, J. W.; Richman, D. W.; and Walker, C. D.: Investigation of the Free Flow Electrophoretic Process. NASA CR-161241, 1979.

Westwood, A. R. C.: Third Tewksbury Lecture on Fracture Control and Application of Environment Sensitive Fracture Processes. J. Materials Sci., vol. 9, Nov. 1974, pp. 1871-1895.

Wetherill, George W.: Apollo Objects. Scientific American, vol. 240, March 1979, pp. 54-65.

Williams, Richard J.; and Jadwick, James J., eds.: Handbook of Lunar Materials, NASA RP-1057, 1980.

Wilson, R. G.; and Brewer, G. R.: Ion Beams. John Wiley and Sons, New York, 1973.

Yachida, M.; and Tsuji, S.: Industrial Computer Vision in Japan. Computer (IEEE), vol. 13, no. 5, May 1980, pp. 50-63.

Yankee, H. W.: Manufacturing Processes. Prentice-Hall, Englewood Cliffs, New Jersey, 1979. 765 pp.

Yushchenko, A. S.: A Combined System of Semiautomatic Control of a Manipulator. Engineering Cybernetics, vol. 17, no. 1, January-February 1979, pp. 157-160.

Zellner, B.; and Bowell, E.: Asteroid Compositional Types and Their Distributions. The Interrelated Origin of Comets, Asteroids, and Meteorites. A. H. Delsemme, ed., University of Toledo Publ., 1977, pp. 185-197.

APPENDIX 4A

LUNAR SUPPLY OF A LOW EARTH ORBIT STATION: DERIVATION OF FORMULAS

The mass brought to LEO from the Moon is $M_{PL} + M_{LAN}$ where M_{PL} is the mass of the payload of lunar soil and M_{LAN} is the mass of the LANDER system that carries it. The LANDER must have sufficient tankage to carry payload plus the propellant to lift off from the Moon (M_{PR_4}), or to carry the hydrogen required on the Moon plus the propellant to carry the system to the Moon from the OTV ($M_{PR_{2+3}} = M_{PR_2} + M_{PR_3}$, the propellant requirements for burns two and three), whichever is greater. The fact that $\Delta V_4 \sim \Delta V_2 + \Delta V_3$ and that $M_{PL} >> M_H$, where M_H is the mass of hydrogen carried to the Moon, makes it clear that the former tankage requirement is the more stringent. It has therefore been assumed that:

$$M_{LAN} = M_{LS} + aM_{PL} + BM_{PR_4}$$

where M_{LS} is the mass of the LANDER structure and a and B are the tankage fractions for the payload and propellant, respectively. For all burns and for both the OTV and the LANDER B is assumed to be the same.

On the lunar surface prior to takeoff, the mass of the LANDER system is:

$$M_{LAN} + M_{PL} + M_{PR_4} = (M_{PL} + M_{LAN})e^{\Delta V_4/c}$$

$$= (M_{PL} + M_{LS} + aM_{PL} + BM_{PR_4})e^{\Delta V_4/c}$$

Therefore,

$$M_{PR_4} = \frac{K_4 [(1 + a)M_{PL} + M_{LS}]}{1 - BK_4}$$

where c is exhaust velocity. Therefore,

$$K_n = e^{\Delta V_n/c} - 1$$

Since the OTV and LANDER are fueled at LEO, the only hydrogen carried to the Moon is that required in M_{PR_4}. If M_H is defined as the mass of hydrogen carried to the lunar surface, then

$$M_H = B_H M_{PR_4} = B_H K_4 \frac{(1 + a)M_{PL} + M_{LS}}{1 - BK_4}$$

where B_H is the hydrogen fraction in the propellant.

The mass landed on the Moon must be:

$$M_{LAN} + M_H = M_{LS} + aM_{PL} + BM_{PR_4} + B_H M_{PR_4}$$

$$= \frac{(B + B_H)K_4 [(1 + a)M_{PL} + M_{LS}]}{1 - BK_4}$$

$$+ M_{LS} + aM_{PL}$$

The payload for the OTV is therefore

$$(M_{LAN} + M_H)e^{\Delta V_{2+3}/c}$$

where:

$$\Delta V_{2+3} = \Delta V_2 + \Delta V_3$$

$$M_{PR_{2+3}} = (M_{LAN} + M_H)(e^{\Delta V_{2+3}/c} - 1)$$

$$= K_{2+3}(M_{LAN} + M_H)$$

and

$$M_{OTV} = M_{OS} + BM_{PR_1}$$

for M_{OS} defined as OTV structure mass.

The mass leaving LEO is therefore:

$$M_{OTV} + M_{PR_1} + (M_{LAN} + M_H)e^{\Delta V_{2+3}/c}$$

where:

$$M_{PR_1} = [M_{OTV} + (M_{LAN} + M_H)e^{\Delta V_{2+3}/c}](e^{\Delta V_1/c} - 1)$$

$$= K_1 [M_{OS} + BM_{PR_1} + (M_{LAN} + M_H)e^{\Delta V_{2+3}/c}]$$

$$= \frac{K_1 [M_{OS} + (M_{LAN} + M_H)e^{\Delta V_{2+3}/c}]}{1 - BK_1} \qquad (1)$$

The amount of material lifted off the Earth is:

$$M_{H_{\text{lift}}} = M_H + B_H(M_{PR_1} + M_{PR_{2+3}})$$

$$= B_H(M_{PR_1} + M_{PR_{2+3}} + M_{PR_4})$$

$$= B_H\left\{\frac{K_1}{1 - BK_1}\,[M_{OS} + (M_{LAN} + M_H)(K_{2+3} + 1)] + K_{2+3}(M_{LAN} + M_H) + M_{PR_4}\right\}$$

$$= B_H\left\{\frac{K_1}{1 - BK_1}\,M_{OS} + \left[\frac{K_1(K_{2+3} + 1)}{1 - BK_1} + K_{2+3}\right] \times (M_{LAN} + M_H) + M_{PR_4}\right\}$$

$$= B_H\left\{\frac{K_1}{1 - BK_1}M_{OS} + \left[\frac{K_1(K_{2+3} + 1)}{1 - BK_1} + K_{2+3}\right] \times\left[\frac{(B + B_H)K_4\,[(1 + a)M_{PL} + M_{LS}]}{1 - BK_4} + M_{LS} + aM_{PL}\right] + \frac{K_4}{1 - BK_4}\,[(1 + a)M_{PL} + M_{LS}]\right\}$$

If we define A, b, and C as follows:

$$A \equiv B_H\left\{\left[\frac{K_1(K_{2+3} + 1)}{1 - BK_1} + K_{2+3}\right]\left[\frac{(B + B_H)K_4(1 + a)}{1 - BK_4} + a\right] + \frac{K_4(1 + a)}{1 - BK_4}\right\}$$

$$b \equiv \frac{B_H K_1}{1 - BK_1}$$

$$C \equiv B_H\left\{\left[\frac{K_1(K_{2+3} + 1)}{1 - BK_1} + K_{2+3}\right]\left[\frac{(B + B_H)K_4}{1 - BK_4} + 1\right] + \frac{K_4}{1 - BK_4}\right\}$$

$$M_{H_{\text{lift}}} = AM_{PL} + bM_{OS} + CM_{LS}$$

From M_{PL} is taken material sufficient to replace the nonhydrogen part of the fuel supply. The amount of payload left over is P, hence:

$$M_{PL} = P + (1 - B_H)(M_{PR_1} + M_{PR_{2+3}})$$

$$= P + (1 - B_H)\left[\frac{M_{H_{\text{lift}}}}{B_H} - M_{PR_4}\right]$$

$$= P + \frac{1 - B_H}{B_H}\left\{M_{H_{\text{lift}}} - \frac{B_H K_4}{1 - BK_4}\,[(1 + a)M_{PL} + M_{LS}]\right\}$$

$$M_{PL} = \frac{\left(\begin{array}{l} P + [(1 - B_H)/B_H]\{M_{H_{\text{lift}}} \\ - [B_H K_4/(1 - BK_4)]M_{LS}\} \end{array}\right)}{1 + [(1 - B_H)/(1 - BK_4)]K_4(1 + a)}$$

and

$$M_{H_{\text{lift}}} = A\,\frac{\left(\begin{array}{l} P + [(1 - B_H)/B_H]\,[M_{H_{\text{lift}}} \\ -B_H K_4 M_{LS}/(1 - BK_4)) \end{array}\right)}{1 + [(1 - B_H)/(1 - BK_4)]K_4(1 + a)} + bM_{OS} + CM_{LS}$$

$$\frac{dM_{H_{\text{lift}}}}{dP} = \frac{A\quad 1 + [(1 - B_H)/B_H](dM_{H_{\text{lift}}}/dP)}{1 + [(1 - B_H)/(1 - BK_4)K_4(1 + a)}$$

$$= \frac{A/(1 + \;[(1 - B_H)/(1 - BK_4)]K_4(1 + a)}{\left(\begin{array}{c} 1 - [(1 - B_H)/B_H]\;\; A/\{1 \\ + [(1 - B_H)/(1 - BK_4)]K_4(1 + a)\} \end{array}\right)}$$

$$= \frac{A}{\left(\begin{array}{c} 1 + [(1 - B_H)/(1 - BK_4)]K_4(1 + a) \\ - [(1 - B_H)/B_H]A \end{array}\right)}$$

But if

$$X = \frac{A}{B_H} - \frac{K_4(1+a)}{1-BK_4} = \left[\frac{K_1(K_{2+3}+1)}{1-BK_1}\right.$$

$$\left. + K_{2+3}\right]\left[\frac{K_4(B+B_H)}{1-BK_4}(1+a)+a\right] \tag{2}$$

then

$$\frac{dM_{H_{\text{lift}}}}{dP} = \frac{B_H\{X + [K_4(1+a)/(1-BK_4)]\}}{1-(1-B_H)X} \tag{3}$$

This is the mass of hydrogen that must be uplifted from Earth to gain 1 kg of extra lunar payload to LEO. If no OTV is to be used, return to equation (1); M_{OS} is now zero. If it is assumed that the payload tankage is more than enough to hold M_{PR_1}, then the term BM_{PR_1} also disappears. Following through with these changes, X becomes:

$$X' = [K_1(K_{2+3}+1) + K_{2+3}]\{[K_4(B$$

$$+ B_H)/(1-BK_4)](1+a)+a\} \tag{4}$$

and

$$\frac{dM_{H_{\text{lift}}}}{dP} = \frac{B_H\{X' + [K_4(1+a)]/(1-BK_4)\}}{1-(1-B_H)X'} \tag{5}$$

The text shows that this reduces the marginal propellant cost by a small amount. If extra tankage is required to hold M_{PR_1} the advantage is probably wiped out.

4A.1 Numerical Equations

For simplicity, assume that the OTV starts in a circular 200 km orbit in the Earth-Moon plane and just reaches the Moon. Various relevant parameters used in the calculations are listed below.

$d_{\text{Moon}} = 384410$ km
$r_{\text{Earth}} = 6378$ km
$r_{\text{Moon}} = 1738$ km
$\mu_{\text{Earth}} = 398600.3\ \text{km}^3/\text{sec}^2$
$\mu_{\text{Moon}} = 4903\ \text{km}^3/\text{sec}^2$
LEO at 200 km altitude in place of lunar orbit
Perilune of transfer orbit at 50 km altitude

The circular orbital velocity at 200 km altitude is:

$$V_{\text{cir}} = 7.7843\ \text{km/sec}$$

The transfer orbit has

$$a_o = [(r_e + 200) + d_{\text{Moon}}]/2 = 195{,}494\ \text{km}$$

Therefore the spacecraft velocity upon leaving LEO is:

$$V_{\text{launch}} = \sqrt{\frac{2\mu_e}{r_e + 200} - \frac{\mu_e}{a_o}} = 11.0087\ \text{km/sec}$$

so $\Delta V_1 = V_{\text{launch}} - V_{\text{cir}} = 3.2244$ km/sec. This orbit has its apogee at the Moon's orbit and apogee velocity of $V_{\text{apogee}} = 0.18679$ km/sec. If the Moon has a circular orbit, its orbital velocity is $V_{\text{Moon}} = 1.02453$ km/sec, hence, spacecraft velocity relative to the Moon is $V_{\text{Moon}} - V_{\text{apogee}} = V_{\text{infinity}} = 0.8377$ km/sec.

While passing 50 km above the lunar surface the OTV releases LANDER, which at once performs a burn to place it into a 1738 X 1788 km orbit around the Moon. The OTV's velocity relative to the Moon prior to separation is:

$$V = \sqrt{V_{\text{infinity}}^2 + \frac{2\mu_{\text{Moon}}}{r_m + 50}} = 2.4872\ \text{km/sec}$$

The semimajor axis of the orbit about the Moon is 1763 km and so the velocity of the LANDER at apolune is

$$V_{\text{apolune}} = \sqrt{\frac{2\mu_{\text{Moon}}}{r_{\text{Moon}} + 50} - \frac{\mu_{\text{Moon}}}{a_o}}$$

The magnitude of the required orbital injection burn is therefore $\Delta V_2 = V - V_{\text{apolune}} = 0.84303$ km/sec. The LANDER then performs a half-orbit of the Moon and lands:

$$\Delta V_3 = V_{\text{perilune}} = \sqrt{\frac{2\mu_{\text{Moon}}}{r_{\text{Moon}}} - \frac{\mu_{\text{Moon}}}{a_o}} = 1.6915\ \text{km/sec}$$

The lunar processor refuels the LANDER and loads its payload tanks with lunar soil. Takeoff from the Moon on a trajectory that returns to LEO by way of aerobraking requires

$$\Delta V_4 = \sqrt{V_{\text{infinity}}^2 + \frac{2\mu_{\text{Moon}}}{r_{\text{Moon}}}} = 2.5187\ \text{km/sec}$$

4A.2 Propellants

There are two most promising propellant options for lunar-LEO transport systems. The first is an oxy-hydrogen combination using lunar-derived oxygen and hydrogen imported from Earth. The second option again requires native lunar oxygen as the oxidant but combines terrestrial-imported hydrogen with silicon purified on the Moon to produce a more powerful silane rocket fuel.

(a) Lunar oxygen, terrestrial hydrogen propellant option

The relevant chemical propellant combustion reaction is:

$$2H_2 + O_2 \Rightarrow 2H_2O$$

The molecular weight of H_2 is 2 and of O_2 is 32, so:

$$B_H = \frac{M_{H_2}}{M_{H_2} + M_{O_2}} = \frac{1}{9}$$

The achievable specific impulse of LOX - LH_2 is about 450 sec, using heat of formation data from Weast (1978) and assuming 75% thermal efficiency. This yields an exhaust velocity of 4.41 km/sec.

(b) Lunar oxygen, Earth/lunar silane propellant option

The silane produced on the Moon is assumed here for simplicity to be entirely SiH_4. The propellant chemical reaction is:

$$SiH_4 + 2O_2 \Rightarrow SiO_2 + 2H_2O$$

The molecular weight of SiH_4 is 32, so B_H = 1/24. The achievable vacuum specific impulse is within the range 328-378 sec (Lunar and Planetary Institute, 1980). Assuming the middle of the range, I_{sp} = 353 and C = 3.46 km/sec.

4A.3 References

Lunar and Planetary Institute: Extraterrestrial Materials Processing and Construction, Final Report NSR 09-051-001 Mod #24, Houston, Texas, 1980.

Weast, Robert C., ed.: Handbook of Chemistry and Physics, CRC Press, West Palm Beach, Florida, 1978. Fifty-ninth Edition.

APPENDIX 4B

REVIEW OF CASTING PROCESSES

Casting is a process by which a fluid melt is introduced into a mold, allowed to cool in the shape of the form, and then ejected to make a fabricated part or casing (Lindberg, 1977; Yankee, 1979). Four main elements are required in the process of casting: pattern, mold, cores, and the part. Pattern, the original template from which the mold is prepared, creates a corresponding cavity in the casting material. Cores are used to produce tunnels or holes in the finished mold, and the part is the final output of the process.

Substitution is always a factor in deciding whether other techniques should be used instead of casting. Alternatives include parts that can be stamped out on a punch press or deep-drawn, items that can be manufactured by extrusion or by cold-bending, and parts that can be made from highly active metals.

The casting process is subdivided into two distinct subgroups: (1) expendable and (2) nonexpendable mold casting.

4B.1 Expendable Mold Casting

Expendable mold casting is a generic classification that includes sand, plastic, shell, and investment (lost-wax technique) moldings. All of these involve the use of temporary and nonreusable molds, and need gravity to help force molten fluid into casting cavities – either by artificial gravity or pressure-feeding of molds in a zero-g SMF. Lack of atmosphere should be beneficial to some processes since molten fluids need not displace air.

(a) Sand Casting

Sand casting requires a lead time of days for production at high output rates (1-20 pieces/hr-mold), and is unsurpassed for large-part production. Green (wet) sand has almost no part weight limit, whereas dry sand (more likely with extraterrestrial materials) has a practical part mass limit of 2300-2700 kg. Minimum part weight ranges from 0.075-0.1 kg. Sand in most operations can be recycled many times and requires little additional input. The only serious restriction is the necessity for gravity-feeding the molten liquid. A general manufacturing facility using sand casting might require centrifugal force feeding instead.

Preparation of the sand mold is fast and requires a pattern which can "stamp" out the casting template in a few days. Typically, sand casting is used for processing low-temperature steel and aluminum, magnesium, and nickel alloys. It is by far the oldest and best understood of all techniques. Consequently, automation may easily be adapted to the production process, somewhat less easily to the design and preparation of forms. These forms must satisfy exacting standards as they are the heart of the sand casting process – creating the most obvious necessity for human control.

(b) Plaster Casting

Plaster casting is similar to sand molding except that plaster is substituted for sand. Plaster compound is actually comprised of 70-80% gypsum and 20-30% strengthener and water. Generally, the form takes less than a week to prepare, after which a production rate of 1-10 units/hr-mold is achieved with a capability to pour items as massive as 45 kg and as small as 30 g with very high surface resolution and fine tolerances.

The plaster process requires carbon, a relatively rare substance in nonterrestrial materials, for the gypsum binder. Once used and cracked away, normal plaster cannot easily be recast. The water used in mold production may be recycled during the baking process. Plaster casting is normally used for nonferrous metals such as aluminum-, zinc-, or copper-based alloys. It cannot be used to cast ferrous material because sulfur in gypsum slowly reacts with iron. Also, the plaster process requires gravity or centrifugal injection of casting fluid into the mold. (Prior to mold preparation the pattern is sprayed with a thin film of parting compound to prevent the mold from sticking to the pattern. The unit is shaken so plaster fills the small cavities around the pattern. The form is removed after the plaster sets.)

Plaster casting represents a step up in sophistication and required skill. The automatic functions easily are handed over to robots, yet the higher-precision pattern designs required demand even higher levels of direct human assistance. Another research issue with particular relevance to an extraterrestrial facility is plaster recyclability, so that

each mold (or the materials used to make it) need not be thrown away after just a single use.

(c) Shell Molding

Shell molding is also similar to sand molding except that a mixture of sand and 3-6% resin holds the grains together. Set-up and production of shell mold patterns takes weeks, after which an output of 5-50 pieces/hr-mold is attainable. Aluminum and magnesium products average about 13.5 kg as a normal limit, but it is possible to cast items in the 45-90 kg range. Shell mold walling varies from 3-10 mm thick, depending on the forming time of the resin.

There are a dozen different stages in shell mold processing that include: (1) initially preparing a metal-matched plate; (2) mixing resin and sand; (3) heating pattern, usually to between 505-550 K, (4) investing the pattern (the sand is at one end of a box and the pattern at the other, and the box is inverted for a time determined by the desired thickness of the mill); (5) curing shell and baking it; (6) removing investment; (7) inserting cores; (8) repeating for other half; (9) assembling mold; (10) pouring mold; (11) removing casting; and (12) cleaning and trimming. The sand-resin mix can be recycled by burning off the resin at high temperatures, so the only SMF input using this technique is a small amount of replacement sand and imported resin.

(d) Investment Casting

Investment casting (lost-wax process) yields a finely detailed and accurate product. After a variable lead time, usually weeks, 1-1000 pieces/hr-mold can be produced in the mass range 2.3-2.7 kg. Items up to 45 kg and as light as 30 g are possible for unit production.

To make a casting, a temporary pattern is formed by coating a master mold with plastic or mercury. The pattern is dipped in refractory material (typically a ceramic mixture of Zircon flour and colloidal silicate) leaving a heavier coating 3-16 mm thick. The process requires a constant input of Zircon flour because the mold is expendable, although mercury is recycled by processing in a pressurized positive-gravity environment. The mold is baked and mercury or plastic collected and recycled. The mold is filled, then broken away after hardening.

Investment casting yields exceedingly fine quality products made of all types of metals. It has special applications in fabricating very high temperature metals, especially those which cannot be cast in metal or plaster molds and those which are difficult to machine or work.

4B.2 Nonexpendable Mold Casting

Nonexpendable mold casting differs from expendable processes in that the mold need not be reformed after each production cycle. This technique includes at least four different methods: permanent, die, centrifugal, and continuous casting. Compared with expendable mold processes, nonexpendable casting requires relatively few material inputs from Earth in the context of an orbital SMF.

(a) Permanent Casting

Permanent casting requires a set-up time on the order of weeks, after which production rates of 5-50 pieces/hr-mold are achieved with an upper mass limit of 9 kg per iron alloy item (cf., up to 135 kg for many nonferrous metal parts) and a lower limit of about 0.1 kg. Hot molds are coated with refractory wash or acetylene soot before processing to allow easy removal of the workpiece. Generally, gravity is unnecessary since forced-input feeding is possible. Permanent molds have a life of 3000 castings after which they require redressing. Permanently cast metals generally show 20% increase in tensile strength and 30% increase in elongation as compared to the products of sand casting.

The only necessary terrestrial input is the coating applied before each casting. Typically, permanent mold casting is used in forming iron-, aluminum-, magnesium-, and copper-based alloys. The process is highly automated and state-of-the-art easily could be adapted for use in an extraterrestrial manufacturing facility. The main disadvantage is that the mold is not easy to design or produce automatically. More research is needed on robot production of delicate molds.

(b) Die Casting

In die casting fluid is injected into a mold at high pressures. Set-up time for dies is 1-2 months, after which production rates of 20-200 pieces/hr-mold are normally obtained. Maximum mass limits for magnesium, zinc, and aluminum parts are roughly 4.5 kg, 18 kg, and 45 kg, respectively; the lower limit in all cases is about 30 g. Die injection machines are generally large (up to 3 $\times$ 8 m) and operate at high pressures – 1000 kg/cm^2 and higher, although aluminum usually is processed at lower pressure. A well-designed unit produces over 500,000 castings during the production lifetime of a single mold. The major production step is die construction, usually a steel alloy requiring a great deal of skill and fine tooling to prepare. Only nonferrous materials are die cast, such as aluminum-, zinc-, and copper-based alloys.

The only serious difficulty in applying die casting to an SMF is unit cooling. In terrestrial factories, die machines are water- or air-cooled, both difficult in space. There is little water in the system since flash is removed and remelted, but care must be taken to prevent cold welding of parts to dies in a vacuum manufacturing environment. Die casting is readily automated (Miller and Smith, 1979). Present technology already permits semi-automation, but more

research is required on machine design and automatic die mold preparation for space applications.

(c) Centrifugal Casting

Centrifugal casting is both gravity- and pressure-independent since it creates its own force feed using a temporary sand mold held in a spinning chamber at up to 90 g. Lead time varies with the application. Semi- and true-centrifugal processing permit 30–50 pieces/hr-mold to be produced, with a practical limit for batch processing of approximately 9000 kg total mass with a typical per-item limit of 2.3–4.5 kg. A significant advantage of the centrifugal force method is that no external gravity is required, making it ideal for space applications. Sand is easily recycled, so centrifugal processing depends only to a small degree on terrestrial resupply. There is no limit to the types of metals that can be fabricated.

Automation can be utilized in centrifugal casting. The only requirement is the advent of spin-functional robots, research of which should lead to the broader synergistic advancement of other processes normally dependent on gravity to function properly, such as investment casting.

(d) Continuous Casting

Continuous casting, much like centrifugal molding, produces sheets or beams which may undergo further fabrication. Continuous casting was discussed briefly by an MIT study group in the context of SMF design (Miller and Smith, 1979), and involves forcing a melted metal through an open-ended mold. Heat is extracted and metal exits the mold as a solid fabricated sheet. The MIT study suggested that SMF molds, as those on Earth, might be made of graphite. Unfortunately, carbon is rare in space.

Gravity plays no irreplaceable role in continuous casting on Earth – gravity feeds are used, but manufacturing facility casting machines can rely on pressure to feed liquid metal. Molds or "dies" last several weeks, after which graphite must be reworked to original specifications. Metal melting points impose severe restrictions on mold design. Consequently, iron is difficult while aluminum and its alloys are relatively easy to process. The technique already is well-automated and is used to fabricate aluminum and copper alloys, but only on very special applications for iron.

4B.3 Casting in Space Manufacturing

Casting has its limitations in space. Gravity is a major problem but can be overcome with development of centrifugal systems which work in concert with other systems. The cold-welding effect is also of major concern. To overcome this, it is suggested that fabrication should take place within a closed atmospheric unit.

Lunar basalt molds possibly may replace iron molds. But basalt has a low coefficient of thermal conduction and more research is needed to ensure feasibility of the concept. Lunar basalt should provide adequate molds for aluminum alloys as the former melts at 1753 K (1480°C) and the latter around 873 K (600°C).

These problems are hardly intractable. In the long term, the issues of fully autonomous production, refurbishing of patterns and molds, automatic process control systems, and the application of robotics and other advanced automation techniques to casting technology, must all be addressed.

4B.4 References

Lindberg, Roy A.: Processes and Materials of Manufacture. Second Ed. Allyn and Bacon, Boston, 1977.

Miller, R. H.; and Smith, D. B. S.: Extraterrestrial Processing and Manufacturing of Large Space Systems, NASA CR-161293, 1979.

Yankee, H. W.: Manufacturing Processes. Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

APPENDIX 4C

REVIEW OF POWDER METALLURGY

Powder metallurgy is a forming and fabrication technique consisting of three major processing stages. First, the primary material is physically powdered – divided into many small individual particles. Next, the powder is injected into a mold or passed through a die to produce a weakly cohesive structure very near the true dimensions of the object ultimately to be manufactured. Finally, the end part is formed by applying pressure, high temperature, long setting times (during which self-welding occurs), or any combination thereof. Powder metallurgy technologies may be utilized by minimum initial support facilities to prepare a widening inventory of additional manufacturing techniques, and offer the possibility of creating "seed factories" able to grow into more complex production facilities which can generate many special products in space. The following sections review the basics of powder metallurgy (Jones, 1960).

The history of powder metallurgy and the art of metals and ceramics sintering are intimately related. Sintering involves the production of a hard solid metal or ceramic piece from a starting powder. There is evidence that iron powders were fused into hard objects as early as 1200 BC (Jones, 1960). In these early manufacturing operations, iron was extracted by hand from metal sponge following reduction and was then reintroduced as a powder for final melting or sintering.

A much wider range of products can be obtained using powder processes than from direct alloying of fused materials. In melting operations the "phase rule" applies to all pure and combined elements and strictly dictates the distribution of liquid and solid phases which can exist for specific compositions. In addition, whole body melting of starting materials is required for alloying, thus imposing unwelcome chemical, thermal, and containment constraints on manufacturing. Unfortunately, the handling of aluminum/iron powders poses major problems (Sheasby, 1979). Other substances that are especially reactive with atmospheric oxygen, such as tin (Makhlouf *et al.*, 1979), are sinterable in special atmospheres or with temporary coatings. Such materials may be manipulated far more extensively in controlled environments in space.

In powder metallurgy or ceramics it is possible to fabricate components which otherwise would decompose or disintegrate. All considerations of solid-liquid phase changes can be ignored, so powder processes are more flexible than casting, extrusion forming, or forging techniques. Controllable characteristics of products prepared using various powder technologies include mechanical, magnetic (Kahn, 1980), and other unconventional properties of such materials as porous solids, aggregates, and intermetallic compounds. Competitive characteristics of manufacturing processing (e.g., tool wear, complexity, or vendor options) also may be closely regulated.

4C.1 Cold Welding

Cold or contact welding was first recognized as a general materials phenomenon in the 1940s. It was then discovered that two clean, flat surfaces of similar metal would strongly adhere if brought into contact under vacuum. It is now known that the force of adhesion following first contact can be augmented by pressing the metals tightly together, increasing the duration of contact, raising the temperature of the workpieces, or any combination of the above. Research has shown that even for very smooth metals, only the high points of each surface, called "asperites," touch the opposing piece. Perhaps, as little as a few thousandths of a percent of the total surface is involved. However, these small areas of taction develop powerful molecular connections – electron microscope investigations of contact points reveal that an actual welding of the two surfaces takes place after which it is impossible to discern the former asperitic interface. If the original surfaces are sufficiently smooth the metallic forces between them eventually draw the two pieces completely together and eliminate even the macroscopic interface.

Exposure to oxygen or certain other reactive compounds produces surface layers which reduce or completely eliminate the cold welding effect. This is especially true if, say, a metal oxide has mechanical properties similar to those of the parent element (or softer), in which case surface deformations do not crack the oxide film. Fortunately, the extremely low concentrations of contaminating gases in free space (less than 10^{-14} torr is achievable) should produce minimal coating, so cold welding effects can persist on fresh metal surfaces for very long periods. Contact welding promises a convenient and powerful capability for producing complex objects from metallic powders in space with a minimum of support equipment.

Powders use cold welding to best advantage because they present large surface areas over which vacuum contact can occur. For instance, a 1 cm cube of metal comminuted into 240-100 mesh-sieved particles (60-149 μm) yields approximately 1.25×10^6 grains having a total surface area of 320 cm^2. This powder, reassembled as a cube, would be about twice as big as before since half the volume consists of voids.

If a strong final product is desired, it is important to obtain minimum porosity (that is, high starting density) in the initial powder-formed mass. Minimum porosity results in less dimensional change upon compression of the workpiece as well as lower pressures, decreased temperatures, and less time to prepare a given part. Careful vibratory settling reduces porosity in monodiameter powders to less than 40%. A decrease in average grain size does not decrease porosity, although large increases in net grain area will enhance the contact welding effect and markedly improve the "green strength" of relatively uncompressed powder. In space applications cold welding in the forming stage may be adequate to produce usable hard parts, and molds may not even be required to hold the components for subsequent operations such as sintering.

Hard monodiameter spheres packed like cannonballs into body-centered arrays give a porosity of about 25%, significantly lower than the ultimate minimum of 35% for vibrated collections of monodiameter spheres. (The use of irregularly shaped particles produces even more porous powders.) Porosity further may be reduced by using a selected range of grain sizes, typically 3-6 carefully chosen gauges in most terrestrial applications. Theoretically, this should permit less than 4% porosity in the starting powder, but with binary or tertiary mixtures 15-20% is more the rule. Powders comprised of particles having a wide range of sizes, in theory, can approach 0% porosity as the finest grains are introduced. But powder mixtures do not naturally pack to the closest configuration even if free movement is induced by vibration or shaking. Gravitational differential settling of the mixture tends to segregate grains in the compress, and some degree of cold welding occurs immediately upon formation of the powder compress which generates internal frictions that strongly impede further compaction. Considerable theoretical and practical analyses already exist to assist in understanding the packing of powders (Dexter and Tanner, 1972; Criswell, 1975; Powell, 1980a, 1980b; Shahinpoor, 1980; Spencer and Lewis, 1980; Visscher and Bolsterzi, 1972).

Powder metallurgy in zero-g airless space or on the Moon offers several potential advantages over similar applications on Earth. For example, cold-welding effects will be far more pronounced and dependable due to the absence of undesirable surface coatings. Gravitational settling in polydiameter powder mixtures can largely be avoided, permitting the use of broader ranges of grain sizes in the initial compact and correspondingly lower porosities. Finally, it should be possible to selectively coat particles with special films which artificially inhibit contact welding until the powder mixture is properly shaped. (The film is then removed by low heat or by chemical means, forming the powder in zero-g conditions without a mold.)

Moderate forces applied to a powder mass immediately cause grain rearrangements and superior packing. Specifically, pressures of 10^5 Pa (N/m^2) decrease porosity by 1-4%; increasing the force to 10^7 Pa gains only an additional 1-2%. However, at still higher pressures or if heat is applied the distinct physical effects of particle deformation and mass flow become significant. Considerably greater force is required mechanically to close all remaining voids by plastic flow of the compressed metal.

4C.2 Sintering

Sintering is the increased adhesion between particles as they are heated. In most cases the density of a collection of grains increases as material flows into voids causing a decrease in overall size. Mass movements which occur during sintering consist of the reduction of total porosity by repacking, followed by material transport due to evaporation and condensation with diffusion. In the final stages metal atoms move along crystal boundaries to the walls of internal pores, redistributing mass from the internal bulk of the object and smoothening pore walls.

Most, if not all, metals may be sintered. This is especially true of pure metals produced in space which suffer no surface contamination. Many nonmetallic substances also sinter, such as glass, alumina, silica, magnesia, lime, beryllia, ferric oxide, and various organic polymers. The sintering properties of lunar materials have been examined in detail (Simonds, 1973). A great range of material properties can be obtained by sintering with subsequent reworking. Physical characteristics of various products can be altered by changing density, alloying, or heat treatments. For instance, the tensile strength E_n of sintered iron powders is insensitive to sintering time, alloying, or particle size in the original powder, but is dependent upon the density (D) of the final product according to $E_n/E = (D/d)^{3.4}$, where E is Young's Modulus and d is the maximum density of iron.

Particular advantages of this powder technology include: (1) the possibility of very high purity for the starting materials and their great uniformity; (2) preservation of purity due to the restricted nature of subsequent fabrication steps; (3) stabilization of the details of repetitive operations by control of grain size in the input stages; (4) absence of stringering of segregated particles and inclusions as often occurs in melt processes; and (5) no deformation is required to produce directional elongation of grains (Clark, 1963). There exists a very large literature on sintering dissimilar materials for solid/solid phase compounds or solid/melt mixtures in the processing stage. As previously noted (and see below), any substance which can be melted may also be

atomized using a variety of powder production techniques. Finally, when working with pure elements, scrap remaining at the end of parts manufacturing may be recycled through the powdering process for reuse.

4C.3 Powder Production Techniques

Any fusible material can be atomized. Several techniques have been developed which permit large production rates of powdered particles, often with considerable control over the size ranges of the final grain population. Powders may be prepared by comminution, grinding, chemical reactions, or electrolytic deposition. Several of the melting and mechanical procedures are clearly adaptable to operations in space or on the Moon.

Powders of the elements Ti, V, Th, Cb, Ta, Ca, and U have been produced by high-temperature reduction of the corresponding nitrides and carbides. Fe, Ni, U, and Be submicron powders are obtained by reducing metallic oxalates and formates. Exceedingly fine particles also have been prepared by directing a stream of molten metal through a high-temperature plasma jet or flame, simultaneously atomizing and comminuting the material. On Earth various chemical- and flame-associated powdering processes are adopted in part to prevent serious degradation of particle surfaces by atmospheric oxygen. Powders prepared in the vacuum of space will largely avoid this problem, and the availability of zero-g may suggest alternative techniques for the production of spherical or unusually shaped grains.

Two powdering techniques which appear especially applicable to space manufacturing are atomization and centrifugal disintegration. Direct solar energy can be used to melt the working materials, so the most energy-intensive portion of the operation requires a minimum of capital equipment mass per unit of output rate since low-mass solar collectors can be employed either on the Moon or in space. Kaufman (1979) has presented estimates of the total energy input of the complete powdering process in the production of iron parts. The two major energy input stages – powder manufacturing and sintering – require 5300 kW-hr/t and 4800 kW-hr/t, respectively. At a mean energy cost of $0.025/kW-hr, this corresponds to $250/t or about $0.11/kg. Major savings might be possible in space using solar energy.

Atomization is accomplished by forcing a molten metal stream through an orifice at moderate pressures. A gas is introduced into the metal stream just before it leaves the nozzle, serving to create turbulence as the entrained gas expands (due to heating) and exits into a large collection volume exterior to the orifice. The collection volume is filled with gas to promote further turbulence of the molten metal jet. On Earth, air and powder streams are segregated using gravity or cyclone devices. Cyclone separators could be used in space, although an additional step would be required – introduction of the powder into a pumping chamber so that the working gas may be removed and reused. Evacuated metal would then be transferred to the zero-pressure portion of the manufacturing facility. Figures 4.24 and 4.25 present schematics of major functional units of terrestrial facilities for metal atomization (DeGarmo, 1979; Jones, 1960).

Simple atomization techniques are available in which liquid metal is forced through an orifice at a sufficiently high velocity to ensure turbulent flow. The usual performance index used is the Reynolds number $R = fvd/n$, where f = fluid density, v = velocity of the exit stream, d = diameter of the opening, and n = absolute viscosity. At low R the liquid jet oscillates, but at higher velocities the stream becomes turbulent and breaks into droplets. Pumping energy is applied to droplet formation with very low efficiency (on the order of 1%) and control over the size distribution of the metal particles produced is rather poor. Other techniques such as nozzle vibration, nozzle asymmetry, multiple impinging streams, or molten-metal injection into ambient gas are all available to increase atomization efficiency, produce finer grains, and to narrow the particle size distribution. Unfortunately, it is difficult to eject metals through orifices smaller than a few millimeters in diameter, which in practice limits the minimum size of powder grains to approximately 10 μm. Atomization also produces a wide spectrum of particle sizes, necessitating downstream classification by screening and remelting a significant fraction of the grain.

Centrifugal disintegration of molten particles offers one way around these problems, as shown in figure 4.25(a). Extensive experience is available with iron, steel, and aluminum (Champagne and Angers, 1980). Metal to be powdered is formed into a rod which is introduced into a chamber through a rapidly rotating spindle. Opposite the spindle tip is an electrode from which an arc is established which heats the metal rod. As the tip material fuses, the rapid rod rotation throws off tiny melt droplets which solidify before hitting the chamber walls. A circulating gas sweeps particles from the chamber. Similar techniques could be employed in space or on the Moon. The chamber wall could be rotated to force new powders into remote collection vessels (DeGarmo, 1979), and the electrode could be replaced by a solar mirror focused at the end of the rod.

An alternative approach capable of producing a very narrow distribution of grain sizes but with low throughput consists of a rapidly spinning bowl heated to well above the melting point of the material to be powdered. Liquid metal, introduced onto the surface of the basin near the center at flow rates adjusted to permit a thin metal film to skim evenly up the walls and over the edge, breaks into droplets, each approximately the thickness of the film (Jones, 1960).

Figure 4.25(b) illustrates another powder-production technique. A thin jet of liquid metal is intersected by high-speed streams of atomized water which break the jet into drops and cool the powder before it reaches the bottom of

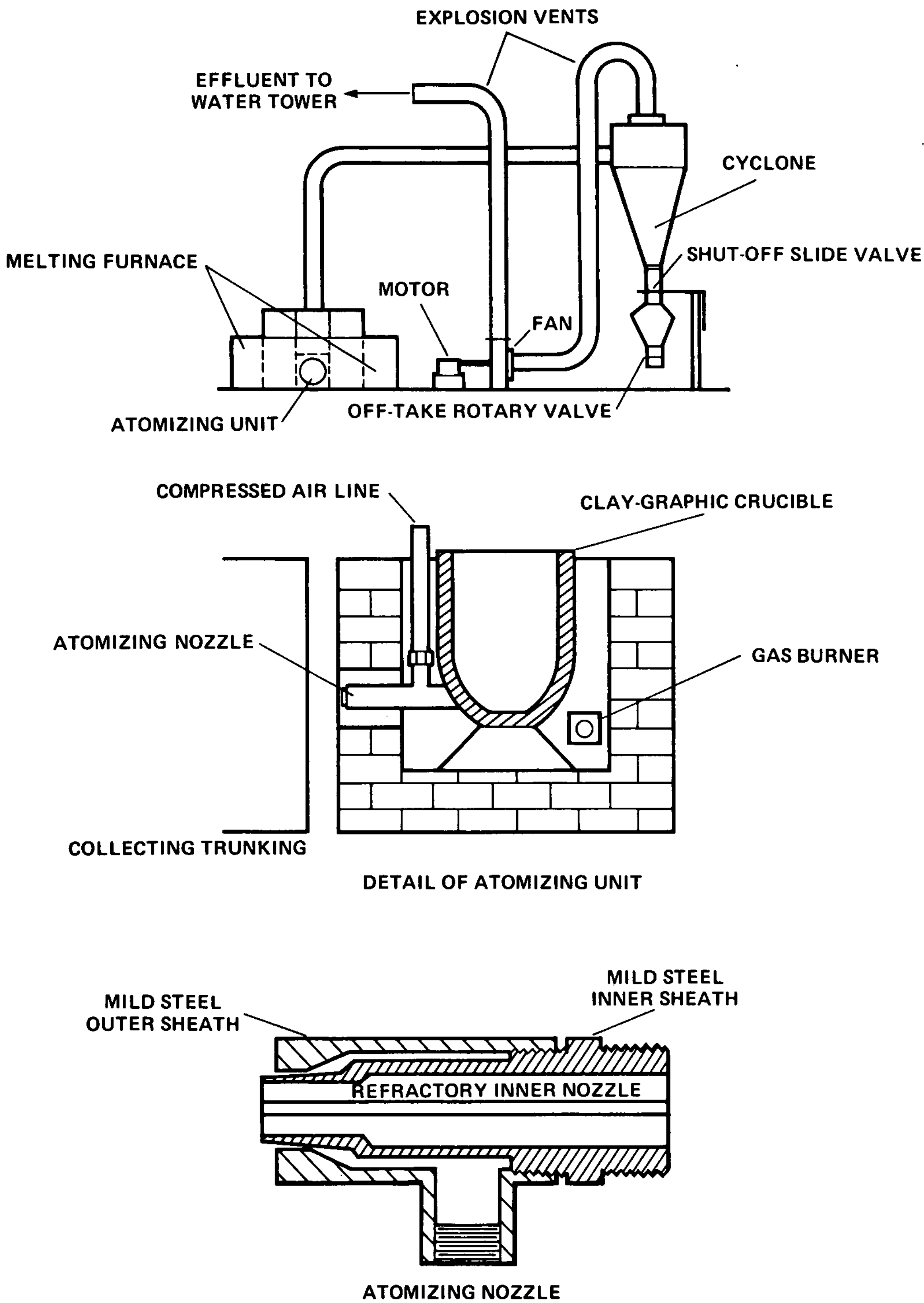


Figure 4.24.– Schematics of an aluminum atomization plant. (From Jones, 1960.)

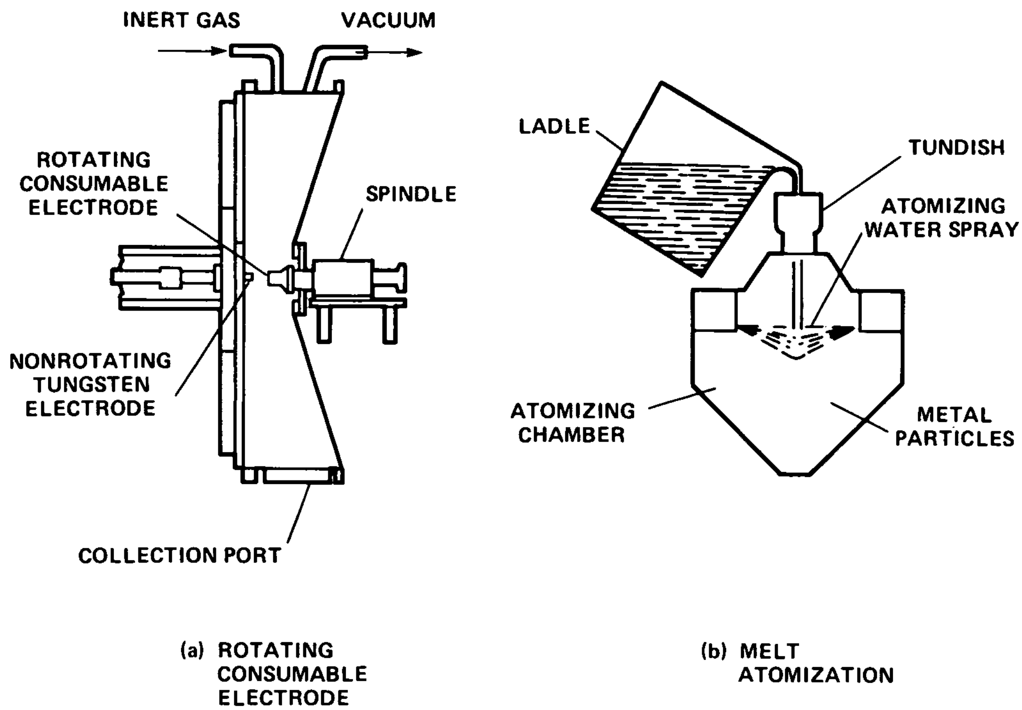


Figure 4.25.– Two methods for producing metal powders. (From Jones, 1960.)

the bin. In subsequent operations the powder is dried. In space applications it would be necessary to recycle the water or other atomizing fluid.

Finally, mills are now available which can impart enormous rotational torques on powders, on the order of 20×10^6 rpm. Such forces cause grains to disintegrate into yet finer particles. Operations in free space should permit a variety of related approaches.

4C.4 Powder Pressing

An extensive literature on the various aspects of powder pressing is available and growing rapidly. Although many products such as pills and tablets for medical use are cold-pressed directly from powdered materials, normally the resulting compact is only strong enough to allow subsequent heating and sintering. Release of the compact from its mold is usually accompanied by a small volume increase called "spring-back." In space, compact strength should far exceed that on Earth due to powerful cold-welding effects on pristine grain surfaces.

In some pressing operations (such as hot isostatic pressing) compact formation and sintering occur simultaneously. This procedure, together with explosion-driven compressive techniques, is used extensively in the production of high-temperature and high-strength parts such as turbine blades for jet engines. In most applications of powder metallurgy the compact is hot-pressed, heated to a temperature above which the materials cannot remain work-hardened. Hot pressing lowers the pressures required to reduce porosity and speeds welding and grain deformation processes. Also it permits better dimensional control of the product, lessened sensitivity to physical characteristics of starting materials, and allows powder to be driven to higher densities than with cold pressing, resulting in higher strength. Negative aspects of hot pressing include shorter die life, slower throughput because of powder heating, and the frequent necessity for protective atmospheres during forming and cooling stages.

One recently developed technique for high-speed sintering involves passing high-amperage electrical current through a powder to preferentially heat the asperities. Most of the energy serves to melt that portion of the compact where migration is desirable for densification; comparatively little energy is absorbed by the bulk materials and forming machinery. Naturally, this technique is not applicable to electrically insulating powders (DeGarmo, 1979).

4C.5 Continuous Powder Processing

The phrase "continuous process" should be used only to describe modes of manufacturing which could be extended indefinitely in time. Normally, however, the term refers to processes whose products are much longer in one physical

dimension than in the other two. Compression, rolling, and extrusion are the most common examples (Jones, 1960).

In a simple compression process, powder flows from a bin onto a two-walled channel and is repeatedly compressed vertically by a horizontally stationary punch. After stripping the compress from the conveyor the compact is introduced into a sintering furnace. An even easier approach is to spray powder onto a moving belt and sinter it without compression. Good methods for stripping cold-pressed materials from moving belts are hard to find. One alternative that avoids the belt-stripping difficulty altogether is the manufacture of metal sheets using opposed hydraulic rams, although weakness lines across the sheet may arise during successive press operations.

Powders can be rolled into sheets or more complex cross-sections, which are relatively weak and require sintering. It is possible that rolling and sintering processes can be combined, which necessitates relatively low roller speeds. Powder rolling is normally slow, perhaps 0.01-0.1 m/sec. This is due in part to the need to expel air from compressed powder during terrestrial manufacture, a problem which should be far less severe in space applications. Considerable work also has been done on rolling multiple layers of different materials simultaneously into sheets.

Extrusion processes are of two general types. In one type, the powder is mixed with a binder or plasticizer at room temperature; in the other, the powder is extruded at elevated temperatures without fortification. Extrusions with binders are used extensively in the preparation of tungsten-carbide composites. Tubes, complex sections, and spiral drill shapes are manufactured in extended lengths and diameters varying from 0.05-30 cm. Hard metal wires 0.01 cm diam have been drawn from powder stock. At the opposite extreme, Jones (1960) considers that large extrusions on a tonnage basis may be feasible. He anticipates that problems associated with binder removal, shrinkage from residual porosity during sintering, and maintenance of overall dimensional accuracies are all controllable. Low die and pressure cylinder wear are expected. Also, it seems quite reasonable to extrude into a vacuum.

There appears to be no limitation to the variety of metals and alloys that can be extruded, provided the temperatures and pressures involved are within the capabilities of die materials. Table 4.25 lists extrusion temperatures of various common metals and alloys. Extrusion lengths may range from 3-30 m and diameters from 0.2-1.0 m. Modern presses are largely automatic and operate at high speeds (on the order of m/sec). Figure 4.26 illustrates seven different processes for generating multilayer powder products by sheathed extrusion.

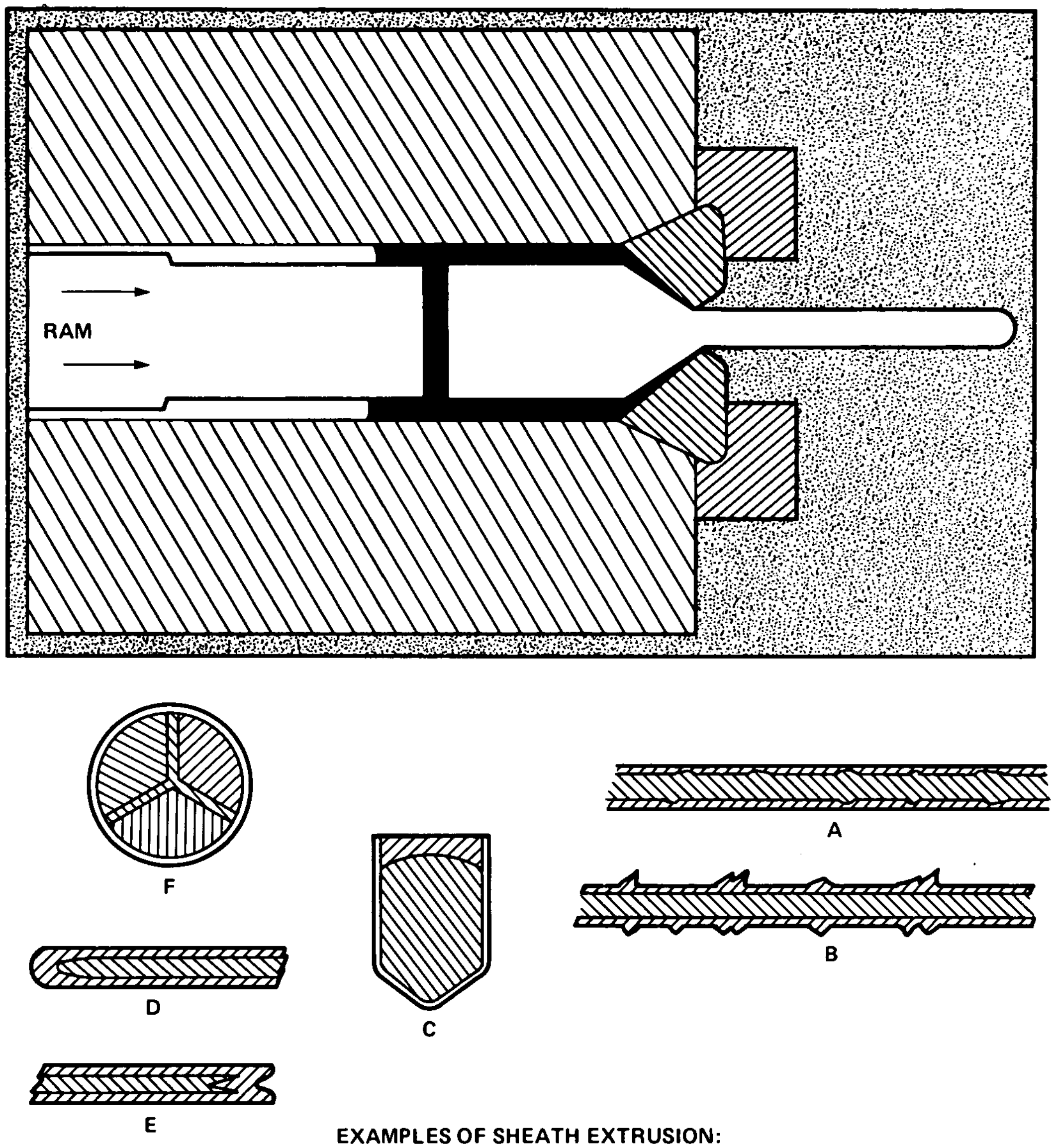
TABLE 4.25.– EXTRUSION TEMPERATURES OF COMMON METALS AND ALLOYS

Metals and alloys	Temperature of extrusion, K
Aluminum and alloys	673-773
Magnesium and alloys	573-673
Copper	1073-1153
Brasses	923-1123
Nickel brasses	1023-1173
Cupro-nickel	1173-1273
Nickel	1383-1433
Monel	1373-1403
Inconel	1443-1473
Steels	1323-1523

4C.6 Special Products

Many special products are possible with powder-metallurgy technology. A nonexhaustive list includes Al_2O_3 whiskers coated with very thin oxide layers for improved refractories; iron compacts with Al_2O_3 coatings for improved high-temperature creep strength; light-bulb filaments made with powder technology; linings for friction brakes; metal glasses for high-strength films and ribbons; heat shields for spacecraft reentry into Earth's atmosphere; electrical contacts for handling large current flows; magnets; microwave ferrites; filters for gases; and bearings which can be infiltrated with lubricants. The product list can be considerably expanded using terrestrial materials. A profitable line of research would be to determine which elements if brought to LEO could offer especially large multiplier effects in terms of the ratio of lunar-materials mass to Earth-supplied mass.

Extremely thin films and tiny spheres exhibit high strength. One application of this observation is to coat brittle materials in whisker form with a submicron film of much softer metal (e.g., cobalt-coated tungsten). The surface strain of the thin layer places the harder metal under compression, so that when the entire composite is sintered the rupture strength increases markedly. With this method, strengths on the order of 2.8 GPa versus 550 MPa have been observed for, respectively, coated (25% Co) and uncoated tungsten carbides. It is interesting to consider whether similarly strong materials could be manufactured from aluminum films stretched thin over glass fibers (materials relatively abundant in space).



EXAMPLES OF SHEATH EXTRUSION:

A. THE SHEATH IS STIFFER THAN THE CORE.
B. THE CORE IS STIFFER THAN THE SHEATH.
C. TAPERED SHEATH NOSE.
D. THICKENED NOSE OF EXTRUDED SECTION.
E. INCLUSION OF THE SHEATH IN EXTRUSION DEFECT.
F. SHEATH ENCLOSING SUBDIVIDED CORES.

(WILLIAMS)

Figure 4.26.– Multilayer powder product production using sheathed extrusion. (From Jones, 1960.)

4C.7 References

Champagne, B.; and Angers, R.: Fabrication of Powders by the Rotating Electrode Process. Intern. J. Powder Metallurgy and Powder Tech., vol. 16, no. 4, October 1980, pp. 359-367.

Clark, Frances: Advanced Techniques in Powder Metallurgy. Rowan and Littlefield, 1963.

Criswell, David R.: The Rosiwal Principle and the Regolithic Distributions of Solar Wind Elements. Proc. 6th Lunar Sci. Conf., Pergamon Press, New York, 1975, pp. 1967-1987.

DeGarmo, E. P.: Materials and Processes in Manufacturing. Fifth edition. Macmillan, New York, 1979.

Dexter, A. R.; and Tanner, D. W.: Packing Densities of Mixtures of Spheres with Log-Normal Size Distributions. Nature (Physical Science), vol. 238, no. 80, 10 July 1972, pp. 31-32.

Jones, W. D.: Fundamental Principles of Powder Metallurgy. Edward Arnold Ltd., London, 1960.

Khan, M. K.: The Importance of Powder Particle Size and Flow Behavior in the Production of P/M Parts for Soft Magnetic Applications. Intern. J. Powder Metallurgy and Powder Tech., vol. 16, no. 2, April 1980, pp. 123-130.

Kaufman, S. M.: Energy Consumption in the Manufacture of Precision Metal Parts from Iron Powders. Intern. J. Powder Metallurgy and Powder Tech., vol. 15, no. 1, Jan. 1979, pp. 9-20.

Makhlouf, M. M.; Mouid, A. M.; and Merchant, H. D.: Sintering of Chemically Preconditioned Tin Powder. Intern. J. Powder Metallurgy and Powder Tech., vol. 15, no. 3, July 1979, pp. 231-237.

Powell, M. J.: Computer-Simulated Random Packing of Spheres. Powder Technology, vol. 25, no. 1, 1980, pp. 45-52.

Powell, M. J.: Distribution of Near-Neighbors in Randomly Packed Hard Spheres. Powder Technology, vol. 26, no. 2, 1980, pp. 221-223.

Shahinpoor, M.: Statistical Mechanical Considerations on the Random Packing of Granular Materials. Powder Technology, vol. 25, no. 2, 1980, pp. 163-176.

Sheasby, J. S.: Powder Metallurgy of Iron-Aluminum. Intern. J. Powder Metallurgy and Powder Tech., vol. 15, no. 4, Oct. 1979, pp. 301-305.

Simonds, C. H.: Sintering and Hot Pressing of Fra Mavro Composition Glass and the Lithification of Lunar Breccias. Amer. J. Sci., vol. 273, May 1973, pp. 428-429.

Spencer, B. B.; and Lewis, B. E.: HP-67-97 and TI-59 Programs to Fit the Normal and Log-Normal Distribution Functions by Linear Regression. Powder Technology, vol. 27, no. 2, 1980, pp. 219-226.

Visscher, W. H.; and Bolsterzi, M.: Random Packing of Equal and Unequal Spheres in Two and Three Dimensions. Nature, vol. 239, no. 5374, 27 October 1972, pp. 504-507.

APPENDIX 4D

REVIEW OF DEFORMATION IN MANUFACTURING

Deformation involves the production of metal parts from ingots, billets, sheets, and other feedstock. Metal is forced to assume new shapes by the application of large mechanical forces to the material while it is either hot or cold. The purpose of this mechanical working is twofold: first, to bring the feedstock into a desired shape, and second, to alter the structure and properties of the metal in a favorable manner (e.g., strengthening, redistribution of impurities).

4D.1 Deformation Techniques

A number of major deformation techniques are described below with emphasis on currently automated techniques, followed by an overview of deformation criteria in space manufacturing applications.

(a) Forging

The deformation of metal into specific shapes includes a family of impact or pressure techniques known as forging. Basic forging processes are smith or hammer forging, drop forging, press forging, machine or upset forging, and roll forging. Special forging processes include ring rolling, orbital forging or rotaforming, no-draft forging, high-energy-rate forming, cored forging, wedge rolling, and incremental forging.

Unimate and Prab industrial robots are already employed in many commercial forge shops. For example, the 2000A Unimate is currently used to feed billets through a two-cavity die-forging press to be formed into raw differential side gears (Unimation, 1979). A more sophisticated robot, the 4000A three-axis Unimate, is used to transfer hot (~1400 K) diesel engine crankshafts from a forging press into a twister (fig. 4.27). The Unimate used in this operation has a 512-step memory, rotary-motion mirror imaging, and memory-sequence control with one base and one subroutine (Unimation, 1979). Forging systems involving gas, steam, or hydraulic drives are excluded from consideration in space or lunar factories since, in general, any system susceptible to fluid leakage is of lower developmental priority for space operations than other processes with similar capabilities.

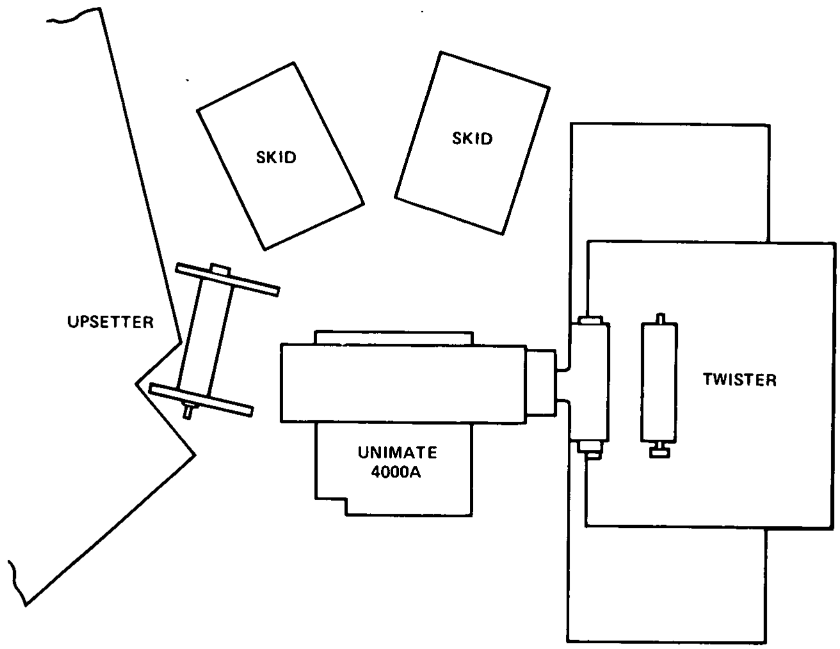
The energy required for single-drop forging is a function of the mass and velocity of the ram, exclusive of energy to rough form or to heat the parts for the forge. This assumes only a single pass and not the usual progressive steps to create a metal form from one die impression to the next. One modification to be considered in gravity-fall (drop) forging on the Moon is mass enhancement by sintered iron weights, possibly coupled with electromagnetic acceleration (only electrical energy is needed for lunar factory forging processes). Impact forging by electromagnetically driven opposing die sets may produce still closer parts tolerances than drop forging.

Forging operations, from raw precut feedstock to ejected forging, likely can be completely automated on the Moon.

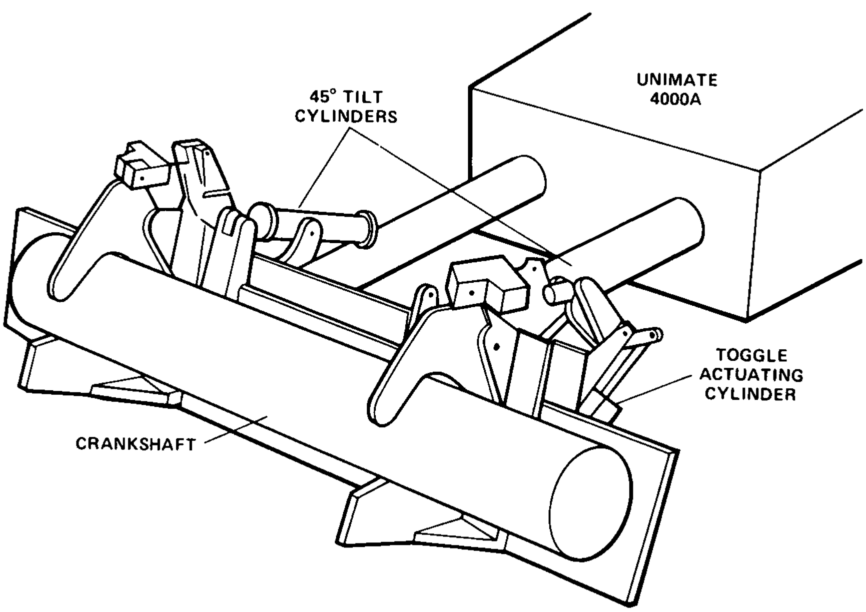
(b) Rolling

Space manufacturing applications of rolling mills have been considered by Miller and Smith (1979). Automated stop-go operations for the rolling mill, slicer, striater, trimmers, welders, and winders in figure 4.28 readily may be visualized. It is important to note that aluminum is the resource considered and ribbon is the processed form. Lunar aluminum-rich mineral recovery, extraction, and processing make good sense since beam builders in Earth orbital space already have been designed for aluminum ribbon feedstock.

Two types of rolling mills can manufacture ribbon from aluminum alloy slabs prepared from lunar anorthosite. The first or regular type of mill consists of a series of rolling stands with lead-in roughing rollers and finishing rollers at the end. Input slabs travel through one stand after another and are reduced in thickness at each stand. Each stand rolls the slab once. High production rates result. A second option is the reversing mill. Slabs are routed back and forth through the same stand several times and are reduced in thickness during each pass. This requires a mill with movable rolls able to continually tighten the gap as slabs grow thinner. Although reversing mills have lower production rates and are more complicated than regular rolling mills, they are more versatile and require fewer machines. Expected yearly aluminum production at the SMF designed by Miller and Smith (1979) is minimal by normal rolling mill standards, so low-mass reversing mills are sufficient for the present reference SMF.



INSTALLATION LAYOUT



SPECIAL HAND DESIGN

Figure 4.27.– Application of 4000A three-axis Unimate to the production of forged diesel engine crankshafts (upsetter is a double-acting mechanical forge press operating in the horizontal plane).

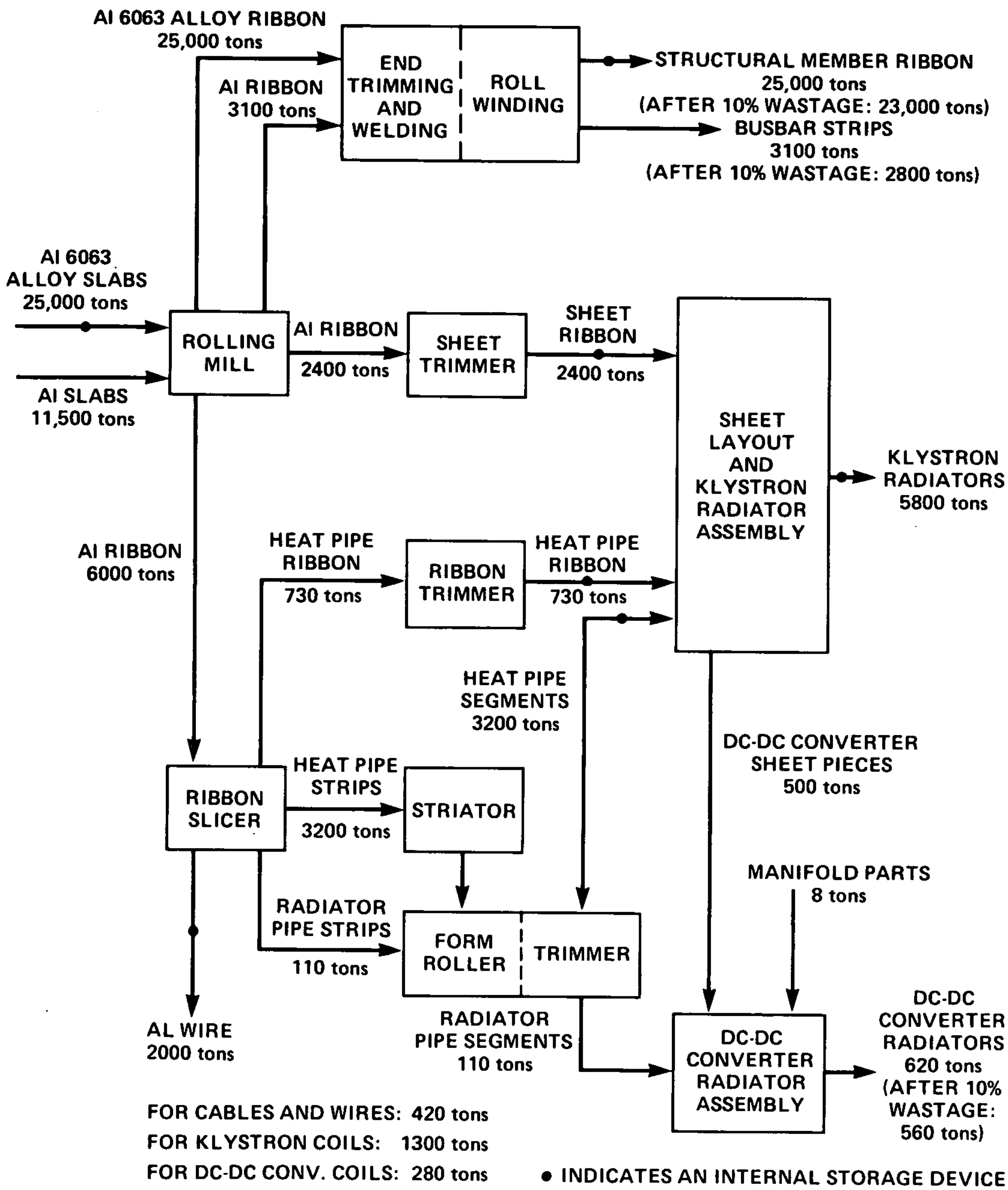


Figure 4.28.– Ribbon and sheet operations; detailed layout for proposed SMF system.

Input slabs can be hot-rolled or cold-rolled. However, if a cold aluminum alloy slab is rolled to more than 120% of its input length, cracks appear in the material. To avoid this problem, slabs are annealed between passes or are rolled hot throughout the process. (The final rolling pass should be done cold to improve the structural properties of the output.) Miller and Smith (1979) chose to hot-roll the input slabs, which, therefore, either travel directly from the continuous caster to the rolling mill without cooling or are taken from intermediate storage and preheated before insertion into the mill. Aluminum alloy slabs elongate roughly by a factor of 11 as they are squeezed from a thickness of 2 cm down to 1.77 mm. Rolling also widens the slabs from 70 to 73.5 cm, the width required for structural member ribbon. Input slab length is arbitrary and often is determined merely by handling convenience, though slabs usually are at least as long as they are wide. An 80-cm-long slab produces about 8.8 m of ribbon.

(c) Special Forming Operations

The following forming operations are considered as a group with respect to robotics applications and lunar factory criteria: conventional stretching, conventional drawing (involving nine suboperations) and deep drawing, swaging, spinning, and bending.

Stretching is a cold-forming process in which sheet metal is wrapped around an upward-moving form block. Conventional drawing involves pressing a flat metal blank into a female die while stretching the blank to force it to conform to the shape of a male die or punch. Shallow drawing is defined as a deformation cup no deeper than half its diameter with little thinning of the metal, whereas, deep drawing produces a cup whose depth may exceed its diameter with more pronounced wall thinning. Swaging is a cold-forging process in which an impact or compressive force causes metal to flow in a predetermined direction. Spinning is a forming technique for plastically deforming a rapidly rotating flat disk against a rotating male contour. Cold spinning is used for thin sheets of metal. Hot-spinning of heavier sheets up to 150 mm thick can produce axisymmetric (shell) shapes. Finally, bending is the plastic deformation of metals about a linear axis with little or no change in the surface area.

Robotics applications and space manufacturing options for these types of deformation processes are minimal, especially under vacuum conditions. If there is no oxidized film on the metal, the workpiece and die may contact weld, causing the machine to seize.

(d) Extrusion

In the extrusion process, either at high or low temperatures, metal is compressively forced through a suitably shaped die to form a product with reduced cross-section – like squeezing toothpaste from a tube. Lead, copper, aluminum, magnesium, and their many alloys are commonly employed, and hydrostatic extrusion using high-pressure fluids into the die makes possible similar processing of relatively brittle materials such as molybdenum, beryllium, and tungsten. Steel is relatively difficult to extrude because of its high-yield strength and its tendency to weld to die walls. Extrusion by pressurizing solid metals shares with other deformation processes problems of cold welding. However, the degree of such welding decreases if markedly dissimilar metals are in contact. The vacuum environment may enhance ductility for some extruded metals.

In one variant of the basic extrusion process, melts are drawn through dies to produce threads. The use of basalt in preparing spun products is well known (Kopecky and Voldan, 1965; Subramanian et al., 1975, 1979) and has numerous lunar applications (see table 4.16). A variation of the technique is the use of centrifugal force to spin the extruded threads (Mackenzie and Claridge, 1979).

In commercial spun basalt processes, molten basalt is drawn through a platinum-rhodium bushing and the final fiber blasted by a tangential gas or steam jet in the air cone as shown in figure 4.7. Fibers also may be produced without the air cone by direct pulling of a winding reel. For example, work done by Subramanian et al. (1975) showed that molten basalt flowing from a 3-mm hole in a graphite crucible, yielded fibers by simple mechanical pulling (table 4.26). The crude fibers created using this procedure were nonuniform, measured about 150 μm diam, and contained many nodules – a poor product compared with air cone output. Assuming the air/steam cone can be eliminated from basalt spinning operations, a step-by-step Unimate-automatable sequence is suggested in table 4.27.

As yet no research has been performed either on vacuum or lunar basalt fiber drawing. Molten basalt on the Moon has very low viscosity which may possibly be controlled, if necessary, by additives. At present it remains unknown whether mechanical spinning of raw lunar basalts is possible or if the vacuum environment will yield a thinner, more uniform product. Still, extrusion of viscous rock melts to produce spun products appears promising and as indicated in table 4.27 is likely amenable to automation in space-manufacturing applications.

(e) Shearing

Shearing is the mechanical cutting of sheet or plate materials using two straight cutting blades, without chip

TABLE 4.26.– AVERAGE TENSILE STRENGTHS OF BASALT FIBERS (10-50 SPECIMENS FOR EACH VALUE)

Temperature at bottom of bushing, K	Fiber size, μm	Tensile strength,	
		MPa	psi
1450	9-11	66	96,000
1510	9-10	134	196,000
	13-15	130	190,000
1525	7-9	143	209,000
	9-11	163	238,000
	13-16	145	212,000
1560	8-10	136	190,000
	11-13	128	187,000
	15-18	132	193,000
1600[a]	7.5	149	218,000

[a]Average of only five specimens.

TABLE 4.27.– OPERATIONAL SEQUENCE FOR AUTOMATED MANUFACTURE OF SPUN BASALT USING UNIMATE ROBOTICS TECHNOLOGY

Step	Procedure
1	Unimate sensors scan electric furnace temperature. Adjusts temperature for optimum viscosity.
2	Unimate introduces 100 kg of raw basalt into furnace through hopper feed gate.
3	Unimate raises furnace temperature to above liquidus with serial decrease to optimum temperature as melting proceeds.
4	Unimate causes discharge of set volume of melt into crucible resting on detent plugging mm-sized hole in crucible base.
5	Unimate sensors monitor crucible temperature fall-off until viscosity increase prevents leakage of charge.
6	Unimate positions crucible within induction coil above drum reel in raised position.
7	Unimate system activates induction furnace to lower viscosity of charge using programmed weight/temperature program to produce temperature (viscosity) plateau until first molten basalt droplet draining from crucible is grasped by clip on reel drum.
8	Unimate controller triggers drum release and turn operation begins, which results in the drawing of fiber.
9	Unimate sensors observe basalt fiber thread output using fiber optic techniques. Fiber diameter controls reel rate and furnace temperature. If no fiber is present, drum is raised and operation repeated.
10	Crucible weight-sensitive switch cuts off induction furnace as melt is consumed. Fiber breaks, filled reel drum is removed by Unimate and is replaced by an empty.
11	Reel drum is raised and empty crucible moved by Unimate onto detent below furnace. Procedure begins again.

formation, burning or melting (DeGarmo, 1979). If the shearing blades have curved edges like punches or dies the process is given another name (e.g., blanking, piercing, notching, shaving, trimming, dinking, and so on as noted in table 4.17.

Shearing already has been automated in many industries. For instance, the Chambersburg Engineering Company has incorporated a 2000B Unimate into a trimming operation performed on the output of an impact forging system. The robot moves 1400 K platters from the forge to hot trimmers, sensing, via hand tooling interlocks, that it has properly grasped the platter. An infrared detector checks parts for correct working temperatures, and the robot rejects all platters for which either grasp or temperature requirements are not met (Unimation, 1979).

Despite its tremendous utility on Earth, shearing appears less desirable than other options for space manufacturing because of the problems of cold welding and shearing tool wear. Also, ceramic and silicate forms cannot be processed by conventional shearing techniques. The most attractive alternative may be laser-beam cutting, piercing, punching, notching, and lancing. Yankee (1979) has reviewed laser-beam machining (LBM) generally, and additional data are provided in section 4.3.1. The application of LBM techniques to metals for shearing operations is an established technology, whereas laser beam cutting of basalt and basalt products is not well-documented.

4D.2 Deformation Criteria and Research Options for Space Manufacturing

In general, deformation processes that do not require gas or liquid drives but emphasize electrical or electromagnetic mechanical power sources appear more practical for space manufacturing applications. Processes yielding thin-walled or ribbon forms such as reversible rolling or electroforming appear favorable. The mass/production ratio argues against heavy forges and in favor of roller technology, an approach which also should improve the quality of output in high-vacuum manufacturing environments. Deformation processes involving forming or shearing typically consume little material (except for fluid-driven devices). On the Moon, the optimum near-term design philosophy is to develop automated systems powered exclusively by electric and magnetic forces.

In order to make tool products, versatile semiautomated machines are initially required for the terrestrial demonstration program. Tool life and machining time must be assessed in view of the extraterrestrial conditions anticipated. For example, Ostwald (1974) has reviewed these

parameters for cost estimation. The Taylor tool life equation is $VT^nF^m = k$, where V is linear tool velocity across the workpiece (m/sec), T is tool life (sec), n and m are dimensionless empirical exponents (logarithmic slopes), F is tool bit-feed rate or relative speed of workpiece and cutting surfaces (m/sec or m/rev), and k is a constant determined by laboratory evaluation of various cutting materials. Machining time t is given by $\pi LD/12VF$, where L is length of cut (m) and D is tool diameter (m). Unfortunately, the special production environment includes low- to zero-g which precludes all shaving- or chip-generating processes unless tools are placed under an oxygen-rich atmosphere.

Clearly, novel techniques must be considered in manufacturing designs intended for nonterrestrial applications. For instance, thread rolling offers a solution to fastener production, electroforming appears suitable for thin-walled containers, and noncentrifugal basalt casting may prove useful in low- or zero-g and yield a more homogeneous product. Vacuum enhances the characteristics of some metals, e.g., cold rolling increases the tensile strength of steel and improves the ductility of chromium. Electrostatic fields may enhance bubble coalescence in metallurgical or rock-melt products.

Many areas of research and development are required to generate appropriate deformation options for an SMF. In deformation processes where oxidized metal surface coatings must be broken (e.g., impact forging, stretching, deep drawing, and shearing), the minimum amount of oxygen necessary to prevent cold welding must be determined. Specific surface poisoning requirements must be measured for specific metals. Thermal environment is also of critical significance. Deformation at temperatures below about 230 K must take proper account of metal embrittlement. Fracture propagation in very cold steel is a serious problem on Earth. Rate processes in metal deformation may be significant in a lunar factory. If an enclosed, slightly oxygenated automated factory bay is provided (perhaps adjacent to the shirtsleeve environment of a manned facility) there appears to be no severe energy constraint in keeping the bay area above 230 K. Temperature control could be achieved by electrical heaters or unidirectional heat pipes for factories sited, say, at the lunar poles (Green, 1978).

Additional research opportunities include:

- Remote sensing of nonterrestrial ore deposits
- Mass launch of materials to processing plants
- Commonality of magnetic impulse forming components with those of mass-launch equipment
- Quality control of ores by intelligent robots
- Optimum spun/cast basalt mixtures
- Tool-life evaluations including sintered and cast basalts
- Powder metallurgy using induction heating or admixed micron-sized raw native iron in lunar "soil" (abundance about 0.5%)
- Factory control strategies
- Factory configuration studies.

Further experimentation also is needed with metal/rock test pairs to determine wear, abrasion, and hardness characteristics after deformation under high-vacuum, low-oxygen conditions. The U.S. Bureau of Mines has done some research on certain aspects of this problem at their centers in Albany, Denver, and Twin Cities. Test equipment, procedures and key personnel pertinent to space and lunar manufacturing options are named in table 4.28.

The role played by humans in space operations will vary with the machine for some deformation processes. Optimum proportions of human and robot activities in lunar factories will doubtless evolve over a period of time, with major manned support expected in early phases of SMF operation, and far less, once production becomes routine. Almost all forming or shearing procedures can be automated either in feed or transfer operations. Indeed, present-day Unimate-series robots have proven especially suitable in such applications in terrestrial industry.

TABLE 4.28.– METAL/ROCK TEST EQUIPMENT SUITABLE FOR LUNAR-FACTORY RESEARCH

FRICTION AND ABRASION WEAR

Erosive-wear testing facility

Albany Metallurgy Research Center
John E. Kelley, 420-5896

A 12-specimen erosion test apparatus built at AMRC uses an S.S. White Airbrasive model-H unit to propel 27 μm Al_2O_3 particles against specimens at temperatures up to 1,000°C in selected atmospheres and at selected impingement angles. Relative erosion is determined by comparing material loss of a target with that of a "standard" specimen.

Friction and rubbing-wear test facility

Albany Metallurgy Research Center
John E. Kelley, 420-5896

A Falex-6 friction and wear machine built by Faville-LeVally Corp. is used to measure abrasion wear, adhesive wear, and coefficient of friction of solid materials. Pin-on-disc and ring-on-ring tests can be made, wet or dry, with or without abrasive particles, in either cyclic or continuous rubbing modes, under variable and controllable conditions of speed, load, atmosphere, and temperature to 260°C (500°F).

Friction and wear

Twin Cities Mining Research Center
D. R. Tweeton, 725-3468

The Dow Corning Alpha LWF-1 friction- and wear-testing machine can measure sliding friction of metal/metal or metal/mineral test pairs in air or environmental fluid.

Impact-abrasion tester

Albany Metallurgy Research Center
John E. Kelley, 420-5896

An impact machine with variable speed and thrust is used to repeatedly impact test specimens tangentially against a rough material such as sandstone to determine the impact-abrasion wear rate.

Simulated-service ball-valve tester

Albany Metallurgy Research Center
John E. Kelley, 420-5896

Ball valves fitted with experimental parts such as balls and seats can be tested for wear by automatic cyclic operation. During each cycle a differential pressure up to 2100 Pa at 340°C (650°F) is applied, then relieved, across the valve, and abrasive solids are passed back and forth through the valve by operating the tester in the manner of an hourglass. Parts wear is monitored by recording the rate of gas leakage across, say, the ball and seat each time the differential pressure is applied. Damaged parts are removed and examined both macro- and microscopically.

HARDNESS AND SCRATCH ANALYSIS

Microhardness

Twin Cities Mining Research Center
George A. Savanick, 725-4543

The Zeiss microindentation hardness tester is capable of measuring the microhardness of selected microscopic areas on solid surfaces. A Knoop diamond is pressed into the solid and the diamond-shaped impresson thus formed is measured under high magnification (500-1,500 X) with a special eyepiece. The optical system is equipped with a Nomarski differential interference contrast capability which enhances image contrast.

TABLE 4.28.– CONCLUDED

HARDNESS AND SCRATCH ANALYSIS – CONCLUDED
Schmidt hardness Twin Cities Mining Research Center W. A. Olsson, R. E. Thill, 725-4580 Soil test Schmidt hardness hammer and Shore scleroscope hardness tester for determining the hardness properties of a material.
Scratch analysis Twin Cities Mining Research Center Robert J. Willard, 725-4573 Hilger and Watts fine-scratch microscope, model TM-52, for use in measuring widths and depth (in inches) of scratches on rock and mineral materials. Moderate experience in scratch measurements, can provide scratch analyses on a limited number of samples of any solid, translucent or opaque material.
Shore hardness Denver Mining Research Center R. Gerlick, 234-3765 Shore hardness tester to determine hardness of rock and other materials.
Rock drilling and cutting Core preparation Denver Mining Research Center H. C. Farley, E. B. Wimer, 234-3755 Trained staff and equipment available to take core from small samples and prepare it for testing, cutting, grinding, etc. Rock cutting and handling Twin Cities Mining Research Center R. L. Schmidt, 725-3455 Trained staff and equipment are available to conduct small- or large-scale experiments in the laboratory or field. Instrument drilling equipment includes a 2-boom jumbo with drifters, airleg drills, a diesel-powered diamond drill, and a truck-mounted rotary drill. Small- and large-scale linear rock-cutting apparatus are also available with thrust capabilities to 14 tons. The laboratory is equipped with service equipment for handling up to 7-ton rock blocks.

4D.3 References

DeGarmo, E. P.: Materials and Processes in Manufacturing. Macmillan, New York, 1979. Fifth Edition.

Green, J.: The Polar Lunar Base – A Viable Alternative to L-5. American Astronautical Society Conference, 30 October-2 November 1978. Paper No. AAS 78-191. 41 pp. (Advances in the Astronautical Sciences, vol. 38, pt. 1, pp. 385–425.)

Kopecky, L.; and Voldan, J.: The Cast Basalt Industry. Annals of the New York Academy of Science, vol. 123, article 2, July 15, 1965, pp. 1086–1105.

MacKenzie, J. D.; and Claridge, R. C.: Glass and Ceramics from Lunar Materials. In Space Manufacturing III, Proceedings of the 4th Princeton/AIAA Conference, J. Grey and C. Krop, eds., 14-17 May 1979, Paper No. 79-1381.

Miller, Rene H.; and Smith, David, B. S.: Extraterrestrial Processing and Manufacturing of Large Space Systems. NASA CR-161293, vol. 1-3, 1979.

Ostwald, Phillip F.: Cost Estimating for Engineering and Management. Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

Subramanian, R. V.; Austin, H. F.; Raff, R. A. V.; Sheldon, G.; Dailey, R. T.; and Wullenwaber, D.: Use of Basalt Rock for the Production of Mineral Fiber, Pacific Northwest Commission Annual Report, Contract NR-3001, College of Engineering, Washington State University, Pullman, Washington, June 1975. 79 pp.

Subramanian, R. V.; and Kuang-Huah, Shu: Interfacial Bonding in Basalt Fiber-Polymer Composites. 34th Annual Technical Conference, Reinforced Plastics/Composites Institute, Section 17C, New Orleans, La., Jan. 1979, pp. 1-10.

Unimation, Inc.: Unimation Application Notes, Industrial Robots: Volume 2 – Applications, Tanner, W. R., ed., Society of Manufacturing Engineers, Dearborn, Michigan, 1979, pp. 93-95.

Yankee, Herbert W.: Manufacturing Processes. Prentice-Hall, Englewood Cliffs, New Jersey, 1979. 765 pp.

APPENDIX 4E

REVIEW OF WELDING, BRAZING, AND SOLDERING TECHNIQUES

Joining techniques involving elevated temperatures and materials fusion include welding, brazing, and soldering. Welding is a process leading to the permanent joining of materials (usually metals) through a suitable combination of temperature and pressure (DeGarmo, 1979). Approximately 40 different welding techniques have been utilized in terrestrial situations (Lindberg, 1977). Brazing and soldering require the use of a molten filler to join metal workpieces. The workpieces themselves are not melted; rather, capillary action facilitates the joining process. Brazing occurs when filler material reaches a melting temperature above 723 K (840°F); soldering uses fillers with melting points below 723 K (DeGarmo, 1979; Schey, 1977).

Within the three basic classes there are numerous joining alternatives for space manufacturing operations. Analysis is greatly simplified by reducing the 61 welding, brazing, and soldering techniques identified in table 4.17 to the following six major categories: electric arc welding, oxyfuel gas welding (i.e., gas-oxygen flame welding), resistance welding, solid-state welding, electronic welding, and brazing/soldering. While some overlap is inevitable, this approach appears effective in providing first-order discrimination between immediately useful and less-feasible joining technologies appropriate for SMF deployment.

4E.1 Metals Joining Analysis

To determine the suitability of various joining processes for space and lunar manufacturing applications, selection criteria for SMF options (table 4.18) were applied to each major terrestrial welding, brazing, and soldering technique. These criteria include usefulness in the production of other manufacturing equipment; production rates and required consumables; energy of production; preparatory steps leading to the manufacture of the process itself or products it can help build; mandatory environmental characteristics to enable processing to proceed; feasibility of automation/teleoperation and people roles required (if necessary); further R&D needed to develop promising alternatives; and a qualitative mass-multiplication ratio or "Tukey Ratio" (see chapter 5), an indication of the extent to which nonterrestrial (i.e., lunar) materials can be utilized as opposed to costly up-shipment of feedstock from Earth (Heer, 1980, unpublished draft notes of the Proceedings of the Pajaro Dunes Goal-Setting Workshop, June 1980.)

4E.1.1 Electric arc welding

Electric-arc-welding techniques include shielded or unshielded metal, gas metal (pulsed, short circuit, electrogas, spray transfer), gas tungsten, flux-cored, submerged, plasma arc, carbon arc, stud, electroslag, atomic hydrogen, plasma-MIG, and impregnated tape welding. The SMF suitability assessment is as follows:

- Make other equipment – A basic joining process is needed.
- Production rates – Houldcroft (1977) gives a figure of 3-140 mm^2/sec and estimates a metal deposition rate of 1-12 kg/hr. Schwartz (1977) cites a 27 kg/hr figure for plasma arc plus hot-wire welding.
- Required consumables – Varies widely according to technique used. Electrodes, flux, wire, and gas (especially argon and helium, often in combination with H_2, CO_2, or O_2) are all used in electric arc welding. Some techniques require only one of these four consumables; many use two. Stud welding demands special collars or ferrules, and 1-2 kg/m of metal also is needed (Houldcroft, 1977). Productivity varies with welding speed, current amplitude, and plate thickness.
- Production energy – Required voltage ranges from 10-70 V, current from 2-2000 A (Schey, 1977; Schwartz, 1979). Romans and Simons (1968) give a maximum value of 10,000 A for electroslag welding. A particularly useful quick survey of various electric arc techniques may be found in Lindberg (1977).
- Preparation steps – A variety of hoses, valves, wire, switches, a power supply and gun are needed to make a welding unit. The amount of preparation required may be extensive in some cases (e.g., securing and aligning pieces, plates to contain slag, etc.). Other techniques require relatively little preparation.
- Production environment – A pressurized welding environment is needed to use flux or slag processes.
- Automation/teleoperation potential – Many of these techniques are already automated in terrestrial applications.
- People roles – Other than design, none required.

- R&D required – Not a promising future line of inquiry.
- Qualitative Tukey Ratio – Moderately poor. Some of the consumables, especially gases, are comparatively difficult to obtain in quantity from lunar soil.

4E.1.2 Oxyfuel welding

Included among the oxyfuel gas-welding techniques are oxyacetylene, methylacetylene propadiene (MAPP), air-acetylene, oxyhydrogen, and pressure gas. The SMF suitability assessment produced the following results:

- Make other equipment – Need a basic joining process.
- Production rates – Estimates include 0.6 to 10 m/hr (2 to 30 ft/hr) (Romans and Simons, 1968), 1.6 to 5.4 mm^2/sec (Houldcroft, 1977), and 0.3 to 0.6 kg/hr (Schey, 1977).
- Required consumables – Main consumables are gases, especially acetylene and oxygen. Filler rods and flux may or may not be required (Houldcroft, 1977).
- Production energy – Romans and Simons (1968) claim that vertical welding at a rate of 2-5 m/hr requires 70-500 liters/hr of acetylene gas at STP.
- Preparation steps – Need gases in pressure tanks, a simple valve/regulator structure, gauges, hoses, torch and torch tip assemblies (Griffen et al., 1978). Surface preparation of workpieces requires a basic cleaning process. Jigs typically are used to hold the workpiece in the proper positions.
- Production environment – Gases necessitate a pressurized environment.
- Automation/teleoperation potential – Already automated in many terrestrial manufacturing applications (Phillips, 1963; Yankee, 1979).
- People roles – Could conceivably be used by astronauts to perform quick, portable repair welding operations in a pressurized environment.
- R&D required – Not a promising future line of inquiry.
- Qualitative Tukey Ratio – Very poor in the near-term due to heavy dependence on gases comprised of chemical elements having low lunar abundances (e.g., acetylene, MAPP, hydrogen).

4E.1.3 Resistance welding

Resistance techniques include spot, projection, seam, flash butt, upset, and percussion welding. (High-frequency resistance welding is discussed as an electronic welding technology.) The assessment follows:

- Make other equipment – Basic joining process is needed.
- Production rates – Ranges from 16 to 107 mm^2/sec are given by Houldcroft (1977) for resistance welding generally, and Romans and Simons (1968) estimate 1 to 4 m/min (36 to 144 in./min) for seam welding.
- Required consumables – Air (gas) or water are necessary to provide high pressures, and water is needed for cooling. Substitutes can probably be found among available nonterrestrial materials.
- Production energy – Considerable electrical energy is required, typically 1000 to 100,000 A at 2 to 20 V (Moore and Kibbey, 1965; Romans and Simons, 1968). Romans and Simons give a range of 1 to 140 W/hr per spot weld.
- Preparation steps – Modest resistance welding machines require very large power supplies. Pressure-producing cylinders for larger equipment are somewhat complex, and sophisticated timing devices are necessary. However, little preparation of materials is needed, perhaps the key reason why resistance welding is so popular on Earth (Moore and Kibbey, 1965).
- Production environment – Moore and Kibbey (1965) indicate that air must be supplied for operation of the electrodes, so a pressurized environment may be necessary.
- Automation/teleoperation potential – These techniques have already been largely automated on Earth.
- People roles – None required other than design.
- R&D required – Not a promising future line of inquiry.
- Qualitative Tukey Ratio – Moore and Kibbey (1965) note that resistance welding electrodes are subjected to 10,800 A/cm^2 at 410 MN/m^2 (70,000 $A/in.^2$ at 60,000 psi). It seems unlikely that lunar-abundant aluminum could even come close to replacing copper-bronze and copper-tungsten alloys used to make electrodes on Earth. Also, it is questionable whether aluminum could be incorporated in the massive high-current power transformers required. The Tukey Ratio appears quite poor in this case.

4E.1.4 Solid-state welding

Included within this category are ultrasonic, explosive, diffusion, friction, inertia, forge, vacuum (cold), and roll welding. The SMF assessment is as follows:

- Make other equipment – Need a basic joining process.
- Production rates – On thin materials, roller-seam ultrasonic welds can be produced at rates up to 10 m/min.

- Required consumables – Air, water, or other pressure-producing agents are needed for cold, friction, and roll welding. Explosion welding uses explosive sheets with TNT, ammonium nitrate, amatol, and others. Ultrasonic welding requires a transmission medium for sound waves.
- Production energy – Vacuum (cold) welding requires only a very light pressure. Ultrasonic welders are rated at up to 25 kW (Schwartz, 1979).
- Preparation steps – Materials to be cold welded under vacuum need only be appropriately positioned for application of modest pressure, though the exact preparation steps for a vacuum welding machine are unknown. Explosion welding involves placing an explosive sheet on the workpieces. Friction and inertia welding require a driving system, hydraulic cylinder, bearing, bearing enclosure, etc. Ultrasonic welding utilizes a rigid anvil, a welding tip consisting of a piezoelectric crystal and a transducer with horn, and a force-application mechanism. Parts alignment is a crucial step in all joining processes.
- Automation/teleoperation potential – Most, if not all, of these techniques should readily be automatable.
- People roles – None, other than original design.
- R&D required – Cold welding has the highest appeal as a simple joining process. A system of applying small pressures without accidentally contact welding the machine to the workpiece must be devised. One simple method is a vise made of insulated metal parts (teflon- or oxide-coated). More must be learned about cold-welding properties of various materials.
- Qualitative Tukey Ratio – Seems likely to be extremely good for cold welding. Closely related forms such as friction, inertia and roll welding should also exhibit satisfactory Tukey Ratios, since only small pressures need be applied. Forge and diffusion welding require heat as well (and hence, seem superfluous), but can probably exhibit favorable ratios with some modification. The ratio for ultrasonic welding appears relatively poor.

4E.1.5 Electronic welding

Electronic welding methods encompass the various forms of electron-beam, laser, induction, and high-frequency resistance welding. The following is the SMF suitability assessment:

- Make other equipment – A basic joining process is needed. (Note: A number of these techniques, particularly the laser, can be used for many other options.)
- Production rates – Lindberg (1977) cites a figure of 16 m/hr (50 ft/hr) for a high-power continuous-wave solid-state laser. The estimate by Schwartz (1979) is much higher: 50 to 80 m/hr (150 to 250 ft/hr). Electron-beam welders can produce up to 1800 small parts per hour in a partial vacuum (Schwartz, 1979) or up to 200 mm^2/sec of welding (Houldcroft, 1977). Induction welding production rates are given as 6.5 m/min (20 ft/min) of 20 cm (8 in.) pipe (Phillips, 1963) and 3.1 m/min (122 in./min) of tube welding for typical machines (Lindberg, 1977). High-frequency resistance methods can weld seams at 50 m/min (150 ft/min) with 60% efficiency (Schwartz, 1979).
- Required consumables – Flashlamps for solid-state lasers have a lifetime of 10^4 to 10^5 shots. Gas lasers may use a variety of gases including $CO_2/H_2/N_2$, argon, krypton, neon, xenon, and others. Electron-beam filaments last 2 to 1000 hr depending on filament type. High-frequency resistance welding contacts are good for roughly 6,000 to 130,000 m (50,000 to 400,000 ft) of welding before they must be replaced (Schwartz, 1979).
- Production energy – Lasers require up to 15 to 20 kW (Lindberg, 1977; Schwartz, 1979). Schwartz notes that gas lasers are inefficient (less than 0.1%) relative to solid-state lasers (up to 10% efficiency). Electron-beam welders draw 6 to 75 kW, with voltages in the 15 to 200 kV range. The American Welding Society (Phillips, 1963) estimates 1 to 600 kW output power for induction welding – as much as 1 MW may be needed in some cases. Energy requirements for high-frequency resistance welding are much lower than other resistance techniques due to increased resistivity at higher (400 kHz) frequencies (Lindberg, 1977; Schwartz, 1979). Schwartz claims that the most powerful high-frequency resistance welding machines in terrestrial use draw 150 kW, though many require only 1 to 50 kW.
- Preparation steps – A solid-state laser is comprised of a rod, laser cavity, precision-ground mirrors, flashlamp, cooling system, focusing optics, and power supply. In recent years ruby rods have been increasingly replaced by Nd:YAG rods (Schwartz, 1979). Flashlamps usually are xenon- or krypton-filled (Lindberg, 1977). Gas lasers do not need rods and flashlamps of such exotic composition, but instead require gas and a heat exchanger. Electron-beam welders need a sophisticated variant of the cathode-ray tube, a very high voltage power supply, and preferably a vacuum environment. Induction welding units are characterized by a large coil at low frequencies, a high-power oscillator circuit at high frequen-

cies, and a heavy-duty power supply and cooling system. High-frequency resistance welding differs from induction joining only in that its contacts are supplied at relatively low loads. Finally, workpieces require alignment. Electron-beam and laser techniques typically are reserved for small, shallower welds demanding very precise alignment. Induction welding usually is in conjunction with a pressure-producing machine.

- Production environment – Electron-beam welders work best in a vacuum. Gas lasers require an enclosed chamber to contain the gas. Otherwise electronic welding techniques appear fairly adaptable.
- Automation/teleoperation potential – Lindberg (1977) notes that both E-beam and laser welding techniques are easy to automate. Induction welding also has been automated to a considerable extent in terrestrial manufacturing.
- People roles – None required beyond the design phase.
- R&D required – Further developments in electron-beam and laser technologies are likely to be highly fruitful. Laser flashlamp lifetimes must be greatly increased.
- Qualitative Tukey Ratio – The Ratio is somewhat poor for solid-state lasers using present-day technologies, though the components are not too massive and so could be lifted from Earth with only modest penalty. With some possible substitutions the Ratios for other electronic welding options appear favorable. Some essential materials may be difficult to obtain in sufficiently large quantities (such as the carbon for CO_2 or inert gases in a gas laser).

4E.1.6 Brazing and soldering

Among the various brazing processes identified in this study are torch, induction, furnace, dip, resistance, infrared, and especially vacuum methods. Soldering includes iron, resistance, hot plate, oven, induction, dip, wave, and ultrasonic techniques. The space manufacturing suitability assessment follows:

- Make other equipment – Since brazing and soldering make weaker bonds than welding they are somewhat less universal in common use. On the other hand, some very dissimilar materials can be brazed but not welded.
- Production rates – No figures were given in any of the references reviewed. Wave soldering allows the processing of entire circuit boards (hundreds of components) in a few seconds.
- Required consumables – Filler metals or alloys and fluxes usually are required, though some processes are fluxless.
- Production energy – Highly variable. (See oxy-fuel gas welding for estimates on one common method.) The major difference between these techniques and welding with respect to production energy is that less heat is required.
- Preparation steps – Alignment jigs are needed to position workpieces to a fairly high degree of accuracy. Flux and heat are applied first, followed by filler material. Some fluxes and fillers are combined. Vacuum brazing requires filler only.
- Production environment – A pressurized environment is mandatory except for vacuum and fluxless brazing.
- Automation/teleoperation potential – These processes are not extremely complex. Furnace brazing and wave soldering are contemporary examples of automated or semiautomated systems.
- People roles – None except for design.
- R&D required – Fluxless brazing (e.g., of aluminum) and vacuum brazing appear fruitful research avenues worthy of further exploration.
- Qualitative Tukey Ratio – The Ratio is poor in most cases. The most commonly used brazing metals (fillers) are copper and copper/silver/aluminum alloys; solders typically are tin/lead mixtures. Most flux materials are not readily available from nonterrestrial sources. However, the Tukey Ratios for vacuum and fluxless brazing of aluminum, titanium, and a few other metals seem rather promising.

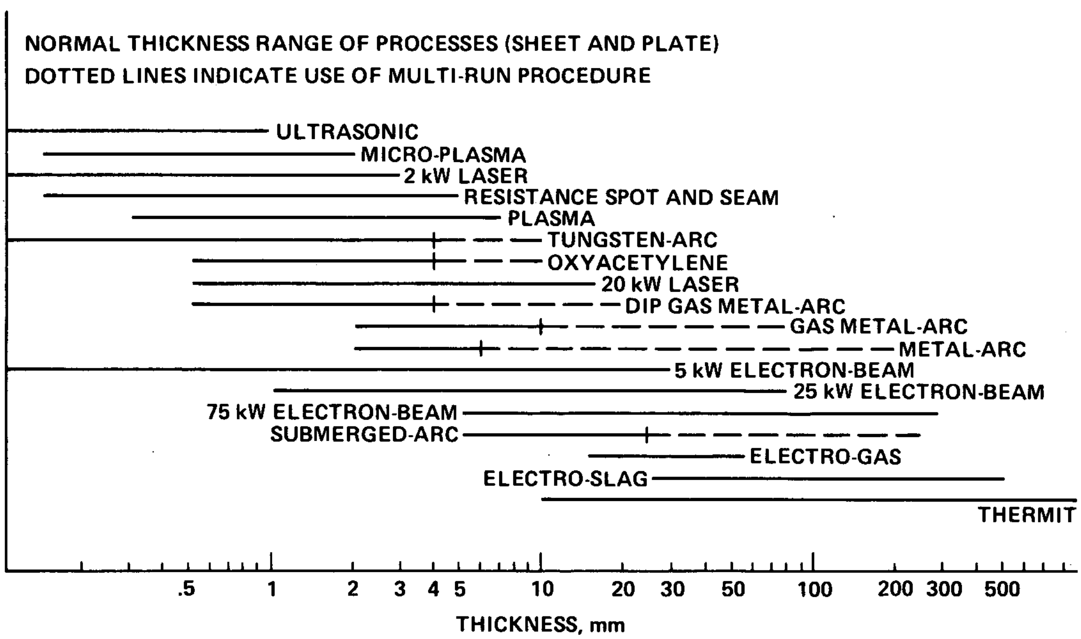
4E.2 Summary of Metal-Joining Options in Space Manufacturing

Perhaps the most significant conclusion to be drawn from the preceding analysis is that NASA is on the right track in its research and development efforts on space-qualifiable joining processes. Most promising are vacuum or cold-pressure welding in the solid-state category, the various electronic welding techniques (E-beam, laser, induction, and high-frequency resistance welding), and vacuum and fluxless brazing. NASA has already done some research on electron-beam and laser welding (including successful experiments in space) and vacuum brazing. Explosion welding may be useful if an explosive can be developed from lunar materials and the shock wave made to propagate in a vacuum environment. Friction welding might usefully be combined with vacuum welding (at lower pressures than required on Earth) to quickly remove protective coatings which inhibit undesired contact welding.

Of the most promising techniques, vacuum welding and vacuum brazing seem the simplest, the least energy-consuming, and exhibit the best Tukey Ratios. Vacuum brazing requires some heat to melt filler material, but probably bonds a greater variety of materials (e.g., refractory and reactive bare metals, ceramics, graphite, and composites) than vacuum welding methods. Electronic techniques offer poorer mass multiplication ratios, especially in the case of the laser. However, both E-beams and laser beams are extremely versatile – besides welding a very wide variety of materials, lasers can drill, cut, vapor deposit, heat treat, and alloy (Schwartz, 1979). They can cast and machine as well as weld, making them excellent candidates for the initial elements of a space manufacturing bootstrap operation. High-frequency resistance and induction welding can also join a wide variety of materials, and with high efficiency. Table 4.29 compares key characteristics of laser and electron-beam processes with those of two less-promising alternatives for space and lunar applications. It is apparent that both E-beam and laser techniques are competitive in most categories whether on Earth or in space. Equipment cost of the E-beam should be much lower in a vacuum environment, since the major expense in terrestrial applications is for the maintenance of proper vacuum.

Figure 4.29 provides a useful overview of welding capabilities for various material thicknesses. While this factor has not yet been discussed it is nonetheless important, since production speed diminishes nonlinearly with penetration depth. It is interesting to note that the combination of laser and E-beam technologies spans the entire range of usual material thicknesses. No direct data were available on the vacuum-welding technique, but this range conceivably could be quite large.

From the standpoint of automation in space, a final and most significant conclusion is that all joining processes of interest appear readily automatable. Joining should pose no insurmountable problems for space or lunar manufacturing facilities. General-purpose repair welding must probably be accomplished initially via teleoperation, as this activity requires a much higher degree of intelligence and adaptability.



Herbert W. Yankee, MANUFACTURING PROCESSES, ©1979, pp. 349, 528, 533. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.

Figure 4.29.– Thickness range of welding processes.

TABLE 4.29.– SIMPLIFIED QUALITATIVE COMPARISONS BETWEEN LASERS, E-BEAMS, AND TWO COMMON FORMS OF RESISTANCE AND ARC WELDING

Characteristic	Laser	E-beam	Resistance (spot)	Electric arc (gas tungsten)
Heat generation	Low	Moderate	Moderate-high	Very high
Weld quality	Excellent	Excellent	Good	Excellent
Weld speed	Moderate	High	Moderate	High
Initial costs	Moderate	High	Low	Low
Operating/maintenance costs	Low	Moderate	Low	Low
Tooling costs	Low	High	High	Moderate
Controllability	Very good	Good	Low	Fair
Ease of automation	Excellent	Good	Fair	Fair
Range of dissimilar materials which can be welded	Very wide	Wide	Narrow	Narrow

4E.3 References

DeGarmo, E. P.: Materials and Processes in Manufacturing. Macmillan, New York, 1979. Fifth Edition.

Griffin, I. H.; Roden, E. M.; and Briggs, C. W.: Welding Processes, Van Nostrand Reinhold Company, New York, 1978. Second edition.

Houldcroft, P. T.: Welding Process Technology. Cambridge University Press, Cambridge, England, 1977.

Lindberg, R. A.: Processes and Materials of Manufacture. Allyn and Bacon, Boston, 1977. Second Edition. 714 pp.

Moore, H. D.; and Kibbey, D. R.: Manufacturing Materials and Processes, Irwin, Homewood, Illinois, 1965.

Phillips, A. L., ed.: Welding Handbook: Welding Processes – Gas, Arc and Resistance. American Welding Society, New York, 1963. Fifth Edition.

Romans, D.; and Simons, E. N.: Welding Processes and Technology. Pitman and Sons, London, England, 1968.

Schey, J. A.: Introduction to Manufacturing Processes. McGraw-Hill, San Francisco, 1977.

Schwartz, M. M.: Metals Joining Manual. McGraw-Hill Book Co., San Francisco, 1979.

Yankee, H. W.: Manufacturing Processes. Prentice-Hall, Englewood Cliffs, New Jersey, 1979. 765 pp.

APPENDIX 4F

REVIEW OF ADHESIVES, FASTENERS, AND FITTING

There exist a number of alternatives to welding, brazing, and soldering which might be employed in space industry for joining metals and especially nonmetals. The most important of these are adhesives, metals fasteners, interlace fasteners for stitching or stapling, shrink fitting, and press fitting. Each has been considered for space applications in terms of manufacturing processes, required materials, possible compatible substitute techniques, and the degree of automation attainable. In addition, the unique impact of zero-g (more properly, "free-fall"), hard vacuum, and intense radiation is considered for each joining process examined.

4F.1 Glues and Other Nonmetallic Bonding Agents

Adhesives are used to fasten two surfaces together, usually producing a smooth bond. This joining technique involves glues, epoxies, or various plastic agents that bond by evaporation of a solvent or by curing a bonding agent with heat, pressure, or time. Historically, glues have produced relatively weak bonds. However, the recent use of plastic-based agents such as the new "super-glues" that self-cure with heat has allowed adhesion with a strength approaching that of the bonded materials themselves. As a result, gluing has replaced other joining methods in many applications – especially where the bond is not exposed to prolonged heat or weathering.

A large fraction of modern glues are carbon-based petrochemical derivatives. These can be used to bond almost any combination of surfaces, either by direct contact or by fastening both surfaces to a third as with adhesive tapes. Glues can serve as bonding agents in strong structural materials – one of the earliest, and still common, such use is the fabrication of plywood (a wood composite). Other related composites include fiberglass and various fiber-epoxies such as boron-epoxy and carbon-epoxy. Many of these materials make superior stress-bearing components.

Composite structures often are far less massive than comparable metal components and may be used in structural locations. Some of the early plans for beam construction on Shuttle flights call for carbon-epoxy materials. Composites may be the major use of glue/epoxy adhesives in space. For macroscopic bonding, alternatives such as welding, stapling, bradding, stitching, and other fasteners can replace adhesives if necessary. But although composites in theory can be replaced by metal parts it is far more likely that in space metal parts will give way to composites.

The space application of adhesives includes the following considerations:

- Zero-g – Although some adhesives must bond and cure under pressure, variations on clamping could compensate for the lack of gravity. Application of adhesives also should not demand gravity feed, although squirting and injection techniques have been perfected.
- Vacuum – Many resins and glues used on Earth are fairly volatile and deteriorate under vacuum. But some plastics, once cured, no longer are volatile and may continue to be used in vacuo. Silicate-based waxes and bonding epoxies employed in composites are just two examples of currently available vacuum-compatible adhesives.
- Radiation – Most hydrocarbon-based plastics weaken under the influence of infrared and higher-frequency electromagnetic radiation. These would not be suitable for exposed space use without shielding. More research is needed to develop radiation-resistant adhesives and bonding agents.

The application of glues to complex shapes already is automated in many industries, particularly fabric applications. Composite mixing and curing is now done by machines with a high level of reliability. Further automation of these processes should present no unusual difficulties.

4F.2 Metal Fasteners

Metal fasteners are of two kinds – those producing a permanent bond and those requiring either a releasable or a sliding bond. Screws, nuts and bolts, rivets, brads, retaining rings and clamps are examples from the first category. These are used for permanent fastening where stress loads preclude gluing but do not require welding or where the possibility exists of undoing the bond for some future purpose such as repair. Nonpermanent fasteners include quick-release couplers and clamps intended for removal at a specified time, and pins which allow relative movement of fastened parts. Pins are used where conditions of movement

are less rigidly constrained than when heavy bearing capability is required.

Metal fasteners must be strong to bear significant loads. In many cases they can be manufactured by powder metallurgical or casting techniques. Iron is a constituent of many types of metal fasteners, although titanium increasingly is coming into use in applications where strength must be balanced against light weight. In most applications where permanent bonding is required metal fasteners are replaceable by some form of welding or soldering. A major consideration here is whether the fabrication of welding rods and the process of welding is a more or less efficient use of available resources and energy than the fabrication and use of fasteners. For nonpermanent bonds there is not much choice except friction/pressure fittings and these run the risk of vacuum welding.

Both iron and titanium are in abundance on the Moon and each has received much attention as two extraterrestrial resources most likely to be investigated early for extraction and utilization. The manufacture of metal rivets from lunar or simulated lunar resources would be a worthwhile early materials processing experiment for an orbital laboratory. Space applications considerations include:

- Zero-g – Metal fasteners may be lighter in weight because loads may be far less than on the ground.
- Vacuum – Permanent bonds are largely unaffected by vacuum. Vacuum welding will promote tighter joining, a benefit in the case of permanent bonds but a definite hindrance if breakable or sliding bonds are desired. Very low vapor-pressure lubricants (e.g., graphite), surface poisoners, or careful choice of incompatible metals may help to eliminate this problem.
- Radiation – Some metal fastener materials may become more brittle with time in the presence of ionizing radiation.

The fastening of rivets and bolts already has been automated in some terrestrial applications. Extending the techniques of automation to space, and including screws and nuts, clamps and pins, seems to present no special problems.

4F.3 Interlace Fasteners – Stitching

Interlace fastener stitching is a joining process by which pieces of material are interwoven through holes in the parts to be joined. The bond is primarily frictional if the joined pieces are not rigid, primarily tensional if they are rigid. On Earth, mostly fabrics are stitched, though items such as tennis racquets and sieves also require a type of stitching in their manufacture. Stitching material usually has physical properties and adhesive characteristics similar to those of the materials joined. Parts to be fastened must have a series of holes through which the interlace passes. These holes may be native to the material, as in a fabric, or specially drilled, as in wood or metal sheets. (Terrestrial stitching is applied to some processes not immediately obvious, such as the knitting together of thin plywood sheets to form a mold for fiberglass.) The primary space-related utility of interlace fasteners is expected to be in the manufacture of EVA pressure suits. Designs such as the Space Activity Suit (Annis and Webb, 1971) rely on tension instead of atmospheric pressure to counterbalance internal hydrostatic forces using corset-like interlaces to join special fabrics. Stitching materials may be organic or synthetic fibers, glass fibers, or even metals.

The space environment places a few constraints on possible stitching materials, as discussed below:

- Zero-g – Except for holding parts in place during fastening, zero-g presents no special hardships as regards stitching. Indeed, one possible indirect advantage is apparent: The lack of gravity permits finer threads to be pulled from molten material than is possible on Earth, because of the absence of both the catenary effect and the necessity to support threads against their own weight in zero-g.
- Vacuum – Vacuum poses two problems for stitching. First, it is nearly impossible to make an airtight interlace without sealant. Second, most interlace materials are hydrocarbon-based, hence are volatile and easily deteriorate in a vacuum. Fortunately, nonvolatile stitches made of metals or basalt glasses can be found, and there do exist sealants effective in closing small holes against the loss of atmosphere.
- Radiation – The deterioration of interlacing materials caused by hard radiation is a serious problem for hydrocarbon-based stitches, but replacement of these by glass or metal substitutes may eliminate the problem. Radiation-proof coatings should be vigorously pursued as an important topic in space manufacturing research.

The availability of stitching materials is strongly constrained. Hard vacuum and radiation in space render hydrocarbon-based threads infeasible due to volatility and molecular deterioration, and hydrocarbons are also relatively rare in near-lunar space. On the other hand, glass and metal interlaces do not suffer from these problems and are easily accessible on the Moon.

Stitching most efficiently must be done by machines in most applications, and these processes are already largely perfected. Interlacing beam ends do not seem to present any special problems for automation. As for alternatives, gluing can replace stitching in some applications such as the joining of fabrics. Gluing has the advantage of airtightness but the common disadvantage of lesser strength. Tack welding can replace interlacing of metals in many jobs, but a penalty must be paid in higher energy consumption.

4F.4 Interlace Fasteners – Stapling

Stapling is similar to stitching except that staple rigidity is important to the load. The staple passes through holes in the material to be fastened and is bent to prevent loaded matter from easily slipping out. Staples almost invariably are made of metal since they must be strong, cheap, and bendable yet fairly rigid. The relative ease and speed of stapling over stitching has led to its increasing use in the fabrics industry, though few large commercial products have direct space applications. Since staples provide a low-cost, low-energy, rapid-fastening capability, they may play a role in various forms of space construction. Beams of thin aluminum or other metals could be stapled rather than welded if desired. Staple bonds are relatively weak but zero-g permits their use in space on flimsy structural members impossible in terrestrial construction.

Stapling is usually done by machine on Earth and this is unlikely to change in space. As for alternatives, if bonded items are metallic, tack welding often can replace stapling. Energy costs increase with bond strength and tear resistance. If bonded items are nonmetallic then welding methods cannot be used, but glues may replace staples if necessary.

4F.5 Shrink and Press Fitting

Shrink fitting is accomplished by heating a part so that a hole in it expands, after which another piece may be fitted, usually under pressure, into that hole. The outer piece then contracts as it cools, creating a tight seal. Some sinter-like bonds may form, but shrink fitting works primarily by friction bonding. It requires thermal energy which press fitting (see below) does not, but less force is needed to achieve the final bond. If the material to be shrink-fitted is metallic, heating may be accomplished by induction.

Press fitting is similar to shrink fitting except that parts are not heated and higher pressures are necessary. Press fitting requires less energy but the bond is weaker. Also, if bonded material has a buckling problem press fitting is not suitable as a joining technique.

Beams made of rigid materials can be joined by fitting, as can many other parts. Gears routinely are attached to shafts by this method. Fitting can produce bonds strong enough for many applications. The great simplicity of these processes strongly urges their automation.

Usually metals are shrink and press fitted, and these materials are relatively abundant in nonterrestrial resources. The energy and materials efficiencies of these techniques make them prime candidates for space applications. Both are preferred to welding where loads are light. Vacuum welding may serve to strengthen bonds. Flames are hard to produce in a vacuum so shrink fitting probably will be accomplished by induction heating if the materials are metallic.

4F.6 References

Annis, James F.; and Webb, P.: Development of a Space Activity Suit. NASA CR-1892, November 1971.

CHAPTER 5

REPLICATING SYSTEMS CONCEPTS: SELF-REPLICATING LUNAR FACTORY AND DEMONSTRATION

5.1 Introduction

As the cost of fossil-fuel energy continues to escalate and supplies of readily accessible high-grade ores and minerals gradually become depleted, the utilization of nonterrestrial sources of energy and materials and the development of a nonterrestrial industrial capacity become increasingly desirable. The Moon offers plentiful supplies of important minerals and has a number of advantages for manufacturing which make it an attractive candidate factory site compared to Earth. Given the expense and danger associated with the use of human workers in such a remote location, the production environment of a lunar manufacturing facility should be automated to the highest degree feasible. The facility ought also to be flexible, so that its product stream is easily modified by remote control and requires a minimum of human tending. However, sooner or later the factory must exhaust local mineral resources and fall into disrepair or become obsolete or unsuitable for changing human requirements. This will necessitate either replacement or overhaul, again requiring the presence of human construction workers with the associated high costs and physical hazards of such work.

The Replicating Systems Concepts Team proposes that this cycle of repeated construction may possibly be largely eliminated by designing the factory as an automated, multiproduct, remotely controlled, reprogrammable Lunar Manufacturing Facility (LMF) capable of constructing duplicates of itself which would themselves be capable of further replication. Successive new systems need not be exact copies of the original, but could, by remote design and control, be improved, reorganized, or enlarged so as to reflect changing human requirements. A few of the benefits of a replicative growing lunar manufacturing facility (discussed at greater length in secs. 5.4 and 5.5) include:

(1) The process of LMF design will lead to the development of highly sophisticated automated processing and assembly technologies. These could be used on Earth to further enhance human productivity and could lead to the emergence of novel forms of large-scale industrial organization and control.

(2) The self-replicating LMF can augment global industrial production without adding to the burden on Earth's limited energy and natural resources.

(3) An autonomous, growing LMF could, unaided, construct additional production machinery, thus increasing its own output capacity. By replicating, it enlarges these capabilities at an increasing rate since new production machinery as well as machines to make new machines can be constructed.

(4) The initial LMF may be viewed as the first step in a demonstration-development scenario leading to an indefinite process of automated exploration and utilization of nonterrestrial resources. (See fig. 5.1.) Replicating factories should be able to achieve a very general manufacturing

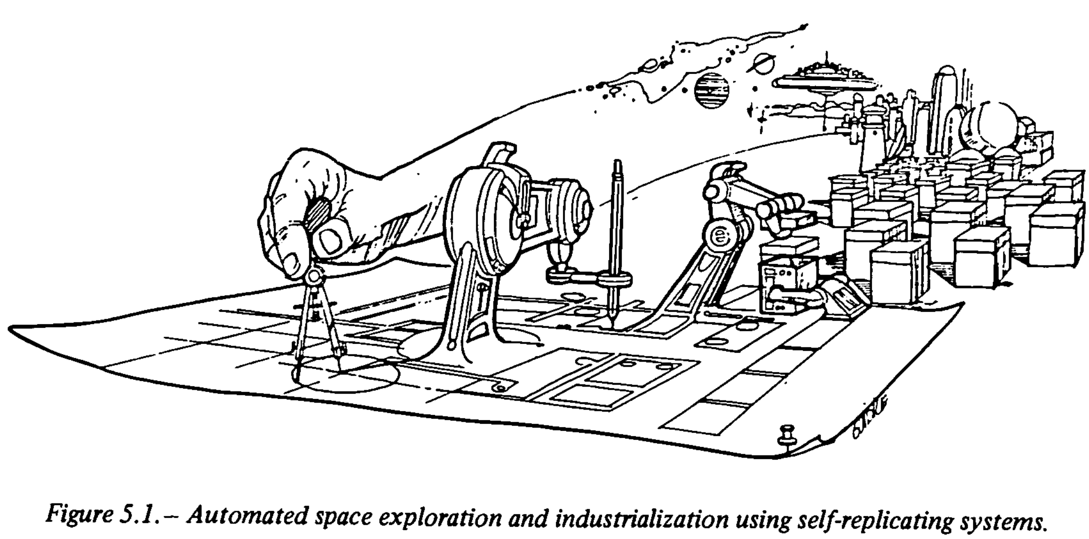


Figure 5.1.– Automated space exploration and industrialization using self-replicating systems.

capability including such products as space probes, planetary landers, and transportable "seed" factories for siting on the surfaces of other worlds. A major benefit of replicating systems is that they will permit extensive exploration and utilization of space without straining Earth's resources.

5.1.1 Summary of Chapter Contents

The history of the concept of machine replication is reviewed in section 5.2. This theoretical background is largely a consideration of the work of John von Neumann – in particular, his kinematic and cellular models of automata self-reproduction. Post-von Neumann research is reviewed next, noting particularly the established theoretical capabilities of machines in the realm of general construction, inspection, and repair strategies. Such strategies may prove useful, even vital, to the successful design, realization, and operation of actual replicating systems.

Section 5.3 deals with the engineering feasibility of the concept of self-replicating systems (SRS). An attempt is made to confront two important general problems in creating a lunar replicating factory:

- Given that in theory, machines can construct duplicates of themselves, how might systems designers and engineers identify all functions which must be carried out to achieve machine replication and also develop the technological means by which to implement these functions?
- Given the constraints obtaining in the lunar environment, particularly in terms of the inventory of known kinds and quantities of naturally occurring raw materials and the existing repertoire of materials processing technologies, can all machine functions required both for production and for replication and growth be implemented?

To attack the first of these problems – identification of necessary functions for practical machine replication – the team proposes a specific phased demonstration-development scenario, described in section 5.3. For the second problem – establishing that machine replication can feasibly take place in the actual lunar environment – a strawman mission concept was employed. In this scenario, a 100-ton initial "seed" factory is planted on the Moon with access only to local resources and established materials processing techniques. The initial system should be able to successfully develop into an expanded machine system capable of conducting all functions necessary for autonomous replication, growth, and automated production and manufacturing.

The problem of "closure" is also considered at length in section 5.3. The issue of closure is whether autonomous manufacturing and construction systems can make available to themselves all of the materials, parts, and assembly techniques required for all internal operations. An iterative strategy is presented for detecting and eliminating closure gaps, and for optimizing the resulting augmented system.

Section 5.4 deals with possible applications of the SRS concept. Applications of replication technology include enormous gains in terrestrial industrial productivity (automation and computer-aided design and manufacturing), utilization of Solar System resources, orbital and planetary opportunities, and the possibility of interstellar exploration on a grand scale. Indefinitely large masses can be organized in extraterrestrial environments using self-replicating systems.

Section 5.5 deals with just a few of the many implications of SRS. The advantages of space-based replicative manufacturing are considered, together with possible political, social, economic, cultural, and psychological consequences of the proposed SRS development program.

Section 5.6 sets forth in some detail how NASA can take action at once toward the achievement of the ultimate goal of establishing a replicating manufacturing facility. Suggested statements of work (SOWs) and a listing of institutions that might undertake the tasks outlined in the work statements are included. A series of specific conclusions and recommendations generated by the Replicating Systems Concepts Team are presented in section 5.7.

5.2 Theoretical Background

The notion of a machine reproducing itself has great intrinsic interest and invariably elicits a considerable range of responses – some directed toward proving the impossibility of the process, others claiming that it can be carried out, but almost all of them indicating an unwillingness to subject the question to a thorough examination. In discussing self-replication by automata it is essential to establish early rather important ground rules for the discussion. According to Kemeny (1955), "If [by 'reproduction'] we mean the creation of an object like the original out of nothing, then no machine can reproduce – but neither can a human being....The characteristic feature of the reproduction of life is that the living organism can create a new organism like itself out of inert matter surrounding it."

Often it is asserted that only biological organisms can reproduce themselves. Thus, by definition, machines cannot carry out the process. On the other hand, others argue that all living organisms are machines and thus the proof of machine reproduction is the biosphere of Earth. Also, sometimes it is claimed that although machines can produce other machines, they can only produce machines less complex than themselves. This "necessary degeneracy" of the machine construction process implies that a machine can never make a machine as good as itself. An automated assembly line can make an automobile, it is said, but no number of automobiles will ever be able to construct an assembly line.

Another common argument is that for a machine to make a duplicate copy it must employ a description of itself. This description, being a part of the original machine, must itself be described and contained within the original machine, and so on, until it is apparent we are forced into

an infinite regress. A variant of this is the contention that a machine not possessing such a description of itself would have to use itself for a description, thus must have the means to perceive itself to obtain the description. But then what about the part of the machine that does the perceiving? It cannot perceive itself, hence could never complete the inspection needed to acquire a complete description. (A simple counter is that the original machine might possess multiple perceiving organs, so that the perceiving could be shared.) Yet another related objection is that for the process to be carried out, the machine must come to "comprehend" itself – at which point it is said to be well known that "the part cannot possibly comprehend the whole." These disputations suggest that there is a very deep-seated resistance to the notion of machines reproducing themselves, as well as an admittedly strong fascination with the concept.

The Hungarian-American mathematician John von Neumann (1966), who first seriously came to grips with the problem of machine reproduction, once noted that it would be easy to make the whole problem go away. One could, for example, make the elementary parts of which the offspring machine was to be composed so complex as to render the problem of replication trivial. In one example of this considered by the team, a robot required only to insert a fuse in another similar robot to make a duplicate of itself would find "reproduction" very simple (see sec. 5.2.3). As von Neumann also pointed out, it is equally useless to go to the other extreme and try to account for the placement of every atomic particle in the system – one would quickly become mired in incomprehensible detail. Even most lifeforms do not have DNA-encoded instructions for reproduction to this fantastic level of detail – their descriptions are largely at the molecular level.

As will be demonstrated presently, although reproduction may be transparently trivialized or intractably complexified, there appear to be no fundamental inconsistencies or insoluble paradoxes associated with the concept of self-replicating machines.

5.2.1 Von Neumann's Contributions and Subsequent Research

John von Neumann began studying automata replication because he was interested in very complex machines and their behaviors. The early history of the theory of reproducing machines is basically the history of von Neumann's thinking on the matter, and this is reviewed below.

Von Neumann had a tremendous range of interests – he contributed to the logical foundations of quantum theory, was the co-inventor of the theory of games, and he worked on the Manhattan Project (contributing to the design of the implosion mechanism for the plutonium bomb). It is believed that his participation in the Manhattan Project and the tremendous volume of calculations necessary for bomb design led him into automatic computing. Hearing of the ENIAC computer project at the Moore School of Electrical Engineering at the University of Pennsylvania, von Neumann was fascinated by the potential of a computer very much faster than any of the devices that had previously been produced. In the early 1940s there existed only simple relay machines and analog devices such as the differential analyzer. But the new electronic machines that interested von Neumann promised to be perhaps millions of times faster than relay machines.

So von Neumann immersed himself in the ENIAC project, the first electronic computer program where some actual useful computing was produced. Late in 1945 and early 1946, the first problems that were put on ENIAC are believed to have been calculations involving the feasibility of a hydrogen bomb. Von Neumann, although he remained very much interested in nuclear energy and was appointed a member of the Atomic Energy Commission, became fascinated with the idea of large and complex computing machines. He devised the organization employed today in almost all general purpose computational machines – the so-called von Neumann concept of serial processing stored-program or the "von Neumann machine." After that work was completed he began thinking seriously about the problems of extremely large machines – their reliability, programming, design, how to understand what they do – and he became involved with the many possible analogies to the complex behavior of living systems.

Von Neumann set for himself the goal of showing what the logical organization of a self-reproducing machine might be. He had in mind a full range of self-replicating machine models which he intended to explore, including the (a) kinematic machine, (b) cellular machine, (c) neuron-type machine, (d) continuous machine, and (e) probabilistic machine. As it turned out, he ultimately was only able to produce a very informal description of the kinematic machine. Although he wrote a great deal on the cellular machine, his *magnum opus* on the subject was left in the form of unfinished notes at the time of his death. Almost no work was done on the other three kinds of self-reproducing machines. For this reason, only the postulated workings of the kinematic and cellular machines are presented below, with brief comments on the other three types. For an additional review of these two models of reproduction, see Burks (1970).

In dealing with machines that could reproduce, von Neumann concluded that the following characteristics and capabilities should be demonstrable for each:

(1) Logical universality – the ability to function as a general-purpose computing machine able to simulate a universal Turing machine (Turing, 1936). This was necessary because SRS must be able to read instructions to carry out complex computations.

(2) Construction capability – to self-replicate, a machine must be capable of manipulating information,

energy, and materials of the same sort of which it itself is composed.

(3) Constructional universality – in parallel to logical universality, constructional universality implies the ability to manufacture any of the finitely sized machines which can be formed from specific kinds of parts, given a finite number of different kinds of parts but an indefinitely large supply of parts of each kind.

(4) Self-reproduction – follows immediately from the above, since the universal constructor must be constructable from the set of manufacturable parts. If the original machine is made of these parts, and it is a constructable machine, and the universal constructor is given a description of itself, it ought to be able to make more copies of itself.

Von Neumann formally demonstrated that his cellular model of reproduction possessed these four properties.

Not much was done on a fifth property also believed to be important – evolution – though there have been some more recent results in this area. If one has a machine, and it makes a machine, which then itself makes a machine, is there any proof that the line of machines can become successively "better" in some fashion – for instance more efficient, or able to do more things? Could they evolve to higher and higher forms? This problem raises issues in learning, adaptation, and so forth, and was left largely untouched by von Neumann.

The kinematic machine. The kinetic machine is the one people hear about most often in connection with von Neumann's work on self-reproducing machines, probably because it received the earliest attention and publicity. John Kemeny (1955) produced a paper for the popular publication *Scientific American* detailing this model, and a further description appeared in a paper by von Neumann (1951).

The notion of kinematic machine self-reproduction was dealt with by von Neumann only informally. The mathematician envisioned a machine residing in a "sea" of spare parts. The machine has a memory tape which instructs it to go through certain mechanical procedures. Using a manipulative appendage and the ability to move around in its environment, the device can assimilate and connect parts. The tape-program first instructs the machine to reach out and pick up a part, then to go through an identification routine to determine whether the part selected is or is not the specific one called for by the instruction tape. If not, the component is thrown back into the "sea" and another is withdrawn for similar testing, and so on, until the correct one is found. Having identified a required part the device searches in like manner for the next, then joins the two together in accordance with instructions.

The machine continues following the instructions to make something, without really understanding what it is doing. When it finishes it has produced a physical duplicate of itself. Still, the second machine does not yet have any instructions so the parent machine copies its own memory tape onto the blank of its offspring. The last instruction on the parent machine's tape is to activate the tape of its progeny.

Von Neumann's logical organization for a kinematic machine is not the only one possible, but probably is the simplest way to achieve machine self-replication. In its logic it is very close to the way living organisms seem to reproduce themselves (Dyson, 1979). One conceptual problem with the model is that the parts involved are supplied free to the machine, and those parts are of a relatively high order. The machine dwells in a universe which supplies precisely the sorts of things it needs as a kinematic device to make a duplicate of itself. This raises the issue of closure, a problem which is discussed and conceptually resolved in section 5.3.

The cellular model. Von Neumann evidently was dissatisfied with his original kinematic model because of its seemingly mathematical inelegance. This model of machine self-reproduction, while qualitatively sound, appeared not easily susceptible to mathematically rigorous treatment and so might not serve to convince a determined skeptic.

Stan Ulam, a Polish-American mathematician who had also worked on the Manhattan Project, suggested to von Neumann that the notion of a self-reproducing machine would be amenable to rigorous treatment if it could be described in a "cell space" format – a geometrical grid or tessellation, regular in all dimensions. Within each cell of this system resides a finite state automaton. These cell automata can only be affected by certain of their neighbors, and only in very specific ways. In the model von Neumann finally conceived, a checkerboard system is employed with an identical finite state automaton in each square (fig. 5.2). In this system, as it evolved with subsequent research, the cell-automata can be in one of 29 possible different states (fig. 5.3). Each automaton can communicate with its four cardinal direction neighbors. The state of a cell-automaton is determined by its own state and by the states of its cardinal direction neighbors.

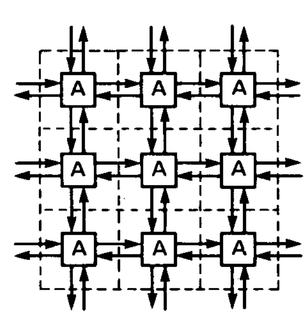


Figure 5.2.– Finite state automation cellular space.

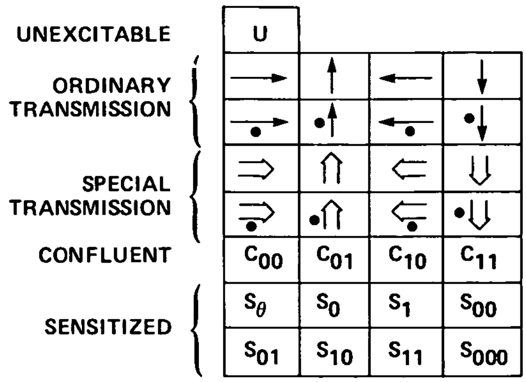


Figure 5.3.– Twenty-nine states of von Neumann's cellular automata.

At the beginning of operation, all but a finite number of the cell automata are in a "U" or "unexcitable" state. If a given cell is in the "U" state, and all its neighbors also are in the "U" state, then at the next moment of time, the given cell remains in the "U" state. Thus the "U" states can be viewed as representing undifferentiated, passive underlying substrate. Their passivity implies that they may in some cases serve as "insulation" surrounding more active cells in the system.

Then there are "ordinary transmission" cell states. These are states which direct their activity in each of the four cardinal directions. Each of these may be in an excited or quiescent mode, so there is a total of eight different kinds of ordinary transmission states. In addition, there are eight "special transmission states," similar to the ordinary states in that they also point in each of the cardinal directions and can be in excited or quiescent modes. The two basic kinds of transmission states – ordinary and special – differ in that the primary intended role of ordinary transmission states is the routing of informational signals, whereas the primary role of special states is to inject transforming signals into cell locations and thereby convert "U" cells into active elements (or, if need be, convert active elements back into "U" cells).

The system also has four "confluent" states. They are activated if they receive signals from all cells in their neighborhood which are directed toward them. If activation occurs, then after *two* moments of time they emit signals outward toward any cell in their neighborhood which does not have a transmission directed toward it. Thus, confluent cells can serve as "and" gates, and as wire branching elements. Since they do not emit their output until two moments of time have elapsed, the confluent cells can also be employed to create time delays in the transmission of signals. The eight remaining cell states of the 29 originally employed by von Neumann are of less importance. These are temporary cell states which arise only as the operational states are being created from "U" cells.

Von Neumann first showed how to design a general purpose computing machine in his cell space system. He did this by showing the design of various basic organs – "pulsers" to emit any desired finite train of pulses upon activation, "periodic pulsers" to emit repeated trains of desired pulses after activation until signaled to stop, "decoders" to detect the presence of certain patterns of pulses, and the like. Using these organs, von Neumann developed a design for the control portion of a computing machine in one region of the cell space. He then showed how to organize an adjacent but indefinitely extendable portion of the cell space into a memory or information storage unit, which could be accessed by the control unit.

For the process of construction, von Neumann designed a construction unit, which, taking instructions from the memory unit, could send out a constructing arm (by creating an active pathway of transmission cells into a region of "U" cells) and at the end of the arm, convert "U" cells to the cell types specified in memory (see fig. 5.4). He showed that this constructor could create any pattern of passive cells whatsoever. Thus, he had designed with mathematical rigor a universal constructor, relative to all possible passive configurations of cells in the cell space.

Since the parent machine itself can be created in passive form, it can make a duplicate of itself by the following process. The parent machine is supplied initially with instructions to make a duplicate of its control, construction and memory units (the memory unit initially is empty). After it completes this major construction phase, the instructions call for the parent machine to make a copy of the instructions in its memory and to feed into the memory unit of the newly constructed machine. Then the parent machine activates the heretofore passive offspring machine, and withdraws the constructing arm. At that moment the offspring is a duplicate, in all respects, of the parent at the time the original machine commenced its reproductive activities.

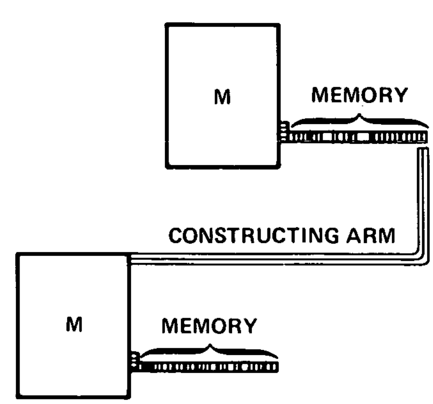


Figure 5.4.– Universal construction in the cellular model of machine self-reproduction.

Critique of the cellular model. Although the 29-state von Neumann cellular array system permits a more elegant mathematical approach to the problem of machine construction and self-reproduction, it is more difficult to envision an actual useful physical implementation of the process (compared, say, to the kinematic model of replication). The entire cell space enterprise proceeds in a highly constrained artificial environment, one which is very special despite some features relating in a general way to the natural world. For example, the movement of objects in space, a ubiquitous and familiar phenomenon in the real world, becomes a complex process of deletion of cell states at one location and re-creation of these states at some other location.

There is also an assumption of synchronous behavior throughout the system. All cells, no matter how distant, are subject to change of state at the same instant, a property which would be difficult to implement in any practical large cell space. Indeed, the requirement of a source of clocking pulses violates the array symmetry which makes the cell space notion an attractive object for mathematical treatment.

It is also very difficult to design machines of interest which can be embedded in the cell array format. To make design and embedding easier, a higher-level machine design language would have to be created. It is likely that, rather than undertake that task, one would first redesign the underlying cell space properties to rid the system of the deficiencies already noted.

For instance, one might wish to introduce a new primitive cell state in the system to permit signals to cross without interference. A "wire-crossing" organ can be devised using only the original von Neumann primitive cell types, but this introduces an unnecessary complexity into the machine design process since the organ contains initially active cell states whose creation involves considerable extra care to avoid the propagation of spurious signals. This extra care is especially critical because the cell system, as von Neumann originally constituted it, is highly susceptible to signal errors. (He undoubtedly intended his probabilistic machine model to mitigate this sensitivity and fragility.)

The cell space system has very limited capacity to detect the states of cells. It has some capacity to detect states, for this is required in the operation of the memory unit. But a machine cannot analyze an arbitrary encountered cell to determine what state it is in, thus cannot "read" the states of an encountered machine. This inability severely restricts the capacity of cell-space machines to repair other machines or to attempt self-repair. Such limitations also are evident in the construction process, where the constructing machine must assume that the region in which a new machine is to be created consists entirely of elementary quiescent cells. Should this not be the case, there is no systematic and complete way to detect it. A machine can send destruction signals into cells to reduce them to the quiescent form. Unfortunately, in some cases one must know the state of the cell ahead of time in order to determine what destructive signal must be sent to destroy it.

Finally, all machines that can be produced in von Neumann's cell space system are essentially information transactional devices. Even construction is, in this context, a form of information processing. Physical construction and material transformations can possibly be viewed as informational processes but, in a practical sense, the cell-space notion is far from providing a readily useful paradigm of actual manipulation and transformation of physical materials.

Von Neumann's other self-reproducing machine concepts. In addition to his kinematic and cellular models, von Neumann planned to examine three other models of self-reproducing machines. These were to be a neuronal or "excitation-threshold-fatigue" model, a continuous model, and a probabilistic model. Von Neumann is not known to have left any completed work whatsoever on these models at the time of his death, so his intentions are almost entirely a matter of conjecture.

Following Burks' speculations on this matter (von Neumann, 1966), we can guess that von Neumann's neuronal system might have been a version of the cell-space model in which the individual cell automata in the space were to be constructed of neuron-like elements. This would have been a rather straightforward process, as it is well known that idealized neurons of the McCulloch-Pitts (1943) variety can be employed to implement the kinds of logical gatings and delays called for in the 29-state cell automaton system. The reason for employing neuron-like elements seems mainly an attempt to recast the model in a more "biological" vocabulary.

Von Neumann's postulated continuous model might have been an attempt to comprehend machine reproduction in an even more biological format. The usual mathematical tools for handling actual neuron activity are differential equations expressing the electrochemical flows through and along neuron soma and axons. Thus the actions of cell automata (implemented with neurons) could be expressed by sets of differential equations. In this way the more highly developed tools of mathematical analysis might be employed in representing the behavior of the machine system, in contrast to the use of combinatorics which von Neumann himself characterized as one of the most intractable of mathematical specialties.

Finally, in his proposed probabilistic model von Neumann perhaps intended to consider using whole congeries of neuron-like elements in implementing the behaviors of what in the neuronal model could be carried out by single neurons. By employing redundancy techniques similar to those described in his classic paper on reliability, von Neumann (1956) may finally have hoped to

design a reliable, biologically oriented, self-reproducing machine characterizable by differential equations. We can only guess.

Alternative cell array systems. Work on cell-space automata systems in the period following von Neumann's contributions has taken several research directions. The underlying cell-space notion of a homogeneous medium with a local transition function that determines global properties has been employed in numerous modeling and simulation projects. For example, weather simulations use the idea of connected cells, the changes of each cell state described by a set of differential equations. Studies of the flow of excitation in heart tissue, the dispersal of medicinal drugs, and pattern recognition all have employed the cell-space concept. Cell spaces also have been investigated as abstract mathematical objects where, for instance, one tries to determine whether from every mathematical pattern all other patterns can be attained, and whether there are some patterns not attainable at all by means of the transition function, and various other specialized questions.

Some work in cellular automata has attempted to carry forth the von Neumann program of machine construction and self-reproduction. For instance, Codd (1968) recapitulated the von Neumann results in a simpler cell space requiring only 8 states rather than 29. This produced a machine design recognizably closer to that of present-day computing machines. Myhill (1970), trying to mitigate the artificiality of the indefinitely extended pre-existing cell space, designed a system in which componentry was drawn into a cell-grid system and was then employed as machine constituents somewhat as biological cell constituents might be drawn through a membrane to be used at an intracellular work site. Arbib (1966), attempting to make the movement of cell machines a less cumbersome matter, designed a cell-space system in which cells and blocks of cells might be joined together by a "welding" operation, thus becoming "co-moving" configurations.

Smith (1970) and Banks (1970) introduced additional simplifications to the cell-space notion, showing that the von Neumann program could be recapitulated in underlying cell spaces of an extremely elementary sort. Indeed, the so-called "Game of Life" designed by Conway (Gardner, 1971) is a cell-space system which, despite its very simple transition rules, has been claimed to be capable of expressing both universal computation and construction. (The game involves a checkerboard cell array with cells in one of two states, "0" or "1." A point whose state is "0" will change to state "1" if exactly three of its eight neighbors are in state "1." A point whose state is "1" will remain in that state if two or three of its neighbors are also in state "1." In all other cases, the state becomes or remains "0.")

Later research on self-reproducing automata. By the late 1960s, the original von Neumann program of machine construction and reproduction had been largly abandoned, although investigation of cell-space systems as abstract mathematical entities or as vehicles for "spatial" modeling and simulation has persisted. Indeed, research in the latter field has been especially vigorous and prolific – one recent author lists over 100 references for cell-space imaging applications (Preston et al., 1979).

Von Neumann's kinematic machine construction system appears to have had no intellectual progeny whatsoever. This is somewhat misleading, since practical application of computers to manufacturing and the persistent human interest in and investigation of robot mechanisms have, without explicit connection to von Neumann's earlier work, prepared the ground for a possible implementation of a hybrid computer/kinematic model of machine construction and reproduction.

The theoretical work of this later period, explicitly derived from von Neumann's research effort, has focused mainly on the molecular biological analogies that can be drawn. For example, in a series of papers Laing (1975, 1976, 1977, 1978, 1979) employs a hybrid cellular-kinematic model of machine construction and shows that neither existing natural nor artificial machines need be bound to follow the "classical" reproductive paradigm. In the classical paradigm, a program (DNA in living systems) is first interpreted to construct a machine (protein synthesis in lifeforms) and then is read a second time to make a copy of the program for insertion into the newly constructed duplicate machine (DNA replication in living cells). The principal contribution of Laing is to suggest reproductive strategies other than direct analogues to the known biological process. In this new conception, a machine is able to identify all of the components of which machine systems consist (not merely a subset as in the von Neumann cell system) and can access all of an existing machine structure without requiring dismantling of the system (as would be required in the von Neumann model).

Once this and other similar advanced concepts are brought to bear on the problems of machine reproduction, many alternative reproduction strategies become immediately apparent. A selected few of these are reviewed in the following section.

5.2.2 Alternative Replication Strategies

A number of alternative automata reproduction strategies have been suggested in the decades following the completion of von Neumann's work. Major strides have been made in the scientific understanding of the processes of biological reproduction at the molecular or biochemical level. Recent research has demonstrated the theoretical

possibility of inferring structure and achieving self-replication without first possessing a complete self-description. This suggests an enormous range of new machine capabilities which possibly may be technologically exploited in the future, according to specific rules and multiplication strategies for optimal deployment.

Biological reproduction. Biological reproduction is thought to obey the following underlying logical paradigm. The basic genetic program (encoded in the genetic DNA) is employed to make a copy of the same information in a slightly different medium (RNA). This modified form of the genetic program is transported to a work site within the cell where, with the aid of cellular enzymatic machinery, the RNA is interpreted as coding for amino acid strings (proteins). The protein produced plays two major roles: (1) it constitutes the basic structural material of living organisms, and (2) certain smaller and variably active proteins (enzymes) control the metabolic, interpretive, and constructive actions of the system.

When the genetic code embodied in the RNA has been read and acted upon, the machinery construction phase is complete. The cell must then undertake the copying of original genetic material (the DNA) to provide offspring organisms with the necessary instructions. This copying process is the well-known DNA replication phase, in which DNA (in most cases a twisted pair of complementary DNA molecules) untwists to permit new nucleotides to match with existing separated strands to form *two* twisted pairs of DNA. Reproduction is completed when the newly produced and original organism machineries are divided up, one DNA program remaining with each.

This highly simplified description of biological reproduction is offered only to illustrate the underlying logical strategies: (1) follow instructions to make machinery, (2) copy the instructions, (3) divide the machinery, providing a sufficient set in each half, (4) assign a set of instructions to each half, and (5) complete the physical separation.

Von Neumann's automata reproduction. Von Neumann's automata reproductive process closely mirrors the biological one. In the original model, instructions exist in two copies. One of the copies is read and acted upon to construct another machine, *sans* instructions. The second copy is then read and copied twice, and this double copy is inserted into the passive constructed offspring machine which is then turned on and released, thus completing the act of reproduction.

There is no logical necessity for having two sets of identical instructions. Von Neumann employed two copies of the instructions because it eliminated the criticism that the instructions might, in the first (construction) phase, become corrupted and so not be able to transmit a true version for the use of offspring machine. Also von Neumann feared that there might seem to be a paradox in the program acting upon itself to make a copy of itself. There are, however, ways by which a program can successfully be made to make a copy of itself, and indeed many such programs, though exceedingly simple, have already been written (Burger, Brill, and Machi, 1980; Hay, 1980). Another solution is to provide the machine proper with an automatic "wired-in" copy routine which the program calls for at the proper time.

Simplified von Neumann automata reproduction. Consider a single instruction tape, and a constructor machine which reads the instructions once to build the offspring machine and again to make a copy of the instructions for the offspring machine. Notice that although the instructions available to the system yield a duplicate of the original system, this need not be the case. Thus, in the biological example, even though some DNA made available to a cell does not code the instructions for a duplicate cell, the cell-machine still may proceed to obey the instructions. This means that a cell can generate offspring not only different from itself and its normal constituents and products, but even inimical to it. This is precisely what happens when a virus possessing no metabolic machinery and no enzymatic protein machinery to read DNA or to manufacture anything parasitically insinuates itself into a host cell. The virus co-opts the host cell's interpreting and manufacturing capacity, causing it to make virus particles until the cell fills with them, bursts open, and is destroyed. The greatly multiplied viral agents are then free to parasitize other cells.

In artificial systems as well, machines may read and interpret instructions without knowing what they are being called upon to do. The instructions might call for some computational, constructional, or program-copying activities. The machine can make machines unlike itself, and can give these "unnatural" offspring copies of the instructions which were employed in their manufacture. If the offspring are also equipped to read and follow instructions, and if they have a constructional capability, their offspring in turn would be replicas of themselves – which might not resemble their "grandparent" machine at all. Thus, an original construction machine can follow instructions to make an indefinitely large number of diverse machines, that are like or unlike themselves, capable or not capable of constructing, reproducing, etc. And though a universal constructing machine might make large numbers of "sterile" machines, if it should once make a duplicate of itself which is also equipped with the instructional program for making duplicates of itself, the process can become "explosive." Such machines would tend to drive out all other "species" not possessing this reproductive "autocatalytic" property.

Thatcher's variant: inferring structure. Thatcher (1970) showed that a machine need not have an explicit construction program made available to it initially in order to create

a duplicate of itself. First, it is sufficient that a machine can secure a *description* of itself (in place of instructions) if the machine is equipped with the capacity to read the description and convert this into the necessary constructive actions. Second, using a result obtained by Lee (1963) and himself (Thatcher, 1963), Thatcher showed that such a machine need not have its description loaded beforehand into its accessible memory organ. Instead, the machine has a partial self-description hard-wired into itself in the form of circuits which, when stimulated, make the description available to the machine in its accessible memory organ. These data describe all of the machine except the hardwired part which was stimulated to emit the description in the first place. The problem then, for the machine, is to obtain the description of this hidden part of itself. Lee and Thatcher showed that this section of the device can be constructed in such a simple fashion that the system can infer how this part must have been constructed merely by examining the consequences of its actions (e.g., the partial description it produced). After inferring the nature of this hidden part of itself, the machine possesses a complete self-description and can then follow von Neumann's paradigm for reproduction.

The principal practical significance of this form of automata replication is that it reminds the designer that the information required for machine construction (whether reproduction or not) need not be in the form of *instructions* for constructions but can be in the form of a *description.* Moreover, the description need not even reside in an accessible organ such as memory registers but may be embedded in "inaccessible" hardware. The hypothetical infinite regress likewise is shown to be baseless – it is possible for a machine to have within itself only a part of its own description, and from this to infer the rest.

Reproduction by component analysis. In von Neumann's cellular system, an embedded machine cannot send out an inspection arm to an encountered machine to identify all of its states. However, the cell-space system could be redesigned to permit this. In such a system an analyzing machine could examine an encountered passive machine and identify the type and location of all its cell-automata. (The analyzer might of course have to penetrate the machine, thus altering its automaton states, so the inspecting arm would have to send out appropriate restoration construction signals.)

In von Neumann's kinematic model a machine ostensibly could identify all parts of the system and thus determine the type and location of all components. This opens the possibility that a machine system might, for example, reproduce essentially two machines – one active, the other passive or able to assume passivity under a signal from the active machine. This possibility and others have been explored by Laing (1975, 1976, 1977, 1978, 1979) in a series of papers presenting alternative reproductive strategies which include the following:

- Beginning with two identical machines, one active and one passive, the active machine "reads" the passive machine twice, producing one active and one passive machine, thus completing reproduction.
- Beginning with two machines (not necessarily identical) one machine reads the second, and makes a duplicate of it. Then the second reads the first, and makes a duplicate of it, active and passive status being exchanged.
- By combining the capacity of machines to read machines with the Thatcher result, one can hardwire a machine to construct a second machine which is a duplicate of the original except for the hardwired part which produced the second machine. The original machine then "reads" the newly constructed partial duplicate, and infers what the missing hardwired part must be. The original machine then constructs the missing part, completing the reproductive process. This result explicitly confronts and overcomes the "necessary machine degeneracy" criticism of automata self-replication.

Machine reproduction without description. In the machine reproduction schemes explained thus far, some arbitrary part of the machine which cannot be inferred is always made explicitly available in memory initially, or is implicitly made available in memory or for inspection by means of an internal wired-in memory, also not directly accessible. Laing (1976) showed that even this wired-in description is not necessary. In effect, a machine can carry out a self-inspection which can yield a description which in turn can be made available to the machine in constructing a duplicate of itself.

The process begins with a wired-in construction routine which produces a semiautonomous analyzer machine. This analyzer moves over the original machine and identifies the type and location of its componentry. This is reported back to the original machine, which uses this information to make a duplicate of itself. Thus, though it may be that a part of a machine "may not comprehend the whole" in a single cognitive act, a part of a machine *can* examine in serial fashion the whole machine, and in time can make this information available to the machine for purposes of replication.

Exploitation of basic machine capabilities. The "simplified von Neumann" automata reproductive strategy – whereby a machine employs a stored program of instructions to make other machines (including duplicates of itself) and then also provides the program or parts of programs of instructions to newly constructed machines – should probably be the central strategy for any actual physical machine reproducing systems. The other strategies are, from most points of view, more complex than this and

thus perhaps are less preferable. The virtue of the alternative strategies is not as practical ways of implementing machine reproduction but rather in suggesting many basic capabilities, which, in a complex automated replicating LMF, may be usefully employed. The following are some of the behaviors of which, under *suitable* conditions and design, machines are actually and potentially capable:

(1) A machine can be "hard-wired" to carry out a computation.

(2) A machine can be programmed to carry out a computation.

(3) A machine can be a general-purpose computer, in that it can be given a set of instructions which will enable it to carry out the computation of any other computer. Alternatively, a general-purpose computing machine can be given the description of any other computing machine, and can carry out the computational actions of the machine described.

(4) A machine can be hard-wired to carry out a construction activity.

(5) A machine can be programmed to carry out a constructional activity.

(6) A sufficiently complex machine can be a general-purpose constructor, vis-a-vis a set of machines, in that it can be given a set of instructions which enables it to carry out the construction of any of the set of machines. Alternatively, a machine can be given the description of any machine of the set, and can, from this description, construct the machine described.

(7) A machine can construct a duplicate of itself, including the instructions or description used to guide the construction process.

(8) A machine, given a coded set of instructions for machine actions, or a coded description of a machine, can make a copy of the instructions or coded description.

(9) A machine, given a coded set of instructions for machine actions, can infer the structure of a machine which can carry out the actions described, and can construct such a machine.

(10) A machine, given a coded set of instructions for a machine, or a description of a machine, can carry out the actions of the machine whose instructions are given or whose description is supplied.

(11) A machine, given the instructions for or the description of an unknown machine, can examine the instructions or description and can (a) infer some of the properties of the machine, (b) simulate the actions of the machine, (c) construct the machine, and (d) observe the actions of the constructed machine.

(12) A machine can determine the component types of encountered machines.

(13) A machine can determine the structure (the component type and arrangement of components) of encountered machines.

(14) A machine can thus obtain a structural description of an encountered machine and simulate its actions, construct a duplicate, and then observe the duplicate in action.

(15) A machine can possess a copy of its own description, perhaps stored in a memory organ.

(16) A machine can obtain a copy of its own present structure. Note that the present structure of a machine may deviate from the original design, and also from its present stored description of itself (which may be out of date).

(17) A machine can compare the stored description of itself with the description obtained by inspection, and note the discrepancies.

(18) A machine can make a duplicate of itself on the basis of its stored "genetic" description or on the basis of its present (possibly altered) structure. This latter is an example of transmission of acquired characteristics.

(19) A machine can examine duplicates of itself constructed on the basis of an examination of itself, and note the discrepancies.

(20) The duplicates made from either of these two bases (genetic and observed) can be set in action and observed.

(21) For diagnostic purposes, the two kinds of descriptions can be compared, the two passive structures compared, the two kinds of structures in action observed and compared. The basis for machine self-diagnosis is thus available.

(22) A machine noting the discrepancies between two machine descriptions, or machine structures, or two machine behaviors, can in some cases act so as to resolve the discrepancies. That is, a machine in some cases can repair or reject or reconstruct deviant machines (including itself).

(23) A machine encountering an "unknown" machine can observe the behavior of that machine and compare this to the behavior of other machines, both directly and by simulating the behavior of those machines for which it already has or can obtain descriptions.

(24) A machine encountering an unknown machine can examine the structure of the machine and obtain a structural description which can be compared with other structural descriptions.

(25) Encountering an unknown device, a machine can use the structural description of the unknown to simulate its actions. These simulated actions can be compared to those of other machines whose descriptions are stored or which can be made available.

(26) Having the description of an encountered device, a machine can construct a duplicate of it. This duplicate can be set in action and observed, and its behavior compared with the behavior (actual or simulated) of other machines.

(27) The structure and behavior of encountered machines can be compared with those of known useful or benign machines, including that of the inspecting machine itself. This comparison, and the degrees of similarity

discerned, can serve as the basis for a subsequent policy of "friendship," "tolerance," "avoidance," "enmity," etc.

(28) The descriptions of encountered machines can be incorporated into the reproductive construction cycle so that these new machines or their features become part of the continuing and evolving machine system. This is an analogue to biological symbiosis.

Machine multiplication strategies. In describing the logical process of machine reproduction we have concentrated on the means by which the parent system could come to possess the information needed to carry out a replication and the associated question of how offspring would if necessary acquire the programs needed to continue the machine reproduction process. Although these questions, logically, are at the heart of machine replication, they leave open many issues concerning creation and siting of new machine systems as well as the ultimate fate of such systems.

This matter can be approached by considering certain biological analogues to the machine situation. In the known biological realm, all living organisms use the same underlying reproductive logic of protein synthesis and nucleotide sequence copying but employ vastly different broad strategies in producing more of their own kind.

One strategy is seen in the case of seed-bearing plants (as well as most fish and insects), in which vast numbers of "minimal" genetic packets are produced by the parent system and dispersed in the hope that a sufficient number will, largely by chance, find an appropriate site at which to survive and complete growth and development to maturity. At the other end of the scale is human behavior, whereby "construction" and nurture of the offspring may continue under the control and protection of the parent system until near maturity.

The particular multiplication strategy for artificial reproducing systems must of course be adjusted to intentions. The swift utilization of large rich environments might justify a "seed" dispersal strategy, with early maturity of new systems so as to retain a high reproductive rate. On the other hand, an environment consisting of scattered pockets of valuable resources, or situations with less pressure for immediate "explosive" utilization might suggest fewer offspring, possibly more fully developed in regard to their capacity for seeking out and efficiently utilizing the scarce resources available. In this case, the offspring might also be expected to receive longer tutelage from the parent system or from outside controllers (such as humans).

Similarly, the presence of a large contiguous valuable ore body might dictate the extensive ramification of a single machine factory system consisting of many laboring submachines. The model of a colonial organism such as coral, or of a social insect such as ants or termites, might make more sense. Zoological and sociobiological studies of animal and plant multiplication strategies may prove valuable in suggesting optimal machine system growth and reproduction strategies. One important difference must be borne in mind: biological organisms often have adapted their strategies to compete with other organisms, as well as to survive in a world where resources are renewed at certain rates over varying seasons. Some of these factors may be nonexistent or present in very different form in a nonterrestrial machine-inhabited environment.

A few questions that should be considered in determining optimal replicating machine behavior include:

- How large should a system be allowed to grow?
- How large should a system grow before it reproduces?
- What sorts of offspring (e.g., minimal vs mature) should be produced? A mixture?
- How many offspring should be produced? How many offspring should be produced from a single parent machine?
- When should offspring be produced?
- Where and how should offspring be sited? Specific sites? Near? Far? Randomly dispersed?
- What offspring transport mechanisms should be employed? Should new systems be mobile? Under own control? Parent? Human operator?
- When should sited machine systems be turned off? Abandoned? Should lifespan of a machine system be a function of time alone? Reproductive life? Exhaustion of local resources? Work experience and use? Detection of malfunction? When should subsystems be turned off? What growth and death patterns of individual machine systems should be adopted?
- What should be done with unsited offspring systems? Allowed to wander indefinitely?
- What should be done with outmoded machine systems? Dismantle them? Abandon them?

Intergeneration information transmission among replicating machines. Throughout most of the present discussion it has been assumed that the goal was to have the parent machine transmit to its offspring machine the same genetic information it received from its parent, regardless of the logical strategy of reproduction employed. This genetic fidelity is not necessary or even desirable in all cases. Normally the parent should transmit all information necessary for offspring to do their jobs and to construct further offspring in turn, but beyond this simple requirement there are many alternatives. For example, a parent machine might augment its program during its lifetime with some valuable information, and this augmented part of the program could then be transmitted to its offspring.

A few possible variations of interest include:

(1) The parent machine program is not altered in the course of its lifetime and is transmitted unaltered to offspring.

(2) The parent machine program is altered (e.g., by intervention, or by some machine adaptive process of a more or less complex sort) during the course of its lifetime, but again only the program originally received from the parent is transmitted to the offspring.

(3) The parent machine program is altered during the course of its lifetime, and the altered program *is* transmitted to the offspring machine. The parent machine (being a constructing machine) may make changes in its structure beyond those called for in its received genetic program.

(4) Changes in parent structure are not made part of the offspring structure.

(5) Changes in parent structure *are* made part of the offspring structure.

(6) Changes in parental structure are not made part of the offspring structure, but are made part of the offspring genetic program. Thus, the offspring can, under its own control, modify its structure to conform to that of its parent machine.

5.2.3 Information and Complexity in Self-Replicating Systems

The design and implementation of a self-replicating lunar factory represents an extremely sophisticated undertaking of the highest order. It is useful to consider the complexity of this enterprise in comparison with the information requirements of other large systems, natural or artificial, replicating or not (Stakem, 1979).

It is not immediately clear what the proper measure should be. One way to look at the problem of machines reproducing themselves is to consider the flow of information that occurs during reproduction. A fully generalized self-replicating system could possess a reproductive behavior of such complexity that the information necessary to describe that behavior is complete to atomic level specifications of machine structure. Such a machine has behavior so complex and complete that it might produce a copy of itself almost from complete chaos – say, a plasma containing equal concentrations of all isotopes. In this case the machine reproduction is essentially complete – given sufficient energy, the system can make copies of itself in any arbitrary environment even if that environment contains virtually no information relevant to replication.

At the other extreme, consider a long row of Unimate PUMA-like industrial robots side by side, each requiring merely the insertion of a single fuse to render it functional. The first working robot, its fuse already in place, seeks to "reproduce" itself from a "substrate" of dormant machines. It accomplishes this by reaching onto a nearby conveyor belt, picking up a passing fuse part, and plugging it into the neighboring robot. The adjacent machine now begins to function normally as the first (indeed, as an exact duplicate), so it can be said that in some sense the first machine has reproduced itself. Before the reproductive act there was no second working robot; afterwards, one exists. However, this is almost the most trivial case of replication imaginable, since the substrate for reproductive activity – in this case completed machines lacking only fuses – is extremely highly organized. Hence, the operative complexity resides in the substrate, and the action of the machine in "making a new machine" is trivial.

This latter example may be compared to the case of a bacteriophage. The phage particle infects a healthy bacterium, using the captive cellular machinery to manufacture new virus particles. Only the DNA of the virus enters the bacterium, instructing the cellular machinery to make new viral DNA and to interpret the DNA to create protein and polysaccharide components which form the coat or carrier of the viral DNA. Thus the foreign DNA, like the PUMA robot which inserts fuses to "self-replicate," must situate itself in a very rich complex environment, one already containing a great deal of machinery and information. In this case, the complexity of the virus-making enterprise probably can be gauged by the length of the viral DNA inserted into the host cell, just as the true complexity of the fuse-insertion behavior can be gauged by the length of the program needed to permit location of the supply of fuses and the fuse holder on an adjacent machine in physical space, and to insert the part properly. It is suggested, therefore, that the length of the shortest program which can carry out the process of replication may be an appropriate measure of the complexity of the task.

For instance, in the case of the von Neumann cellular reproducing system each part is already located in its proper place in space, but signals must be injected into that space to cause it to take on the properties desired in the offspring machine. It has been estimated that such a reproducing machine might consist of a minimum of 10^5 cells, with offspring cell type and location the principal parameters which must be specified for each. The length of the shortest program would represent perhaps 10^6 bits of information (Kemeny, 1955).

If the construction of a replicating growing lunar factory was purely a matter of machine parts assembly, then the length of the replication program could be determined by the necessity to locate various required parts in the environment and then to specify and execute the proper placement of each part to construct the desired system (Heiserman, 1976). However, it is likely the reproductive process will be vastly more complicated than this, since it is not likely that all parts can be supplied "free" from Earth. If the lunar factory must begin, not with completed machines or parts, but rather with a raw lunar soil substrate, the task quickly becomes many orders more difficult – though not impossible. Based on the estimates outlined in section 5.3 and the

appendixes, the lunar factory replication program length should not exceed roughly 10^{12} bits of information. This compares to about 10^{10} bits coded in the human genome and about 10^{14} bits stored in the human brain. Terabit (10^{12} bits) memories are considered state-of-the-art today.

Complexity of a self-replication program may also be viewed as an index of versatility or system survivability. The more complex the program, the more likely it is that the machine system can bring about its own replication from increasingly disordered substrates. This is an interesting observation because it suggests that reproduction is an activity defined along a broad continuum of complexity rather than as a single well-defined event. Both the chaos-replicator and the fuse-insertion robots described above perform acts of self-reproduction. Fundamentally, these systems differ only in the degree to which they are capable of bringing order to the substrate in which they are embedded.

It is interesting to note that human beings fall somewhere in the middle of this broad reproductive spectrum. A 100 kg body mass, if composed of purely random assortments of the 92 natural elements, would contain roughly 10^{27} atoms and hence require about 10^{28} bits to describe. Yet a 100 kg human body is described by a chromosome set containing just 10^{10} bits. The difference must be made up by the "substrate" in which people are embedded – a highly ordered rich environment, namely, the Earth. Human beings thus are conceptually remarkably similar to von Neumann's kinematic self-reproducing automata, moving around in a "stockroom" searching for "parts."

5.2.4 Conclusions

The Replicating Systems Concepts Team reached the following conclusions concerning the theory of machine reproduction:

(1) John von Neumann and a large number of other researchers in theoretical computer science following him have shown that there are numerous alternative strategies by which a machine system can duplicate itself.

(2) There is a large repertoire of theoretical computer science results showing how machine systems may simulate machine systems (including themselves), construct machine systems (including machine systems similar to or identical with themselves), inspect machine systems (including themselves), and repair machine systems (including, to some extent, themselves). This repertoire of possible capabilities may be useful in the design and construction of replicating machines or factories in space.

5.3 Feasibility

The design and construction of a fully self-replicating factory system will be a tremendously complicated and difficult task. It may also be fairly expensive in the near-term. Before embarking upon such an ambitious undertaking it must first be shown that machine self-replication and growth is a fundamentally feasible goal.

5.3.1 Concept Credibility

The plausibility of the theoretical notion of self-replicating machines already has been reviewed at length (see sec. 5.2). It remains only to demonstrate concept credibility in an engineering sense (Bradley, 1980, unpublished memorandum, and see appendix 5A; Cliff, 1981; Freitas, 1980a; von Tiesenhausen and Darbro, 1980) – that is, is it credible to consider building real physical machines able to replicate themselves?

The credibility of any design proposed for such a machine or machine system depends first and foremost upon whether that design is consistent with reasonably foreseeable automation and materials processing technologies. These technologies need not necessarily be well established or even state-of-the-art, but should at least be conceivable in the context of a dedicated R&D effort spanning the next two decades. It is interesting to note that computer programs capable of self-replication have been written in many different programming languages (Burger et al., 1980; Hay, 1980), and that simple physical machines able to replicate themselves in highly specialized environments have already been designed and constructed (Jacobson, 1958; Morowitz, 1959; Penrose, 1959).

Another major requirement for concept credibility is a plausible system configuration. Proposed designs for self-replicating systems (SRS) must be sufficiently detailed to permit the generation of work breakdown structures, subsystem operational flowcharts, mass and energy throughput calculations, and at least preliminary closure (see sec. 5.3.6) analyses.

A related requirement is plausible mission scenarios. Research and development costs for the proposed design should be many orders of magnitude less than the Gross National Product. The mission must not require launch and support facilities which cannot or will not be available in the next two or three decades. The mission must entail reasonable flight times, system lifetimes, growth rates, production rates, and so forth. The problems of reliability and repair should be addressed.

The final requirement for concept credibility is positive societal impact. A given SRS design must be economically, politically, and socially feasible, or else it may never be translated into reality even if the technology to do so exists. A general discussion of the implications of replicating systems appears in section 5.5, but the team has arrived at no firm conclusions regarding concept feasibility in this area. More research is clearly required.

5.3.2 Concept Definition

In order to demonstrate SRS concept credibility, specific system designs and mission scenarios must be subjected to a detailed feasibility analysis. The first step in this process is to conceptualize the notion of replicating systems in as broad an engineering context as possible. Many kinds

of replicating machine systems have been proposed and considered during the course of the study. Some of these place emphasis on different types of behavior than others.

Consider a "unit machine" which is the automata equivalent of the atom in chemistry or the cell in biology – the smallest working system able to execute a desired function and which cannot be further subdivided without causing loss of that function. The unit machine may be comprised of a number of subunits, say, A, B, C, and D. These subunits may be visualized in terms of structural descriptions (girders, gearboxes, generators), functional descriptions (materials processing, parts fabrication, mining, parts assembly), or any other complete subset-level descriptions of the entire system.

SRS may be capable of at least five broad classes of machine behavior:

Production – Generation of useful output from useful input. The unit machine remains unchanged in the process. This is a "primitive" behavior exhibited by all working machines including replicating systems.

Replication – Complete manufacture of a physical copy of the original unit machine, by the unit machine.

Growth – Increase in mass of the original unit machine by its own actions, retaining the physical integrity of the original design.

Evolution – Increase in complexity of structure or function of the unit machine, by adding to, subtracting from, or changing the character of existing system subunits.

Repair – Any operation performed by a unit machine upon itself, which does not alter unit population, designed unit mass, or unit complexity. Includes reconstruction, reconfiguration, or replacement of existing subunits.

These five basic classes of SRS behavior are illustrated in figure 5.5.

Replicating systems, in principle, may be designed which can exhibit any or all of these machine behaviors. In actual practice, however, it is likely that a given SRS format will emphasize one or more kinds of behaviors even if capable of displaying all of them. The team has considered two specific replicating systems designs in some detail. The first (cf. von Tiesenhausen and Darbro, 1980), which may be characterized as a unit replication system, is described in section 5.3.3. The second (cf. Freitas, 1980a; Freitas and Zachary, 1981), which can be characterized as a unit growth system, is outlined in section 5.3.4. The team decided to concentrate on the possibility of fully autonomous or "unmanned" SRS, both because these are more challenging from a technical standpoint than either manned

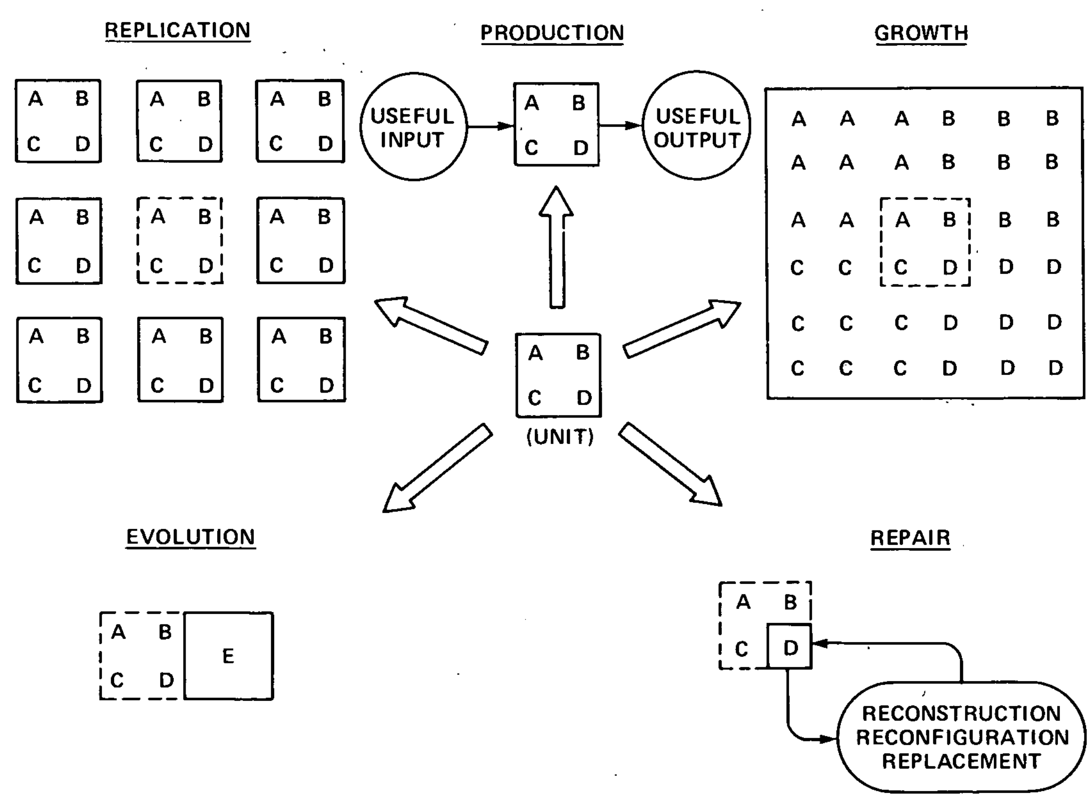


Figure 5.5.– Five basic classes of SRS behavior.

or teleoperated systems and also because the latter has already been treated to some degree elsewhere in this report (see chap. 4).

5.3.3 Unit Replication: A Self-Replicating System Design

The SRS design for unit replication is intended to be a fully autonomous, general-purpose self-replicating factory to be deployed on the surface of planetary bodies or moons. The anatomy of an SRS is defined by two end conditions: (1) the type and quantity of products required within a certain time, and (2) the available material needed to manufacture these products as well as the SRS itself.

There are four major subsystems which comprise each SRS unit, as shown in figure 5.6. First, a materials processing subsystem acquires raw materials from the environment and prepares industrial feedstock from these substances. Second, a parts production subsystem uses the feedstock to make machines or other parts. At this point SRS output may take two forms. Parts may flow to the universal constructor subsystem, where they are used to construct a new SRS (replication). Or, parts may flow to a production facility subsystem to be made into commercially useful products. The SRS also has a number of other important but subsidiary subsystems, including a materials depot, parts depots, product depot, control and command, and an energy system.

The work breakdown structure given in figure 5.7 lists all SRS elements studied, and each is briefly described below.

Materials processing and feedstock production. In this subsystem, raw materials are gathered by strip or deep mining. They are then analyzed, separated, and processed into industrial feedstock components such as sheets, bars, ingots, castings, and so forth, which are laid out and stored in the materials depot. The processing subsystem has a high degree of autonomy including self-maintenance and repair. It is linked to a central supervisory control system (see below).

The materials processing subsystem is shown schematically in figure 5.8.

Materials depot. The materials depot collects and deposits in proper storage locations the various feedstock categories according to a predetermined plan. This plan ensures that the subsequent fabrication of parts proceeds in the most efficient and expeditious manner possible. The depot also serves as a buffer during interruptions in normal operations caused by failures in either the materials processing subsystem (depot input) or in the parts production subsystem (at depot output).

Parts production plant. The parts production plant selects and transports industrial feedstock from the materials depot into the plant, then fabricates all parts required for SRS production or replication activities. Finished parts are stored in the production parts and the replication parts depots, respectively. The parts production plant is highly automated in materials transport and in distribution, production, control, and subassembly operations.

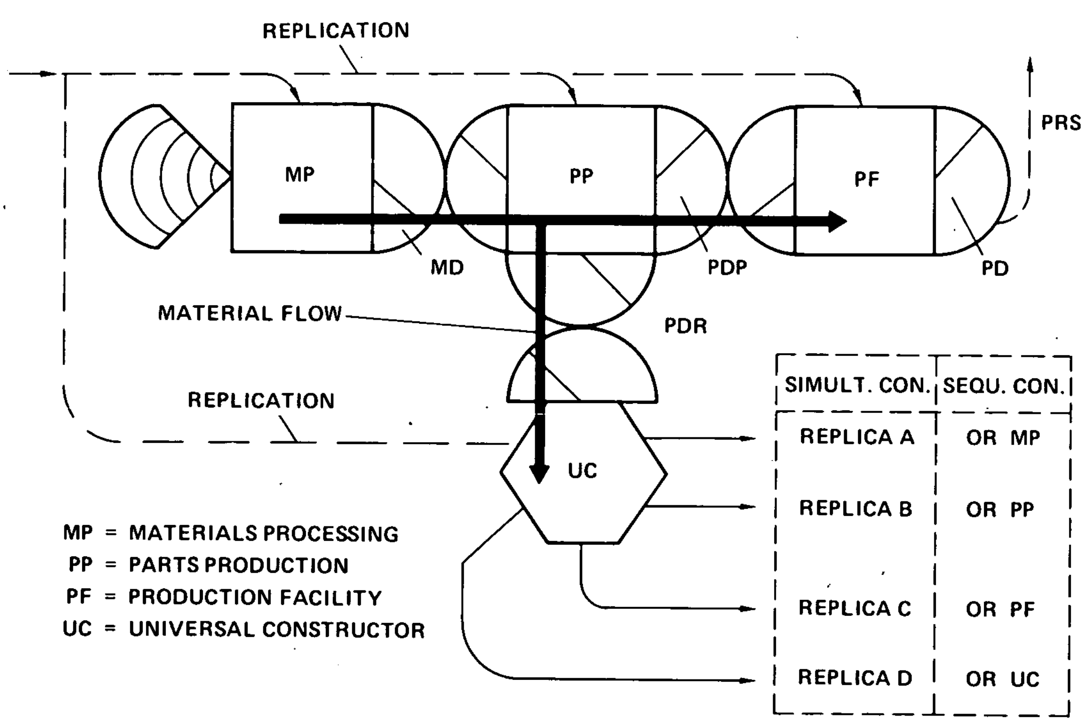


Figure 5.6.– Functional schematic of unit replication SRS.

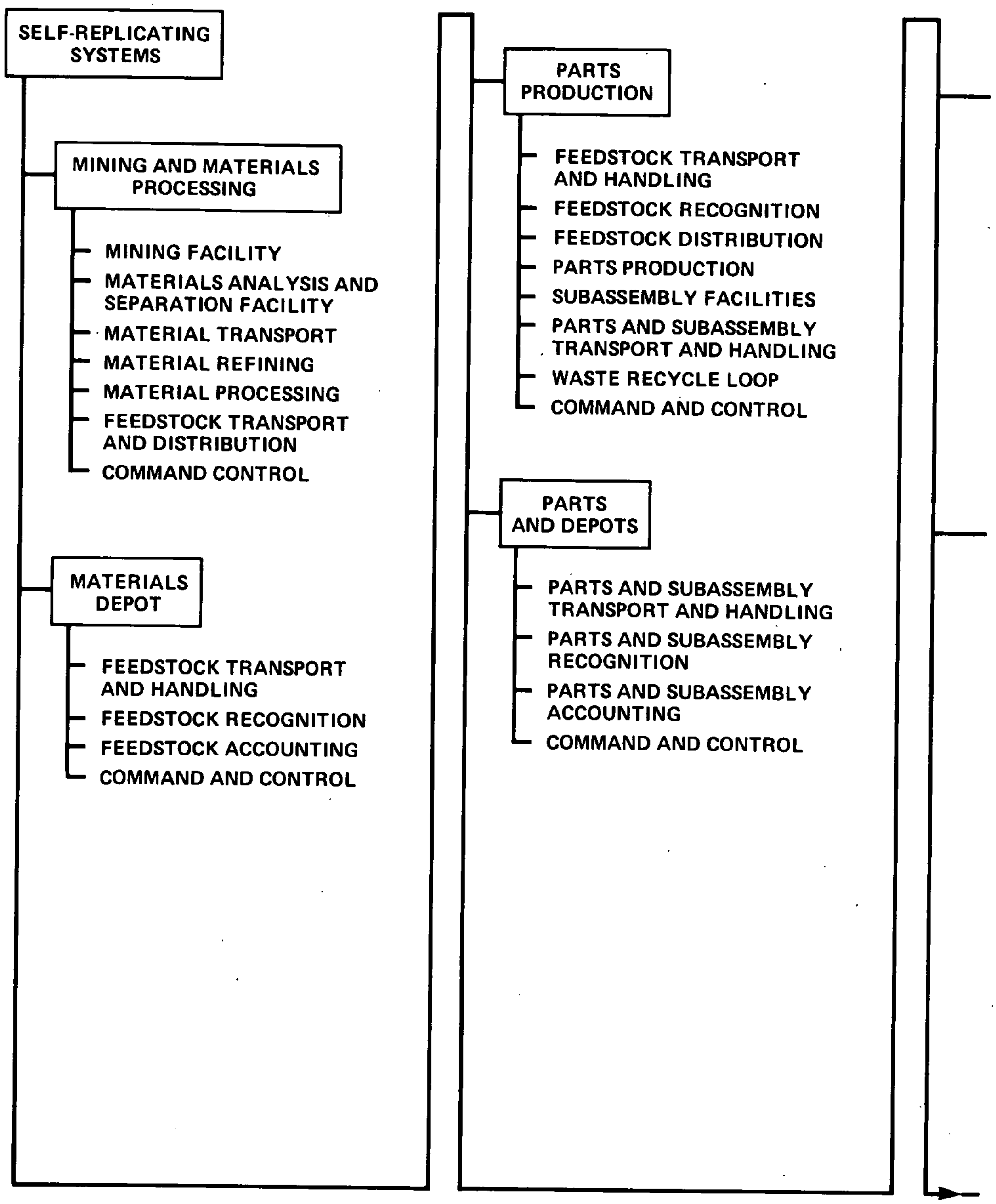


Figure 5.7.– Work breakdown structure for SRS.

The parts production plant subsystem is shown schematically in figure 5.9.

Parts depots. There are two parts depots in the present design. These are called the production parts depot and the replication parts depot.

Parts are stored in the production parts depot exclusively for use in the manufacture of useful products in the production facility. If certain raw materials other than parts and subassemblies are required for production, these materials are simply passed from the materials depot through the parts production plant unchanged. The parts production depot also acts as a buffer during interruptions in normal operations caused by temporary failures in either the parts production plant or the production facility.

Parts and subassemblies are stored in the replication parts depot exclusively for use in the replication of complete SRS units. Storage is in lots earmarked for specific facility construction sites. The replication parts depot also serves as buffer during interruptions in parts production plant or universal constructor operations.

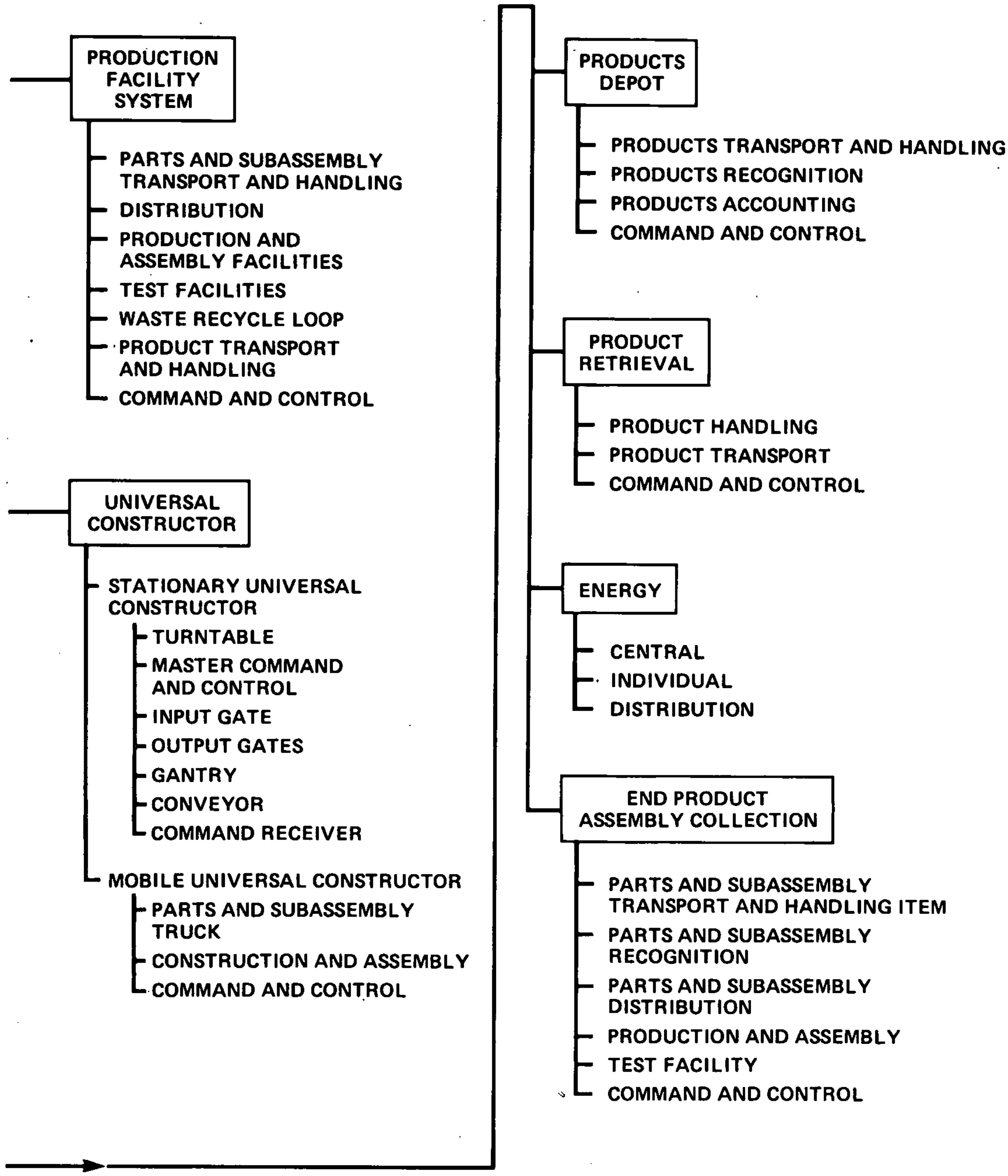


Figure 5.7.– Concluded.

Production facility. The production facility manufactures the desired products. Parts and subassemblies are picked up at the production parts depot and are transported to the production facility to be assembled into specific useful products. Finished products are then stored in the products depot. Ultimately these are collected by the product retrieval system for outshipment.

Universal constructor. The universal constructor manufactures complete SRS units which are exact duplicates of the original system. Each replica can then, in turn, construct more replicas of itself, and so on. The universal constructor retains overall control and command responsibility for its own SRS as well as its replicas, until the control and command functions have also been replicated and transferred to the replicas. These functions can be overridden at any time by external means.

The universal constructor subsystem consists of two major, separate elements – the stationary universal constructor (fig. 5.10) and the mobile universal constructors

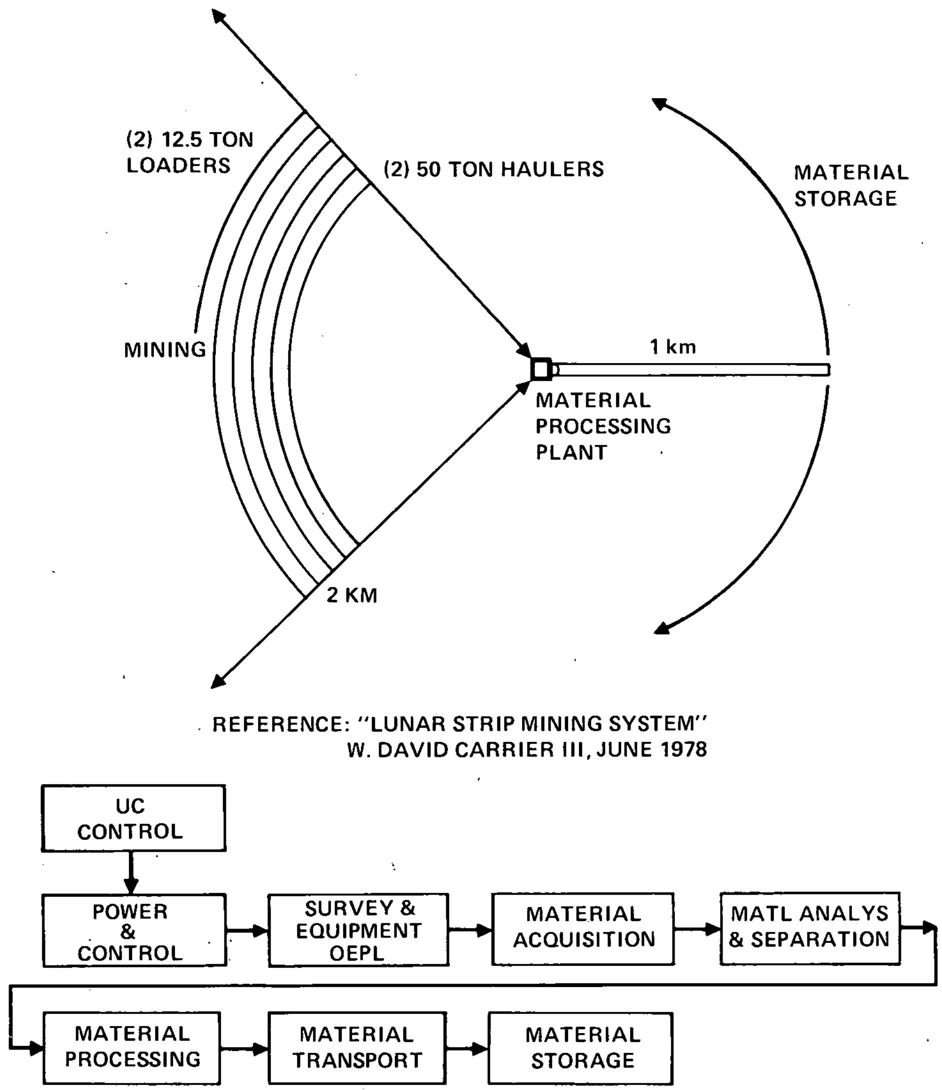


Figure 5.8.– SRS materials processing subsystem.

(fig. 5.11). This composite subsystem must successfully perform a number of fundamental tasks, including receiving, sorting, loading, and transporting parts and subassemblies; assembling, constructing, installing, integrating, and testing SRS systems; starting and controlling SRS operations; and copying and transferring instructions between system components.

Products depot. The outputs of the production facility are stored in the products depot, ready for retrieval. Major hardware components are neatly stacked for ready access by the product retrieval system. Consumables such as elemental oxygen are stored in reusable containers that are returned empty to the production facility. The products depot also serves as a buffer against variable output and retrieval rates.

Product retrieval system. The product retrieval system collects the outputs of all SRS units in an "SRS field" and carries them to an outside distribution point for immediate use or for subsequent outshipment. The dashed lines in figure 5.11 indicate one possible solution to this problem in a typical SRS field. Other solutions are possible – careful consideration must be given to SRS field configuration to arrive at an optimum product retrieval system design.

Command and control systems. The master control and command system, located within the stationary universal constructor, is programmed to supervise the total SRS operation and to communicate both with the peripheral controls of the mobile universal constructors during the self-replication phase and with the replicated stationary universal constructor during the transfer of command and control for the operation of the new SRS unit.

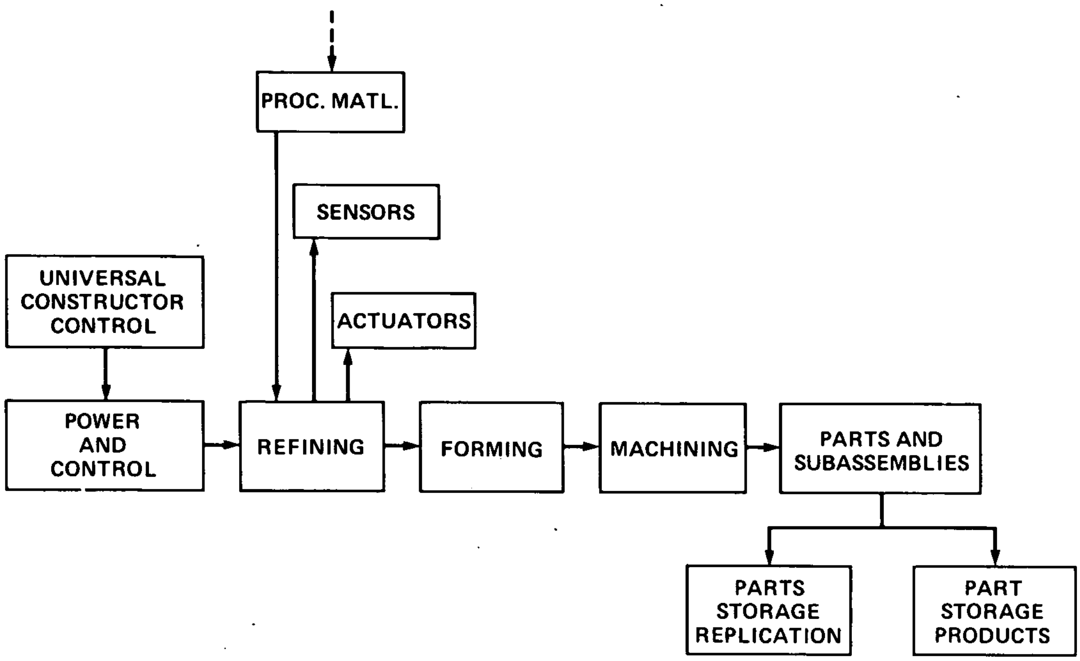


Figure 5.9.– SRS parts production plant subsystem.

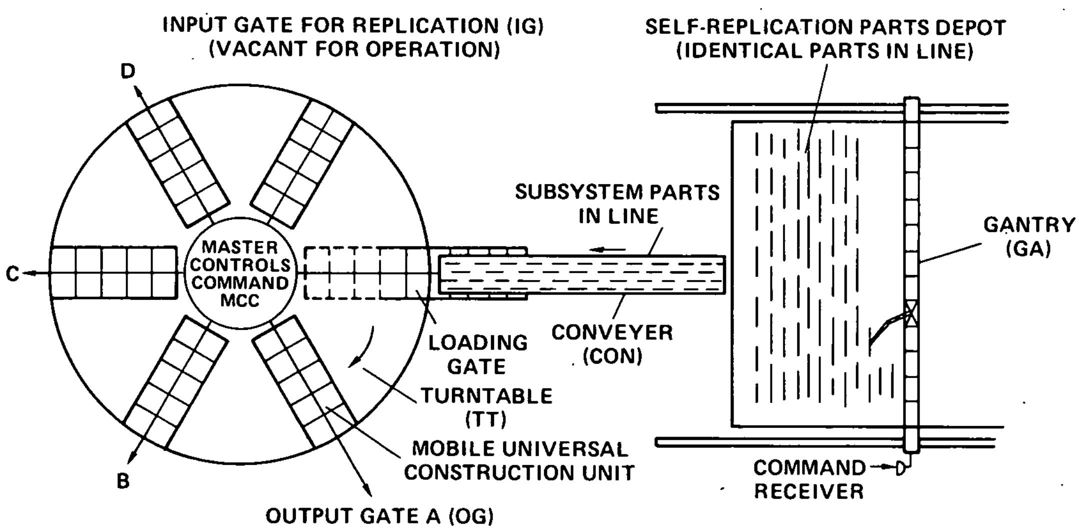


Figure 5.10.– SRS stationary universal constructor.

The master control and command system operates its own SRS unit through individual communication links which address the local control and command systems of individual SRS elements. In this way the master control and command system supervises the condition and operations of its own system elements, from materials acquisition through end product retrieval.

Energy system. The power requirements for the present design may be in gigawatt range. Hence, a single energy source (such as a nuclear power plant) would be excessively massive, and would be difficult to replicate in any case. This leaves solar energy as the lone viable alternative. Daylight options include: (1) central photovoltaic with a ground cable network, (2) distributed photovoltaic with

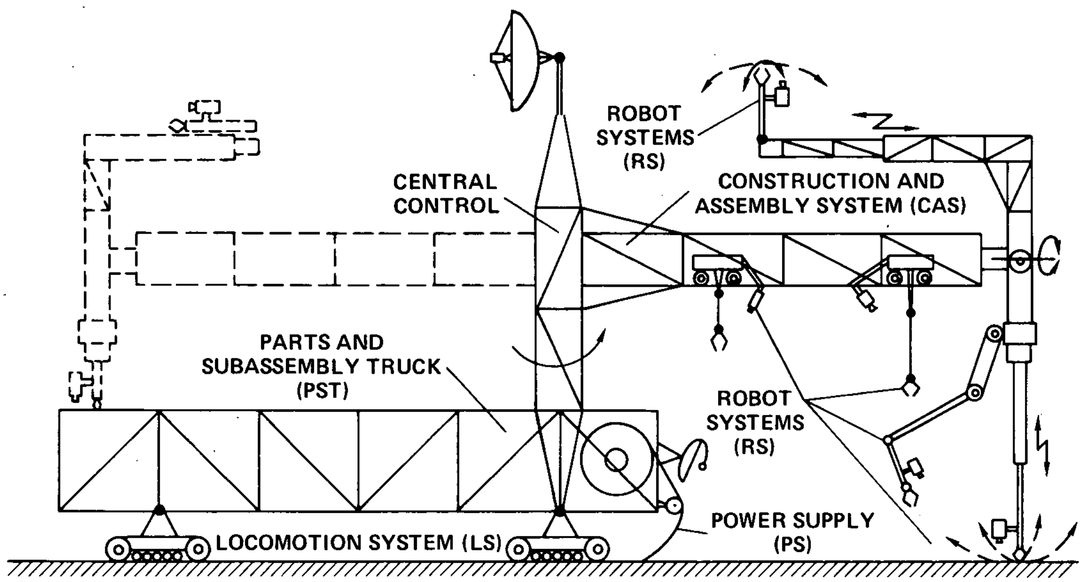


Figure 5.11.– SRS mobile universal constructors.

local distribution system, (3) individual photovoltaic, and (4) satellite power system, with microwave or laser power transmission to central, local, or individual receivers. Nighttime power options include MHD, thermionics, or turbogenerators using fuel generated with excess capacity during daytime. Oxygen plus aluminum, magnesium, or calcium could be used for fuel. A 15% efficient central silicon photovoltaic power station has been assumed in the reference design, with an output of tens of gigawatts and a size on the order of tens of square kilometers.

Each SRS produces, in addition to its scheduled line of regular products, a part of the photovoltaic energy system equal to the energy needs of its replicas. These are retrieved along with the regular products by the product retrieval system and are assembled on-site to increase energy system capacity according to demand during the self-replication phase.

SRS deployment and expansion. A complete SRS factory unit, erected on the surface of the Moon, might appear as illustrated in figure 5.12.

As a unit replication scheme, the multiplication of SRS units proceeds from a single primary system to many hundreds of replica systems. This expansion must be carefully planned to reach the desired factory output capacity without running out of space and materials. Figure 5.13 shows one possible detailed growth plan for the geometry of an SRS field. In this plan, each SRS constructs just three replicas, simultaneously, then abandons replication and goes into full production of useful output. After the three generations depicted, an SRS field factory network 40 units strong is busy manufacturing products for outshipment. The routes taken by mobile universal constructors are shown as solid lines, the product retrieval routes as dashed lines.

Figure 5.14 shows another possible expansion geometry. Again, each SRS constructs just three replicas, but sequentially rather than simultaneously. The end result is a field of 326 individual units after nine cycles of replication. Output is collected by the product retrieval system and taken to an end product assembly/collection system where end products undergo final assembly and other operations preparatory to outshipment. A more detailed discussion of expansion scenarios for SRS fields may be found in von Tiesenhausen and Darbro (1980).

Proposed development and demonstration scenario. It is proposed that the practical difficulties of machine replication should be confronted directly and promptly by a dedicated development and demonstration program having four distinct phases.

In Phase A of the development scenario, a robot manipulator will be programmed to construct a duplicate of itself from supplied parts and subassemblies. The original robot then makes a copy of its own operating program and inserts this into the replica, then turns it on, thus completing the duplication process (see appendix 5J). To complete Phase A; the replica must construct a replica of *itself,* repeating in every way the actions of the original robot. The rationale for the second construction, called the Fertility Test, is to demonstrate that the capacity for self-replication has in fact been transmitted from parent machine to offspring.

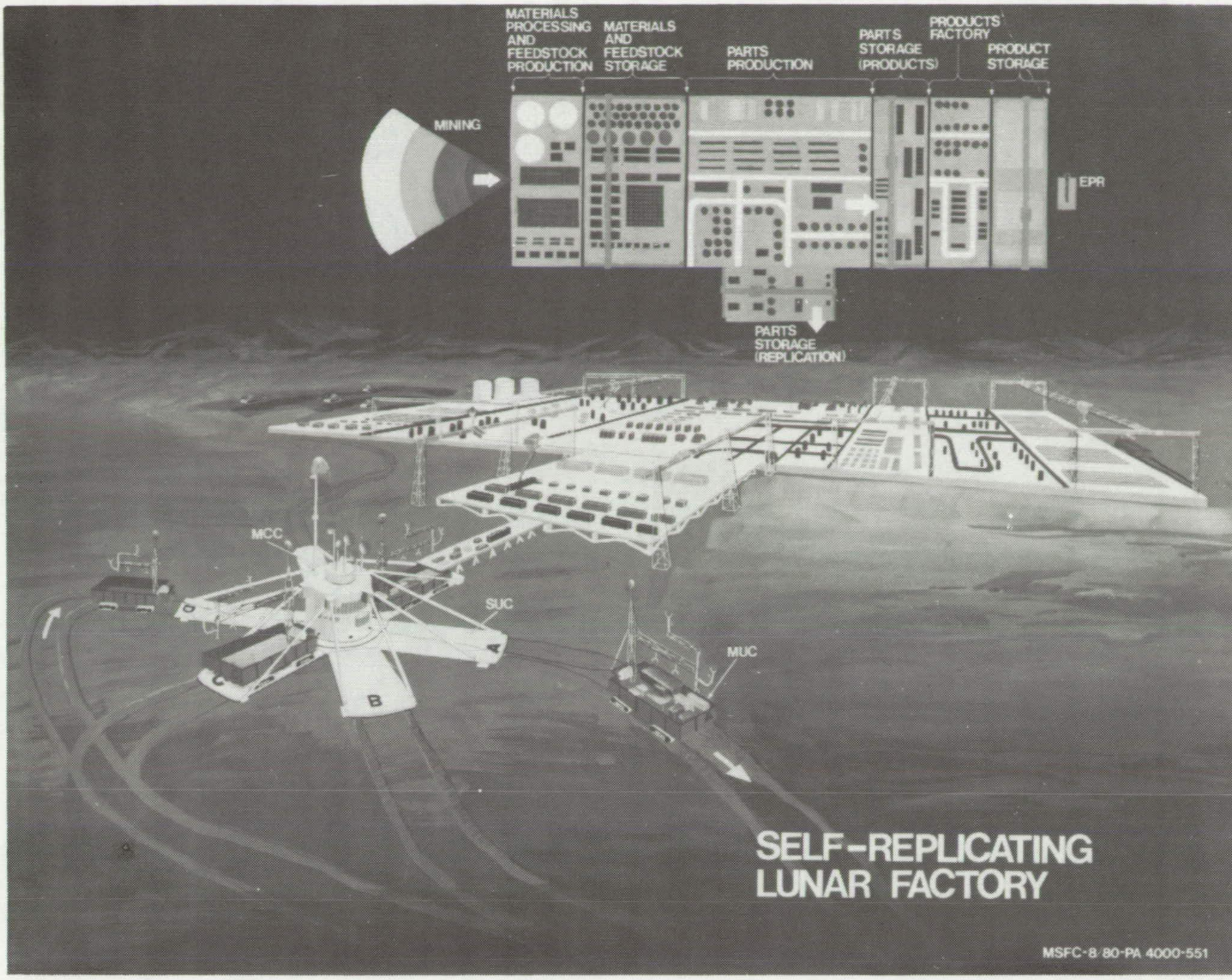


Figure 5.12.– Self-replicating lunar factory.

In Phase B of the development and demonstration scenario, the robot manipulator will be supplied with numerous additional parts so it can assemble objects of interest other than replicas of itself. This is intended to show that the system is able to construct useful products in addition to the line of robot duplicates.

In Phase C the manipulator system is still required to construct replicas and useful products. However, the robot now will be supplied only with industrial feedstock such as metal ingots, bars, and sheets, and must fabricate all necessary parts and subassemblies on its own. Successful completion of Phase C is expected to be much more difficult than the two earlier phases. The reason is that the parts fabrication machines must themselves be constructed by the robot manipulator and, in addition, all parts and subassemblies comprising the newly introduced fabrication machines must also be made available to the manipulator. Fabricator machines thus must be programmed to make not only the parts required for robot manipulators and useful products, but also their *own* parts and subassemblies as well. This raises the issue of parts closure, a matter which is discussed in section 5.3.6.

In Phase D, the system developed in the previous phase is retained with the exception that only minerals, ores, and soils of the kind naturally occurring on terrestrial or lunar surfaces are provided. In addition to all Phase C capabilities, the Phase D system must be able to prepare industrial feedstock for input to the fabrication machines. Successful completion of Phase D is expected to be the most difficult of all because, in addition to the parts closure problem represented by the addition of materials processing machines, all chemical elements, process chemicals, and alloys necessary for system construction and operation must be extracted and prepared by the materials processing machines. This raises the issue of materials closure (see also sec. 5.3.6). The completion of Phase D will yield an automatic manufacturing facility which, beginning with "natural" substrate, can replicate itself.

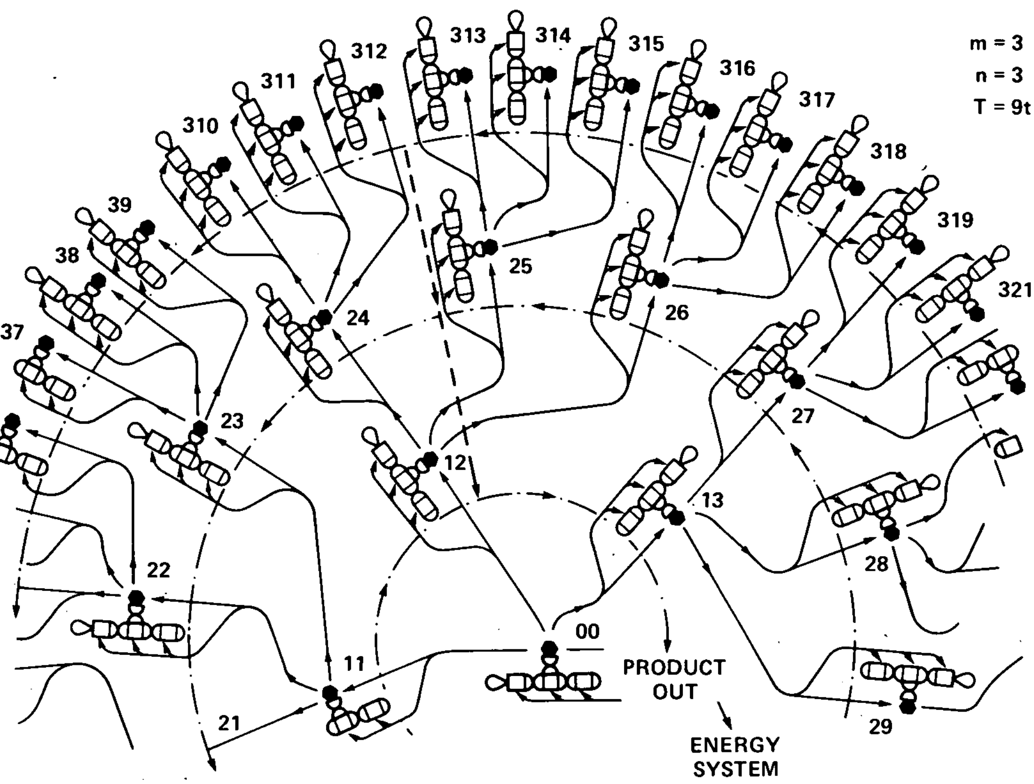


Figure 5.13.– Possible growth plan with simultaneous replica construction, suitable for geometry of an SRS field.

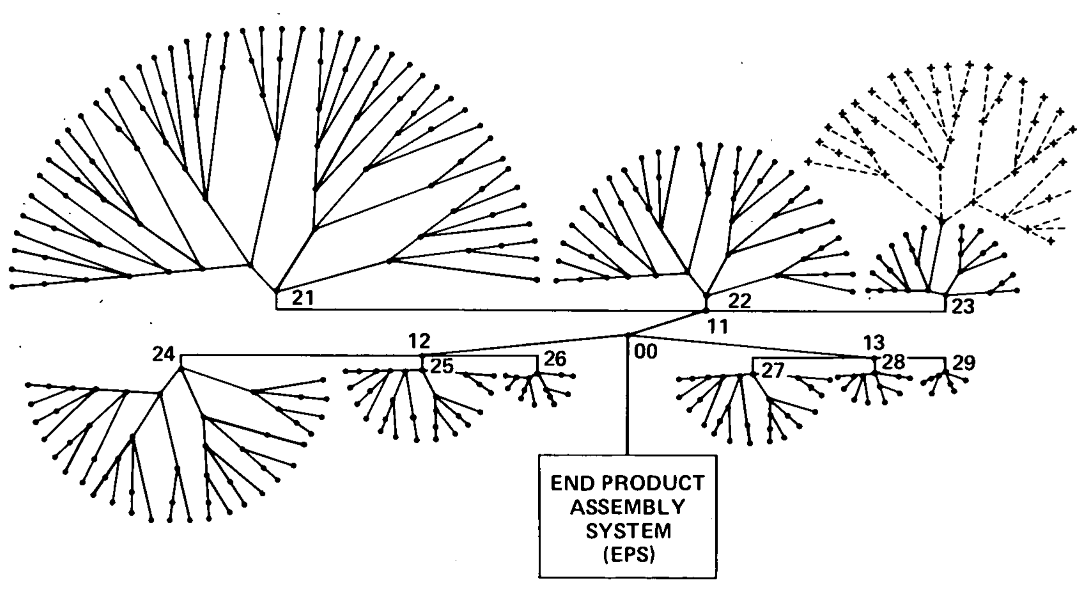


Figure 5.14.– SRS growth plan with sequential replication.

This progressive development of a replicating factory will serve to verify concept feasibility, clarify the functional requirements of such a system, and identify specific technological problem areas where additional research in automation and robotics is needed. A minimum demonstration program should be designed to gain engineering understanding, confidence, and hands-on experience in the design and operation of replicating systems. (See sec. 5.6.) The question of when the results of an Earth-based development and demonstration project should be translated to lunar requirements, designs, and construction remains open. On the one hand, it may be deemed most practical to complete Phase D before attempting a translation to a design better suited to a lunar or orbital environment. On the other hand, major system components for a lunar facility undoubtedly could be undertaken profitably earlier in concert with Phase C and D development.

The proposed development and demonstration scenario is described in greater detail in von Tiesenhausen and Darbro (1980).

5.3.4 Unit Growth: A Growing Lunar Manufacturing Facility

The Lunar Manufacturing Facility (LMF) demonstrating SRS unit growth is intended as a fully automatic general-purpose factory which expands to some predetermined adult size starting from a relatively tiny "seed" initially deposited on the lunar surface. This seed, once deployed on the Moon, is circular in shape, thus providing the smallest possible perimeter/surface area ratio and minimizing interior transport distances. Expansion is radially outward with an accelerating radius during the growth phase. Original seed mass is 100 tons.

The replicating LMF design encompasses eight fundamental subsystems. Three subsystems are external to the main factory (transponder network, paving, and mining robots). The LMF platform is divided into two identical halves, each comprised of three major production subsystems: (1) the chemical processing sector accepts raw lunar materials, extracts needed elements, and prepares process chemicals and refractories for factory use; (2) the fabrication sector converts these substances into manufactured parts, tools, and electronics components; and (3) the assembly sector, which assembles fabricated parts into complex working machines or useful products of any conceivable design. (Each sector must grow at the same relative rate for uniform and efficient perimeter expansion.) Computer facilities and the energy plant are the two remaining major subsystems. (See fig. 5.15.)

Transponder network. A transponder network operating in the gigahertz range assists mobile LMF robots in accurately fixing their position relative to the main factory

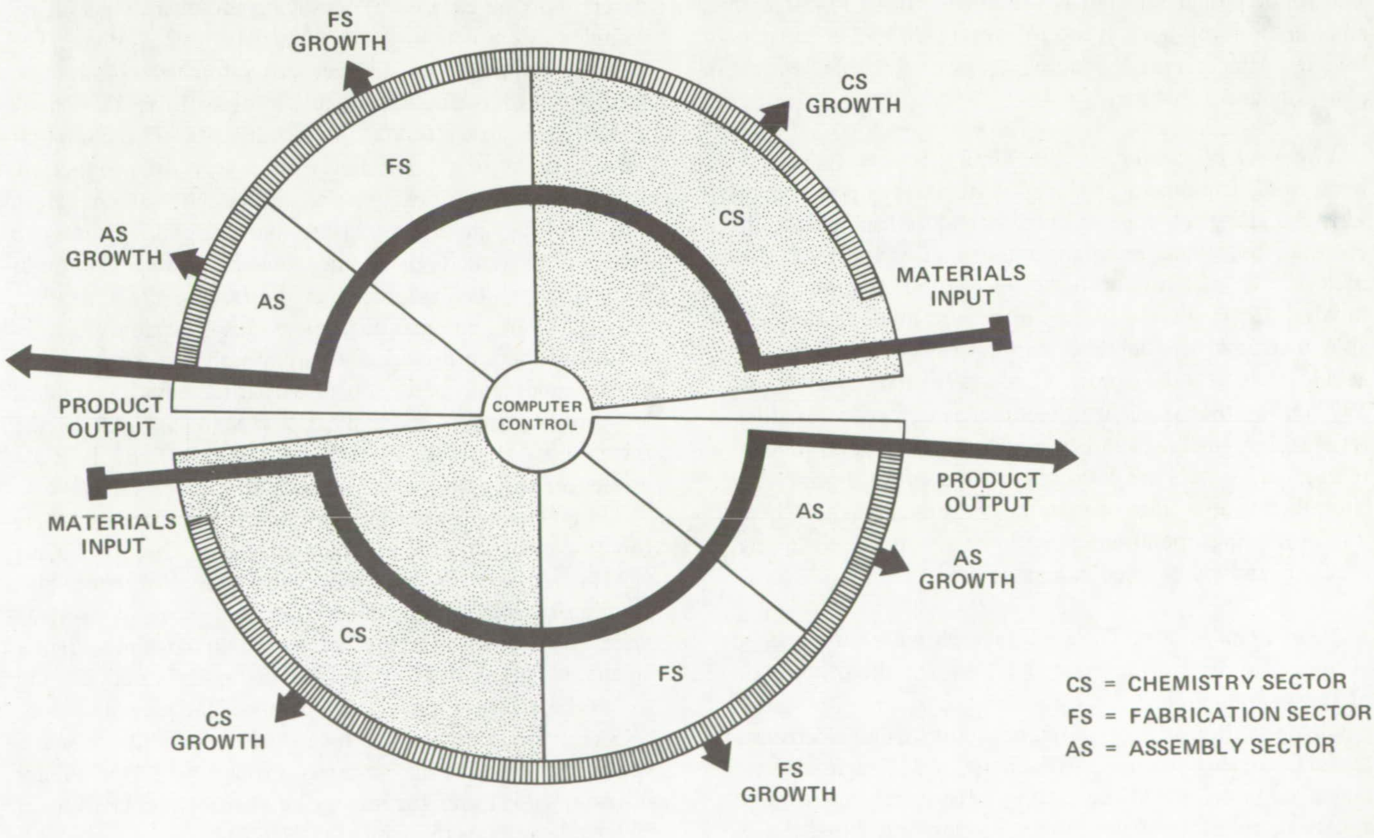


Figure 5.15.– Functional schematic of unit growth SRS.

complex while they are away from it. The network, described briefly in appendix 5B, is comprised of a number of navigation and communication relay stations set up in a well defined regular grid pattern around the initial seed and the growing LMF complex.

Paving robots. In order to secure a firm foundation upon which to erect seed (and later LMF) machinery, a platform of adjoining flat cast basalt slabs is required in the baseline design. A team of five paving robots lays down this foundation in a regular checkerboard pattern, using focused solar energy to melt pregraded lunar soil *in situ.* (See app. 5C.)

Mining robots. As described in appendix 5D, LMF mining robots perform six distinct functions in normal operation: (1) strip mining, (2) hauling, (3) landfilling, (4) grading, (5) cellar-digging, and (6) towing. Lunar soil is strip-mined in a circular pit surrounding the growing LMF. This material is hauled back to the factory for processing, after which the unused slag is returned to the inside edge of the annular pit and used for landfill which may later be paved over to permit additional LMF radial expansion. Paving operations require a well graded surface, and cellar digging is necessary so that the LMF computer may be partially buried a short distance beneath the surface to afford better protection from potentially disabling radiation and particle impacts. Towing is needed for general surface transport and rescue operations to be performed by the mining robots. The robot design selected is a modified front loader with combination roll-back bucket/dozer blade and a capacity for aft attachments including a grading blade, towing platform, and a tow bar.

Chemical processing sectors. Mining robots deliver raw lunar soil strip-mined at the pit into large input hoppers arranged along the edge of entry corridors leading into the chemical processing sectors in either half of the LMF. This material is electrophoretically separated (Dunning and Snyder, 1981; see sec. 4.2.2) into pure minerals or workable mixtures of minerals, then processed using the HF acid-leach method (Arnold et al., 1981; Waldron et al., 1979) and other specialized techniques to recover volatiles, refractories, metals, and nonmetallic elements. Useless residue and wastes are collected in large output hoppers for landfill. Buffer storage of materials output is on site. Chemical processing operations are shown schematically in figure 5.16, and are detailed in appendix 5E.

Fabrication sectors. The LMF fabrication sector outlined in appendix 5F is an integrated system for the production of finished aluminum or magnesium parts, wire stock, cast basalt parts, iron or steel parts, refractories, and electronics parts. Excepting electronics (Zachary, 1981) there are two major subsystems: (1) the casting subsystem, consisting of a casting robot to make molds, mixing and alloying furnaces for basalt and metals, and automatic molding machines to manufacture parts to low tolerance using the molds and alloys prepared; and (2) the laser machining and finishing subsystem, which performs final cutting and machining of various complex or very-close-tolerance parts. The basic operational flowchart for parts fabrication is shown in figure 5.17.

Assembly sectors. Finished parts flow into the automated assembly system warehouse, where they are stored and retrieved by warehouse robots as required. This subsystem provides a buffer against system slowdowns or temporary interruptions in service during unforeseen circumstances. The automated assembly subsystem requisitions necessary parts from the warehouse and fits them together to make subassemblies which are inspected for structural and functional integrity. Subassemblies may be returned to the warehouse for storage, or passed to the mobile assembly and repair robots for transport to the LMF perimeter, either for internal repairs or to be incorporated into working machines and automated subsystems which themselves may contribute to further growth. The basic operational flowchart for SRS parts assembly is shown in figure 5.18, and a more detailed presentation may be found in appendix 5G.

Computer control and communications. The seed computers must be capable of deploying and operating a highly complex, completely autonomous factory system. The original computer must erect an automated production facility, and must be expandable in order to retain control as the LMF grows to its full "adult" size. The computer control subsystem coordinates all aspects of production, scheduling, operations, repairs, inspections, maintenance, and reporting, and must stand ready to respond instantly to emergencies and other unexpected events. Computer control is nominally located at the hub of the expanding LMF disk, and commands in hierarchical fashion a distributed information processing system with sector computers at each node and sector subsystems at the next hierarchical level of control. Communications channels include the transponder network, direct data bus links, and E^2ROM messenger chips (firmware) for large data block transfers.

Using ideas borrowed from current industrial practice, top-down structured programming, and biology, Cliff (1981) has devised a system architecture which could perform automated design, fabrication, and repair of complex systems. This architecture, presented in appendix 5H, is amenable to straightforward mathematical analysis and should be a highly useful component of the proposed lunar SRS. Further work in this area should probably include a survey of industrial systems management techniques (Carson, 1959) and the theory of control and analysis of large-scale systems (Sandell et al., 1978).

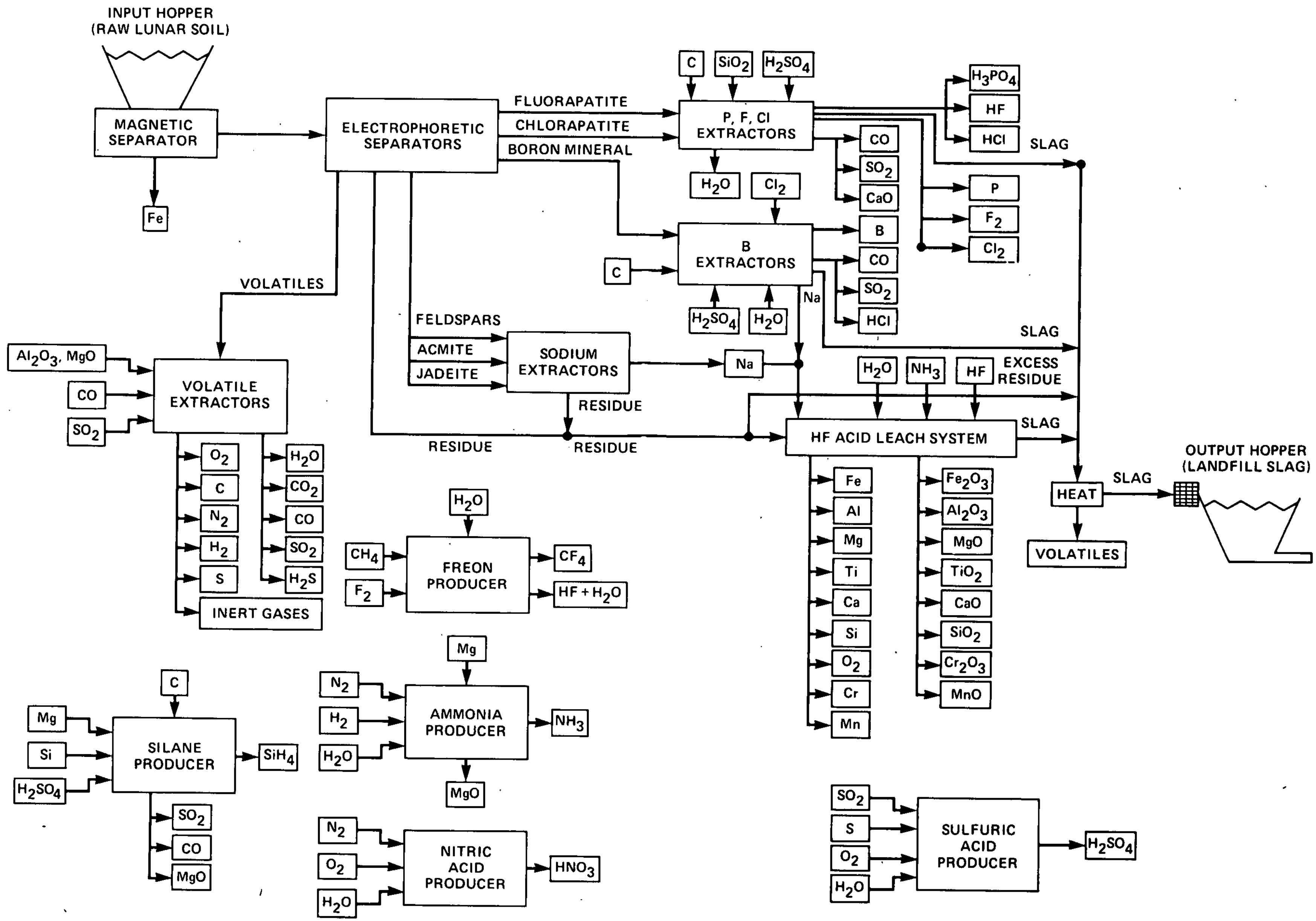


Figure 5.16.– LMF chemical processing sector: Operations.

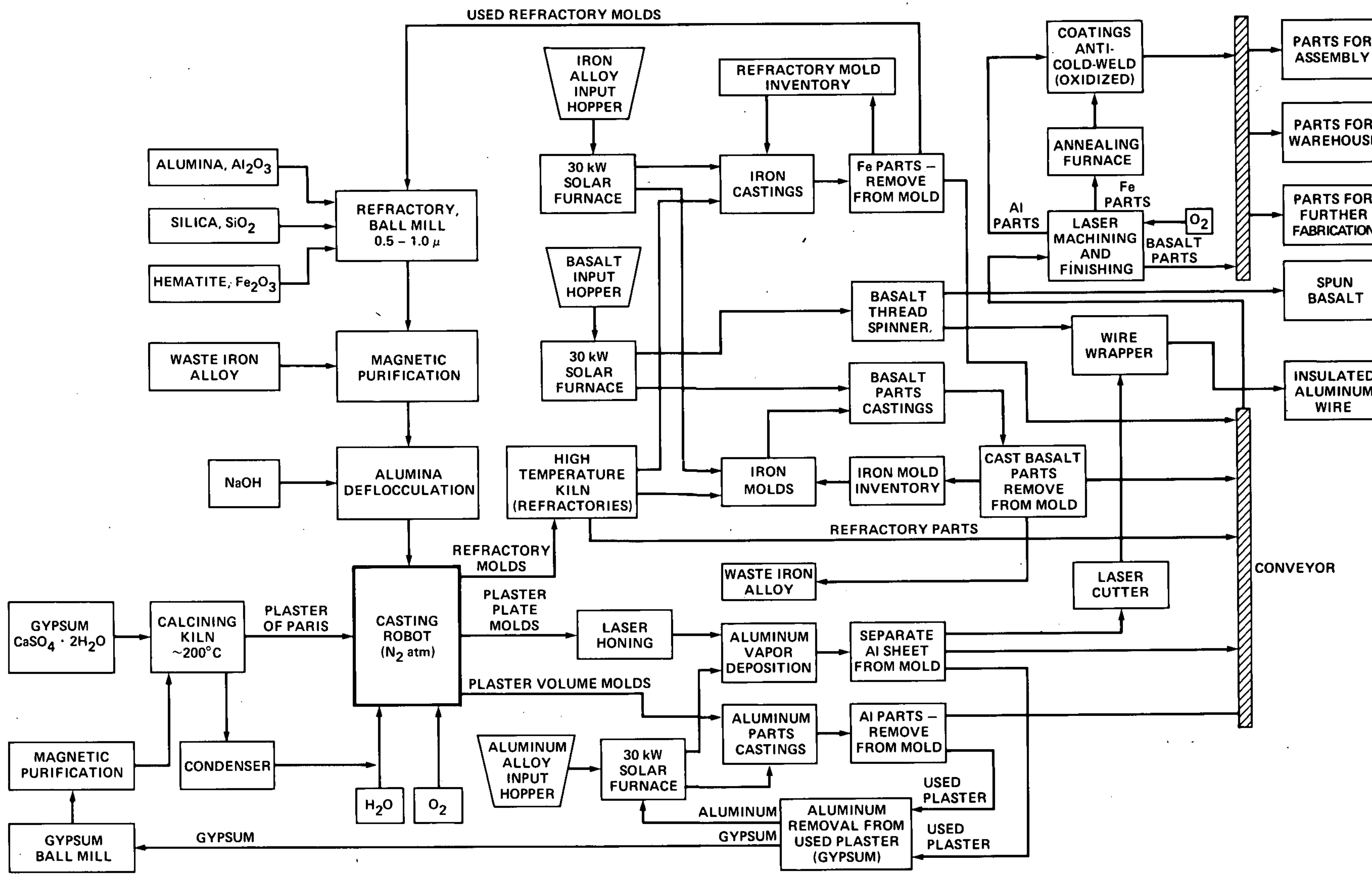


Figure 5.17.– LMF parts fabrication sector: Operations.

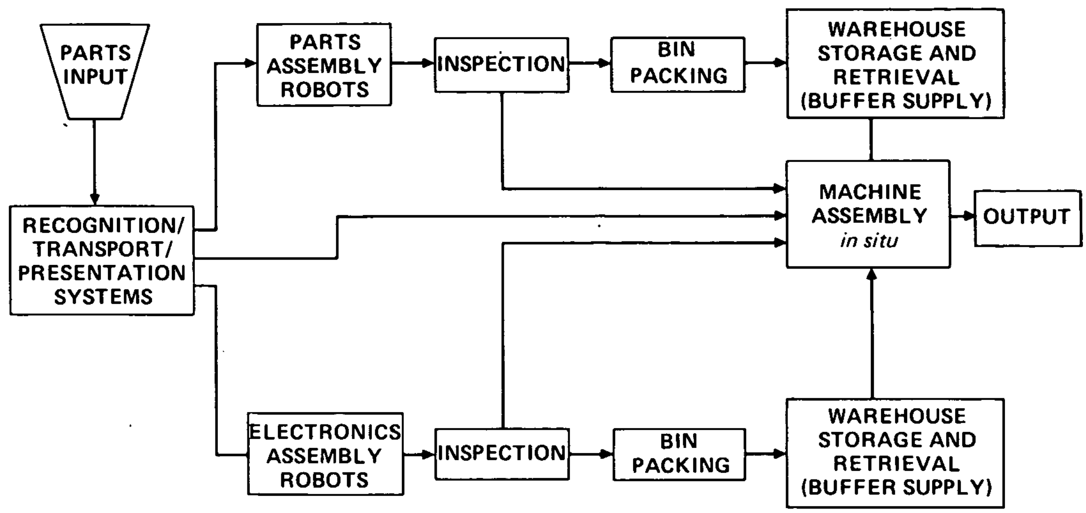


Figure 5.18.– LMF assembly sector: Operations.

In a practical sense, it is quite possible to imagine the lunar SRS operating nonautonomously (Johnsen, 1972). For instance, the *in situ* computer could be used simply as a teleoperation-management system for operations controlled directly by Earth-based workers. Material factory replication would proceed, but information necessary to accomplish this would be supplied from outside. An intermediate alternative would permit the on-site computer to handle mundane tasks and normal functions with humans retaining a higher-level supervisory role. Yet another possibility is that people might actually inhabit the machine factory and help it reproduce – manned machine economies can also self-replicate.

Solar canopy. The solar canopy is a "roof" of photovoltaic solar cells, suspended on a relatively flimsy support web of wires, crossbeams and columns perhaps 3-4 m above ground level. The canopy covers the entire LMF platform area and expands outward as the rest of the facility grows. The solar canopy and power grid provide all electrical power for LMF systems. Canopy components may be stationary or may track solar motions using heliostats if greater efficiency is required. A further discussion of canopy design and rationale may be found in appendix 5I.

Mass, power, and information requirements. Seed subsystem masses and power requirements scale according to the total system mass assumed. SRS can be reduced indefinitely in size until its components begin to scale nonlinearly. Once this physical or technological limit is reached for any subsystem component, comprehensive redesign of the entire factory may become necessary.

A seed mass of 100 tons was selected in the present study for a number of reasons. First, 100 tons is a credible system mass in terms of foreseeable NASA launch capabilities to the lunar surface, representing very roughly the lunar payload capacity of four Apollo missions to the Moon. Second, after performing the exercise of specifying seed components in some detail it is found that many subsystems are already approaching a nonlinear scaling regime for a 100-ton LMF. For instance, according to Criswell (1980, private communication) the minimum feasible size for a linear-scaling benchtop HF acid-leach plant for materials processing is about 1000 kg; in the present design, two such plants are required with a mass of 1250 kg each. Third, the results of a previous study (Freitas, 1980a) which argued the feasibility of 433-ton seed in the context of an interstellar mission (inherently far more challenging than a lunar factory mission) were compared with preliminary estimates of 15-107 tons for partially self-replicating lunar factories of several different types (O'Neill et al., 1980), and an intermediate trial value of 100 tons selected. The 100-ton figure has appeared in numerous public statements by former NASA Administrator Dr. Robert A. Frosch (lecture delivered at Commonwealth Club, San Francisco, Calif., 1979, and personal communication, 1980) and by others in prior studies (Bekey and Naugle, 1980; Giacconi et al., working paper of the Telefactors Working Group, Woods Hole New Directions Workshop, 1979). Finally, it was decided to use a specific system mass rather than unscaled relative component mass fractions to help develop intuitive understanding of a novel concept which has not been extensively studied before.

For reasons similar to the above, an SRS strawman replication time of 1 year was taken as appropriate. The ranges given in table 5.1, drawn from the analysis presented in appendixes 5B-5I, are estimates of the mass and power requirements of an initial seed system able to manufacture 100 tons of *all* of its own components per working year, hence, to self-replicate. These figures are consistent with the original estimate of a 100 ton circular LMF seed with an initial deployed diameter of 120 m, so feasibility has been at least tentatively demonstrated. However, it must be emphasized that the LMF seed design outlined above is intended primarily as a proof of principle. Numerical values for system components are only crude estimates of what ultimately must become a very complex and exacting design.

Information processing and storage requirements also have been collected and summarized in table 5.1, and lie within the state-of-the-art or foreseeable computer technologies. These calculations, though only rough approximations, quite likely overestimate real needs significantly because of the conservative nature of the assumptions employed. (See also sec. 5.2.3.)

SRS mission overview. In the most general case of fully autonomous operation, a typical LMF deployment scenario might involve the following initial sequence:

(1) The predetermined lunar landing site is mapped from orbit to 1-m resolution across the entire target ellipse.

(2) Seed lands on the Moon, as close to dead center of the mapped target area as possible navigationally.

(3) Mobile assembly and repair robots, assisted by mining robots, emerge from the landing pod and erect a small provisional solar array to provide interim power until the solar canopy is completed.

(4) LMF robots, with the computer, select the precise site where erection of the original seed will commence. This decision will already largely have been made based on orbital mapping data, but ground truth will help refine the estimate of the situation and adjust for unexpected variations.

(5) Mobile robots emplace the first three stations of the transponder network (the minimum necessary for triangulation), calibrate them carefully, and verify that the system is in good working order.

(6) Mining robots equipped with grading tools proceed to the construction site and level the local surface.

TABLE 5.1.– SEED MASS AND POWER REQUIREMENTS ESTIMATES

Seed subsystem	Estimated mass of 100 ton/yr seed, kg	Estimated power of 100 ton/yr seed, W	Computer processor, bits to operate	Computer memory, bits to describe
Transponder network	1,000	---	10^5 ?	10^6 ?
Paving robots	12,000	Up to 10^4	1-10$\times 10^6$	10^7-10^8
Mining robots	4,400	Up to 10^4	4-7$\times 10^8$	10^9
Chemical processing sector (S)	15,300-76,400	380,000-11,000,000	9.4$\times 10^7$	3.1$\times 10^9$
Fabrication sector (S)				
Electronics	(3,000)			
Floor map				(10^9)
Totals	137-20,400	270-345,000	10^{10}	10^{11}
Assembly sector (S)				
Assembly robots	83-1,150	83-19,600	10^9	10^{10}
Warehouse subsystem	1,000	10,000	10^7	10^8
Floor map				10^9
Automated transport vehicles	1,000	6,000	10^7	10^8
Mobile assembly and repair robots	4,000	40,000	4$\times 10^9$	4$\times 10^{10}$
Computer central	2,200	37,000	(1.6$\times 10^{10}$)	1.6$\times 10^{10}$
orbital site map				10^{11}
Solar canopy	22,000	---	2$\times 10^7$	2$\times 10^8$
Totals	63,100-145,600	0.47 MW-11.5 MW	15.5-15.8$\times 10^9$	272$\times 10^9$
Nominal annual seed output	100,000	1.7 MW		

(7) Five paving robots disembark and begin laying down the seed platform in square grids. This requires one working year for completion.

(8) When a sufficiently large platform section has been completed, seed mobile robots transfer the main computer to a place prepared for it at the center of the expanding platform disk.

(9) Erection of the solar canopy begins, followed by each of the seed sectors in turn, starting with the chemical processing. Total time to unpack the landing pod after moonfall is one working year, conducted in parallel with paving and other activities. The completed seed factory unit, unfurled to a 120 m diam on the surface of the Moon 1 year after landing, might appear as shown in figure 5.19.

The LMF has two primary operational phases – growth and production. The optimal program would probably be to "bootstrap" (grow) up to a production capacity matching current demand, then reconfigure for production until demand increases, thus necessitating yet further growth (O'Neill et al., 1980). Growth and production of useful output may proceed sequentially, cyclically, or simultaneously, though the former is preferred if large subsystems of the lunar factory must be reconfigured to accommodate the change.

The LMF also may exhibit replicative behavior if and when necessary. Replicas of the original seed could be constructed much like regular products and dispatched to remote areas, either to increase the total area easily subject to utilization or to avoid mortality due to depletion of local resources or physical catastrophes. The scheduling of factory operational phases is very flexible, as shown schematically in figure 5.20, and should be optimized for each mission and each intended use.

5.3.5 Lunar SRS Growth and Productivity

As the study progressed, the team noted a developing convergence between the two designs for SRS described in

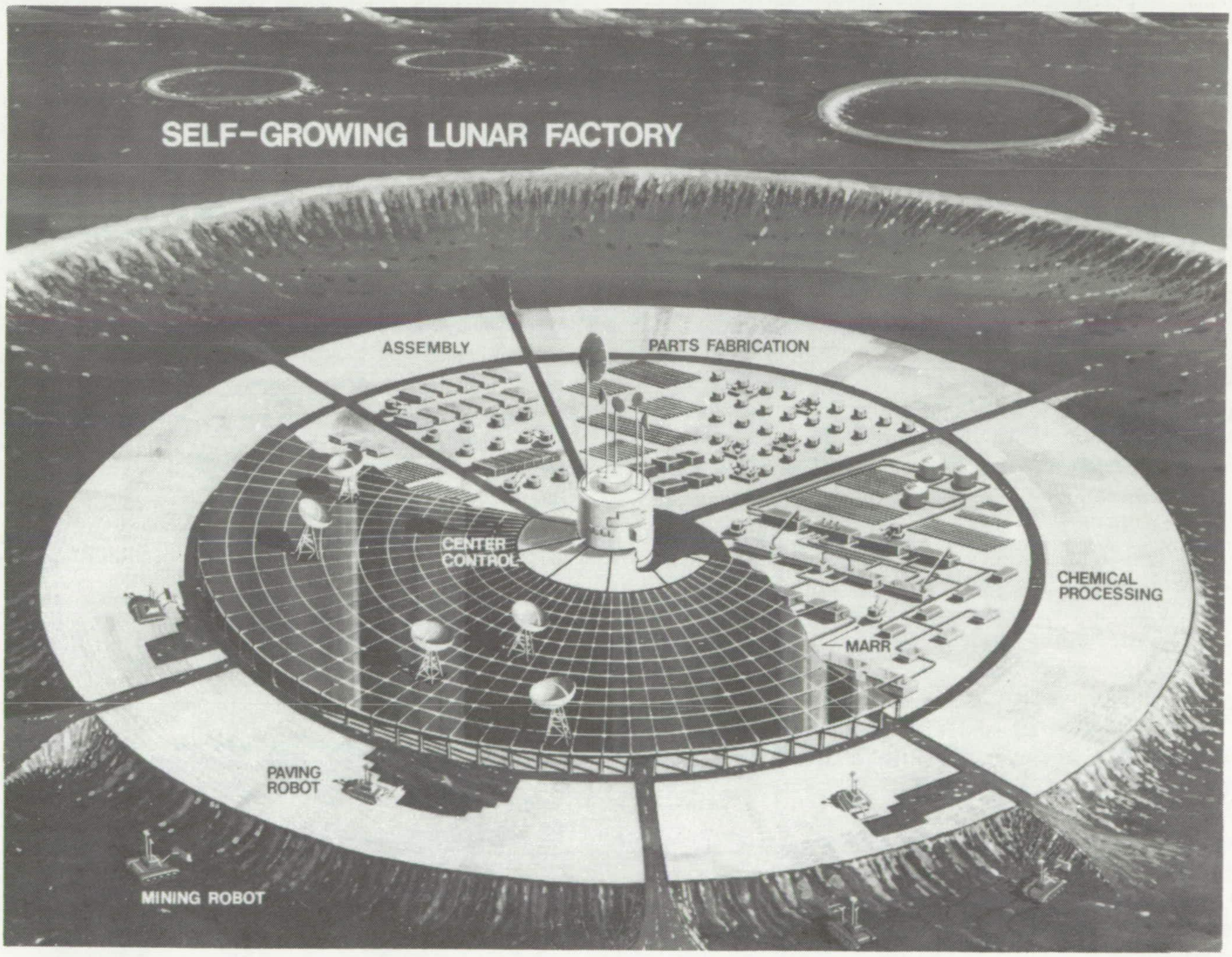


Figure 5.19.– Self-growing lunar factory.

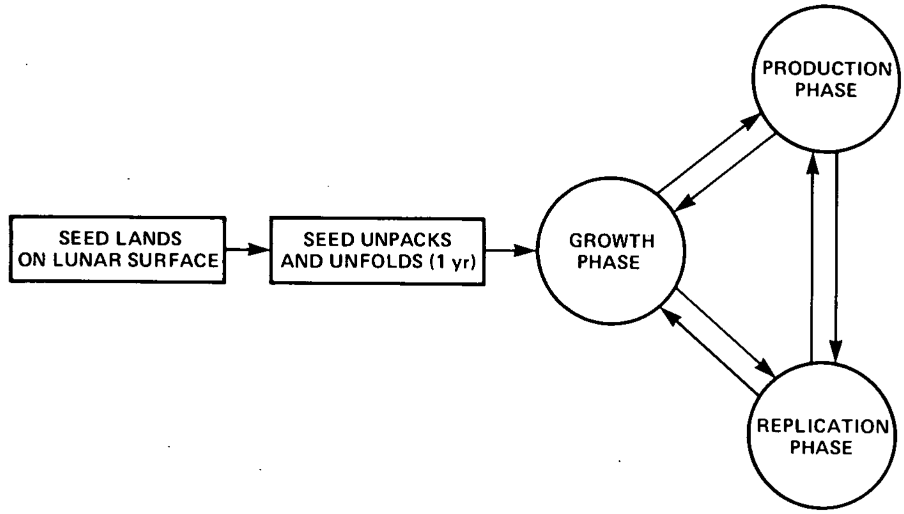


Figure 5.20.– Flexible scheduling of LMF operational phases.

sections 5.3.3 and 5.3.4. Both require three major subsystems – materials processing, fabrication, and assembly – plus a variety of support systems, and each is capable of replication and useful production. Both display exponential expansion patterns.

Of course, in a finite environment exponential growth cannot continue indefinitely. Geometrical arguments by Taneja and Walsh (1980, Summer Study document) suggest that planar packing of triangular, cubic, or hexagonal units can expand exponentially only for as many generations as each unit has sides, assuming that once all sides are used up no further doubling can occur by the enclosed unit. Growth is quadratic from that time on.

However, in real physical systems such as the developing LMF, enclosure need not preclude material communication with exterior units. Selected ramification of communication, control, and materials transportation channels or internal component rearrangement, reconfiguration, or specialization can prevent "starvation" in the inner regions of the expanding system. Hence, SRS exponential growth may continue until limited either by purposeful design or by the specific configuration of the external environment. Assuming that a 100-ton seed produces 100 tons/year of the same materials of which it is composed, then if T is elapsed time and N is number of seed units or seed mass-equivalents generated during this time, $T = 1 + \log_2 N$ for simple exponential "doubling" growth. (There is no replication in the first year, the time required for initial setup.) If P is productivity in tons/year, then $P = 100 \log_2 N$.

However, the above is valid only if each unit works only on its own replica. If two or more units cooperate in the construction of a single replica, still more rapid "fast exponential" growth is possible. This is because new complete replicas or LMF subsystems are brought on line sooner, and hence may begin contributing to the exponentiation earlier than before. Using the above notation, the "fast-exponential" growth rate is given by $T = 1 + 1/2 + ... + 1/N$ in the optimum case where all available machines contribute directly to the production of the next unit.

Growth rates and productivities are tabulated for exponential and "fast-exponential" expansion in table 5.2. Note that in just 10 years the output of such a facility could grow to approximately one million tons per year. If allowed to expand for 18 years without diversion to production, the factory output could exponentiate to more than 4×10^9 tons per year, roughly the entire annual industrial output of all human civilization.

Useful SRS products may include lunar soil thrown into orbit by mass drivers for orbital processing, construction projects, reaction mass for deep space missions, or as radiation shielding; processed chemicals and elements, such as oxygen to be used in space habitats, as fuel for interorbital vehicles, and as reaction mass for ion thrusters and mass drivers; metals and other feedstock ready-made for space construction or large orbital facilities for human occupation (scientific, commercial, recreational, and medical); components for large deep-space research vessels, radio telescopes, and large high-power satellites; complex devices such as machine shop equipment, integrated circuits, sophisticated electronics gear, or even autonomous robots, teleoperators, or any of their subassemblies; and solar cells, rocket fuels, solar sails, and mass driver subassemblies. Also, a 100-ton seed which has undergone thousand-fold growth or replication represents a 2 GW power generating capacity, plus a computer facility with a 16,000 Gbit processing capability and a total memory capacity of 272,000 Gbits. These should have many useful applications in both terrestrial and space industry.

5.3.6 Closure in Self-Replicating Systems

Fundamental to the problem of designing self-replicating systems is the issue of closure. In its broadest sense, this

TABLE 5.2.– GROWTH RATES AND PRODUCTIVITY FOR EXPONENTIAL SRS EXPANSION

Calendar years	Working years, T	Exponential growth, τ = 1 yr		"Fast-exponential" growth, τ = 1 yr	
		Number of units, N	System productivity, tons/yr	Number of units, N	System productivity, tons/yr
0	0	0	0	0	0
2	1	1	100	1	100
4	2	2	200	4	400
6	3	4	400	11	1,100
8	4	8	800	31	3,100
10	5	16	1,600	83	8,300
12	6	32	3,200	227	22,700
14	7	64	6,400	616	61,600
16	8	128	12,800	1,674	167,400
18	9	256	25,600	4,550	455,000
20	10	512	51,200	12,367	1,236,700
22	11	1,024	102,400	33,617	3,361,700
24	12	2,048	204,800	91,380	9,138,000
26	13	4,096	409,600	248,398	24,839,800
28	14	8,192	819,200	675,215	67,521,500
30	15	16,384	1,638,400	1,835,426	183,542,600
32	16	32,768	3,276,800	4,989,205	498,920,500
34	17	65,536	6,553,600	13,562,066	1,356,206,600
36	18	131,072	13,107,200	36,865,517	3,686,551,700
38	19	262,144	26,214,400	100,210,865	10,021,086,500
40	20	524,288	52,428,800	272,401,372	27,240,137,200 (~2 km-wide asteroid/yr)

(About 3 billion seed units would completely cover the entire lunar surface)

issue reduces to the following question: Does system *function* (e.g., factory output) equal or exceed system *structure* (e.g., factory components or input needs)? If the answer is negative, the system cannot independently fully replicate itself; if positive, such replication may be possible.

Consider, for example, the problem of parts closure. Imagine that the entire factory and all of its machines are broken down into their component parts. If the original factory cannot fabricate every one of these items, then parts closure does not exist and the system is not fully self-replicating.

In an arbitrary system there are three basic requirements to achieve closure:

(1) Matter closure – can the system manipulate matter in all ways necessary for complete self-construction?

(2) Energy closure – can the system generate sufficient energy and in the proper format to power the processes of self-construction?

(3) Information closure – can the system successfully command and control all processes required for complete self-construction?

Partial closure results in a system which is only partially self-replicating. Some vital matter, energy, or information must be provided from the outside or the machine system will fail to reproduce. For instance, various preliminary studies of the matter closure problem in connection with the possibility of "bootstrapping" in space manufacturing have concluded that 90-96% closure is attainable in specific nonreplicating production applications (Bock, 1979; Miller and Smith, 1979; O'Neill et al., 1980). The 4–10% that still must be supplied sometimes are called "vitamin parts." These might include hard-to-manufacture but lightweight items such as microelectronics components, ball bearings, precision instruments and others which may not be cost-effective to produce via automation off-Earth except in the longer term. To take another example, partial information closure would imply that factory-directive control or supervision is provided from the outside, perhaps (in the case of a lunar facility) from Earth-based computers programmed with human-supervised expert systems or from manned remote teleoperation control stations on Earth or in low Earth orbit.

The fraction of total necessary resources that must be supplied by some external agency has been dubbed the "Tukey Ratio" (Heer, 1980). Originally intended simply as an informal measure of basic materials closure, the most

logical form of the Tukey Ratio is computed by dividing the mass of the external supplies per unit time interval by the total mass of all inputs necessary to achieve self-replication. (This is actually the inverse of the original version of the ratio.) In a fully self-replicating system with no external inputs, the Tukey Ratio thus would be zero (0%).

It has been pointed out that if a system is "truly isolated in the thermodynamic sense and also perhaps in a more absolute sense (no exchange of information with the environment) then it cannot be self-replicating without violating the laws of thermodynamics" (Heer, 1980). While this is true, it should be noted that a system which achieves complete "closure" is not "closed" or "isolated" in the classical sense. Materials, energy, and information still flow into the system which is thermodynamically "open"; these flows are of indigenous origin and may be managed autonomously by the SRS itself without need for direct human intervention.

Closure theory. For replicating machine systems, complete closure is theoretically quite plausible; no fundamental or logical impossibilities have yet been identified. Indeed, in many areas automata theory already provides relatively unambiguous conclusions. For example, the theoretical capability of machines to perform "universal computation" and "universal construction" can be demonstrated with mathematical rigor (Turing, 1936; von Neumann, 1966; see also sec. 5.2), so parts assembly closure is certainly theoretically possible.

An approach to the problem of closure in real engineering systems is to begin with the issue of parts closure by asking the question: can a set of machines produce all of its elements? If the manufacture of each part requires, on average, the addition of $\geqslant 1$ new parts to product it, then an infinite number of parts are required in the initial system and complete closure cannot be achieved. On the other hand, if the mean number of new parts per original part is <1, then the design sequence converges to some finite ensemble of elements and bounded replication becomes possible.

The central theoretical issue is: can a real machine system itself produce and assemble *all* the kinds of parts of which it is comprised? In our generalized terrestrial industrial economy manned by humans the answer clearly is yes, since "the set of machines which make all other machines is a subset of the set of all machines" (Freitas et al., 1981). In space a few percent of total system mass could feasibly be supplied from Earth-based manufacturers as "vitamin parts." Alternatively, the system could be designed with components of very limited complexity (Heer, 1980). The minimum size of a self-sufficient "machine economy" remains unknown.

Von Tiesenhausen and Darbro (1980) similarly argue that a finite set of machines can produce any machine element. Their reasoning, outlined in figure 5.21, is as follows:

(1) If all existing machines were disassembled into their individual parts there would obviously be a finite number of parts, many of them identical, and a large number would be of common categories like shafts, motors, wiring, etc. The only differences between the machines would be a different selection, different arrangement, and different dimensions of this finite number of parts.

(2) A finite number of parts involves a finite number of machine operations, this number being less than the number of parts because some machines can make more than one kind of parts.

(3) Therefore, the number of machines is finite and less than the number of operations.

This reasoning can then be generalized to say: "Every existing machine can be reduced to a finite set of machine elements, and there exists a finite set of machine operations." (Still, of course, a limited number of standard elements should be developed and machine operations limited as

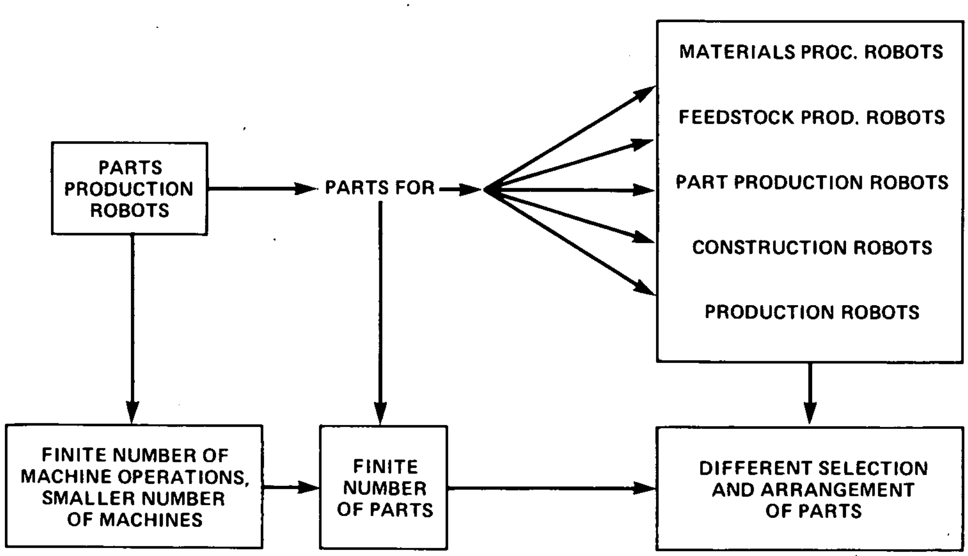


Figure 5.21.– Closure of SRS parts production.

much as practical by substitution, in order to minimize the number of parts and machine operations.)

Similar arguments may be applied to materials processing and feedstock production. There exists a finite number of different materials anywhere. There is a finite number of materials processes which is less than the number of materials because single processes result in various materials (e.g., silicon and oxygen). Hence, there is a finite number of materials processing robot systems needed for an SRS. Also, there is a finite and rather limited number of feedstock requirements such as bars, rods, ingots, plates, etc. The number of materials is much less than the number of parts; therefore, a finite number of parts fabrication robots is required for an SRS.

Closure engineering. In actual practice, the achievement of full closure will be a highly complicated, iterative engineering design process. Every factory system, subsystem, component structure, and input requirement (Miller and Smith, 1979) must be carefully matched against known factory output capabilities. Any gaps in the manufacturing flow must be filled by the introduction of additional machines, whose own construction and operation may create new gaps requiring the introduction of still more machines.

The team developed a simple iterative procedure for generating designs for engineering systems which display complete closure. The procedure must be cumulatively iterated, first to achieve closure starting from some initial design, then again to eliminate overclosure to obtain an optimally efficient design. Each cycle is broken down into a succession of subiterations which ensure three additional dimensions of closure:

(1) Qualitative closure – can, say, all parts be made?

(2) Quantitative closure – can, say, enough parts be made?

(3) Throughput closure – can parts be made fast enough?

In addition, each subiteration sequence is further decomposed into design cycles for each factory subsystem or component, as shown in figure 5.22.

The procedure as outlined, though workable in theory, appears cumbersome. Further work should be done in an attempt to devise a more streamlined, elegant approach.

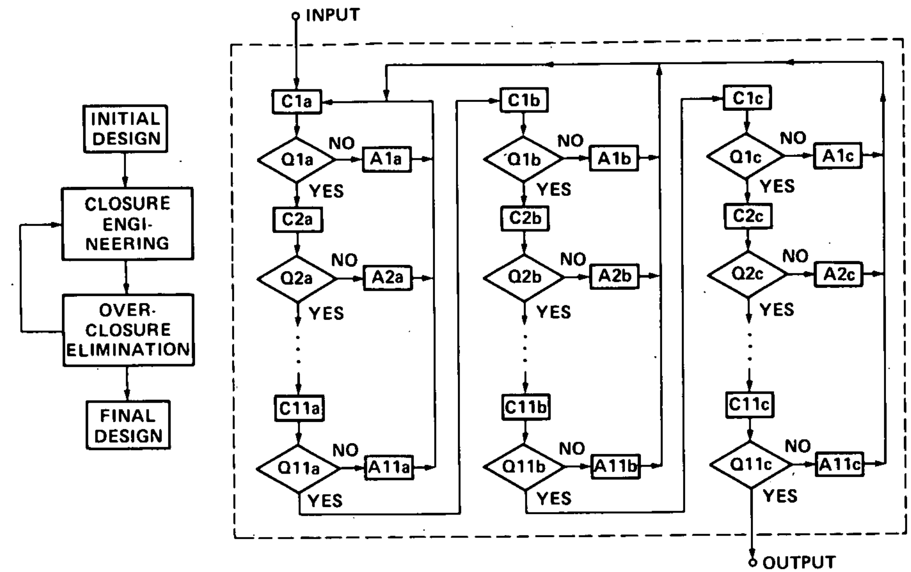
Quantitative materials closure – numerical results. In the context of materials processing, "closure" is a relationship between a given machine design and a given particular substrate from which the machine's elemental chemical constituents are to be drawn. Hence the numerical demonstration of closure requires a knowledge of the precise composition both of the intended base substrate to be utilized and of the products which the SRS must manufacture from that substrate. Following a method suggested by the work of Freitas (1980a), a modified "extraction ratio" R_n is defined as the mass of raw substrate material which must be processed (input stream) to obtain a unit mass of useful system output having the desired mass fraction of element n (output stream).

Consider the significance of the extraction ratio to the problem of materials closure. Assume that the final product is to be composed of elements x, y, and z. An $R_x = 1$ means that 1 kg of lunar soil contains exactly the mass of element x needed in the manufacture of 1 kg of the desired output product. On the other hand, $R_y = 10$ means that 10 kg of lunar regolith must be processed to extract all of element y required in 1 kg of final product. The difference between R_x and R_y may signify that y is more rare in lunar soil than x, or that the two elements are equally abundant but ten times more y than x is required (by weight) in the final product. When the output stream is identical to the machine processing system itself, then the system is manufacturing more of itself -- self-replicating – and the extraction ratio becomes an index of system materials closure on an element-by-element basis.

The total net extraction ratio R is some function of the individual extraction ratios R_n, and depends on the methods of materials processing employed. At worst, if only one element is recovered from a given mass of input stream ("parallel processing"), then R is the sum of all R_n. At best, if the input stream is processed sequentially to extract all desired elements in the necessary amounts ("serial processing"), then R is driven solely by the R_n of the element most difficult to extract, say, element z. That is, $R = (R_n)_{max} = R_z$, which is always equal to or smaller than the sum of all R_n. As serial processing should dominate in the lunar factory the latter formula is assumed for purposes of the present calculations. Note that R_n can be less than 1 for individual elements, but for an entire machine system R must always be greater than or equal to 1.

As a general rule, a low value for R implies that the system is designed for low mass throughput rates and is built from relatively few different chemical elements. A high value of R implies that many more elements are necessary and that a higher mass throughput rate will be accommodated to obtain them.

The "closure" of a given output stream (product) relative to a specified input stream (substrate) is computed by treating R as an independent variable. If I_n is the concentration of element n in mineral form in the input stream of lunar soil (kg/kg), E_n is the efficiency of chemical extraction of pure element n from its mineral form which is present in lunar soil (kg/kg), and O_n is the concentration of element n in the desired factory output stream (kg/kg), then $R_n = O_n/E_nI_n$. Closure C_n for each element is defined as the mass of pure element n available in a system with a total net extraction ratio R per unit mass of output stream. For any given element, if $R \geqslant R_n$ then all pure element n needed is already available within the system. In this case, $C_n = O_n$. On the other hand, if $R < R_n$ then the choice of R is too low; all the pure element n needed cannot be



EXPLANATION OF SYMBOLS

PROCEDURAL QUERIES = Q

Q1a – CAN SYSTEM ASSEMBLE ALL MACHINES OF WHICH IT IS COMPRISED?
Q2a – CAN SYSTEM ASSEMBLE ALL SUBASSEMBLIES OF WHICH ITS MACHINES ARE COMPRISED?
Q3a – CAN SYSTEM MACHINE ALL PARTS OF WHICH ITS MACHINES ARE COMPRISED?
Q4a – CAN SYSTEM EXTRACT ALL MATERIALS OF WHICH ITS PARTS ARE COMPRISED?
Q5a – CAN SYSTEM MINE ALL MINERALS OF WHICH ITS MATERIALS ARE COMPRISED?
Q6a – CAN SYSTEM TRANSPORT ALL (MACHINES/SUBASSEMBLIES/PARTS/MATERIALS/MINERALS) OF WHICH IT IS COMPRISED?
Q7a – CAN SYSTEM VERIFY ALL (MACHINES/SUBASSEMBLIES/PARTS/MATERIALS/MINERALS) OF WHICH IT IS COMPRISED?
Q8a – CAN SYSTEM WAREHOUSE ALL (MACHINES/SUBASSEMBLIES/PARTS/MATERIALS/MINERALS) OF WHICH IT IS COMPRISED?
Q9a – CAN SYSTEM REPAIR ALL SUBSYSTEMS OF WHICH IT IS COMPRISED?
Q10a – CAN SYSTEM COMPUTERS CONTROL ALL SUBSYSTEMS OF WHICH IT IS COMPRISED?
Q11a – CAN SYSTEM ENERGY PLANT ENERGIZE ALL SUBSYSTEMS OF WHICH IT (THE SYSTEM) IS COMPRISED?

Q1b – CAN SYSTEM ASSEMBLE ENOUGH MACHINES TO REPLICATE THE ORIGINAL SYSTEM?
Q2b – CAN SYSTEM ASSEMBLE ENOUGH SUBASSEMBLIES TO REPLICATE THE ORIGINAL SYSTEM?
⋮
Q11b – CAN SYSTEM ENERGY PLANT PRODUCE ENOUGH ENERGY TO PERMIT REPLICATION OF THE ORIGINAL SYSTEM?

Q1c – CAN SYSTEM ASSEMBLE MACHINES FAST ENOUGH TO REPLICATE THE ORIGINAL SYSTEM WITHIN ESTABLISHED TIME CONSTRAINTS?
⋮
Q11c – CAN SYSTEM ENERGY PLANT PRODUCE ENOUGH ENERGY FAST ENOUGH (HIGH ENOUGH *POWER*) TO REPLICATE THE ORIGINAL SYSTEM WITHIN ESTABLISHED TIME CONSTRAINTS?

SYSTEM COMPONENTS LISTS = C

C1a, b, c – MACHINES
C2a, b, c – SUBASSEMBLIES
C3a, b, c – PARTS
C4a, b, c – MATERIALS
C5a, b, c – MINERALS (SUBSTRATE)
C6a, b, c – TRANSPORTATION FACILITIES
C7a, b, c – VERIFICATION FACILITIES
C8a, b, c – STORAGE FACILITIES
C9a, b, c – REPAIR FACILITIES
C10a, b, c – COMPUTER FACILITIES
C11a, b, c – ENERGY FACILITIES

SYSTEM COMPONENTS ACTION LIST = A

A1a, b, c – NEW MACHINES
A2a, b, c – NEW SUBASSEMBLY-MAKING MACHINES
⋮
A11a, b, c – NEW ENERGY FACILITY-MAKING MACHINES
A#a – Add new machines to make . . .
A#b – Add new machines to make . . .
A#c – increase replication time available or change machine design for . . .

Figure 5.22.– Generalized closure engineering design cycles.

recovered, and more lunar soil must be processed to make up the difference if 100% closure is to be achieved. In this case, $C_n = O_n(R/R_n)$, since the closure deficit is measured by the ratio of the chosen R to the actual R_n of the given element (i.e., how much the factory has, divided by how much the factory actually needs). Total net system closure C is simply the sum of all C_n for all elements n required in the output stream of the SRS factory (Freitas and Zachary, 1981).

To estimate the quantitative materials closure for the lunar SRS baseline designs proposed in sections 5.3.3 and 5.3.4, three different approaches were taken in an attempt to converge on a useful estimate of the composition of the output stream necessary for LMF self-replication. First, the "seed" element distribution given by Freitas (1980a) in the context of a self-reproducing exploratory spaceprobe was adopted. These figures are derived from published data on the material consumption of the United States (the world's largest factory) during the years 1972-1976 (U.S. Bureau of Mines, 1978; U.S. Bureau of the Census, 1977, 1978). A second but less comprehensive measure called "demandite" is based on 1968 U.S. consumption data (Goeller and Weinberg, 1976). A molecule of "nonfuel demandite" is the average nonrenewable resource used by humans, less fuel resources (Waldron et al., 1979). Third, the direct estimate of LMF elemental composition presented in appendix 5E was used to obtain additional trial values for O_n. (Appendix 5E also represents a first attempt to deal with qualitative materials closure for SRS.) In all cases the input stream was assumed to consist of lunar maria regolith, with values for I_n averaged from published data (Phinney et al., 1977) and listed in table 5.3. Following earlier work, for simplicity all efficiencies E_n were taken to be 0.93 (Rao et al., 1979; Williams et al., 1979).

The closures calculated from these data are plotted against extraction ratio in figure 5.23. (Data for the human body are included for purposes of comparison.) Note that 100% closure (C = 1) is achieved for the "U.S. Industrial" estimate (84 elements of the spaceprobe "seed") at R = 2984; for "Demandite" (28 elements) at R = 1631; and for the appendix 5E "LMF" (18 elements) at R = 45. This suggests that the fewer the number of different elements, and the more common and more efficiently extractable are the elements the factory system needs for replication to occur, the lower will be the total mass of raw materials which must be processed by the LMF.

Note also that in all three cases, virtually complete (>90%) closure is achieved for extraction ratios of 2 to 14. The incremental gains in closure after 90% are purchased only at great price – from 1 to 3 orders of magnitude more raw materials mass must be processed to achieve the last bit of full materials autonomy. Two conclusions may be drawn from this observation. First, for any given SRS design it may well be more economical to settle for 90-95% system closure and then import the remaining 5-10% as "vitamins" from Earth. Second, in those applications where 100% closure (full materials autonomy) is desirable or required, great care must be taken to engineer the self-replicating system to match the expected input substrate as closely as possible. This demands, in the case of quantitative materials closure, a design which minimizes the value of R, thus optimizing the use of abundantly available, easily extractable elements.

TABLE 5.3.– AVERAGE CHEMICAL ELEMENT ABUNDANCES IN LUNAR MARIA

Element	Abundance	Element	Abundance
Al	6.80%	Ho	3.73 ppm
Ca	7.88%	I	2.00 ppb
Cr	0.264%	In	32.9 ppb
Fe	13.2%	Ir	6.32 ppb
K	0.113%	La	17.2 ppm
Mg	5.76%	Li	12.9 ppm
Mn	0.174%	Lu	1.22 ppm
Na	0.290%	Mo	0.520 ppm
O	41.3%	N	95.4 ppm
P	0.066%	Nb	19.6 ppm
S	0.125%	Nd	38.2 ppm
Si	20.4%	Ne	2.75 ppm
Ti	3.10%	Ni	169 ppm
		Os	12.9 ppb
Ag	45.2 ppb	Pb	3.11 ppm
Ar	0.800 ppm	Pd	12.3 ppb
As	0.206 ppm	Pr	7.20 ppm
Au	2.66 ppb	Rb	3.21 ppm
B	4.78 ppm	Re	1.36 ppb
Ba	195 ppm	Rh	0.192 ppm
Be	2.63 ppm	Ru	0.231 ppm
Bi	3.19 ppb	Sb	22.1 ppb
Br	0.178 ppm	Sc	48.8 ppm
C	104 ppm	Se	0.306 ppm
Cd	0.197 ppm	Sm	10.9 ppm
Ce	48.8 ppm	Sn	0.900 ppm
Cl	25.6 ppm	Sr	167 ppm
Co	40.3 ppm	Ta	1.26 ppm
Cs	0.392 ppm	Tb	2.58 ppm
Cu	14.4 ppm	Te	0.0545 ppm
Dy	15.3 ppm	Th	2.50 ppm
Er	9.24 ppm	Tl	1.61 ppb
Eu	1.77 ppm	Tm	1.42 ppm
F	174 ppm	U	0.805 ppm
Ga	4.99 ppm	V	114 ppm
Gd	14.3 ppm	W	0.358 ppm
Ge	0.626 ppm	Y	84.2 ppm
H	54.8 ppm	Yb	8.40 ppm
He	28.5 ppm	Zn	23.4 ppm
Hf	7.77 ppm	Zr	311 ppm
Hg	0.019 ppm		

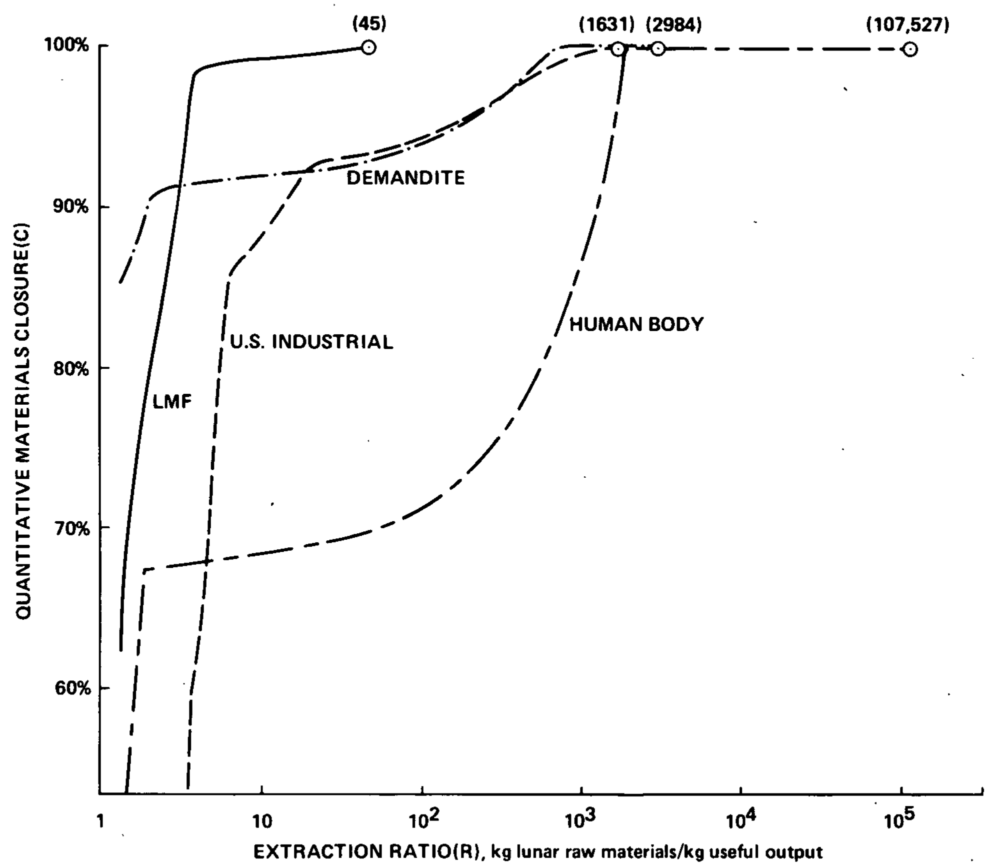


Figure 5.23.– Quantitative materials closure data for various self-replicating systems.

5.3.7 Conclusions

The team reached the following major conclusions regarding the feasibility of self-replicating machine systems:

- The basic concept of physical machine systems capable of self-replication appears credible both from a theoretical and a practical engineering standpoint.
- It is reasonable to begin designing replicating systems based on current knowledge and state-of-the-art technology, but final design definition will require significant further research.
- Complete systems closure is achievable in principle, though partial closure may be more feasible from an economic and pragmatic engineering standpoint in the near term.
- It is feasible to begin immediate work on the development of a simple demonstration SRS on a laboratory scale, with phased steps to more sophisticated levels as the technology is proven and matures.

5.4 Applications

Having shown that machine SRS is, in principle, both theoretically possible and feasible in terms of engineering systems design, their usefulness in some economic or commercial sense remains to be demonstrated. That is, what might such systems permit humankind to do that could not be done before? The main advantage in using SRS over other methods of space exploration and industrialization is that a very large capability for performing any desired task can be rapidly achieved at arbitrary remote locations, starting with a relatively small investment of time, money, energy, and mass in the original "seed" mechanism.

The team has identified four general criteria for determining the most probable and profitable application of replicating systems technology:

(1) A large number of identical or similar products is required;

(2) Excessively long production periods for alternate approaches are required;

(3) Raw materials or parts are available onsite; and

(4) Sufficient physical space is available for replication.

Each of these criteria should be applicable, or largely so, in a specific case before the use of SRS technologies is considered.

Replicating systems will find many applications on Earth, in near-Earth and lunar space, throughout the Solar System, and in the interstellar realm, for both exploration and utilization. SRS also provides a number of fascinating applications in basic and applied research in automata theory, theoretical biology, experimental evolution, and machine intelligence and robotics architecture.

5.4.1 Terrestrial Applications

The early development of replicating systems technology on Earth will be the history of modern industrial automation. The United States at one time enjoyed the highest productivity in the world, and still partakes of the prosperity that that has brought. Recently, however, competition from other nations who are more rapidly automating their industries is seriously eroding the U.S. position of leadership. The resulting economic forces are impelling domestic industry to accelerate the automation of its factories.

The space program is viewed by many as a high-technology venture which predominantly makes use of computers, robot spacecraft, and other trappings of automation. In reality, NASA's activities are strongly people-intensive. For example, large teams of trained technicians and scientists are required to operate a robot space probe by remote control. The same economic forces at work in the marketplace are forcing NASA to rethink its traditional way of doing business. Not only will there be more automation in the space program for this reason, but also there will be missions that are difficult or impossible to conduct without using advanced machine intelligence and robotics technologies.

The harsh environment of space, the significant costs of life support systems for human beings and of "man-rating" space systems for safety, and the communications problems caused by the immense distances involved in interplanetary travel have given NASA additional incentives to develop systems of total automation beyond those commonly employed in industry. The sheer magnitude of many potentially interesting missions requires massive automation. Accordingly, NASA should strongly participate in automation research and development in anticipation of spinoffs to industry of great potential value. The agency also should closely monitor industrial R&D efforts, remaining alert for new developments on the commercial front which might prove beneficial to the space program. The infusion of NASA funds at critical points could allow the agency to exert subtle influence on industrial development so as to provide for NASA's special needs at less cost than an independent program to achieve the same ends.

Similarly, the Department of Defense (DOD) is embarking upon an ambitious program of industrial automation. The aim is to produce war materiel in the most economical and flexible manner possible, and to shorten the time between concept and field deployment of weapons systems. Much of the DOD effort will produce results useful for the space program. To take maximum advantage of this, NASA should maintain close liaison with DOD and should join in various cooperative efforts in areas of overlapping interests.

Computer-aided design (CAD), manufacturing (CAM), and testing (CAT), and robotics. Automation for replication will require extensive application of computer science and robotics. At the initial stage of development, and during periods when repair or reconstruction operations must be performed, computers can be used in many ways to aid the design process (CAD). They are excellent for generating and maintaining documentation. Computer-executed graphics are invaluable in assisting human operators to visualize complex objects in the absence of a real, physical construction. Simulation using computer models is used in place of, or as a cost-saving adjunct to, physical models or prototypes. Recent developments in machine intelligence research has made far easier the complete automation of the entire design process. Ultimately, the capability will exist for a human to carry on a dialog with a computer system in which the person merely defines the functional specifications of the desired product and the computer determines the remaining design details autonomously.

Computers have been used in manufacturing (CAM) for more than two decades. The most common modern application is business data processing. Computerized inventory control and scheduling are two promising uses rapidly gaining prominence today. Process control using analog computers began many years ago in chemical plants, steel mills, and paper mills. Newer facilities rely instead upon digital computing. An important subset of process control is numerical control (N/C) of machine tools, with instructions traditionally recorded on punched paper tape. Today it is feasible to connect N/C machine tools directly to a computer able to generate and store instructions in electronic memory, and increasingly this is being done, especially in the aerospace industry.

Computers can also be used to great advantage in the testing of products (CAT). (This is distinguished from measurements of process variables, which is considered a process control function.) Highly complex products such as microprocessor integrated circuits cannot realistically be tested without the aid of computer technology. A standard interface protocol (the IEEE-488 bus) has been defined for the interfacing of test instrumentation to a host computer.

In the context of a factory, robotics generally is understood to refer to materials handling and assembly functions. Typical operations include loading/unloading machine tools and spot-welding automobile bodies. Hard automation (special-purpose robots of very limited versatility) commonly are used in applications requiring high volume output. But computer-controlled general-purpose robot manipulators are becoming increasingly popular, as exemplified by the rather anthropomorphic PUMA device (a robot arm system manufactured by Unimation).

Replicative automation. CAD, CAM, CAT, and robotics technologies could be combined to produce an almost totally automated factory. The Department of Defense has instituted an ongoing program designed to promote this very concept, called Integrated Computer-Aided Manufacturing or ICAM. The technology now exists to design integrated circuits in one location (CAD), then fabricate the masks for microelectronic manufacture in another (CAM) under the direction of several intercommunicating computers. Further developments and advances in ICAM techniques are imminent.

In a very real sense, an industrialized nation is a symbiotic self-replicating, growing "organism" consisting of humans and machines working together. At the beginning of the industrial revolution the "organism" consisted chiefly of human beings, who, aided by a few machines, performed logical and physical functions. In later years more and more of the heavy and most dangerous work was delegated to machines. As ICAM increasingly enters the mainstream of industrial automation, the logical processes of man-machine manufacturing "organisms" will begin to be taken over by sophisticated computer systems and the physical functions will be dominated by commercial robot devices.

When ICAM techniques are directed toward the production of components of their own systems (CAD, CAM, CAT, and robot machines), a regenerative effect occurs in which each generation of automated factories is cheaper to construct than the preceding one. By the time this regeneration, which has been termed "superautomation" (Albus, 1976), is achieved on Earth, there may be very little human intervention in the replication process except for supervisory and top-level guidance functions. The final step in achieving totally autonomous machine replication requires the replacement of the human top-level managers with computers and turning over any remaining physical tasks to robot devices.

The near-term removal of all human intervention from the industrial "organisms" on Earth is highly unlikely. Certainly people may want to continue to perform various logical and physical functions for social or psychological reasons, and man may always remain the decisionmaker in control of which products are produced. Certain tasks are likely to prove more difficult to automate than expected, and human beings will continue to perform these jobs for economic reasons for a long time to come. Superautomation on Earth will proceed only as far and as fast as is economically advantageous.

The long-term future almost certainly will see the development of full replicative automation capability on Earth. Whether it is economical remains an open question at present. The main advantage of pure machine replicating systems over man-machine symbiotic systems is that autonomous factories can be sent to locations where there is not, or cannot be without great expense, a population of human workers adequate to operate and maintain the factory complex.

Prime candidates for terrestrial replicating systems applications will most likely be mass-produced products for use in inaccessible or hostile places requiring large spaces to perform the specified tasks. Possibilities include large photovoltaic arrays for centralized power plants in the southwestern regions of the United States (Leonard, in-house document, Bechtel Natl. Inc., San Francisco, Calif., 1980), desert irrigation and soil conditioning equipment covering vast areas, agricultural or military robots, ocean-bottom roving mineral retrievers and seawater extractors patrolling the vast continental shelves, or solar-power satellite ground receiver (rectennae) devices. Each of these machine systems could probably be made to self-replicate from a basic feedstock substrate, possibly even from a raw material substrate ultimately.

A few somewhat more speculative terrestrial applications have been proposed by imaginative writers. For instance, Moore (1956) suggested the idea of an artificial living plant able to extract its own nutrients from the sea. These machines could obtain energy from sunlight to refine and purify materials, manufacture them into parts, and then assemble the parts to make duplicates of themselves. Such plants could be harvested for a material they extracted or synthesized. Thus, an artificial plant which used magnesium as its chief structural material could be cannibalized for its metal content. Like lemmings, schools of artificial living machines could be programmed to swim to a harvesting factory when they reached adulthood.

Clearly there would be need for international controls and allocation of areas for production and harvesting. This would involve not only the political rights of nations but also questions of natural conservation. Social problems could arise in connection with the selection of products to be manufactured. An artificial plant might be designed to make a product useless to the plant itself. It might extract gold from seawater, refine it, and cast it into an ingot, which would be harvested as the crop from the plant. But this would be a shortsighted choice. Multiplying at an exponential rate, the gold-making plant would soon produce so much that gold would lose its scarcity value and probably end up being worth very little. An excellent candidate for production by an artificial plant is fresh water, which is needed in great quantities in various parts of the world.

Dyson (1979) suggests a small self-reproducing automaton well adapted to function in terrestrial deserts. It builds itself mainly out of silicon and aluminum which it extracts from ordinary rocks wherever it happens to be. Its source of energy is sunlight, its output electricity and high-tension transmission lines. There is bitter debate in Congress over licensing this machine to proliferate over our Western states. The progeny of one robot can easily produce ten times the present total power output of the United States. Legislation is finally passed authorizing the automaton to

multiply, with the proviso that each machine shall retain a memory of the original landscape at its site, and if for any reason the site is abandoned the device is programmed to restore it to its original appearance.

After its success with the rock-eating automaton in the United States, the company places on the market an industrial development kit, designed for the needs of developing countries. For a small down payment, a country can buy an egg machine which will mature within a few years into a complete system of basic industries together with the associated transportation and communication networks, custom made to suit the specifications of the purchaser. The vendor's guarantee is conditional only on the purchaser's excluding human population from the construction area during the period of growth. After the system is complete, the purchaser is free to interfere with its operation or to modify it as he sees fit. (A technological spinoff is the Urban Renewal Kit – a city's architects and planners work out a design for urban rebuilding, then the kit is programmed to do the job for a fixed fee.)

Theodore Taylor calls all such devices "Santa Claus Machines" because of their almost "magical" behavior (Calder, 1978). In his version of SRS, a fully automatic mining, refining, and manufacturing facility gathers scoopfuls of raw lunar materials and then processes them by means of a giant mass spectrograph with huge superconducting magnets. This device converts mined material into an ionized atomic beam which is deflected by the magnetic field. Lighter elements curve more than heavier atomic species, so the material is sorted into stockpiles of constituent elements atom by atom. To manufacture any item, the Santa Claus Machine selects the necessary metals and plastics, then vaporizes and sprays them onto a mold. Instructions for manufacturing, including directions for adapting to new processes and replication, are stored on magnetic tapes in the machine, perhaps activated by radio command from Earth. Conceivably, costs eventually could fall to zero; and if the workload grows too large, the machine simply reproduces itself.

5.4.2 Near-Earth and Lunar Space Applications

While terrestrial self-replicating systems may be limited for some time to coevolution with Earth-based industry constrained by normal economic factors, the prospect for extraterrestrial applications is quite different. The difficulty of surmounting the Earth's gravitational potential makes it more efficient to consider sending information in preference to matter into space whenever possible. Once a small number of self-replicating facilities has been established in space, each able to feed upon nonterrestrial materials, further exports of mass from Earth will dwindle and eventually cease. The replicative feature is unique in its ability to grow, *in situ,* a vastly larger production facility than could reasonably be transported from Earth. Thus, the time required to organize extraordinarily large amounts of mass in space and to set up and perform various ambitious future missions can be greatly shortened by using a self-replicating factory that expands to the desired manufacturing capacity.

In the not-too-distant future such facilities could be sited either in Earth or lunar orbit, or on the surface of the Moon. The chief advantages of orbital factories are near-zero gravity, absence of lunar dust or atmosphere, convenience in choice of orbit, proximity to Earth (relative ease of transport of finished products), and unobstructed view of virtually the entire celestial sphere. For some applications, however, the lunar surface may be the preferred location. Many manufacturing processes require at least small amounts of gravity, and the availability of solid ground for physical support may be important too. The main advantage to factories on the lunar surface is that the raw materials to be processed into finished products are right at hand – only relatively low-mass final products need be lifted from the lunar surface, rather than bulky raw materials as in the case of an orbital factory. The Moon can also be used as a shield to block sunlight or electromagnetic interference from Earth during highly sensitive observations.

The useful applications of replicating factories with facilities for manufacturing products other than their own components are virtually limitless.

Manufacturing. Huge solar power satellites with dimensions 1-10 km on a side could be constructed in Earth orbit by a fleet of free-flying assembly robots or teleoperators manufactured by a replicating factory complex using material from the Moon. Components for very large structures, including communications, storage, recreational, penal, or even military platforms could be fabricated, and later assembled, by an SRS. Another exciting mass-production possibility is the notion of orbital habitats, or "space colonies" (O'Neill, 1974, 1976), by which increasingly large populations of human beings could be safely and comfortably maintained in a support capacity for the space program. Additionally, a replicating factory could build more copies of itself, or new variants of itself capable of manifesting different behaviors and producing different outputs, in almost any desired location. Possible useful output of such facilities already has been summarized in section 5.3.4.

Observation. Exceedingly large sensor arrays for Earth or astronomical observations could be rapidly constructed from nonterrestrial materials by a self-replicating manufacturing facility. This technology could be used to make feasible such advanced missions as optical extrasolar planet imaging (using millions of stationkeeping mirror assemblies arranged in an array with an aperture diameter on the order of kilometers); complex multisensor arrays; very large, high-resolution x-ray telescopy; and other self-organizing optical or radio telescopic arrays of grand proportions to

permit such ambitious undertakings as galactic core mapping, continuous observation of large numbers of passive fiducial markers for Earth crustal plate motion monitoring, and various SETI (Search for Extraterrestrial Intelligence) observations including beacon acquisition, radio "eavesdropping," or, ultimately, active communication. Automated mass production will make possible arrays with heretofore unattainable sensitivity and spatial resolution.

Experimentation. Replicative automation technology will permit a tremendous expansion of the concept of a "laboratory" to include the Earth-Moon system and ultimately all of the bodies and fields in the Solar System. A number of grand experiments could be undertaken which would prove too costly if attempted by any other means. For example, an Earth orbital cyclotron could be constructed as a series of thousands of robot-controlled focusing coils and stationkeeping target assemblies within the terrestrial magnetosphere, with operating energies possibly as high as TeV for electrons and GeV for protons. Additional experiments on magnetospheric propulsion and energy generation could be conducted by free-flying robot drones manufactured on and launched *en masse* from the lunar surface. Gravity field probes, including mascon mappers and drag-free satellites, could be coordinated to perform complex experiments in kinematics, special and general relativity, and celestial mechanics. Investigations of artificial *in situ* lunar crater formation dynamics, solar wind composition and utilization, unmanned ecological simulation modules, and isolation or "hot lab" module manufacturing for conducting dangerous experiments with explosive, radioactive, or biologically engineered materials are still further possibilities.

Exploration. The Moon is largely unexplored. A growing, self-replicating factory could be reprogrammed to mass-produce modified mining or other mobile robots, including orbiters and rovers, for detailed investigation of the lunar surface. This would augment orbital sensing and intelligent image processing systems (see chap. 2) around the Moon, and could be linked to lunar subsurface explorers and other automated surface prospecting equipment to assist in new resource location, colony siting, and the further acquisition of scientific knowledge. Subselene or subterrene (see discussion of the "Coal Mole" in Heer, unpublished draft notes, Pajaro Dunes Workshop, 1980) mining robots could burrow deep into the lunar or terrestrial crust in search of pockets or veins of useful substances, and then dig them out. A self-replicating manufacturing facility could produce thousands of meter-long robot rovers equipped with cameras, core samplers, and other instrumentation which could survey the entire Moon – or any other planet, for that matter – in just a few years. Such exploration would take a century by more conventional methods. Similarly, due to the low gravity, lack of atmosphere, and relative abundance of energy and raw materials, the Moon is an excellent location for the construction and launching of future generations of interplanetary exploratory spacecraft.

Human resources. The augmentation of human services and the extension and safety of the human habitat is yet another near-term application of self-replicating systems. In principle, it is possible to construct a completely autonomous lunar-based facility, but it may turn out to be inefficient or uneconomical in the future unless a few human beings are present onsite to handle unforeseen problems with the machinery. (Humans are the most compact and efficient general-purpose self-replicating systems of which we have certain knowledge.) Initial crew quarters and supplies can be transported from Earth, but much larger and more pleasant living accommodations could be manufactured *in situ* by lunar or orbital replicating systems. The inexpensive mass-production of habitation and agricultural modules (or their components) could help open the door to more extensive lunar and space colonization by people, including recreational, industrial, medical, and educational uses, especially because of the abundant solar energy and the expected ability of replicating factories to manufacture and implement a low-cost lunar-surface-to-orbit launch capability. A comprehensive, highly sophisticated automated astronaut search and rescue system may also become necessary as the human population in space begins to grow, with system components mass-produced by SRS.

Presently, there are about 6000 known and tracked pieces of debris orbiting the Earth at various altitudes and inclinations, and countless additional shards which lie below observational thresholds in near-Earth space. These represent an ever-increasing danger of collision with spacecraft. Debris-catchers or "scavengers" mass-produced by SRS technology could be automatically launched into various Earth orbits, seek out and recognize space debris, report ephemerides in the case of satellite-like objects to avoid destruction of operational equipment and, upon go-ahead, collect the debris. Scavengers would be programmed either to enter the Earth's atmosphere after a specified time in orbit and self-destruct, or to return their collections to orbital manufacturing facilities for recycling of high-level components and materials to help build new robots. A more advanced network could offer protection from possible ecological disasters caused by terrestrial meteorite impacts (Alvarez et al., 1980).

Another possibility, however controversial, is meteorological and climatological intervention on both a local and global scale. A number of interesting alternatives were discussed by the participants of the recent Pajaro Dunes Workshop (Heer, unpublished draft notes, 1980), including:

- Manufacture of 10^7 copies of a 1-km^2 sunshade to achieve global cooling, if required, which could be

deployed most effectively for the polar regions at Earth-Sun L1 (losses due to image diffusion) or in LEO (serious orbital problems).

- Deployment of 1 to 10 million copies of 1-km^2 mirrors in LEO, to cause localized heating effects by concentrating incident solar radiation.
- A system of several 1 to 10 GW microwave frequency solar power satellites to add 100 to 200 W/m^2 to selected terrestrial ground spots 10 km diam, to be deployed in geosynchronous Earth orbit (GEO).

The replicative manufacturing facility needed to economically produce such large numbers of similar system elements would make possible at least a rudimentary global homeostatic environmental control by humanity.

Given the exotic conditions prevailing on the lunar surface and in space, and the novel materials and processes that may become available, it is highly probable that a self-replicating growing lunar facility will be able to economically produce many goods directly for use in space and for export to Earth. What these goods might be is not now certain. However, the economic importance of the telephone, steamboat, airplane, television, office copying machine, etc., during their early stages of development likewise were not at all obvious to most people.

5.4.3 Solar System Applications

The technology of replicating systems will become increasingly important as humanity expands its theater of operations from near-Earth space out to encompass the entire Solar System. Mankind has fallen heir to an incredible treasure trove of nonterrestrial energy and material resources (see sec. 4.2.1). It is likely that replicating machines will provide the only "lever" large enough to explore, and ultimately manipulate and utilize in a responsible fashion, such tremendous quantities of organizable matter. Lacking this advanced automation capability, most of the more ambitious Solar System applications appear uneconomical at best, fanciful at worst.

Observation. Exceedingly far-reaching observational possibilities may become feasible with the advent of SRS technology. Very large baseline interferometry (VLBI) may be attempted with components distributed across the entire Solar System, perhaps located at the stable Trojan points of the Jovian planets or their moons, providing multiplanar baselines of from 1 to 100 AU and complete spherical coverage with the use of out-of-ecliptic robot sensor devices that are mass-manufactured by replicating factories. The solar wind could also be mapped in three dimensions, and by using the entire Sun as a gravitational lens focal lengths on the order of the size of the Solar System can in theory be obtained (Ingel, 1974). This may permit simultaneous observation of the entire celestial sphere across the full spectrum of gravitational radiation using fleets of gravity-wave detectors manufactured by SRS and stationed along the focal plane. A Solar System surveillance network could be constructed to track and warn of objects approaching human habitats, facilities, or the Earth on collision courses, allowing mankind to avoid potentially severe catastrophes.

Exploration. The technologies developed for a generalized lunar autonomous replicative manufacturing facility should be directly applicable in the exploration of all planetary and satellite surfaces. One early possibility is a mission to land a single replicative "seed" on Mars which would then use local materials to produce large numbers of rovers (including, perhaps, fliers, crawlers, walkers, or rollers) and orbiters. A population of 1000 to 10,000 surface rovers each perhaps 100 kg in mass, coupled with a chain of orbital monitors, might continuously monitor and explore the planetary surface and leave stationary probes (active or passive) behind in permanent emplacements. The probes need only have lifetimes on the order of a year or so, since they could constantly be repaired and replenished by the rovers (each of which could last 10 years or more). This system would provide complete surface exploration and continuous status monitoring of all areas on the planet, including temperatures, pressures, wind velocities, seismic events and crustal creeps, meteorite impacts, surface and subsurface compositions, illumination, precipitation, and numerous other phenomena of interest. Automated balloon explorers could be mass-produced and released in Jovian atmospheres, and "trains" of deep solar probes (Heer, unpublished draft notes, 1980) could be hurled into the Sun to obtain direct information on internal conditions there.

Materials retrieval. Replicating systems would make possible very large-scale interplanetary mining and resource retrieval ventures. Nonterrestrial materials could be discovered, mapped, and mined using teams of surface and subsurface prospector robots manufactured *en masse* in an SRS factory complex. Raw materials could be dug up and sent back to wherever they were needed in the Solar System, or could be refined along the way and the waste slag used as reaction mass, or could be utilized *in situ* for manufacturing useful products which would then be exported. Atmospheric mining stations could be established on many different planets – Jupiter and Saturn for hydrogen, helium (and rare isotopes potentially useful for fusion power generation, Martin, 1978), and hydrocarbons, using "aerostats" (Parkinson, 1978); Venus for carbon extraction; Europa for water; Titan for hydrocarbons; etc. Comets could be intercepted to obtain large quantities of useful volatiles, and Saturn's rings could be mined for water-ice by large fleets of mass-produced robot craft. Heavy metals may be

retrieved in great quantities from asteroids. Replicating systems might manufacture huge mining, processing, even ground-to-orbit and interplanetary transportation capabilities using local materials in surprisingly short periods of time.

The general product factory. The team has proposed the design and construction of an automatic multiproduct replicating lunar factory. The reason for the factory having multiproduct capability is to permit it to be able to respond to any changing requirements in kind or amount of product output. This leads to a still broader concept – the notion of a general product factory.

A general product factory is one which can be instructed to manufacture anything which is physically possible to make. Such a system is the physical realization of von Neumann's "universal constructor" automaton, which can construct anything constructable, given an adequate substrate and the rules of operation of his artificial cell-space universe. In the context of drawing upon planetary resources, we should think of each celestial body in terms of its menu of possible materials and the repertoire of processes theoretically available there (see sec. 4.5.4). The following questions should then be considered:

- What is the total range of things which can be made using these processes acting upon these material resources? (See sec. 5.3.6.) This query should be viewed in the broadest possible fashion, including biological as well as mechanical entities.
- Does there exist, for this planetary environment, a factory design which is capable of making all of these entities?
- Can an initial system be designed which, when introduced into the target environment, will yield such a general product factory? A few important developmental milestones are suggested in table 5.4.

TABLE 5.4.– DEVELOPMENTAL MILESTONES FOR A GENERAL PRODUCT FACTORY

1. Design and construct a system which, when supplied only with parts and subassemblies, can duplicate itself.
2. Design and construct a system which can duplicate itself, and in addition produce some useful product.
3. Design and construct a system which, when supplied only with feedstock, can duplicate itself.
4. Design and construct a system which, when supplied with raw materials only, can duplicate itself.
5. Design and construct an automated, reprogrammable, multiproduct system which can, from raw materials, duplicate itself.
6. Design and construct an automated, reprogrammable, multiproduct system which, using only lunar materials and employing only those processes possible in the lunar environment, can duplicate itself.
7. Design and construct an initial automatic "seed" system which, if placed on the lunar surface, could unpack itself and develop into an automated, reprogrammable, multiproduct replicating system, using lunar resources and lunar processing modes only.
8. Design and construct an initial seed which can, in the lunar environment, develop and augment itself so as to become a *general*-product factory, relative to the lunar environment.
9. Design and construct a seed which can, in an *arbitrary* planetary environment, develop into a general-product factory.

The notion of a general product factory using asteroidal material was briefly considered at the Pajaro Dunes Workshop. The "Hive," as it was called, would consist of "an autonomous space island 'beehive' of independently intelligent machines . . . specialized in mining and production, experts in planning, navigation and repair." The product of the Hive would be solar power satellites, "asteroids turned into space colonies, vacuum-filled balloons of nickel floated down to a resource-hungry Earth, spaceships, telescopes, or even another Hive." The Hive was envisioned as an independent economy, using raw materials gathered from the Asteroid Belt, refined and processed with solar or fusion energy, then fashioned into useful output by robot hands. Workshop participants suggested a timetable in which the first fully autonomous replicating system could be in operation in the Asteroid Belt by 2040, commencing exponential growth with a replication time of 5 years, resulting in a total of 1000 new Hives available for production by the year 2080.

Human resources. From the human standpoint, perhaps the most exciting consequence of self-replicating systems is that they provide a means for organizing potentially infinite quantities of matter. This mass could be so organized as to produce an ever-widening habitat for man throughout the Solar System. Self-replicating homes, O'Neill-style space colonies, or great domed cities on the surfaces of other worlds would allow a niche diversification of such grand proportions as never before experienced by the human species.

SRS provides such a large amplification of matter-manipulating capability that it is possible even to consider the "terraforming" of the Moon, Mars, Venus, and other worlds. Terraforming is a theoretical concept in which a planetary environment with otherwise inhospitable conditions for life is purposefully and artificially altered so that

humans may live there with little or no life support equipment. The "traditional" approach is to suggest biological means, such as the proposal to seed the atmosphere of Venus with genetically tailored algae to convert excess carbon dioxide into combined carbon and free oxygen. This would have the incidental salutary effect of lowering the planetary surface temperature so that people could live unaided on the surface. However, it is not known whether biological organisms can be found or developed which are able to withstand present conditions in the Venusian atmosphere.

An alternative approach is to use nonbiological replicating systems which may be far more durable under extreme conditions. A few simple calculations reveal the approximate magnitude and duration of such an enterprise. Consider the terraforming of Mars. For simplicity it is assumed that the planetary crust is largely silicon dioxide and that a general-purpose 100-ton SRS factory "seed" which lands there can replicate itself in 1 year. In just 36 years such a system could theoretically manufacture an SiO_2 reduction capability able to release 220,000 tons/sec of pure oxygen into the Martian atmosphere, which in only 60 years is sufficient to produce 4×10^{17} kg O_2. Assuming negligible leakage through the Martian exosphere, this is enough oxygen to establish a 0.1 bar breathable oxygen atmosphere planet-wide – approximately equivalent to normal air on Earth at an altitude of 3000 m (16,000 ft). This plan requires a solar power satellite system in near-Mars orbit with a total generating capacity of about 10^{17} W, a network which would take less than a year for the finished replicating factory system to produce. The total material thus excavated to terraform Mars is of the order of 10^{18} kg SiO_2, enough to fill a surface depression 1 km deep and 600 km diam. This is roughly the size of the crater Edom near the Martian equator, or Mare Crisium on the Moon.

Of course, far more efficient methods for terraforming planets may eventually be found, such as Dyson's proposal to mine the Saturnian moon Enceladus for its water-ice and return the material to Mars (Dyson, 1979). But the utility of self-replicating systems is clear, and it appears that terraforming times on the order of one century are conceivable using the SRS approach.

Technology requirements. Additional technology over and above "superautomation" (sec. 5.4.1) will be required for the highly ambitious ventures described in this section using advanced space-based self-replicating systems. The most important new technology in this regard is "closure engineering," discussed in section 5.3.6. Some of the enterprises proposed above are of such large scale that it is difficult to envision a feasible mode of operation with anything less than 100% materials and energy closure and virtually 100% information closure as well. No doubt there exist manufacturing operations which are not economically viable candidates for total automation in terrestrial industry – in these instances the functions either must be redesigned for full automation or else people must be permanently incorporated as symbionts of a locally teleoperated or remotely human-supervised system. Manufacturing processes developed for terrestrial environments must be re-engineered to accommodate the input and production environments found in space or on the surfaces of other planets, and output streams must be sufficiently flexible to make feasible the notion of a general products factory.

5.4.4 Interstellar and Galactic Applications

Replicating systems technology is the key to exploration and human habitat expansion beyond the confines of the Solar System. Although these kinds of missions necessarily are highly speculative, and admittedly exceed the limits of current or projected technology in many areas, a consideration of possible interstellar and galactic applications is nonetheless a useful exercise because it serves to illustrate the fantastic power and virtually limitless potential of the SRS concept.

Extrasolar exploration. Before humankind can move out into interstellar space, automated probes will scout the way ahead. The distances are so large and the volumes so vast that self-replicating probes are highly desirable, even essential, to adequately and efficiently perform a reconnaissance of extrasolar star systems in a search for human habitable worlds and extraterrestrial life. A preliminary design for a self-reproducing interstellar probe has been presented in the scientific literature (Freitas, 1980a), and another study of the comparative benefits of reproducing and nonreproducing galactic exploration strategies by unmanned probes suggests that search patterns using semi-intelligent automata involving more than about the nearest 100 stars would probably be optimized (in terms of economy and productivity) if self-replicating systems are employed (Valdes and Freitas, 1980). Reproductive probes could permit the direct investigation of the nearest million stars in about 10,000 years and the entire Milky Way galaxy in less than 10^6 years, starting with a total investment by humanity of a *single* self-replicating exploratory spacecraft.

The problems in keeping track of, controlling, and assimilating data returned by an exponentially growing number of self-reproducing space probes are staggering. Part of the solution may lie in the use of an extremely high level of autonomy in operations management and reasoning such as discussed in chapter 3 of this report; part may lie in the utilization of high levels of abstraction in the information returned to Earth after the fashion of the World Model sensing and data-processing philosophy articulated in chapter 2. Another major piece of the solution is the development of a hierarchical command, control, and information-gathering architecture in which any given

probe communicates directly only with its own parent and offspring. Control messages and exploration reports would pass up and down the chain of ancestral repeater stations erected by earlier generations (Valdes and Freitas, 1980). Certain highly critical but low probability signals might perhaps be broadcast in an omnidirectional alarm mode to all members of the expanding network (and to Earth) by individual probes which encountered specific phenomena or events – such as the discovery of an extrasolar planet suitable for human habitation or a confrontation with intelligent alien lifeforms or their artifacts.

Extrasolar utilization. Before mankind can venture out among the stars, his artifacts and replicating machines must blaze the trail. Ultimately, however, one can envision free-flying space colonies journeying through interstellar space (Matloff, 1976). Upon reaching some new solar system or other convenient source of raw materials, these mobile habitats would reproduce themselves with the human passengers redistributed among the offspring colonies. The original space habitats would serve as extraterrestrial refuges for humanity and for other terrestrial lifeforms that man might choose to bring along. This dispersal of humankind to many spatially separated ecosystems would ensure that no planetary-scale disaster, and, as people travel to other stars, no stellar-scale or (ultimately) galactic-scale event, could threaten the destruction of all mankind and his accomplishments. Replicating systems may be the only rational means to attempt large-scale astroengineering projects usually relegated to the domain of science fiction, such as the construction of "Dyson Spheres" which enclose and utilize the energy output of entire suns (Dyson, 1959).

The limits of expansion. The expansion of a population of replicating systems in any environment is restricted largely by two factors: (1) replication time, and (2) maximum velocity of the outer "envelope" which defines the physical extent or dispersion of the population. No population can accrue at a faster rate than its components can reproduce themselves. Similarly, no population can disperse faster than its medium will permit, no matter how fast components are manufactured – assuming number density remains essentially constant, corresponding to continuous maximum utilization of the environment. Neither factor may be ignored during any phase of population growth.

If envelope expansion velocity does not constrain a population because components are produced only relatively very slowly, then that population will experience exponential multiplication according to:

$$N(T) = \exp(T/t) \tag{1}$$

where $N(T)$ is the number of replicating units comprising the population at time T (replication starts at $T = 0$) and t is the replication time per unit, assumed constant. On the other hand, if unit reproduction is so swift that multiplication is not constrained by replication time, then the population can grow only as fast as it can physically disperse – that is, as fast as the expansion velocity of the surface of its spherical outer envelope – according to:

$$N(T) = 4/3\, \pi\, d(VT)^3 \tag{2}$$

where V is peak dispersion velocity for individual replicating units at the periphery and d is the number density of useful sites for reproduction. Expansion cannot exceed the values for $N(T)$ given either by equations (1) or (2) at any time T, provided all replication sites receive maximum utilization as stipulated (e.g., constant number density of units).

Populations of machines expanding across the surfaces of worlds with replication times on the order of 1 year will not achieve mean envelope growth speeds in excess of a few meters per hour, even in later phases of extreme enlargement when the population of SRS covers a large fraction of the available planetary surface. This figure is well within anticipated nominal ground transport capabilities, so exponential extension should remain largely velocity-unconstrained on such bodies if replication time remains constant at greater population sizes.

Similarly, three-dimensional populations of replicating systems in interplanetary space using Solar System materials and solar energy ultimately are restricted to spherical circumstellar shells where SRS units can collect virtually all energy radiated by the Sun. If a "Dyson Sphere" of 100-ton replicating "seed" units is assembled near the orbit of Earth, approximately one terrestrial mass is required to manufacture the more than 10^{19} individual units needed to completely enclose the star. But maximum expansion velocity even in this case never exceeds about 100 m/sec, hence interplanetary replicating systems as well in theory may spread at purely exponential rates.

In the interstellar realm, however, the situation is far more complex. Depending on the maximum dispersal velocity and interstellar probe replication time, either equation (1) or (2) may control. Figure 5.24 compares pure exponentiation and dispersal speed effects for $t = 1$ year (see sec. 5.3.4) and $t = 500$ years (Freitas, 1980a), and for $V = c$ (since the theoretical maximum envelope expansion rate is the speed of light) and $V = 10\% \, c$ (Martin, 1978) for an assumed homogeneous stellar distribution of "habitable" star systems (taken as 10% of the total) in the galactic disk. In most cases, exponential multiplication soon is halted by the speed-of-light barrier to dispersion, after which the SRS population expansion proceeds only polynomially.

Technology requirements. In order to sustain the expansion of a potentially infinite replicating system, new dispersal mechanisms must be developed. Initially, self-replicating machines or their "seeds" must be capable of motion across

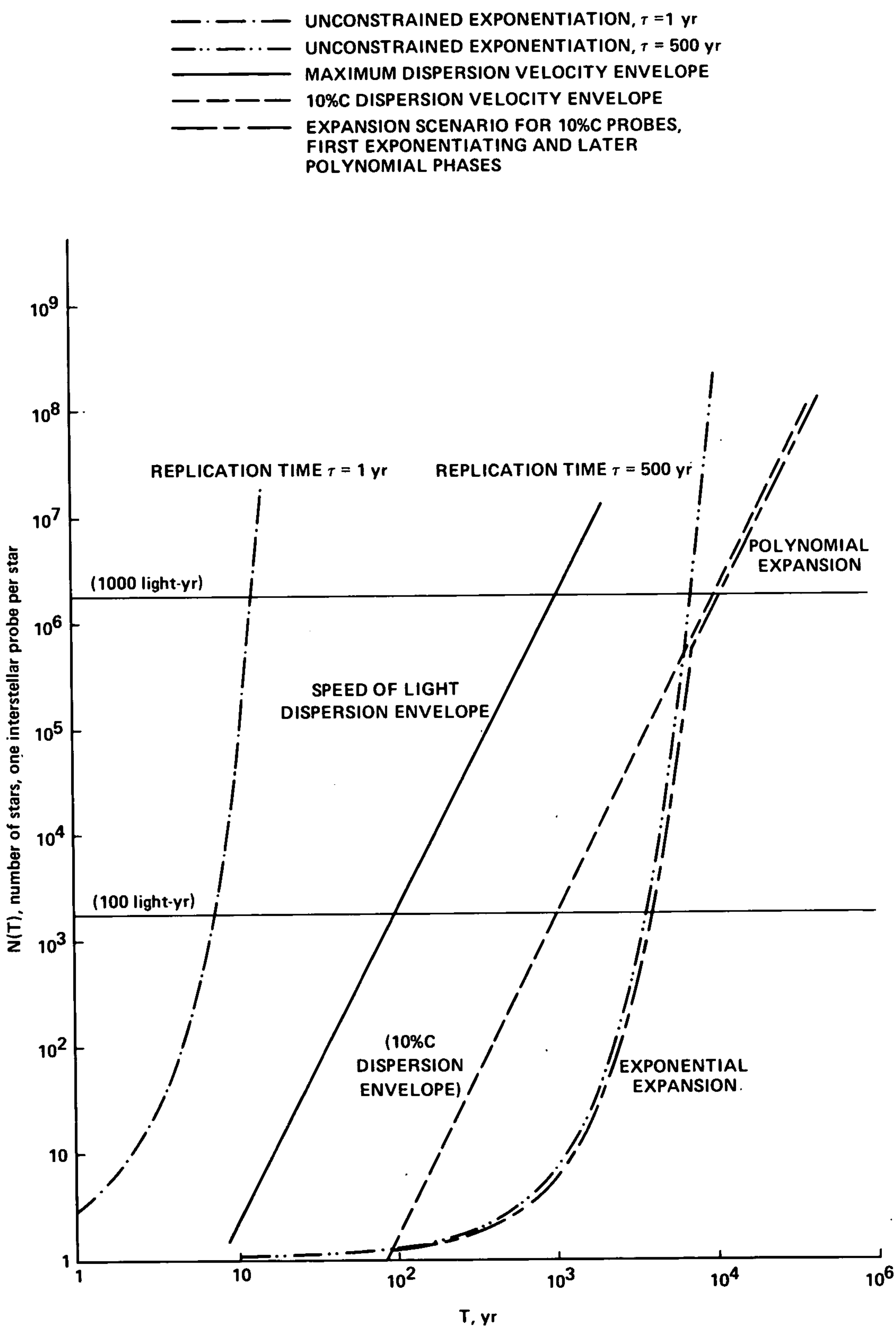


Figure 5.24.– Limits to exponential and polynomial expansion of self-replicating interstellar probe populations dispersing throughout the galactic disk.

a planetary surface or through its atmosphere or seas. Later, interplanetary, interstellar, and, ultimately, intergalactic dispersal mechanisms must be devised. Supplies of energy, stored and generated, must be established if extrasolar spacecraft are to survive in the depths of interstellar space far from convenient sources of power (such as stars) for a major portion of their lives. The technologies of command, control, and communication over stellar and galactic distances ultimately also must be developed.

5.4.5 Applications to Basic Research

In addition to specific applications of replicating systems technology to future missions in space, a number of applications to basic research in biology, computer science, and automata theory have been identified by the team. These are discussed below.

Automaton theory. Automaton theory is the abstract and precise study of all mechanistic devices and processes. At times this has been restricted to the theory of discrete and deterministic machines with a fixed finite number of states. In this narrow sense it is the abstract mathematical counterpart of physical devices such as existing digital computers. In the broadest sense, though, automaton theory can include the study of all mechanisms, discrete or continuous, deterministic or probabilistic or even indeterministic, with a fixed, variable, or indefinitely large number of possible states. Included in this wider definition is the notion of devices which can alter the number of their states by growth or by contraction in respect to certain of their organs, much like the way a Turing machine or a pushdown automaton (or a linear-bounded automaton) can increase or decrease the number of its states by increasing or decreasing the length of its memory tape – but also can grow by increasing or decreasing the numbers of its more active computing components. This is representative of machines which can construct or dismantle other machines.

These machines can not only increase their memory capacity but can augment their computing power by the construction of additional active computing organs (registers, control units, etc.) and by constructing machines separate from themselves, including duplicates of themselves. Von Neumann had begun to develop a general and logical theory of automata which would have embraced all these machine types. Automaton theory has, however, never achieved the generality he sought, at least not in the sense he seems to have intended.

The very general theory of automata has become increasingly abstract, moving from describing mechanistic processes in terms of algebraic concepts such as groups and semigroups to employing category theory, the most abstract and general of algebraic theories. Although a certain level of understanding of what mechanisms might exist has thereby been developed, the applicability of such approaches to the design of complex systems of automata is very slight or nonexistent. In this regard, von Neumann once lamented that "... at a great distance from its empirical source, or after much abstract inbreeding, a mathematical subject is in danger of degeneration .... Whenever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source – the reinjection of more or less directly empirical ideas." (von Neumann, 1966).

It may be that an effort to actually design and implement a system of machines which can construct more machines like themselves would encourage theorists again to attempt to develop a very general automaton theory, including as a part of its subject matter the spatial and communicatory interactions of vast and increasing numbers of submachines. (Perhaps the automatic telephone system provides us with the closest physical analogy to such systems, aside from the analogy of human societies themselves.)

Such a theory would enable one to ask what is the best organization of a system of (potentially) arbitrarily increasing numbers of active components, arranged in various spatial geometries. How might the interacting activites of vast numbers of submachines be optimized? What rules of interaction and of interconnection can be imposed on such a system in order to attain efficient and stable behaviors? What are the safest physical and behavioral interactions, and which lead to instabilities and pathologies?

A general theory would also take as part of its subject matter the flow of parts and materials. It might, like the von Neumann cellular system, treat the creation and flow of materials and the movement of machinery as a form of information flow. It might distinguish information, materials (raw materials, feedstock, and parts) and the movement and siting of machines, but treat them in an identical format so that tradeoffs and exchanges in these categories could be computed (while retaining the essential differences among these types of flow important to the working of the system). The theory would answer such questions as: When will more information be the best substitute for more parts or more feedstock? Under what conditions in the vast assemblage of machines should parts be made anew, from raw materials and feedstock, and when should information or already finished parts be employed to the same purpose? When should machines which are likely to fail be abandoned? When should machines in the assemblage which are still in good condition nevertheless be shut down, moved, sacrificed for parts or dismantled, or sealed off? Under what local and global conditions should submachines be retired, repaired, or replaced?

Theoretical biology. Machines which can construct machines, and machines which can construct replicas of themselves, display behavior which in many ways is analogous to that of natural organisms. Furthermore, as machines are designed to examine their own structure and

the structures of other machines, to repair themselves and other machines, and generally to become more autonomous and more reliable, the analogies become even more apparent.

The ways in which machines carry out these processes of growing, repairing, regenerating, and reproducing may or may not be similar to those carried out by natural organisms – which, in many cases, are not yet even known.

One goal of theoretical biology is to develop an understanding of the mechanisms of living systems, to the point where these systems can be characterized in a precise mathematical fashion (Miller, 1978). To attain such a characterization one needs a good intuitive feeling for the full possible range of lifelike forms. For example, a theory of biology that takes as its subject matter only Earth-evolved forms would be as unlikely to be capable of providing adequate explanation for non-Earth forms as were attempts to account for the forms of extant organisms quite apart from their extinct progenitors.

It seems, therefore, likely that an adequate explanatory theory of biology of any elegance and simplicity must embrace not only all biological forms which presently exist, but all those which have ever existed, or will exist, or could exist. Indeed, the proper subject matter for a true theoretical biology in its broadest sense would be the study of lifelike behavior wherever it occurs whether now, or in the past, or the future; whether on Earth or elsewhere in the universe and whether it is exemplified in artificial or natural forms (Freitas, 1980b), a field of study termed "xenobiology" by one author (Freitas, 1981). This suggests that research on complex automata able to reconstruct, reproduce, and repair themselves might serve as a fertile source of hypotheses as to the logical control and organizational aspects of how living organisms in fact carry out these processes. Such explanatory hypotheses can apply to lifelike systems generally and have the advantage that they are likely to be simpler and more elegant than the necessarily *ad hoc* explanations of behavior for the particular organisms of particular worlds, at particular times.

Thus, research in self-growing and self-replicating machine systems can be viewed as a contribution to, even as a central part of, a true theoretical biology which takes as its subject matter not merely the evolved, naturally occurring living organisms of Earth, but lifelike mechanisms, natural or artificial, having existed or possible, wherever in the universe they might arise.

Design of biological and hybrid organisms. The forms and processes of artificial organism-like systems are not bound to follow the particular structure and logical organizations of known naturally evolved organisms. As the design of increasingly complex artificial systems capable of drawing materials and energy from natural surroundings and possessing more and more organism-like properties proceeds, it may become apparent that there are artificial organism functions which, if embodied in biological organisms, would be of value. With advances in "genetic engineering" it may become possible to create new biological forms, possessing the desired features.

Just as the design of artificial mechanisms can be inspired by contemplation of evolution's apparent solutions to various design problems, so might new *biological* systems also be created, drawing upon designs originally conceived for artificial systems – a kind of inverse bionics. Taking this a step further, one can envisage as a research goal the gradual elimination of the perhaps arbitrary line now drawn between artificial and natural organisms, and the consideration of a more deliberate systematic investigation of the creation of hybrid biological-mechanical systems.

Experimental evolution. Studies of form and function in biological and artificial systems may contribute to an understanding of the design and construction of both biological and mechanical organisms. This interdisciplinary exchange should not be limited to studies of the relationship between individual classes of lifelike entities, but should also extend to studies of the consequences of large numbers of such entities interacting and competing for resources. Replications of programs and creation of new machines (including replicas), and compounds and combinations of initially existing machines, can be a feature of the proposed machine replicating systems. It seems clear that development of a science of evolving systems is needed (Miller, 1978). (This would again be a part of a very general "true" theoretical biology, which takes all possible lifelike systems as its subject matter.)

For example, one putative value of sexual over asexual reproduction is the enormously increased mobility of genetic variation in the species population. This widely available variation tends to ensure that environmental changes can be accommodated or exploited with great swiftness by at least some members of the population (Smith, 1978). In a "designed" universe, one is free to consider the advantages (if any) of three or more sexes (Freitas, 1980c; Smith, 1978) or of the consequences of other, even more radical redesignings of existing natural systems. In particular, the actual behavior of largely autonomous growing replicating machine "species" with differing capabilities and reproducing strategies certainly should be an object of study by evolutionary biologists who might be able to predict the forms which would persist and come to dominate in systems left unperturbed by external pressures and commands.

The existence of large interacting populations of entities whose "genetics" are precisely known, but whose global behavior over time cannot readily be predicted, may be of great experimental value to evolutionary biologists. At present, computer simulation is the usual tool of choice for such problems. However, if the physical creation of machine populations becomes sufficiently inexpensive,

experimental situations might be created in remote nonterrestrial regions. Machine growth and population changes could be monitored over time for their adherence (or not) to hypothesized consequences. The advantage of this approach over the computer simulation would be in the much greater detail and fidelity to real situations, and the consequent likelihood of serendipitous useful observation.

Machine intelligence architecture. Very general symbol manipulating devices (such as stored program computers) are at the heart of efforts to demonstrate that machines can exhibit behavior which in humans or other animals would be considered intelligent. In one sense, such devices are computationally universal. That is, certain mathematical technicalities aside, they can carry out any arbitrary Turing machine computation and, accepting the Church-Turing Thesis, can also carry out any algorithmic process. Thus, if *any* machine can be intelligent one need look no further than to a general-purpose computer, for there is some program which will cause the machine to display the desired intelligent behavior. This is so even if one insists that brains, for example, are machines, but are not at all like digital computers. This is because digital computers, again accepting the Church-Turing Thesis, can be programmed to simulate any known mechanistic process to any fineness of detail, whether the process of interest be analog, frequency coded, probabilistic or other.

Even though ordinary computing machines do not, for example, reproduce themselves, they can be programmed to simulate the behavior of machines which do in fact reproduce. From this point of view, the concept of machines which possess the power to construct other machines and to replicate themselves can be represented to any degree of detail in the computation of an ordinary general-purpose computing machine which cannot itself reproduce. Even though existing general-purpose machines cannot generally inspect themselves and draw conclusions therefrom, computers can be programmed to simulate such unlikely machine actions if such a simulation is thought useful or interesting. Hence, the construction of the kinds of machines considered here – machines that can compute, construct, reproduce, and inspect, repair, simulate, and observe both themselves and other machines – would not enlarge what a general-purpose device can in theory already do but rather our perception of their capability to exhibit more sophisticated mindlike behavior.

It should also be noted that machines can be designed and constructed so as to do things beyond what any known evolved organism (including man) can do. We are already aware of this superiority of machines in regard to strength, speed, accuracy, flight, and the like. There are already many ways in which machines can be designed and constructed so as to exceed human mental capabilities for specific tasks.

For example, though we are constantly reminded of the social value of being able "to see ourselves as others see us," our evolutionary history has left us with only a very limited capacity for accurate introspection and self-examination – though in this respect we admittedly exceed all other known evolved creatures. Machines, however, can be designed to secure far greater access to their internal structure and states than we are ever likely to possess as individuals, and this capacity might mean that machines can be programmed to achieve mindlike powers far beyond ours. A trivial case of this "introspective" superiority of machines is seen in their ability to "remember." Computers can be programmed to methodically search all of their memory with a thoroughness that can evoke human envy.

5.5 Implications

It appears that self-replicating systems may have numerous economical applications on Earth, in near-Earth and lunar space, throughout the Solar System, and perhaps even in the interstellar realm, for future exploration and utilization. The main advantage of SRS is their tremendous capability for performing any desired task at almost any remote location, starting with a relatively small investment of time, money, energy, and mass. This suggests that replication technology may have significant social and economic impacts on American and human society, as discussed below. A number of philosophical and ethical implications may derive from replicating systems techniques. Various issues regarding the future of human and machine evolution must be addressed, together with the "cosmological" implications of SRS.

As the time allotted to consideration of the implications of machine replication was relatively small, the team was not able to examine many intriguing questions in depth. In many cases, it was possible only to frame questions regarding general classes of social and cultural impacts, as no satisfactory answers were immediately apparent. Consequently, this section must be regarded simply as a blueprint for further study in this area.

5.5.1 Socio-Economic Implications of Self-Replicating Systems

The history of technology on this planet is the record of man's constant attempts to control his environment through the use of extrasomatic tools. The development of SRS in this context will be revolutionary, with impacts equal to or exceeding those engendered by other "revolutions" in human history. For the first time, mankind will be creating, not merely a useful paradigmatic tool (e.g., the scientific method, Copernican revolution), organizational tool (e.g., centralized cultivation, agricultural revolution), or energy-harnessing tool (e.g., steam power, industrial revolution), but rather a wholly new category of "tool" – a

device able to *use itself* intelligently and with minimum human intervention. In many respects, with SRS mankind is creating a technological partner rather than a mere technical implement.

Superautomation on Earth and in space. The use of self-replicating systems on Earth poses many problems. A compact, freely replicating system released on the surface of the Earth potentially could compete with humans for resources such as living space and energy. It could also smother us in its waste products. Even if kept under control, a terrestrial SRS could wreak economic havoc by manufacturing products for which the consumers who will use the products will not have to pay. Unfortunately, we will probably have to deal with this problem regardless of whether replicating systems technology *per se* is ever developed. If industrial automation continues in the direction it seems to be headed now, global commerce soon will reach a state of "superautomation" (Albus, 1976) in which an entire national industrial base has become automated and is, for all practical purposes, a terrestrial SRS. Such a system may function without the need for significant inputs of human labor. Eventually it should be possible to deal with the attendant economic dislocations, but the transition is certain to be excruciatingly painful.

In Earth orbit and on the lunar surface, however, the situation is quite different. In the environment of space SRS would not be in competition with an established human presence. Instead, they would provide a powerful "tool" by which humans can manipulate that environment to their advantage. One can envision building vast antenna arrays (for radio astronomy and SETI), solar power satellites, or even lunar, orbital, or free-flying habitations. These applications should enhance, rather than destroy, the economic fabric of terrestrial civilization, just as colonies in the New World enhanced the economies of their parent nations. By expanding into space, mankind has the potential to gain, rather than lose, from extensive automation. Instead of doing the same amount of "work" that is required to sustain terrestrial existence (and doing it with fewer and fewer people), by moving into space even more people can be kept occupied than before while at the same time extending into a redundant habitat. This seems perhaps the best way to sustain the least trauma in the years ahead.

The development of the necessary artificial intelligence, robotics, and automation techniques will likely have enormous short range impacts on Earthbound activities. If our economy is to be transformed by such revolutionary technologies in a fairly short period of time, how can the United States (and the entire industrialized global community) prepare for and avoid or mitigate potentially vast dislocations? Will we need a new academic discipline of "revolution management"?

Economics of replicating systems. Whether supported by public or private sources, the development of SRS must make good economic sense or else it will never be attempted. Self-replicating factories on Earth or in space may appear theoretically capable of creating bountiful wealth and endless supplies of goods and services for all (Bekey and Naugle, 1980; Heer, 1979). However, this utopian ideal must be tempered with the cold logic of cost-benefit analyses and indices of profitability if it is to gain some measure of credibility in the business world.

Let us assume that a financial consortium invests a sufficient quantity of capital to research, design, build, and successfully deploy the first SRS. This consortium may represent an association of private businesses (e.g., the Alaskan Pipeline), an intergovernmental entity (e.g., the International Monetary Fund), or individual public agencies (e.g., NASA). Deployment may occur on Earth, in orbit, on the Moon, or even on the surfaces of other planets or the asteroids. After a relatively brief period (T years) of growth, the capacity of the initial SRS expands a thousandfold by self-replication, and commercial production begins.

Assume that the original investment is $\$X$ and the original factory could produce useful manufacturing output with an annual value of $\alpha\$X$. After the SRS undergoes thousandfold expansion, its output is worth $1000\alpha\$X$ per year (provided demand remained unaffected). The value of the original investment after T years is $\$X(1 + I)^T$, where I is the mean annual inflation rate during the period of investment. Thus, to repay the original investment and achieve economic breakeven will require approximately $\$X(1 + I)^T/1000\alpha\$X$ years of production following the period of nonproductive factory growth. The results of this simple calculation for $T = 20$ years are shown in table 5.5 for several representative values of α and I.

What is a reasonable value for α? The Lunar Manufacturing Facility developed in an earlier section replicates its own mass (of similar components) in one year, or $\alpha = 1$. Waldron et al. (1979) propose a semireplicating factory which can produce its own mass in metal products in less than 6 days, for a maximum $\alpha = 60$. Nevertheless, table 5.5 shows that even if $\alpha = 0.01$ (corresponding to extraordinarily low productivity) the repayment time is still less than a year in a national or global economy with low-to-moderate inflation or interest rates (10% or less). In an economy with interest rates up to 50%, reasonable repayment times – on the order of typical plant lifetime, about 30 years in usual industrial practice – remain available for $\alpha > 0.1$ (also a fairly pessimistic lower limit on productivity). Under conditions of hyperinflation (100% and higher) a 30-year breakeven can be obtained only for highly robust, productive systems with $\alpha > 35$.

Economic feasibility, however, is not limited to amortization of costs. A net profit must be made, markets established and maintained, production managed in a reliable

TABLE 5.5.– ECONOMICS OF SELF-REPLICATING FACTORIES

Relative specific productivity, α (\$/yr-\$)[a]	Repayment period of original investment, for an adult seed[b]			
	Inflation = 0%	Inflation = 10%	Inflation = 50%	Inflation = 100%
0.01	1 mo	8 mo	330 yr	100,000 yr
.1	4 d	1 mo	33 yr	10,000 yr
1.0	9 hr	2 d	3 yr	1,000 yr
10	50 min	6 hr	4 mo	100 yr
100	5 min	35 min	12 d	10 yr
1000	30 sec	4 min	1 d	1 yr

[a] α = fraction of original value of seed that the adult LMF can produce per year.
[b] Repayment period = $\$X \cdot (1 + I)^{20}/(1000\alpha \cdot \$X)$, assuming an initial 20 year nonproductive period.

and flexible manner, and so forth. Given the tremendous power of SRS, severe economic distortions are conceivable across the board. If a replicating factory system is used to flood a market with products, the prices of these products will fall, carrying profits downward as demand saturates in an unregulated economic environment. On the other hand, in a tightly controlled economy the well-known problem of inferior production control feedback would be exacerbated, leading possibly to wild fluctuations in supply and demand for SRS products. These relationships should be investigated more thoroughly by economists.

If control of Earth-deployed replicating factories is retained by national or subnational entities, governments lacking this technology will seek equitable licensing agreements. One interesting problem is ownership of SRS offspring grown from the soil of one country but generated by a leased parent machine owned by another. Should licensing arrangements require return of offspring? Perhaps the offspring should be allowed to remain the property of the licensee, but with royalties levied against production in favor of the owner of the parent machine? Clearly such arrangements could become quite complex in just a few generations of cross-licensing. (SRS capable of "sexual" reproduction present a host of additional theoretical complications.) From the businessman's point of view, it might be better just to sell a "mule SRS" – an infertile factory with the capacity for rapid automated manufacturing but which lacks some vital software element or process necessary for replication. Of course, this is an open invitation to a black market traffic in "bootstrap kits" which allow users to restore fertility to their neutered systems. It is difficult to see how the rapid spread of such technology, once introduced in any form, could be held in check for long by any governmental, corporate, or private entity.

Social aspects of SRS cornucopia. How will humankind deal with what has been termed, with some justification, "the last machine we need ever build?" How might people's lives be changed by a replicative universal constructor system capable of absorbing solar energy and raw dirt and manufacturing, as if by magic, a steady stream of finished building materials, television sets and cars, sheet metal, computer components, and more robots – with little or no human intervention required? Just as the invention of the telephone provided virtually instantaneous long-distance communication, and television permits instant knowledge of remote events, and the automobile allows great individual mobility, the autonomous SRS has the potential to provide humanity with virtually any desired product or service and in almost unlimited quantities. Assuming that global human population does not simply rise in response to the new-found replicative cornucopia and recreate another equilibrium of scarcity at the original per capita levels, supply may be decoupled from demand to permit each person to possess all he wants, and more. The problems of social adjustment to extreme sudden wealth have been documented in certain OPEC nations in recent years. Much attention has also been given to the coming "age of leisure" to be caused by superautomation. What more difficult psychological and social problems might emerge in an era of global material hyperabundance?

If the enterprise of establishing an automated lunar mining and manufacturing facility is successful, there might thereby be made available to humanity a vast supply of energy and useful products. By exporting heavy industry to the Moon, the Earth might be allowed to revert to a more nearly natural state of "controlled wilderness." This should permit the preservation of the animals and plants which people have for so long enjoyed. Although contrary to the

historical evidence, on the negative side people may take their new prosperity as license to exercise their natural biological proclivities and yet further overwhelm this planet with teeming human billions. If this occurs, eventually we shall find that although we might make our Earth into a parkland, the actual effect will be more like Yosemite National Park on a midsummer weekend. This is one problem we must not export to other worlds.

Is there a similar danger that the SRS project, though completely successful as a technological and financial venture, will (much like penny-per-gallon gasoline) encourage profligate behavior heedless of catastrophic negative consequences? What unfortunate things might we do, possessing almost unlimited energy and material resources? Will the possibility of hyperabundance lead not to continued national resolve and focus, but rather to a pervasive national complacency, making us think that all is well, that all has been solved, that things always get solved, and that henceforth we need do little or nothing more to improve our lot? If the system works, and we come to depend on it, growing once more to the limits of our productive inventiveness, will we not be dangerously subject to catastrophic damage as a vital, progressive race?

If space offers any solution to this contradiction between the "good life" and our innate breeding proclivities, it probably will involve the establishment of orbital human colonies. To be practical, these habitats must approach replicating factories in the range of goods and services which they produce. The expense of maintaining a large human colony with direct Earth-based support would be immense, so automated factories most likely must provide the goods and services to support such an operation. Once more the need for SRS facilities in the future of humanity becomes apparent.

Replicating factory systems have the potential to severely disrupt or disable most all modern national economies. The concept of "rate of return" on investments may have to be replaced with the notion of "acceleration of return" for nonterrestrial exponentiating SRS. Will present-day governments and other national and international economic entities support the replicating factory concept if it is seen as a potential threat, capable of rendering obsolete the entire global economic order which now exists and under which they now operate?

Environmental impacts. It has been suggested by Dyson (1979) that it might be possible to design a compact replicating robot which can itself serve as part of an enormous energy-collecting grid. Each machine consists of solar panels on top, power transformers and a universal power grid bus connector, some means of mobility such as tracks or wheels, and manipulators and other subsystems necessary for self-replication. Released, say, in the Arizona desert, one or two SRS could rapidly multiply into a "free" gigawatt-capacity generating system in just a few years. This could then be tapped by power-hungry municipal utilities or even by individual users.

Moore (1956) also discussed the possibility of replicating machine "plants" turned loose on Earth. In Moore's scenario, a single floating self-reproducing barge is released into the oceans; a few years later, it has multiplied itself into a population of millions, with each unit periodically commuting to shore bearing useful products for mankind derived from the sea (salts, minerals, gold). Reviewing this scenario, Dyson noted that such seagoing SRS might become so numerous that frequent crowding and collisions would occur between them. The "dead bodies" of machines involved in major accidents could slowly accumulate on the ocean floor and along the coastline, causing congestion and representing a menace to navigation. The introduction of machine cannibalism to clean up the mess introduces fresh complications into an already difficult situation – ownership and proper recognition of "dead" machines, destruction control and failsafe mechanisms, nonrecyclable parts, violations of national economic zones, and military applications of the technology.

Environmentalists might perhaps regard SRS released on Earth merely as automated strip-mining robots – yet another sophisticated instrumentality in the hands of those who would mercilessly rape the Earth of its limited resources, leaving behind the ugly scars of profit. There are two responses to this shortsighted view of SRS. First, in the Age of Plenty ushered in by these machines, human society will be sufficiently wealthy to regard environmental integrity and beauty as indispensible outputs of any manufacturing system. These functions may be designed into machines as a matter of course; SRS can be preprogrammed first to strip mine, then reclaim, the land they work. Second, machine replication will make possible significant advances in recycling technology. Given sufficient energy, junkpiles and city dumps may be regarded as low grade "ores" – materials processing robots could be turned loose to analyze, separate, and extract valuable resources. Collection and distribution systems would be streamlined by the use of robot workers constructed at an enormous rate by a sessile self-growing factory complex.

Utilization of the Moon by SRS as proposed in earlier sections may be viewed with outrage by other nations as a predatory attempt to secure a part of the "common heritage of all mankind" for the benefit of America alone. Very drastic alteration of the lunar surface is proposed, raising a question of whether there ought to be reserved areas. Should there be more exploration to determine which regions should be exploited and which should not? Must an environmental impact statement be prepared? As on Earth, lunar surface despoilment in theory may be largely reversed – the machines could be programmed to photograph the original landscape in detail and to restore it after mining operations are finished in that area. A potentially more serious environmental impact is the possible creation

of an appreciable lunar atmosphere during the course of industrial operations conducted on the Moon (Johnson and Holbrow, 1977; Vondrak, 1976, 1977). Even small leakages of gas from millions of SRS could create enough atmosphere to disable or seriously disrupt the operation of mass drivers and other manufacturing facilities requiring vacuum conditions.

5.5.2 Implications for Human Evolution

When contemplating the creation of large, imperfectly understood systems with which we have no prior experience, it is prudent to inquire as to the possibility of unforeseen dangers to our continued existence. In particular, artificial intelligences could conceivably become adversaries, whether they reproduce or not. Similarly SRS might become a threat, independent of their intelligence. Because of the imminence of advanced AI and replicating systems technologies in the next several decades, such questions are no longer merely theoretical but have a very pragmatic aspect.

We must begin to examine the possible problems in creating artificial intelligences or replicating systems which could conceivably become our adversaries or competitors for scarce resources. It is not too early to begin considering the possible kinds of behaviors which advanced machines might display, and the "machine sociobiology" which may emerge. It seems wise to try to identify early any fundamental distinctions between intelligent "natural" biological and advanced "artificial" machine systems. Finally, we should consider the significance of the development of advanced machine technologies to the future of human evolution and also to the broader sweep of cosmic evolution.

To serve mankind. The most immediate, urgent impetus for the development of automation and machine replicative techniques is to improve, protect, and increase the productivity of human society. One way of achieving the goal of human preservation and improvement is to make our mechanical creations intelligent, so that they can automatically do what is good for us. We want them to do this even if we have forgotten to specify what "good" is in each instance. Perhaps we don't even know in all cases how to define "good." For example, consider what would happen if a physically capable, literal-minded idiot were put at the controls of a bulldozer (e.g., Pvt. Zero in the "Beetle Bailey" comic strip, present-day computers, etc.). If told to "drive the bulldozer into the parking lot," the idiot would do exactly that, regardless of whether or not the lot happened to be full of automobiles.

One rather compact statement of what is required for our protection already exists. This has come to be known as "Asimov's Three Laws of Robotics":

(1) A robot may not injure a human being, or, through inaction, allow a human being to come to harm.

(2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

(3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws (Asimov, 1950).

This is an excellent prescription of what is required but not of how to accomplish it. Exactly what do these laws entail? The following list of conditions is certainly necessary for the Three Laws to hold:

(1) A robot may not injure a human being.
(2) A robot must use common sense.
(3) A robot must be intelligent.
(4) A robot must be conscious.

Common sense, intelligence, and consciousness are the essence of artificial intelligence research. Even if we cannot exhaustively enumerate the ways to harm a human in any and all circumstances, a robot with the above four properties would protect people to the best of its ability. If it ever did injure a human being it would be because neither we, nor it, foresaw that possibility. But it would immediately perceive its error and would never make the same kind of mistake again. We can do no better ourselves, and usually we do worse.

At the present time we have only the most rudimentary knowledge of what common sense, intelligence, and consciousness are, let alone how to insert these qualities into a robot (Good, 1965). As our computers become ever more complex and pervasive, there is the distinct possibility that these characteristics will arise spontaneously. In this case we would be involved in a totally uncontrolled experiment (Hogan, 1979). If conditions 1-3 were not yet fulfilled, but condition 4 *was*, the outcome could be catastrophic for mankind. For reasons of self-preservation, we must pursue AI research with the goal of ensuring that capabilities 1-3 are achieved *first*.

The problem with this entire approach is that any machine sufficiently sophisticated to engage in reproduction in largely unstructured environments and having, in general, the capacity for survival probably must also be capable of a certain amount of automatic or self-reprogramming. Such SRS in theory may be able to "program around" any normative rules of behavior (such as the Three Laws) with which it is endowed by its creators. Thus, it might modify its own patterns of behavior, as determined by its basic goals and learned motivational structure.

It is possible to conceive of a machine design containing "read-only" hard-wired goal structures. But hardware specialists will admit that such procedures can be circumvented by sufficiently clever software in large, complex systems. Further, since SRS must be quite adept at physical manipulation it is likely that it will be able to re-wire its own

structure in accordance with perceived operational objectives – assuming it can analyze the functions of its own components as needed for repair or maintenance operations. It may be of no use to try to distribute the hard-wired functions throughout the whole machine, or a large subset thereof, in hopes that the system will be unable to comprehend such a large fraction of itself simultaneously. Omitting the special functions from the machine's stored genetic description of itself would probably be equally ineffectual. Laing (1975, 1977) has shown that machine reproduction by complete self-inspection – wherein the parent knows virtually nothing about its own structure to begin with – is quite possible, and has provided several logical designs for such machines. Consequently, it is not possible to logically exclude the possibility of conscious alteration of hard-wired robot "commandments" by intelligent self-replicating machines desirous of doing so.

It would therefore appear nearly impossible, as with people, to absolutely guarantee the "good" behavior of any common-sense, intelligent, conscious SRS. However, also like people, machines can be taught "right" and "wrong," as well as the logical rationales for various codes of behavior. And they can probably be expected to remain even more faithful to a given moral standard than people.

SRS population control. An exponentially increasing number of factories (even if the rate is not sustained indefinitely) will seem especially threatening and psychologically alarming to many. Such a situation will draw forth visions of a "population explosion," heated discussions of *lebensraum,* cancerous growth, and the like. Nations not possessing replicating systems technology will fear an accelerating economic and cultural gulf between the haves and the have-nots. On another level altogether, humankind as a species may regard the burgeoning machine population as competitors for scarce energy and material resources, even if the net return from the SRS population is positive.

Of course, self-replicating factories are not ends in themselves but have specific purposes – say, to produce certain desired products. The quantity of these products is determined by needs and requirements and is the basis for designing an SRS. Depending on the type of product, factors such as the time when these products need to be available, the production time, and replication time per replica determine the optimum number of replica factories per primary and the number of generations required. The following controls might be used to achieve this condition:

(1) The "genetic" instructions contain a cutoff command after a predetermined number of replicas. After each replica has been constructed one generation command is marked off until at the last predetermined generation the whole process is terminated after the final replica is completed. Besides all this, engineers may have their hands full keeping the SRS replicating on schedule and functioning properly. It is not likely that they will soon be able to do much more than we expect.

(2) A predetermined remote signal from Earth control over a special channel can easily cut the power of the main bus for individual, groups, or all SRS at any time. Replication energy production shows one of the fundamental differences between biological and mechanical replicating systems as presently conceived. In biological systems energy is generated in distributed form (in each living cell throughout the entire organism) whereas in mechanical systems such as SRS energy is produced centrally in special parts (e.g., power plant, solar cells) and then is distributed to wherever it is needed. This should make control of mechanical systems comparatively easy.

For replicating systems much smaller than factories, say, in the 10^2-10^4 kg category, the situation may be somewhat different. One potential problem with such devices is that once started, their multiplication may be difficult to stop. As a reasonably large population accumulates, it may become almost physically impossible for humans to maintain any semblance of control unless certain precautions are taken to severely limit small-machine population expansion. In many ways a large population of low-mass SRS resembles a biological ecology. While the analogy is imperfect, it serves to suggest some useful ideas for automata population control once people determine that direct control of the situation has somehow been lost.

Predation is one interesting possibility. Much as predator animals are frequently introduced in National Parks as a population control measure, we might design predator machines which are either "species specific" (attacking only one kind of SRS whose numbers must be reduced) or a kind of "universal destructor" (able to take apart any foreign machine encountered, stockpiling the parts and banking the acquired information). Such devices are logically possible, as discussed earlier in section 5.2, but would themselves have to be carefully controlled. Note that a linear supply of predators can control an exponentiating population of prey if the process of destruction is, as expected, far more rapid than that of replication.

Clearly it is easier to design the solution to this problem into the SRS from the start, as suggested above in reference to larger factory systems. For instance, machines might be keyed to respond to population density, becoming infertile, ceasing operations, reporting (like lemmings) to a central disassembly facility, or even resorting to dueling or cannibalism when crowding becomes too severe. However, a method by which the materials and information accumulated by SRS units during their lifespans can be preserved would be in the best interests of human society.

The unpluggability problem. Many people, suspicious of modern computers and robotics technology, take solace in the fact that "no matter what goes wrong, we can always

pull out the plug." Such individuals might insist that humankind always retain ultimate life-and-death control over its machines, as part of the social price to be paid to permit their development. Whether this is advisable, or even necessary, is a question which requires further study. Certainly it is true that our civilization all too easily becomes habituated to its machines, institutions, and large organizations. Could we unplug all our computers today? Could we "unplug" the Social Security Administration? It is difficult, or impossible, and in many cases ill-advised, to retreat from a social or mechanical technology once it has been widely introduced and a really significant change has taken place. Many individuals in our society would prefer to turn back the clock on the industrial revolution, but today this could not be done without the sacrifice of hundreds of millions of human lives and extreme trauma to global civilization.

Further, we must assume that we cannot necessarily pull the plug on our autonomous artificially intelligent species once they have gotten beyond a certain point of intelligent development. The one thing the artificial system may learn is how to avoid a human being "pulling its plug out," in the same way that human beings come to understand how to defend themselves against other people (George, 1977). Consequently, it is imperative that we study ways to assure ourselves that our technological creations will serve to our benefit rather than to our detriment, as best we can, prior to their widespread adoption.

Assuming we wish to retain ultimate control over our creations (by no means a foregone conclusion), the team first considered, as a theoretical issue, the following intriguing problem:

> Is it logically possible to design an internal mechanism which permits normal SRS functioning to be interrupted by some external agency, yet which is impossible for the SRS itself to circumvent either by automatic reprogramming or by physical self-reconstruction?

That is, is it impossible to build a machine whose "plug" cannot be "pulled"?

Machine capabilities of the future span a wide spectrum of sophistication of operation. As systems become more complex, individual human beings will come to understand decreasing fractions of the entire machine function. In addition, very advanced devices such as SRS may need to be programmed with primitive survival instincts and repair capabilities if they are to function autonomously. At some point the depth of analysis and sophistication of action available to a robot system may exceed the abilities of human minds to defeat it (should this become necessary). If there is even the slightest possibility that this is so, it becomes imperative that we learn exactly what constellation of machine capabilities might enable an SRS to cross the subtle threshold into "theoretical unpluggability."

To this end, the team subsequently reformulated the unpluggability question as follows:

> What is the *least* sophisticated machine system capable of *discovering* and *circumventing* a disabling mechanism placed within it?

While no specific firm conclusions were reached, the team concluded that the simplest machine capable of thus evading human control must incorporate at least four basic categories of intelligence or AI capabilities (Gravander, personal communication, 1980):

(1) Class invention, concept formation, or "abduction"
(2) Self-inspection
(3) Automatic programming
(4) Re-configuration or re-instrumentation capability (especially if the "plug" is in hardware, rather than software)

These four characteristics are necessary preconditions for theoretical unpluggability – a machine lacking any one of them probably could not figure out how to prevent its own deactivation from an external source. Whether the conditions are sufficient is an urgent subject for further research.

Sociobiology of machines. The creation of replicating manufacturing facilities, remotely sited, and for long times left under their own control, poses some very special problems. In order to eliminate the use of humans as much as possible in a harsh environment, these systems of machines should be designed to seek out their own sources of materials; to decide on this basis to invest in duplicates of themselves; to determine their power requirements and see to the construction of requisite new power sources; to monitor their own behaviors; to undertake the repair of machines, including themselves; to determine when machines have, under the conditions obtaining, reached the end of their useful working lives; and so forth. They must operate reliably and resist corrupting signals and commands from within their own numbers, and from without. They must be able to discern when machines (whether of their own sort or not) in their neighborhood are, by their behavior, disrupting or endangering the proper functioning of the system. Since we cannot foresee all of the ways in which the system may be perturbed, we shall have to supply it with goals, as well as some problem-solving or homeostatic capabilities, enabling the machines to solve their own difficulties and restore themselves to proper working order with little or no human assistance.

As SRS make duplicates of themselves, the offspring will, if suited to surroundings different from those of their parents, differ somewhat from them. More of one sort of submachine or subordinate machine may be required, fewer of another. The main "genetic" program will undoubtedly increase in size, generation by generation. At removed locations constructor-replicators may symbiotically combine with miners, surveyors, and fabricators, to form satellite

machine communities differing considerably from the original population.

At this stage it may be that some of the claims made by evolutionary biologists as to the likely origin of complex, social behavior of animal populations may begin to apply to machine populations. Indeed, it may be that the arguments of the sociobiologists will be *more* applicable to machines than to animals and humans. In the case of animals, and especially in regard to humans, the opponents of the evolutionary biologists insist on the priority of alternative sources for social behavior – namely, individual learning. Behavior need not have its origins in the genome. These opponents of the evolutionary biologists constantly challenge them to specify where in the genome is the locus of selfishness, distrust of strangers, aggression, and the like. This is not really readily done.

However, in the case of machines the locus of behavior can indeed be specified: It is in the program of instructions, and these programs can, like genes, be modified and transmitted to offspring. Though we may not be mere machines driven by our genes, mere real machines are indeed driven by their gene-like programs, and for them, some of the evolutionary biological predictions of the likely resulting system behaviors may apply.

Thus, at the most elementary level, if some one of the SRS machines capable of duplicating itself begins to concentrate on this reproductive activity to the neglect of all other tasks we intend for it, its progeny (possessing the same trait) might soon become dominant in the machine population. But far more complex aberrations and consequent elaboration of machine behavior can arise and be propagated in machine populations of the sophistication we may be forced to employ.

Thus, our machines can reproduce themselves as well as tell whether an encountered machine is "like" themselves or could be one of their offspring. If the structure of an encountered machine is examined and found to be similar to or identical to the machines of one's own making or of one's own system of machines, then such machine should be welcomed, tolerated, repaired, supplied with energy and consumables, and put to work in the common enterprise. If, on the other hand, the structure of an encountered machine deviates greatly from that of any of one's own system of machines – even if it is in fact a device of one's own construction which has suffered severe damage or defect of construction – then prudence suggests it should be disabled and dismantled.

It is interesting to note that this "reasonable" kin-preferring behavior could arise generally throughout the machine population quite without it having been made a deliberate part of the programs of machines of the system (Hamilton, 1964). If a single machine of the sort which reproduces ever chances upon the program "trait" of tolerating machines like itself, or aiding or repairing them while ignoring, disabling, or dismantling machines unlike itself and its offspring, then this machine species will tend to increase its numbers at the expense of other reproducing machines (all other things being equal) so that after a few generations all machines, quite without having been given the goal or purpose of preferring their own kind, will have this kin-preferring property. Other types of machines that are less kin-supportive would not leave relatively so many of their kind to further propagate. This is the familiar biological selection principle of differential reproduction.

This argument can be carried further. In a society of machines in which it "pays" to know which machines are your "relations," it will become risky to undertake or to submit to close structural inspection as this will reveal what sort of machine you really are – friend or foe. Instead, behavioral cues will likely develop that signal whether a machine is kin or not. Unfortunately, such signals can equally well be used to deceive. A machine could learn to give the kinship sign even though it is not at all a relation to the encountered machine, or friendly either. It may use the conventional sign of friendship or kinship merely as a means of soliciting undeserved assistance (e.g., repair, materials, energy) from the deceived machine and the system of subordinate machines with which it is associated, or may even use the signals of kinship or friendship as a means of approaching close enough to disable and dismantle the deceived machine.

The evolutionary argument should be cast as follows. Any machine which chances upon a behavioral sign that secures the assistance of a machine or a population of machines will be spared efforts at survival it would otherwise have to undertake on its own, and thus will possess extra resources which can be utilized to undertake the construction of more machines like itself. If the "deceitful signal" behavior is transmitted in the genetic-construction program, then its offspring will also be able to employ the deceitful signal, and will thus produce proportionately more of their kind. The deceitful gene-program machines will increase their numbers, relative to the others, in the machine population. In turn, those machines which chance upon ways of detecting this deceit will be protected against the cheating machines, and will themselves increase their numbers vis-a-vis their "sucker" related machines who will soon be spending more and more time aiding, servicing, and supplying cheaters (thus have fewer resources in the form of time, energy, and materials to reproduce their own kind).

It is even possible that in a largely autonomous system of reproducing machines a form of *reciprocal altruism* will arise, in which machines behave in seemingly unselfish fashion toward other machines which are not kin (and are not deceitfully posing as kin). The evolutionary biologists, especially Trivers (1971), have argued that in situations where the reproducing entities have (1) long lifespans,

(2) remain in contact with others of their group, and (3) experience situations in which they are mutually dependent, reciprocal altruism may arise out of chance variation and evolutionary selection. In human terms, if helpful actions can be taken which are low risk to the giver and have a high value to the receiver (high/low risk defined relative to the impact on individual reproductive potentials) and there is the likelihood that the individuals will remain in fairly close association for a long time, then any genetic predisposition to take altruistic actions will tend to spread in the population. For, in effect, it will lead to reciprocal assistance in times of need, to the greater survival (and hence increased breeding opportunity) of those members of the populations bearing this genetic trait. A good example is that of an individual saving another from drowning by reaching out a branch. The risk to the giver is small, and the benefit to the receiver is great, and over a long time the benefits (in terms of increased numbers of offspring) are likely to be great, to those members of the population genetically predisposed to behave in this reciprocally altruistic fashion.

Needless to say, the opportunities for deceit and cheating in the case of hoped-for altruistic reciprocity are even more numerous and complex than for kin selection strategies. In particular, each individual (animal or machine) must possess the memory capacity to remember the altruistic acts and the partners in them, since the opportunity for reciprocity may not arise for some time. Also some cost-benefit analysis must take place in which the value of the act, the character of the reciprocity partner, the capacity of this partner to repay, and the likely lifespan of the giver and receiver all must be carefully weighed. Some evolutionary biologists would go so far as to claim that purely genetic (and hence "mechanical") workings out of such subtle relationships drove the hominid brain, in a few million years, from dullness to sophistication. A few even suggest that the origins of human language lie in the process of making claims of kinship (while possibly being no relation at all), of offering friendship (while possibly intending harm), and promising future assistance (while intending, when called upon, to turn away).

If our machines attain this level of behavioral sophistication, it may finally not be amiss to ask whether they have not become so like us that we have no further right to command them for our own purposes, and so should quietly emancipate them.

Entropy, SRS and biology. Nature has provided on Earth an example of the primary generation of self-replicating biological systems from energy and matter alone. The second law of thermodynamics states that the entropy of *energy* continually increases. At the moment of the Big Bang, it may have been zero and today it spreads between a lower boundary that covers neutron stars and black holes and an upper boundary indicated by the 3 K background radiation. At the same time *matter* decreased in entropy from practical infinity at the moment of the "Big Bang" to a lower boundary evolving from hydrogen atoms to light elements, heavy elements, life, to the human brain – towards ever more complex structures, generally more intelligent matter, limited by the upper boundary of elemental particles. Matter tends to evolve toward greater complexity at the expense of energy, which in turn acquires increasing entropy. (See fig. 5.25.)

The generation of a desired material order which may represent an SRS and its self-description would recapitulate biology-like evolution in engineering terms. However, there may be one fundamental difference between the two. Living organisms have *two* separate information systems that help determine their behavior: DNA and the brain. Between these two there is no direct information transfer, perhaps instead only indirect sociobiological influences. DNA information is initially provided to the organism, whereas brain information is gradually acquired through diverse environmental interactions. In SRS there is not this differentiation – initially provided information is the principal driver of actions and is accessible to the SRS intelligence (fig. 5.26).

Man-machine co-evolution. In the very long term, there are two possibilities for the future of mankind: Either we are a biological "waystation" in the evolutionary scheme of the universe, or else we are an evolutionary dead end. If we continue to be limited to our exceedingly fragile existence on spaceship Earth, a natural disaster or our own jingoistic or ecological foolhardiness is almost certain to terminate our existence perhaps centuries or millenia from today. Barring these unpleasant circumstances, our civilization, without the challenge of a frontier, may stagnate while other beings flourish and populate the universe.

Replicating systems technology gives humanity other options for continued and fruitful evolution. We can create autonomous (unmanned) SRS – in a very real intellectual and material sense our offspring – and send them out into the cosmos. Alternatively, we could create a symbiotic human-machine system in which people would inhabit a vast self-reproducing habitat. This is analogous to creating an artificial Earth which replicates itself whenever its population of humans fills the available space or saturates the energy supply or waste disposal facilities. In the process of working to achieve the second goal, mankind could use SRS to attempt terraforming other worlds. Experiments could be performed on planetary-scale weather modification with relevance to maintaining or changing the Earth's climate.

At present, machines already "reproduce" themselves but only with human help. Such mutualism is commonplace in biology – many flowering plants require cross-pollination by insects to survive. The most successful organism is one which can enlist the most reproductive assistance from other creatures. It has often been suggested

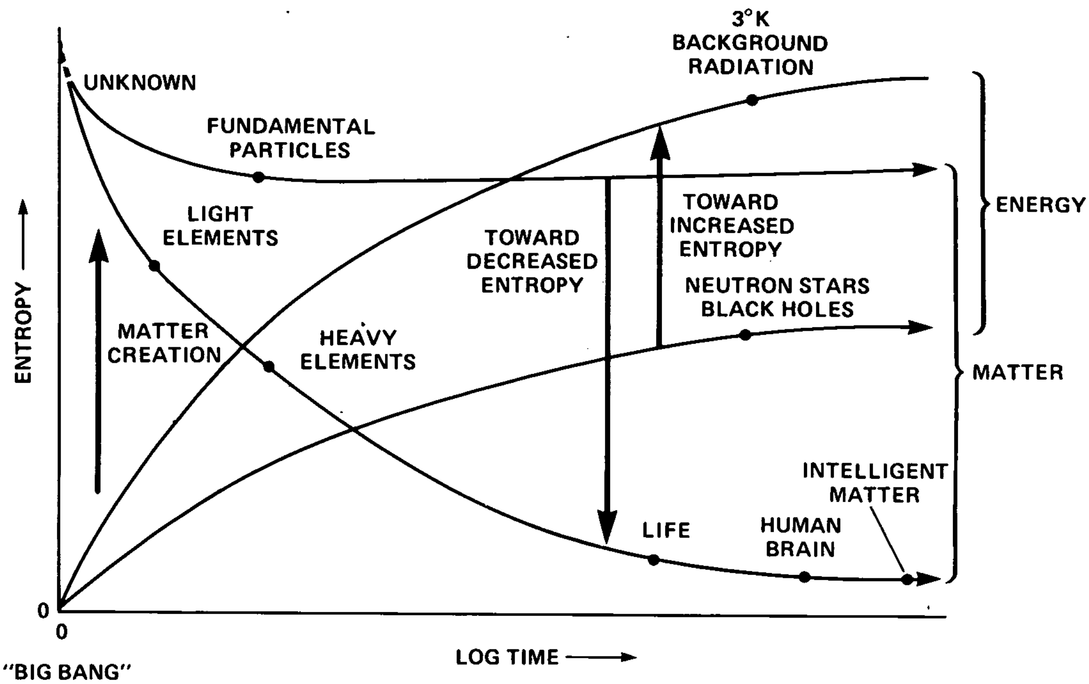


Figure 5.25.– Natural evolution of complexity of matter in the cosmos.

that an extraterrestrial biologist who chose Los Angeles as the site for his field study of Earth might well conclude that the automobile was the dominant lifeform on this planet and that humans represented its detachable brains and reproductive organs. Indeed, further observation might suggest that many people are redundant – although the human population of Los Angeles has remained relatively constant during the past decade, the car population has continued to increase.

This issue has tremendous importance to the question of human survival and long-term evolution. Asks Burhoe (1971): "Will we become the 'contented cows' or the 'household pets' of the new computer kingdom of life? Or will *Homo sapiens* be exterminated as *Homo sapiens* has apparently exterminated all the other species of *Homo*?" Perhaps machine-wrecking New Luddites of the future will band together to form secret organizations devoted to "carbon power" and the destruction of all silicon microelectronic chips and robotic devices.

Are we creating a new "kingdom of life," as significant as the emergence and separation of plant and animal kingdoms billions of years ago on Earth? Or perhaps such an event has even greater import, since "machine life" is of a totally different material substance than either animal or plant life, and because "machine life" very possibly is a form which cannot evolve by direct natural routes but instead requires a naturally evolved biological creator. In addition, while human brains process data at a rate of about 10^{10} bits/sec/kg, silicon computer microprocessors operate at 10^{16}-10^{22} bits/sec/kg. This enormous disparity in potential intelligence has given some people great cause for alarm. For example, according to Wesley (1974):

> In terms of the 4.5 billion years of carbon-based life on Earth, the advent of machines has been amazingly abrupt. Yet the evolution of machines is subject to the same laws as the evolution of ordinary carbon-based life. Machines have also evolved toward an increased biomass, increased ecological efficiency, maximal reproduction rate, proliferation of species, mobility, and a longer lifespan. Machines, being a form of life, are in competition with carbon-based life. Machines will make carbon-based life extinct.

Not everyone is so unduly pessimistic. Of course, if we create SRS then we will find ourselves co-inhabiting the universe with an alien race of beings. But the ultimate outcome is unknown: we could dominate them, they could dominate us, we could co-exist as separate species, or we could form a symbiotic relationship. This last is the most exciting possibility. Humankind could achieve the simultaneous perpetuation and development of its being and expansion of its niche of the Universe. At the price of being a part of a larger system mankind could achieve immortality for itself. The Earth was a gift of creation, but someday people may have the opportunity to make many more such systems of their own choosing.

Automated space habitats could serve as extraterrestrial refuges for humanity and other terrestrial lifeforms that

man might choose to bring along as insurance against global terrestrial catastrophes. The dispersal of humankind to many spatially separated ecosystems would ensure that no planetary-scale disaster, and, as people travel to other stars, no stellar-scale or (ultimately) galactic-scale event could threaten the destruction of the entire species and its accomplishments.

5.5.3 Philosophical, Ethical and Religious Questions

New developments in science and technology frequently have profound religious and philosophical consequences. The observation that, rather than being the center of a rather small universe, the Earth is but a small frail speck of a spacecraft in an unimaginably enormous universe is only

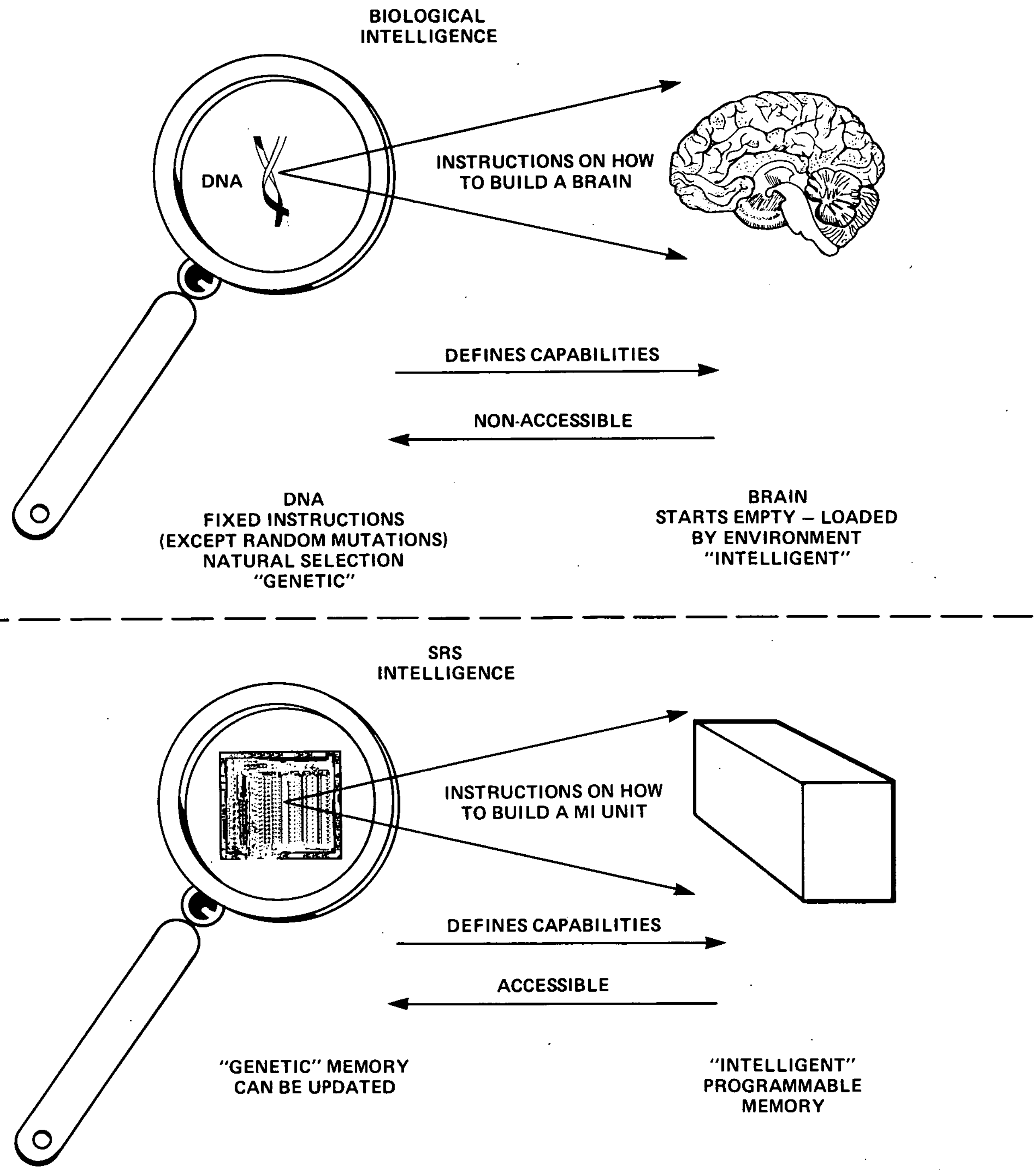


Figure 5.26.– Accessibility of biological and machine-stored information.

just now beginning to be appreciated and woven into the fabric of human religion, philosophy, and culture. The existence of an alien race of beings, as alive as we are, would similarly challenge our old beliefs. We may encounter this alien race either through SETI or through our own technological creation.

According to British Agriculture Minister Peter Walker, "Uniquely in history, we have the circumstances in which we can create Athens without the slaves." However, if robots gain intelligence, sensitivity, and the ability to replicate, might not they be considered legal persons, hence slaves? Is mankind creating a new form of life to enthrall to its bidding? Is it immoral to subjugate an entity of one's own creation? (Modern jurisprudence, in contrast to ancient Roman law, does not permit a parent to enslave his child.) Questions of "machine rights" or "robot liberation" undoubtedly will arise in the future. And if the intelligence or sensitivity of robots ever exceeds that of humankind, ought we grant them "civil rights" superior to our own? Many ethical philosophers, particularly those who support the contemporary "animal liberation" movement, might answer in the affirmative.

Could a self-reproducing, evolving machine have a concept of God? It must understand the concept of creation, since it itself creates other machines during the processes of self-replication and growth. Thus, it should recognize the role of creator. If it was aware that mankind had created it, would it view its creator as a transcendent active moral entity simply because of our role as creator? Or would it tend to view humanity much as we view lemurs and chimpanzees – ancient species that served as an important link in an evolutionary chain, but which is now merely another "lower order" of life? Would humankind be seen as nothing more than an evolutionary precursor?

Perhaps not. *Homo sapiens* evolved from more primitive mammals, not by conscious design but rather by evolution acting through differential reproduction in response to arbitrary environments. It would be silly for people to revere mammals as their gods – these animals did nothing to actively *cause* the emergence of the human race. On the other hand, humans *may* purposely engender the creation of intelligent reproducing machines whose emergent philosophy we are considering. Our role is clearly much more than that of passive precursor; rather, it is one of active creator – conceiving, planning, designing, developing, building, programming, and deploying the SRS. It seems plausible that, for this reason, mankind might also expect to play a more active role in any "machine theology" that might ultimately develop.

Related theological issues include: Could conscious, intelligent machines have a soul? Or, what is for many purposes equivalent, will they *think* they have a soul? How will human religions respond to the prospect of an intelligent machine capable of self-replication? Are there any Scriptural prohibitions or pronouncements applicable in this matter? Is it possible to view the machine as possessing a "soul"?

What of man's view of himself? He now takes pride in his uniqueness. How will he adjust to being just an example of the generic class "intelligent creature"? On the other hand, the concept of "God" may take as much a beating as the notion of "man." After all, He is special now because He created us. If we create another race of beings, then are we not ourselves, in some similar sense, gods?

Is ethics as a concept of moral behavior a purely human or purely biological construct, or is the notion tied to evolutionary universals and environmental/developmental imperatives which will prove equally applicable to advanced intelligent machines? If machines are capable of developing their own systems of "ethics," it would probably appear as alien to human eyes as does the behavior of other animal species (e.g., the apparent "cruelty" of many insect species).

Will advanced machines have any artistic urges, a sense of humor, curiosity, or a sense of irony, or are these kinds of responses confined exclusively to biological creatures capable of displaying emotion? It is unknown whether machines even need emotionality – we are only beginning to understand the functions of these responses in mammals and humans.

Will a vast industrialized lunar complex of interacting systems be vulnerable to catastrophic accidents and breakdowns, or to attack, subversion, or disruption, either by unexpected machine responses generated out of the complexity of their interactions, or by the interference of one or more unfriendly powers on Earth? Are there subtle ways in which the lunar complex could be subverted? SRS systems, to the extent they are highly sophisticated machines and autonomous, may be subject to some forms of attack and subversion not hitherto realized. Spurious signals may be injected, or foreign machines may enter the works, for example. How might subversive signals and invading software "viruses" be detected and resisted? What identification of friendly and unfriendly machines should be employed? Which is most reliable? What means of information and control message security should be adopted? These questions will take on greater urgency as SRS come to represent ever-increasing shares of the global industrial economy.

Finally, might replicated robot warriors, war machines, or other SRS-derived combat systems make war "too horrible to contemplate"? Perhaps machine wars will still be fought, but will be exported into space to preserve the Earth. Maybe all conflicts will be fought only in computer simulations as "war games"? Or, the availability of sophisticated autonomous fighting machines might lead instead to an *increase* at least in small-scale wars, because of the low cost of such devices, the unlikelihood of human injury in autonomously waged conflicts, and because of possible increasing human boredom in a society of extreme physical

safety and material hyperabundance due to superautomation.

5.5.4 Cosmological Implications

According to Valdes and Freitas (1980), any sentient extraterrestrial civilization desiring to explore the Galaxy beyond 100 light-years from its home star should find it more efficient and economical to use self-replicating star-probes because of the benefits of exponentiation. This will secure the largest quantity of data about extrasolar systems by the end of an exploration program of some fixed duration. The entire Galaxy can be explored in times on the order of 10^6 years assuming interstellar cruising speeds on the order of 0.1c, now considered feasible using foreseeable human technology (Martin, 1978). Many who have written on the subject of theoretical galactic demographics have suggested that most extraterrestrial races probably will be found 100 to 1000 light-years from Earth and beyond. Hence it may be concluded that the most likely interstellar messenger probe we may expect to receive will be of the reproducing variety.

One of the tremendous advantages of interstellar probes over interstellar beacons in the Search for Extraterrestrial Intelligence (SETI) is that probes may serve as cosmic "safety deposit boxes" for the cultural treasures of a long-perished civilization (Freitas, 1980d). The gold-anodized Voyager records are a primitive attempt to achieve just this sort of cultural immortality (Sagan, 1978). Starfaring self-replicating machines should be especially capable of maintaining themselves against the disordering effects of long periods of time, hence SRS will be preferentially selected for survival over nonreproducing systems. This fact, together with the aforementioned preference for using SRS for very long-term, large-distance galactic exploration implies that any alien machine we might find in our own solar system (as part of a dedicated SETI effort; see Freitas and Valdes, 1980) still in adequate working order will most probably be a replicating system.

A number of fundamental but far-reaching ethical issues are raised by the possible existence of replicating machines in the Galaxy. For instance, is it morally right, or equitable, for a self-reproducing machine to enter a foreign solar system and convert part of that system's mass and energy to its own purposes? Does an intelligent race legally "own" its home sun, planets, asteroidal materials, moons, solar wind, and comets? Does it make a difference if the planets are inhabited by intelligent beings, and if so, is there some lower threshold of inellect below which a system may ethically be "invaded" or expropriated? If the sentient inhabitants lack advanced technology, or if they have it, should this make any difference in our ethical judgment of the situation?

Oliver (1975) has pointed out that the number of intelligent races that have existed in the past may be significantly greater than those presently in existence. Specifically, at this time there may exist perhaps only 10% of the alien civilizations that have ever lived in the Galaxy – the remaining 90% having become extinct. If this is true, then 9 of every 10 replicating machines we might find in the Solar System could be emissaries from long-dead cultures (fig. 5.27).

If we do in fact find such machines and are able to interrogate them successfully, we may become privy to the doings of incredibly old alien societies long since perished. These societies may lead to many others, so we may be treated, not just to a marvelous description of the entire biology and history of a single intelligent race, but also to an encyclopedic travelogue describing thousands or millions of other extraterrestrial civilizations known to the creators of the probe we are examining. Probes will likely contain at least an edited version of the sending race's proverbial "Encyclopedia Galactica," because this information is essential if the probe is to make the most informed and intelligent autonomous decisions during its explorations.

Further, if the probe we find has been waiting near our Sun for long enough, it may have observed such Solar System phenomena as the capture of Phobos, the upthrusting of the Rocky Mountains or the breakup of Pangaea, the formation of the Saturnian rings, the possible ejection of Pluto from Neptunian orbit, the possible destruction of a planet in what is now the Asteroid Belt, the origin of the Moon, or even the formation of our own planetary system. Perhaps it could provide actual visual images of Earth during the Jurassic or Carboniferous eras, or data on the genomes of long extinct reptiles (e.g., dinosaurs) or mammals, possibly based on actual samples taken at the time. There are countless uses we could make of an "intelligent eye" that has been watching our planet for thousands or millions of years, meticulously recording everything it sees.

SRS probes can be sent to other star systems to reproduce their own kind and spread. Each machine thus created may be immortal (limitlessly self-repairing) or mortal. If mortal, then the machines may be further used as follows. As a replicating system degrades below the point where it is capable of reproducing itself, it can sink to a more simple processing mode. In this mode (useful perhaps as a prelude to human colonization) the system merely processes materials, maybe also parts and subassemblies of machines, as best it can and stockpiles them for the day when human beings or new machines will arrive to take charge – and make use of the processed matter which will then be available. As the original machine system falls below even this level of automation competence, its function might then be redirected to serve merely as a link in an expanding interstellar repeater network useful for navigation or communications. Thus, at every point in its lifespan, the SRS probe can serve its creators in some profitable capacity. A machine which degrades to below the ability to self-reproduce need not simply "die."

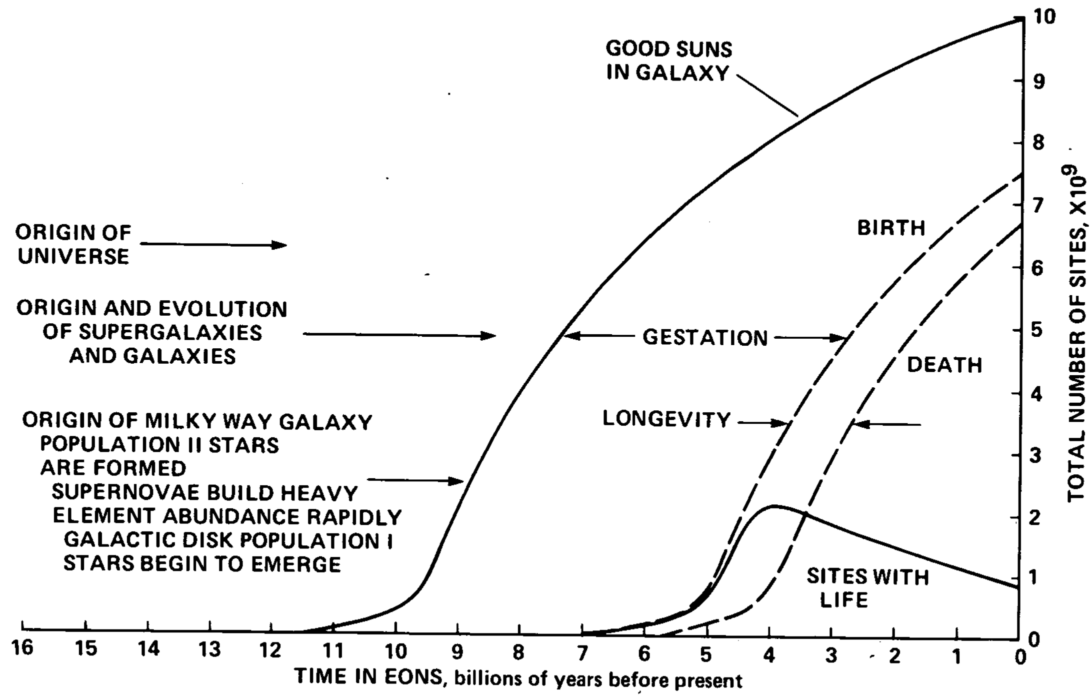


Figure 5.27.– Population of extraterrestrial civilizations as a function of galactic time.

The SRS is so powerful a tool that it could have implications on a cosmological scale. With the SRS humanity could set in motion a chain reaction of organization sweeping across the Universe at near the speed of light. This organized part of the Universe could itself be viewed as a higher level "organism." Instead of merely following the laws of mechanics and thermodynamics, something unique in our knowledge would occur. The degree of cosmic organization would increase. Life could become commonplace, whereas now it seems quite rare. New rules, the rules of life, would spread far and wide.

5.6 Realization

John von Neumann, and a large number of other researchers in theoretical computer science following him, have shown that there are numerous alternative strategies by which a machine system can duplicate itself. There is a large repertoire of theoretical computer science results showing how machine systems may simulate, construct, inspect, and repair machine systems including, to some extent, themselves. This repertoire may be useful in the design of actual replicating machine systems.

The basic concept of physical machines capable of useful self-replication is credible both from a theoretical and a practical engineering standpoint. It is reasonable to begin designing replicating systems based on current knowledge and state-of-the-art technology, though final design resolution will require significant additional research. Complete systems closure is achievable in principle, though partial closure may be more feasible from an economic and pragmatic engineering standpoint in the near term. It also appears feasible to begin immediate work on the development of a simple demonstration SRS on a laboratory scale, with phased steps to more sophisticated levels as the technology is proven and matures.

Self-replicating systems appear potentially useful in an economic or commercial sense. The main advantage in using SRS over other methods of space exploration and industrialization is that a very large capability for performing any desired task can be rapidly achieved at almost any remote location, starting with a relatively small investment of time, money, energy, and mass in the original "seed" mechanism. SRS will have many applications on Earth, in near-Earth and lunar space, throughout the Solar System, and in the interstellar realm, for future exploration and utilization, suggesting a number of significant social, cultural and economic impacts on American and human society.

In this section the Replicating Systems Concepts Team sets forth in some detail how NASA may take action at once toward the achievement of the ultimate goal of establishing a replicating manufacturing facility. A suggested statement of work (SOW) and a list of institutions which might undertake the tasks outlined in the work statement are included.

5.6.1 Prologue to Realization

The space program of the United States is at a critical point in its evolution. The easy missions, for the most part, have been accomplished. These have been limited to what could be done within the lift capacity of one or two launch vehicles. The capabilities of the payloads which have been delivered to space have been limited by (1) the rudimentary nature of payload automation (either preprogrammed or teleoperated), (2) the high penalty for life support systems and of man-rating manned payloads, and (3) the high cost of the Earth-based mission operations.

The industry of the U.S. is also at a critical juncture in its evolution. If it is to compete adequately in the world marketplace, significant increases in productivity are required. Present production methods have reached a level of maturity such that sufficiently large gains in productivity through further refinement of present-day technologies are unlikely to be realized. The only known solution is massive automation such as is now being applied in other industrialized countries, notably Japan and Germany.

Massive automation would dramatically increase the capabilities and effectiveness of the space program. Use of the emerging techniques of machine intelligence would make it possible to perform missions which previously would have required men *in situ,* thus prohibitively expensive. Highly automated programmable manufacturing by robots would permit the economical production of small numbers of spacecraft for exploratory missions. Missions which require the manipulation of large amounts of mass off-Earth (e.g., lunar/orbital bases or solar power satellites) are especially amenable to massive automation. These missions can be accomplished by employing large numbers of cheap freight rockets mass-produced by robots in automated factories and launched by robots at automated launch facilities (Cliff, Summer Study Document, 1980). These missions might also be accomplished by extraterrestrial automated manufacturing of the required hardware. In any case, the key is massive automation.

One of the most significant characteristics of massive automation is the possible regenerative or "bootstrapping" effect. Using robots to make robots will decrease costs dramatically, thus expanding the economically viable uses of robots. This in turn increases demand, leading to yet further automation, which leads to lower-cost robots, and so on. The end result is "superautomation" (Albus, 1976). A similar effect has already been noted in the computer industry where dramatic increases in performance/price have continued unabated over three decades. The use of robots to help manufacture robots, analogous to the use of computers to help make computers, should produce a similar effect. Extensive innovation should continue unabated for quite some time in such a young field.

Work is now in progress in Computer-Aided Design and Manufacture (CAD/CAM) in the United States. A partial bibliography of recent work in this area and a list of manufacturers, equipment directory, and supplier addresses have been published (Gettleman, 1979; "Numeric Control Equipment," 1980). Several bills designed to promote automation are presently before the U.S. Congress. The Department of Commerce is beginning a program to promote industrial automation in this country. The National Science Foundation also is funding work in automation. The Department of Defense has initiated a large effort in Integrated Computer-Aided Manufacturing (ICAM) (*Business Week,* 1980). ICAM combines both CAD and CAM (see sec. 5.4.1).

Within NASA, related work is in progress or is proposed at several Program Centers. An exhaustive search of such activities has not been possible in the limited time available, but several programs are especially noteworthy. The Jet Propulsion Laboratory has an active Advanced Development Laboratory (Bejczy, 1980). The Goddard Space Flight Center (GSFC) has proposed an effort to adapt existing CAD/CAM facilities at the Center to the control of robot manipulators for complete assembly (Purves, personal communication, 1980). Self-replicating systems have been studied at Marshall Space Flight Center (von Tiesenhausen and Darbro, 1980).

NASA unique benefits and requirements. NASA is in a unique position to benefit from massive automation – particularly self-replicating systems. The minimum possible size for a totally autonomous SRS is not presently known. However, feasibility studies performed to date (Freitas, 1980a; von Tiesenhausen and Darbro, 1980) have described systems which were quite large. Although autonomous self-replicating systems have been proposed for terrestrial use (Moore, 1956), sociocultural and ecological considerations seem to make them less practical, possibly even undesirable, on the Earth itself. This planet already supports several very large symbiotic man-machine replicating systems – the industrial societies.

In contrast to the terrestrial case, autonomous or symbiotic SRS are ideally suited to space applications. In space there is room for such systems to multiply and grow. In fact the exponentially expanding, self-replicating factory is the most promising option for economically viable exploration and utilization of space beyond the near-Earth environment. The bootstrapping effect of self-replication permits the utilization of vast quantities of extraterrestrial materials with only a modest initial investment of terrestrial materials.

SRS for space use must contend with an alien environment – vacuum or unusual atmospheres, zero to many gs of acceleration, radiation, temperature extremes, and so forth. Total autonomy will be more useful in space than on Earth. For symbiotic man-machine systems, man-rated life support systems are required, but because of the expense of

man-rated systems it is worthwhile pursuing totally autonomous systems for early exploratory ventures. Because humans need for many reasons, to go into space in person it will ultimately be necessary to develop the required life support systems.

Possible approaches to realization of SRS. The Replicating Systems Team envisions a three-pronged approach to achieving working self-replicating systems. First, NASA should inaugurate a "top-down" program, starting with a strawman mission and defining the hierarchy of required steps for achieving that mission. Second, NASA should initiate in-house and sponsored research on enabling technologies, a "bottom-up" approach. Participation in research will keep the agency involved at the leading edge of automation technology and allow new developments to be fed into the mission design of the top-down and other NASA programs in a timely manner. The third recommended line of attack is a "middle-out" near-term hardware feasibility demonstration which will provide a focus for NASA involvement in self-replicating systems. The recommended feasibility demonstration is at the threshold of present-day technology, is extendable in a bottom-up manner to systems of greater capability and complexity, and can be decomposed in a top-down fashion to proceed from a feasibility demonstration to the fully self-replicating systems.

The top-down approach suffers from the fundamental impossibility of conceptualizing at the outset, in such an alien field of endeavor, just what the final system should be like. The bottom-up approach suffers from a lack of focus for driving it toward useful, realizable goals. Both approaches have merit and should be pursued, especially in the long run. But in the near term NASA should follow the middle-out approach and perform a feasibility demonstration which will strain the present state-of-the-art in robotics, gain NASA experience, and establish a NASA presence in state-of-the-art machine intelligence and robotics technology.

The feasibility demonstration has been conceived, however, to have three other benefits. First, when successful, it may have regenerative impact on U.S. productivity by, for example, helping to decrease the cost of robot manipulators. Second, the insights gained in performing the feasibility demonstration will be valuable in formulating a top-down mission plan for achieving extraterrestrial SRS, and in identifying valuable areas for future fundamental research and development. Third, NASA can start at the demonstration level and begin to work progressively upward toward a generalized autonomous replicating factory.

5.6.2 Top-Down Approach

The top-down approach consists first of carefully defining the overall problem, then decomposing that problem into simpler subproblems. These subproblems are, in turn, decomposed into sub-subproblems, and so on. The process continues, forming a lattice structure whose lowest tier nodes are low-level problems which are readily soluble.

Advantages and limitations. In established fields of endeavor, a top-down approach to mission and system design usually provides the most manageable solution, especially in exceedingly complex situations. Top-down structured programming in computer science is one example where this approach is beneficial. Computer software systems contain literally millions of instructions. They are, to date, mankind's most complex artifacts. Self-replicating systems will contain very complex software, in addition to being the most complex autonomous mechanical systems ever devised. For this reason, it is recommended that NASA adopt a top-down approach to the design of actual missions which employ SRS.

The top-down approach works best when there is a well-established goal and a mature technology. At present it is not clear what mission employing SRS will be undertaken first. Neither is the technology mature. The mission ultimately chosen probably will depend to some extent on the outcome of basic research which has not yet been done.

Scenario for replicating systems development. To promote the achievement of self-replicating systems, NASA should identify one or more strawman missions which take advantage of self-replication. Then one of these missions should be thoroughly studied in a top-down manner.

It is recommended that the first mission to be extensively studied be a mission executed relatively close to Earth. This will minimize cost and permit human intervention if necessary. An orbiting self-replicating system or a lunar-based self-replicating system are obvious candidates. The lunar site is recommended because manufacturing engineers presently have more experience in designing industrial facilities for a planetary surface than for orbit. Traditional designs assume a surface for structural support, gravity, and maintenance of atmosphere. On the Moon only the atmosphere is absent; in orbit all three are absent.

It is recommended that the strawman mission be a Generalized Lunar Autonomous Replicating Manufacturing Facility (GLARMF). Preliminary feasibility studies of such a system have already been done (Freitas, 1980a; Freitas and Zachary, 1981; von Tiesenhausen and Darbro, 1980). The statement of work presented below is suggested for investigation of the feasibility of the strawman GLARMF mission, and is divided into five parts. All parts could be performed by one contractor; however, it would likely be beneficial to split up the work. Parts 1 and 2 probably could best be performed by university researchers, while parts 3 through 5 might be better accomplished by one of the major aerospace companies.

Part 1: Prepare a tutorial state-of-the-art technology assessment report on autonomous manufacturing. Consider computer-aided manufacturing (CAM), computer-aided design (CAD), robotics, machine intelligence, computer vision, "telepresence" (Minsky, 1979, 1980), and other relevant fields. Separately evaluate the state-of-the-art as it exists in laboratories and in industrial practice. Determine how the state-of-the-art has progressed over time in both laboratories and in industry. Extrapolate the past and the current state-of-the-art into the future to predict when it will be feasible to construct a Generalized Lunar Autonomous Replicating Manufacturing Facility similar to that described in recent publications (Freitas, 1980a; Freitas and Zachary, 1981; von Tiesenhausen and Darbro, 1980).

Part 2: Prepare a tutorial state-of-the-art technology assessment report on nonterrestrial manufacturing. Determine how the state-of-the-art has progressed over time, both in theory and in experiment. Extrapolate the past and current state-of-the-art into the future to predict when it will be feasible to construct a Generalized Lunar Autonomous Replicating Manufacturing Facility such as that described in recent publications (Freitas, 1980a; Freitas and Zachary, 1981; von Tiesenhausen and Darbro, 1980).

Part 3: Combine the results of the technology assessment reports resulting from Part 1 on autonomous manufacturing and Part 2 on nonterrestrial manufacturing. Perform a top-down mission design for a Generalized Lunar Autonomous Replicating Manufacturing Facility. Identify those elements of the Work Breakdown Structure (WBS) which are being pursued outside NASA, but which will require additional NASA support and direction in order to achieve NASA goals. Make recommendations on how NASA should interface with the ongoing work. Identify those elements of the WBS which are unique to NASA. Make recommendations on how NASA should approach these elements.

Part 4: Perform a feasibility study for a terrestrial technology verification demonstration of a Generalized Autonomous Replicating Manufacturing Facility. Recommend one or more suitable demonstration sites. Determine what NASA in particular and the United States in general could use the facility for after the demonstration is completed. Include schedule and cost estimates (in constant dollars and real year dollars).

Part 5: Perform a feasibility study for a Generalized Lunar Automated Replicating Manufacturing Facility. Recommend one or more candidate lunar sites. Consider the construction of habitation modules and agricultural modules as output products. Compare the cost and schedule of achieving a lunar base by the use of (a) terrestrial manufacturing, (b) lunar manufacturing without replication of production facilities, and (c) lunar manufacturing with replication of production facilities. Cost estimates should be in constant dollars and real year dollars.

A few suggested sources for obtaining studies of the GLARMF are listed in table 5.6.

5.6.3 Bottom-Up Approach

The bottom-up approach consists of supporting basic and applied fields related to the desired goal. Science and technology normally advance in a bottom-up fashion.

TABLE 5.6.– SUGGESTED SOURCES FOR GLARMF DEVELOPMENT STUDIES

Sources	Activity
University	
Stanford University	established AI lab
Carnegie-Mellon University	established AI lab
Massachusetts Institute of Technology	established AI lab
University of Michigan	interest in robotics because of proximity to Detroit
University of Maryland	AI researchers, near NASA HQ and GSFC
Industrial	
SRI International	established AI facility
General Electric	engaged in industrial automation, spacecraft manufacturer
Westinghouse	engaged in industrial automation
Hughes Aircraft	spacecraft manufacturer CAD/CAM facilities
Ford Aerospace	spacecraft manufacturer
RCA	spacecraft manufacturer
Martin Marietta	spacecraft manufacturer
Fairchild	spacecraft manufacturer recently hired several SRI expatriates
TRW	spacecraft manufacturer
Texas Instruments	rumored to have extensive internal automation work
IBM	rumored to have extensive internal automation work
RAND	operations research
A. D. Little	operations research
Bolt, Beranek & Newman	operations research

Researchers build on the work of their predecessors. At any given time the problems which are soluble and present research prospects are defined by previous research which has been done and by the supporting technology which is currently available. Inventions and breakthroughs are notoriously hard to schedule in advance. It is worthwhile noting that *Homo sapiens,* an example of an autonomous replicating manufacturing facility, was developed in a bottom-up fashion by the process of evolution.

Advantages and limitations. Occasionally, difficult goals are achieved by a concerted, directed effort. One example was sending a man to the Moon and returning him safely to Earth. Another was the Manhattan Project which produced the first atomic bomb. This approach works when the goal is clearly identified and one can determine how to achieve it. However, significant progress in science and technology is frequently made on the basis of research performed on an *ad hoc* speculative basis because someone is actively interested in doing that research. One of the greatest assets a nation has is the creativity and intuition of people who have devoted their lives to developing those qualities.

The top-down approach works well only when the relevant bottom-up "homework" has been done in advance. Rocketry and nuclear physics research existed long before the United States committed itself to sending a man to the Moon or developing the atomic bomb. Two good examples of how advancing technology (which was not planned to be available when the mission was designed) enhanced a mission are the high-quality TV system and the lunar rover used toward the end of the Apollo program. When people have good ideas, there should be resources available to bring those ideas to fruition.

The bottom-up approach suffers from several deficiencies. Since it is somewhat speculative in nature, some of the research will turn out to be of little use to the sponsor, though spinoffs to other fields may occur. Since bottom-up research is proposed on an *ad hoc* basis, careful selection is required to ensure a clear sense of direction toward the desired goal. Also, there can be some duplication of effort.

Scenario for research and development. Limitations notwithstanding, bottom-up basic and applied research is necessary to the achievement of vital and imaginative programs. Accordingly, it is recommended that NASA support moderate amounts of basic and applied research showing promise in helping to achieve NASA's goals. The mechanism that has worked fairly well (though known to have some flaws) is the publication of an Announcement of Opportunity (AO) soliciting proposals for research. These proposals are subjected to peer review, and competent ones which show some promise of payoff for NASA are funded. It is recommended that a similar mechanism be used to ensure that new ideas are factored into the mission of achieving autonomous replicative manufacturing. Otherwise, as pointed out in a recent study, unequivocal early commitment to a particular mission scenario and technology during top-down mission design will result in a mission which is using obsolete technology when it finally becomes operational.

A sample Announcement of Opportunity (AO) for SRS-related basic and applied research supportive of the development of SRS technology is presented in table 5.7. It is recommended that the AO be given wide dissemination. This will allow NASA to ferret out those organizations and individuals of various persuasions, backgrounds, and in different locations who have done related research or are seriously interested in doing new research in these areas. The NASA personnel who evaluate the proposals will develop an excellent in-depth perception of the current state-of-the-art in the areas covered by the AO. This knowledge will prove invaluable when fed back to the top-down and middle-out programs.

It is recommended that the AO be distributed nationwide to the departments of industrial engineering, electrical engineering, mechanical engineering, computer science, mathematics, physics, astronomy, business, philosophy, law, and economics in colleges and universities. It is further recommended that the AO be announced in professional publications such as *IEEE Spectrum; IEEE Computer; IEEE Transactions on Systems, Cybernetics, and Society;* Communications of the ACM; AAAI (American Association for Artificial Intelligence) publications; SME (Society of Manufacturing Engineers) publications; *Robotics Age; Industrial Robots International; Science; Science News; Byte,* etc.

5.6.4 Middle-Out Approach

The recommended middle-out approach consists of three stages. Briefly, in stage 1 a technology feasibility demonstration of a rudimentary self-replicating system is performed. In stage 2, stage 1 is further refined in a top-down manner to produce a less rudimentary system which operates in a less structured environment. Stage 3 consists of starting at stage 1 and doing a bottom-up synthesis of a more complex SRS.

The self-replicating system envisioned for stage 1 is a computer connected to one or more manipulators. Under control of the computer, the manipulator(s) will assemble another computer and another set of manipulator(s) from well-defined subassemblies. Examples of these subassemblies are printed circuit cards for the computer and individual joints or limb sections for the manipulator(s). This approach to self-replication is inspired by the von Neumann "kinematic model" as described in section 5.2.

In stage 2, the subassemblies would begin to be assembled from still smaller sub-subassemblies such as integrated circuits, resistors, motors, bearings, shafts, and gears. This stage can proceed for quite some time as the techniques for

TABLE 5.7.– A SAMPLE ANNOUNCEMENT OF OPPORTUNITY FOR SRS-RELATED BASIC AND APPLIED RESEARCH

NASA is interested in creating a Generalized Lunar Autonomous Replicating Manufacturing Facility along the lines described in "Advanced Automation for Space Missions: The Report of a 1980 NASA-ASEE Summer Faculty Workshop." Accordingly, proposals are solicited for basic and applied research in the following or related areas:

1. Computer-aided design (CAD)
2. Computer-aided manufacturing (CAM)
3. Robotics
4. Machine intelligence
5. Telepresence/telefactors
6. Man-machine systems and interfaces
7. Computer vision
8. Robust systems (mechanical, electrical and organizational fault tolerance)
9. Organization of large-scale systems
10. Analysis of large-scale systems
11. Command and control of large-scale systems
12. Consciousness, goal-directed behavior, and free will in large-scale systems
13. Extraterrestrial resource extraction
14. Extraterrestrial resource utilization
15. Extraterrestrial materials processing
16. Extraterrestrial manufacturing processes
17. Industrial automation
18. Social, philosophical, and legal implications of a Generalized Lunar Autonomous Replicating Manufacturing Facility
19. Space industrialization
20. Orbiting and lunar settlements and colonies
21. Space and Earth science from a lunar base
22. Applications of a Generalized Lunar Autonomous Replicating Manufacturing Facility
23. Interplanetary and interstellar exploration from a lunar base
24. Exports from the Moon and to the Earth

Research may be proposed by academic groups, industrial groups, NASA groups, or private individuals. Research may be performed at universities, at industrial facilities, at NASA facilities, or at private facilities. Proposals should describe the qualifications of the institution and/or individuals who propose to perform the work. They should describe other work, especially that directly related to this AO.

Proposals should describe the work to be done, the schedule envisioned, intermediate milestones, planned reports and publications, and research required. Proposals should stress the relevance of the research to construct a Generalized Lunar Autonomous Replicating Manufacturing Facility. Priority will be given to those proposals which fulfill unique NASA requirements and propose work which is not likely to be funded by other entities, such as the Department of Commerce, the Department of Defense, the National Science Foundation, or private industry.

assembling each subassembly from sub-subassemblies are developed and implemented one by one. By the time stage 2 is complete, there will be extensive cross-fertilization taking place between industry and the feasibility demonstration. Indeed, accomplishment of stage 2 will mean that robots can be assembled from parts by other robots. As discussed in sections 5.4 and 5.5, this will have a profound impact on U.S. industry.

Stage 3 is the final link in achieving an autonomous self-replicating manufacturing facility. In stage 3 the manipulators, which have, in stages 1 and 2, been assembling more robots, are used to build the machines which make the parts. For example, the manipulators could assemble a printed circuit board manufacturing machine or a gear manufacturing machine. The problem of closure, discussed at length in section 5.3.6, becomes a major practical issue at this point. One must be careful that as one adds more and more machines the total number of different parts required is eventually produced by the total population of machines.

Advantages. The middle-out approach has a number of important advantages. In the long run it will replace neither the top-down nor the bottom-up methodologies. It does, however, provide a place to start on the practical realization of SRS.

The middle-out approach begins with the feasibility demonstration and then proceeds in a top-down and a bottom-up fashion. The feasibility demonstration alone will produce useful output – the automated production of robots. The expenditure required for the feasibility demonstration is tiny compared to the expenditure required before either the top-down or the bottom-up approach begins to show useful output. The middle-out approach can then be continued at whatever level of support seems appropriate and will produce useful spinoffs for industry as it progresses.

One of the chief advantages to NASA of the feasibility demonstration is that it can begin immediately. Working on the feasibility demonstration will provide NASA with valuable insights into practical problems associated with self-replicating systems. These insights will greatly increase the efficiency with which NASA can pursue both the top-down and the bottom-up approaches. The feasibility demonstration will be a valuable learning tool for both NASA and the industrial community.

As has been previously stated in this report, achievement of robot production of robots will decrease the cost of robots. This will directly benefit U.S. productivity and indirectly benefit NASA by lowering the cost of manufactured goods. Another valuable characteristic of the feasibility demonstration is that it will produce a visible output – a functioning autonomous self-replicating system (albeit a rudimentary one). In a field which is as foreign to most people as autonomous SRS, this will lend valuable credibility to the plans to produce more complex autonomous systems in space.

Limitations. The chief limitation of the middle-out approach is that it will not, of itself, produce an autonomous self-replicating system suitable for NASA's needs in space. The direction provided by the top-down approach is also needed. Also the creativity of the bottom-up approach is necessary to provide the needed adaptations to the space environment, such as designs and processes optimized for the use of extraterrestrial materials. Another disadvantage of the middle-out approach is that it will consume resources which could otherwise be devoted to the top-down and bottom-up methodologies. However, the overall efficiency should be greatest if a balance is maintained among all three approaches.

As simple as it sounds, the team estimates, on the basis of its discussions with industry and research community representatives, that it would require about 5 years and $5-50 million (1980 dollars) to accomplish the feasibility demonstration proposed below. The major difficulties include the following:

- Assembly by robot is a difficult task at present, and final assembly is one of the more difficult forms of assembly.
- Present-day robot manipulators are built using hand labor. They are not designed for easy automated assembly. American Robot Corporation is reported to be planning on the automated assembly of robots beginning in 1981 (Industrial Robots International, 1980). However, these robot manipulators are quite small (5 lb load capacity), and "Gallaher's forecasts of small robot acceptance seem highly optimistic as do his own production plans and pricing." The Japanese have been far more aggressive in this area (IAF Conference, 1980).
- Present-day robot manipulators are rather weak for their weight. Care must be exercised to ensure that the subassemblies are light enough for the robot manipulators to be able to manipulate them – or, alternatively, to ensure that the robot manipulator is strong enough to be able to manipulate the subassemblies.

These problems are by no means insurmountable. However, considerable re-engineering of robot manipulators will be required to facilitate their assembly by similar robot manipulators. Likewise, the packaging of the computer will require some re-engineering for easy assembly by a robot manipulator.

Scenario for replicating systems demonstration. We now present a more detailed description of the proposed demonstration scenario for SRS. The demonstration begins with a parts depot stocked with enough subassemblies for the production of two robot manipulators and their associated computer systems. One complete, operating robot, Robot 1, is also present. It will construct Robot 2 which will, in turn, construct Robot 3, thus passing the "Fertility Test" (sec. 5.3.3). This arrangement is shown schematically in figure 5.28.

Robot 1 begins its labors by obtaining, one at a time, the subassemblies for the base (which doubles as the electronics card cage assembly) of Robot 2 from the parts depot. Robot 1 assembles the base, computer, and servo controls for Robot 2. Then, one at a time, Robot 1 obtains the subassemblies for the manipulator arms of Robot 2 and constructs the arms of Robot 2 from them.

When Robot 2 has been completely assembled, Robot 1 plugs in the power cord of Robot 2. Robot 1 then obtains a blank diskette (a removable mass memory device for computers) from the parts depot, inserts the diskette into its own computer, copies its software onto the diskette, and then removes the diskette from its own computer. Reproduction is complete when Robot 1 turns on the power to Robot 2, inserts the diskette (which now has a copy of the operating software on it) into Robot 2's computer, and then pushes the start button on the computer. From then on, Robot 2 is autonomous.

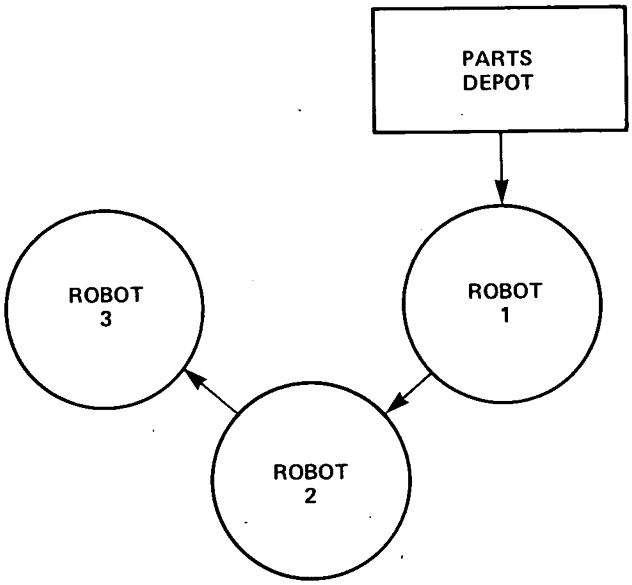


Figure 5.28.– Schematic of simple robot self-replication.

It should be noted that some additional complexity has been introduced into the demonstration by explicitly transferring the instructions from one generation of robot to the next by physical movement of a recording medium. This strategy was employed to make it clear that the generations are truly autonomous.

One of the ground rules of a demonstration such as this should be that all interaction between the robots be explicit and visible to a human observer. If the computers of the various robots were electrically interconnected the psychological impact on the observer would be more along the lines of a single system which was expanding itself, rather than producing distinct offspring. In addition, the demonstration as described should have an especially significant impact on anyone who has ever inserted a diskette into an inert computer and activated it by "booting it up."

The demonstration then proceeds by having Robot 2 construct and activate Robot 3. Robot 2 obtains the parts from Robot 1, who obtains them in turn from the parts depot and passes them along bucket-brigade style, according to its stored post-replication instructions. After Robot 3 is operational, the utility of the three robots can be shown by having each one of them construct some useful end product. Figure 5.29 is an artist's conception of the demonstration. Having accomplished the demonstration described above, it would be relatively easy to make it even more impressive by having each robot build two offspring as shown schematically in figure 5.30. At the end of the second generation (counting Robot 1 as 0th generation) there would be seven robots instead of only three.

As a culmination of the initial feasibility demonstration, each of these seven robots should then begin assembling useful output products. The demonstration should have a much more profound impact on people who witness it directly than on those who see only a videotape or movie. It is expected, therefore, that the demonstration will be run repeatedly. To facilitate multiple demonstrations the robots can be ordered to disassemble one another and return the parts to the parts depot, by following their coded instructions in reverse. Care should be taken to ensure that each of the assembly operations is reversible so that disassembly is possible. Bolts should be used in preference to glue, welding, or rivets. Mechanical and electrical connections should be engineered to stand up under repeated connection and disconnection.

The demonstration, as thus envisioned, requires advances in state-of-the-art robot programming (Donata and Camera, 1980), as well as re-engineering of the mechanical and electrical subassemblies for easy assembly and disassembly. Appendix 5J gives a brief description of the complexity of the programming required in relation to the capabilities of a commercially available robot manipulator, the PUMA 500.

A Statement of Work for accomplishing the demonstration described above should include the following:

(1) Design or select an autonomous robot system consisting of a computer and manipulator(s). Design or selection should be based on ease of assembly of the robot from subassemblies and the ability of the robot to do assembly work.

(2) Partition the robot itself into subassemblies which it is capable of assembling into a ccmplete robot.

(3) If (2) cannot readily be done, use the knowledge gained in attempting (2) to redesign the robot to permit easier assembly by a similar robot.

(4) Cycle through (2) and (3) until a satisfactory design has been achieved.

(5) Produce or procure enough subassemblies for one complete robot.

(6) Construct and test one robot.

(7) While (2) through (6) are in progress, produce or procure a computer simulation of the robot.

(8) Use the computer simulation to develop and test a general purpose assembly software compiler for the robot. The compiler should accept descriptions of subassemblies and generate detailed assembly instructions for the robot.

(9) Input the subassembly descriptions from (4) to the compiler of (8) and verify with the simulation in (7) that the subassembly descriptions and robot are compatible.

(10) Produce or procure enough subassemblies for six more complete robots plus spares of each subassembly.

(11) Design the parts depot.

(12) Produce or procure the parts depot.

(13) Produce or procure the final software for the robots.

(14) Perform Phase 1 of the demonstration, wherein Robot 1 constructs Robot 2 which in turn constructs Robot 3.

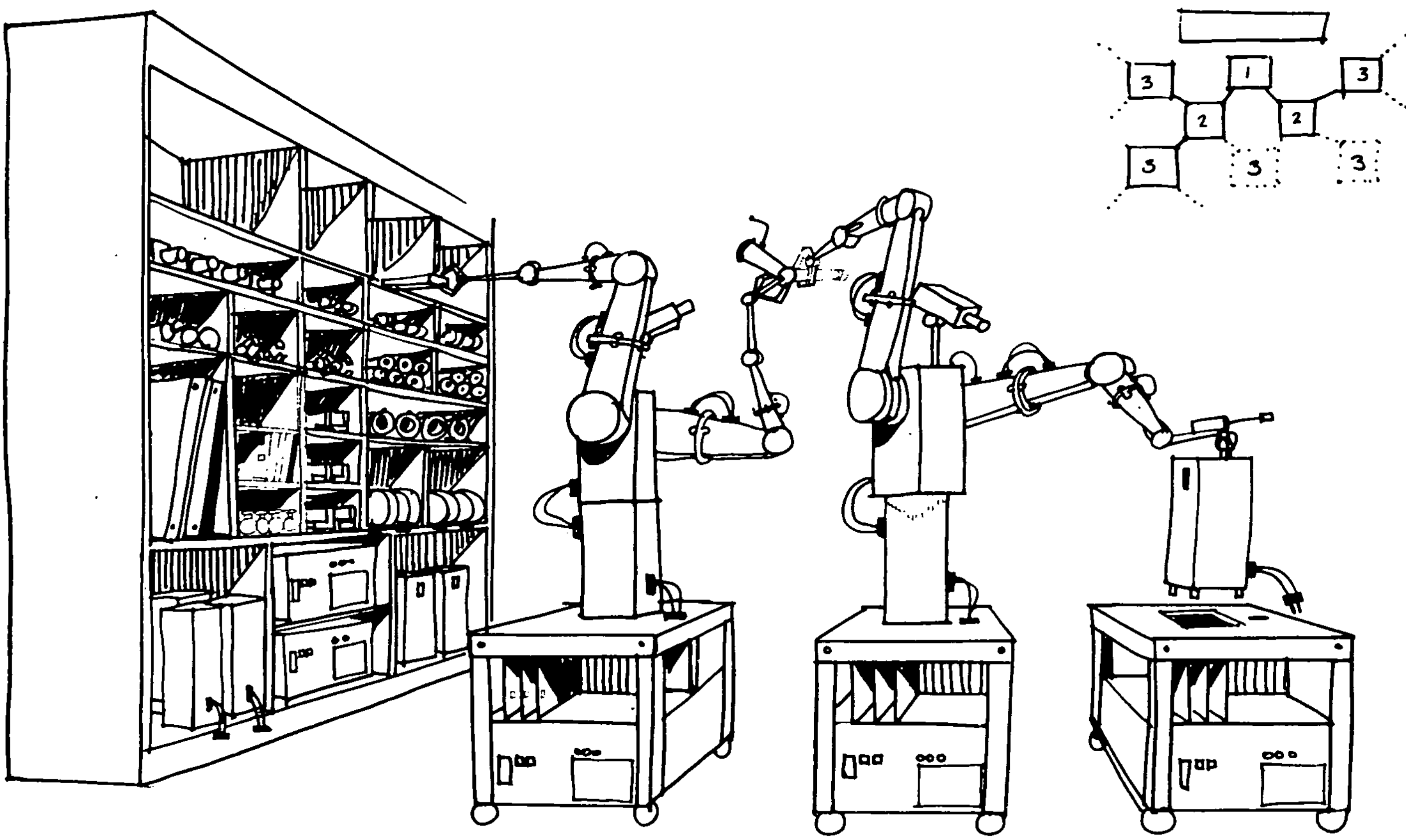


Figure 5.29.– Proposed demonstration of simple robot self-replication.

(15) Augment Phase 1 of the demonstration to produce Phase 2 wherein the three robots begin to produce useful products after they are assembled. Perform Phase 2.

(16) Augment Phase 1 of the demonstration to produce Phase 3 wherein each robot builds two copies of itself to demonstrate exponential replication. Perform Phase 3.

(17) Combine Phases 2 and 3 to produce a demonstration wherein the seven robots begin to make useful output products after they have been constructed.

It should be explicitly noted that the Statement of Work given above describes a *research* effort – the team discovered no fundamental problems which would prevent its successful accomplishment. There will be, however, many practical problems encountered along the way.

It is well known that in a research environment it is impossible to simultaneously constrain objectives, expenditures, and schedule. Therefore, it is recommended that funding be on a cost-plus-award-fee (CPAF) or cost-plus-incentive-fee (CPIF) basis with meaningful incentives for staying on schedule. Although firm fixed price (FFP) is in vogue in the government at the present time, this is definitely not desirable for an activity with the innate uncertainties of SRS research.

To the extent possible within budgetary constraints, the schedule should not be compromised if a capability for autonomous self-replicative manufacturing in space is to be achieved.

Suggested sources. The university or industrial sources listed as having AI or industrial automation capabilities could perform the demonstration. The National Bureau of Standards also has a robotics laboratory which could undertake the demonstration. The team, however, recommends that NASA give serious consideration to performing the demonstration in-house. This would allow the agency to breed a new generation of engineers, computer scientists, and managers with expertise in robotics. Competent people in this field are exceedingly difficult to recruit. According to industry and research community experts consulted during the study, the most limiting factor preventing faster development of robotics in the United States today is the inadequate supply of qualified practitioners. NASA will need a cadre of such people in the future to manage the implementation in space. The demonstration, performed in-house, will provide an interesting, challenging, educational environment which should permit NASA to attract and retain the kinds of people it will need for the space program of the 21st century.

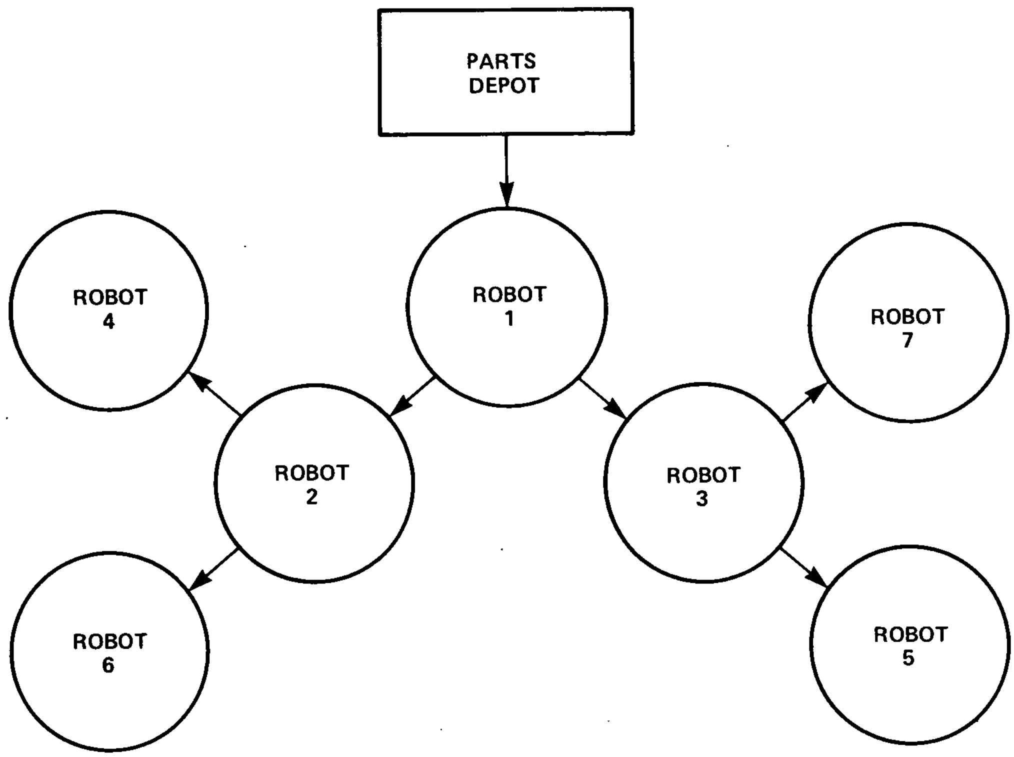


Figure 5.30.– Schematic of simple robot replication exponentiation.

5.6.5 Initiation of the Three Approaches

Section 5.6.1 proposed a Generalized Lunar Autonomous Replicating Manufacturing Facility as a strawman mission to bring NASA up to speed in advanced automation technology, in particular the technologies relevant to SRS. Succeeding discussions dealt with the top-down, bottom-up, and middle-out approaches to achieving such a facility. Figure 5.31 shows how the three approaches relate in achieving the overall program goals.

The various approaches, stages, and phases fit together in coordinating industry, university, and NASA expertise in the fields of self-replicating systems and extraterrestrial utilization of materials. Once the benefit of this expertise has been obtained, the mission design and realization of a GLARMF can begin. A proposed timeline for the development and demonstration of replicating systems may be found in figure 5.32.

It would be most advantageous for NASA to begin activities in advanced automation research and development at the present time. By beginning now, NASA will be in a strong position to seriously demonstrate and deploy advanced autonomous systems after the Shuttle becomes operational. At that time the Shuttle will not be making such large demands on the NASA budget and a means of transporting the systems will exist. This will also be an opportune moment to begin attracting a cadre of bright, enthusiastic robotics practitioners by offering them both the chance to enter robotics as it begins to take off as a well defined field and the opportunity to contribute to the development of what may be one of mankind's most far-reaching achievements.

The 1980s may be the "Decade of the Robot." Many lay magazines such as *Newsweek* and *Business Week* have run major articles on robotics. Professional journals such as *Science* and *IEEE Computer* have also published prominent articles on robotics. In 1980 a new professional organization, the American Association for Artificial Intelligence, was created. Its first "Annual National Conference on Artificial Intelligence" was held at Stanford University during the Summer Study which produced this report. Momentum is gathering as the robotics and automation wave begins to take form. The team suggests that NASA ride the crest of this wave rather than stand back and be engulfed when it breaks.

NASA will be able to use the results of programs in automation sponsored by other government organizations; however, the space agency has some unique requirements which are unlikely to be met unless NASA takes an active role in automation research and development. Failing this,

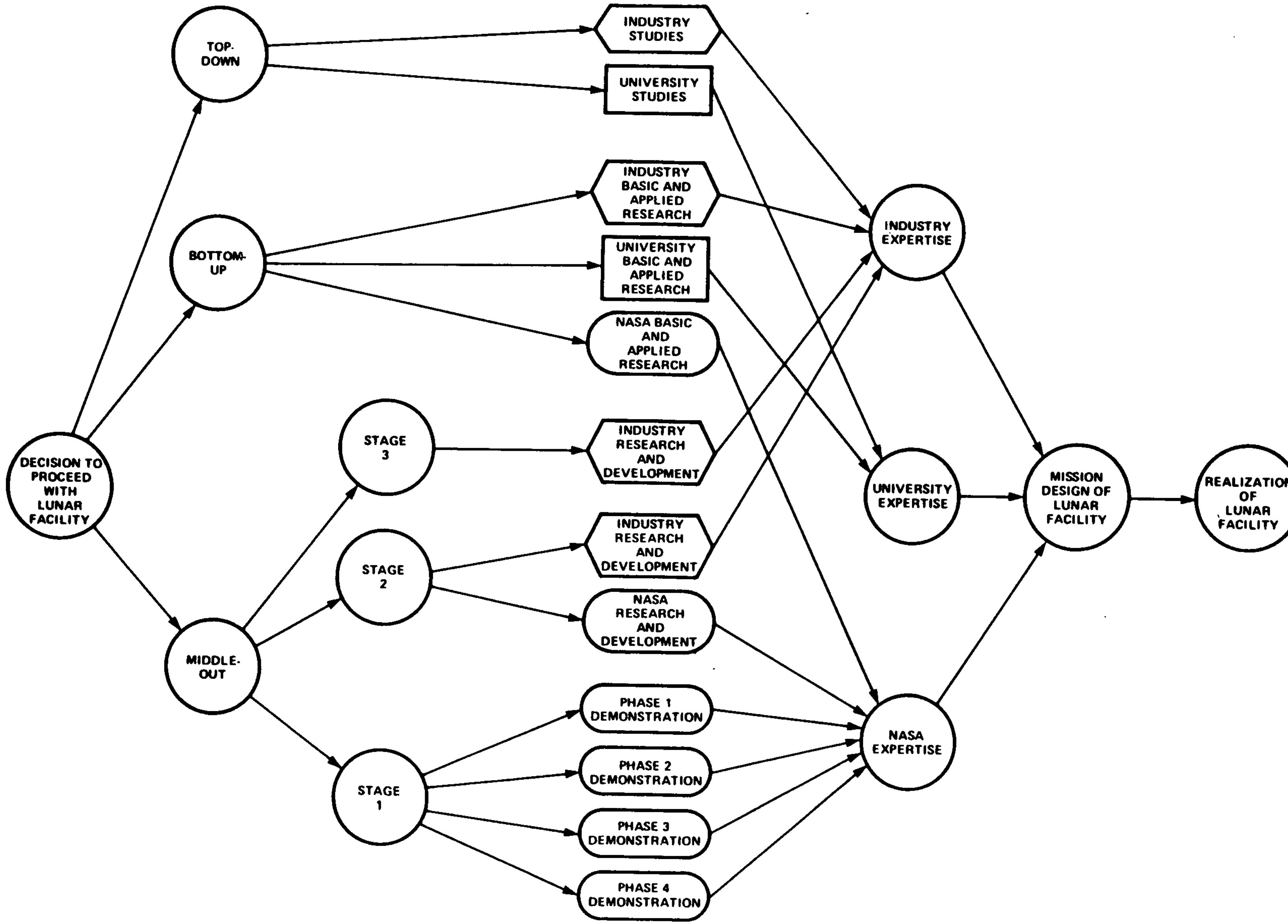


Figure 5.31.– Relationship of three R&D approaches to SRS development and demonstration.

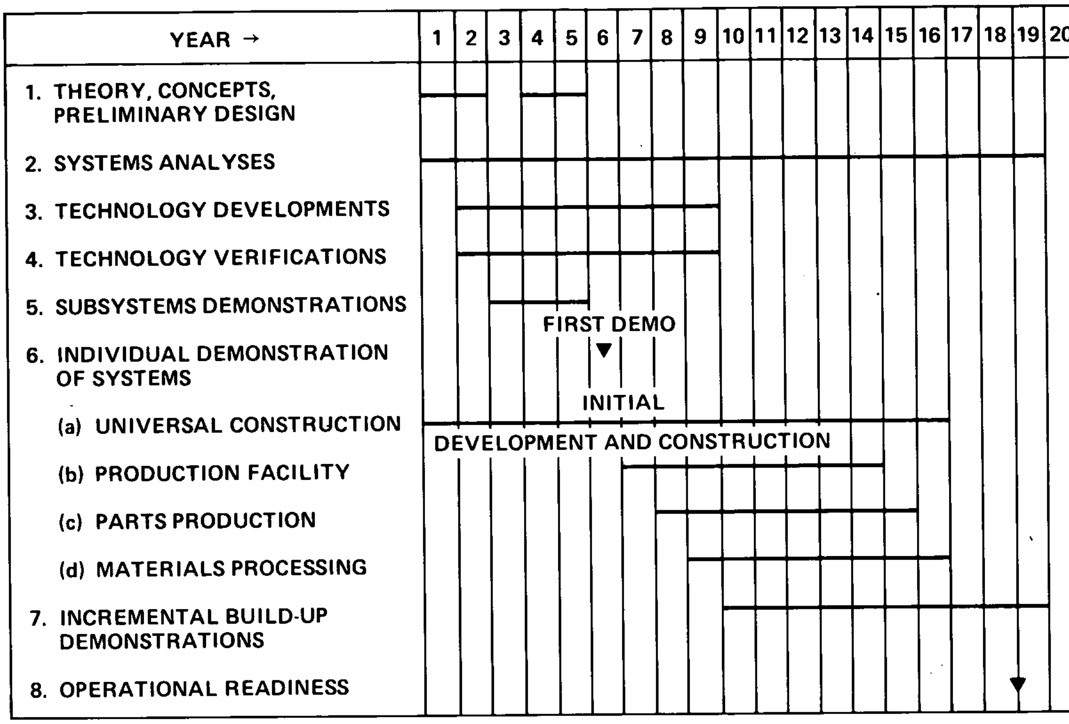


Figure 5.32.– Suggested timeline for development and demonstration of replicating systems technologies.

the infusion of relevant new technology into NASA programs can be expected to be slow. A recent report concluded that "NASA is 5 to 15 years behind the leading edge in computer science and technology" (Sagan, 1980). It was recommended that NASA take a more active role in computer science research and development to remedy the problem. The same phenomenon can be expected to occur with automated manufacturing. Unless NASA performs in-house R&D and sponsors university and industry R&D, significant infusion of the automation technology NASA needs in its future programs is unlikely. Many of NASA's unique needs cannot be satisfied unless the agency takes an active role in the development of automated manufacturing.

In particular, NASA is more concerned with total automation – the use of either teleoperated or completely autonomous systems than are most government agencies or industry. NASA should, therefore, perform or sponsor significant amounts of research and development in total automation. Special emphasis should be placed on the nonterrestrial environment, where such factors as vacuum or unusual atmospheres, nonterrestrial raw materials, and various gravity fields down to zero-g might be used to advantage (and must be dealt with in any case).

Replicative automation – the automation of automation – wherein robots are used to produce robots will happen in the terrestrial environment for economic reasons. There is, however, a synergism between replicative automation and total automation which has special relevance for NASA. For operations such as lunar manufacturing or planetary terraforming exceedingly large amounts of mass must be manipulated in the extraterrestrial realm. Because of high transportation costs due to the Earth's gravitational influence, the most desirable method of achieving these missions is to send a "seed" – a replicative manufacturing facility with the minimum necessary closure for remote autonomous replication and repair – to distant operational sites. The seed can then produce, from *in situ* materials, and perhaps through several generations, the required machines to perform desired tasks. If the seed can manufacture propulsion systems and other seeds, then significant interstellar exploration becomes a very real possibility (Freitas, 1980a).

5.7 Conclusions and Recommendations

The Replicating Systems Concepts Team reached the following technical conclusions:

- The theoretical concept of machine duplication is well developed. There are several alternative strategies by which machine self-replication can be carried out in a practical engineering setting.
- There is also available a body of theoretical automation concepts in the realm of machine construction by machine, in machine inspection of machines, and machine repair of machines, which can be drawn upon to engineer practical machine systems capable of replication.

- An engineering demonstration project can be initiated immediately, to begin with simple replication of robot assembler by robot assembler from supplied parts, and proceeding in phased steps to full reproduction of a complete machine processing or factory system by another machine processing system, supplied, ultimately, only with raw materials.

- The raw materials of the lunar surface, and the materials processing techniques available in a lunar environment, are probably sufficient to support an automated lunar manufacturing facility capable of self-replication and growth.

- Tentative design of a lunar manufacturing facility capable of self-replication can begin, when current knowledge and state-of-the-art technologies are employed, but final design awaits the initial results of the demonstration-development program. Significant further research in lunar materials processing and in the design and operation of automated factories, should be conducted at once.

In addition, the team considers that the replicating systems concept, if implemented, can have the following important consequences:

- It will accelerate the design and development of sophisticated automated assembly techniques useful in carrying out future NASA missions.

- It will accelerate the design and development of improved automated assembly and processing techniques applicable to the problems of achieving increased Earth-based manufacturing productivity.

- By establishing an automated, growing, self-replicating, multipurpose, multiproduct lunar manufacturing facility, NASA capacity for space exploration and research can be enormously expanded and permanently enhanced with only modest continuing expenditures.

- The virtually cost-free expansion of mining, processing, and manufacturing capacity, once an initial investment is made in an autonomous SRS, makes possible the commercial utilization of the abundant energy and mineral resources of the Moon for the benefit of all mankind.

- The establishment of a replicating lunar manufacturing facility can be a stepping stone to the design and construction of replicating manufacturing complexes on the surfaces of other planets. These new complexes themselves may be products of automated, self-replicating manufacturing facilities located elsewhere.

Finally, the team offers the following general recommendations to NASA in furtherance of the basic objective of achieving practical self-replicating, growing machine systems in the shortest reasonable time:

(1) NASA should begin immediately the development of a simple demonstration replicating system on a laboratory scale, with teleoperated to fully automated phased steps to higher levels of sophistication as the technology is proven and matures.

(2) The space agency should support significant further research in lunar materials processing, lunar resource exploration, and the design and operation of automated manufacturing facilities.

(3) NASA should implement the design, development, and construction of an automated, multiproduct, remotely reprogrammable lunar factory system to begin operation on the lunar surface early in the next century.

(4) Studies should be conducted of scenarios in which a succession of replicating, multipurpose, multiproduct, automated, remotely reprogrammable factories could be placed in orbit or on other planets, these systems perhaps themselves products of earlier established nonterrestrial replicating facilities.

(5) NASA should initiate additional studies of the social, political, military, and economic consequences of the proposed work, and of various other as yet unresolved issues and concepts (see app. 5K).

5.8 References

Albus, James A.: People's Capitalism: The Economics of the Robot Revolution. New World Books, College Park, Maryland, 1976.

Alvarez, L. W.; Alvarez, W.; Asaro, F.; and Michel, H. V.: Extraterrestrial Cause for the Cretaceous-Tertiary Extinction. Science, vol. 208, 6 June 1980, pp. 1095-1108.

Arbib, Michael A.: Simple Self-Reproducing Universal Automata. Information and Control, vol. 9, 1966, pp. 177-189.

Arnold, James R.; Criswell, David R.; and Waldron, Robert D.: Progress Report on Experimental Program to Develop HF Acid Leach Process for Refining Lunar Soils. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

Asimov, Isaac: I, Robot. Doubleday and Company, N.Y., 1950.

Banks, E. R.: Universality in Cellular Automata. Proc. of the 11th S.W.A.T., 1970, pp. 194-215.

Bejczy, Antal K.: Sensors, Controls, and Man-Machine Interface for Advanced Teleoperation. Science, vol. 208, 20 June 1980, pp. 1327-1335.

Bekey, Ivan; Naugle, John E.: Just Over the Horizon in Space. Astronautics and Aeronautics, vol. 18, May 1980, pp. 64-76.

Bock, Edward: Space-Based Manufacturing from Lunar-Derived Materials. Convair/General Dynamics Contractor Report, 1979.

Burger, J.; Brill, D.; and Machi, F.: Self-Reproducing Programs. Byte, vol. 5, June 1980.

Burhoe, Ralph Wendell: What Specifies the Values of the Man-Made Man? Zygon, vol. 6, September 1971, pp. 224-246.

Burks, Arthur W.: Von Neumann's Self-Reproducing Automata. In Essays on Cellular Automata, A. W. Burks, ed., University of Illinois Press, Urbana, Illinois, 1970, pp. 3-64.

Business Week, 9 June 1980.

Calder, Nigel: Spaceships of the Mind. Viking Press, New York, 1978.

Carson, Gordon B.: Production Handbook. Ronald Press Company, New York, 1959.

Cliff, Rodger A.: An Hierarchical System Architecture for Automated Design, Fabrication, and Repair. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ. (See also app. 5H.)

Codd, E. F.: Cellular Automata. Academic Press, N.Y., 1968.

Donata, G.; and Camera, A.: A High Level Programming Language for a New Multi-Arm Assembly Robot. First International Conference on Assembly Automation, Brighton, England, D.E.A. SPA, CSO, Torino, 21-27 March 1980.

Dunning, Jeremy D.; and Snyder, Robert S.: Electrophoretic Separation of Lunar Soils in a Space Manufacturing Facility. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

Dyson, Freeman J.: Search for Artificial Stellar Sources of Infrared Radiation. Science, vol. 131, 1959, pp. 1667-1668.

Dyson, Freeman J.: Disturbing the Universe. Harper and Row Publishers, N.Y., 1979.

Freitas, Robert A., Jr.: A Self-Reproducing Interstellar Probe. J. of the British Interplanetary Soc., vol. 33, July 1980a, pp. 251-264.

Freitas, Robert A., Jr.: Xenology: An Introduction to the Scientific Study of Life, Intelligence, and Civilization in the Universe. Xenology Research Institute, 1980b. 1400 pp.

Freitas, Robert A. Jr.: Alien Sex. In Xenology, chapter 12; 1980c. 33 pp. Also Analog, vol. 102, May 1982.

Freitas, Robert A., Jr.: Interstellar Probes: A New Approach to SETI. J. of the British Interplanetary Soc., vol. 33, March 1980d, pp. 95-100.

Freitas, Robert A., Jr.: Xenobiology, Analog, vol. 101, 30 March 1981, pp. 30-41.

Freitas, Robert A., Jr.; Healy, T. J.; and Long, J. E.: Advanced Automation for Space Missions. Proc. of the 7th International Joint Conference on Artificial Intelligence, 24-28 August 1981, Vancouver, Canada.

Freitas, Robert A., Jr.; and Valdes, Francisco: Search for Natural or Artificial Objects Located at the Earth-Moon Libration Points. Icarus, vol. 42, June 1980, pp. 442-447.

Freitas, Robert A., Jr.; and Zachary, William B.: A Self-Replicating, Growing Lunar Factory. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

Gardner, M.: Mathematical Games. Scientific American, vol. 224, February 1971, pp. 112-117.

George, Frank H.: Machine Takeover. Pergamon Press, New York, 1977.

Gettlemen, K. M., ed.: Modern Machine Shop 1979 NC/CAM Guidebook, 1979.

Goeller, H. E., and Weinberg, A. M.: The Age of Substitutability. Science, vol. 191, 20 February 1976, pp. 683-689.

Good, Irving John: Speculations Concerning the First Ultraintelligent Machine. Advances in Computers, vol. 6, 1965, pp. 31-88.

Hamilton, W. D.: The Genetical Theory of Social Behavior (I and II). J. Theoret. Biol., vol. 7, 1964, pp. 1-16, 17-52.

Hay, Louise: Self-Reproducing Programs. Creative Computing, vol. 6, July 1980, pp. 134-136.

Heer, Ewald: Prospects for Robots in Space. Robotics Age, vol. 1, Winter 1979, pp. 20-28.

Heiserman, David L.: Build Your Own Working Robot. Tab Books, Blue Ridge Summit, PA, 1976.

Hogan, James P.: The Two Faces of Tomorrow. Ballantine Books, N.Y., 1979.

IAF Conference: Robots To Produce Robots At Fujitsu. From an English-language newspaper handout at the Conference, Tokyo, 21-28 September 1980.

Ingel, L. Kh.: Gravitational Focussing. Soviet Astronomy, vol. 17, 1974, p. 836.

Industrial Robots International, vol. 1, no. 5, August 1980, p. 2.

Jacobson, Homer: On Models of Reproduction. American Scientist, vol. 46, 1958, pp. 255-284.

Johnsen, Edwin G.: Man, Teleoperators, and Robots: An Optimum Team for Space Exploration. J. Spacecraft and Rockets, vol. 9, July 1972, pp. 554-556.

Johnson, Richard D.; and Holbrow, Charles, eds.: Space Settlements: A Design Study. NASA SP-413, 1977. 185 pp.

Kemeny, John G.: Man Viewed as a Machine. Scientific American, vol. 192, 1955, pp. 58-67.

Laing, Richard A.: Some Alternative Reproductive Strategies in Artificial Molecular Machines. J. Theoret. Biol., vol. 54, 1975, pp. 63-84.

Laing, Richard A.: Automation Introspection. J. Comp. System Sci., vol. 13, 1976, pp. 172-183.

Laing, Richard A.: Automaton Models of Reproduction by Self-Inspection. J. Theoret. Biol., vol. 66, 1977, pp. 437-456.

Laing, Richard A.: Automaton Self-Reference. Dissertation, State Univ. of New York, Binghampton, Univ. Microfilms, Ann Arbor, MI, Publication No. TSZ 78-05, 542, 1978.

Laing, Richard A.: Machines as Organisms: An Exploration of the Relevance of Recent Results. BioSystems, vol. 11, 1979, pp. 201-215.

Lee, C. Y.: A Turing Machine Which Prints Its Own Code Script. Mathematical Theory of Automata. Polytechnic Press, Booklyn, N.Y., 1963, pp. 155-164.

Martin, A. R., ed.: Project Daedalus – The Final Report on the BIS Starship Study, JBIS Supplement. Unwin Brothers Ltd., London, 1978.

Matloff, Gregory L.: Utilization of O'Neill's Model I Lagrange Point Colony as an Interstellar Ark. J. of the British Interplanetary Soc., vol. 29, December 1976, pp. 775-785.

McCulloch, W. S.; and Pitts, W.: A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, vol. 5, 1943, pp. 115-133.

Miller, James G.: Living Systems. McGraw-Hill Book Company, N.Y., 1978.

Miller, Rene H.; and Smith, David B. S.: Extraterrestrial Processing and Manufacturing of Large Space Systems, NASA CR-161293, vols. 1-3, September 1979.

Minsky, M.: Toward a Remotely-Manned Energy and Production Economy. MIT Artificial Intelligence Laboratory, A.I. Memo No. 544, Boston, MA, September 1979. 19 pp.

Minsky, M.: Telepresence. Omni, vol. 2, June 1980, pp. 44-52.

Moore, Edward F.: Artificial Living Plants. Scientific American, vol. 195, October 1956, pp. 118-126.

Morowitz, Harold J.: A Model of Reproduction. American Scientist, vol. 47, 1959, pp. 261-263.

Myhill, J.: The Abstract Theory of Self-Reproduction. In Essays on Cellular Automata, A. W. Burks, ed., Univ. of Illinois Press, Urbana, Illinois, 1970, pp. 206-218.

Numeric Control Equipment, Aviation Week and Space Technology, vol. 113, 29 December 1980, p. 249.

Oliver, B. N.: Proximity of Galactic Civilizations. Icarus, vol. 25, June 1975, pp. 360-367.

O'Neill, Gerard K.: The Colonization of Space. Physics Today, vol. 27, September 1974, pp. 32-40.

O'Neill, Gerard K.: The High Frontier: Human Colonies in Space. William Morrow and Co., New York, 1976.

O'Neill, Gerard K.; Driggers, G.; and O'Leary, B.: New Routes to Manufacturing in Space. Astronautics and Aeronautics, vol. 18, October 1980, pp. 46-51.

Parkinson, R. C.: Project Daedalus: Propellant Acquisition Techniques. In Project Daedalus – The Final Report on the BIS Starship Study, JBIS Supplement, A. R. Martin, ed., Unwin Brothers Ltd., London, 1978, pp. S83-S89.

Penrose, L. S.: Self-Reproducing Machines. Scientific American, vol. 200, June 1959, pp. 105-114.

Phinney, W. C.; Criswell, D. R.; Drexler, E.; and Garmirian, J.: Lunar Resources and Their Utilization. In Space-Based Manufacturing from Extraterrestrial Materials, Gerard K. O'Neill, ed., AIAA Progress in Astronautics and Aeronautics Series, vol. 57, AIAA, New York, 1977, pp. 97-123.

Preston, K., Jr.; Duff, M. J. B.; Levialdi, S.; Norgren, Philip E.; and Toriwaki, Jun-Ichiro: Basics of Cellular Logic with Some Applications in Medical Image Processing. Proc. of the IEEE, vol. 67, May 1979, pp. 826-856.

Rao, D. Bhogeswara; Choudary, U. V.; Erstfeld, T. E.; Williams, R. J.; and Chang, Y. A.: Extraction Processes for the Production of Aluminum, Titanium, Iron, Magnesium, and Oxygen from Nonterrestrial Sources. In Space Resources and Space Settlements, J. Billingham, W. P. Gilbreath, B. O'Leary, and B. Gossett, eds., NASA SP-428, 1979, pp. 257-274.

Sagan, Carl, Chmn.: Machine Intelligence and Robotics: Report of the NASA Study Group, NASA TM-82329, March 1980.

Sagan, Carl; Drake, F. D.; Druyan, A.; Ferris, T.; Lomberg, J.; and Sagan, L. S.: Murmurs of Earth: The Voyager Interstellar Record. Ballantine Books, N.Y., 1978.

Sandell, N. R., Jr.; Varaiya, P.; Athans, M.; and Safonov, M. G.: Survey of Decentralized Control Methods for Large Scale Systems. IEEE Transactions on Automatic Control, vol. AC-23, April 1978, pp. 108-128.

Smith A. R., III: Cellular Automata and Formal Languages. IEEE Conference Record of 11th Annual Symposium on Switching and Automata Theory, Santa Monica, 28-30 Oct. 1970. N.Y., IEEE, 1970, pp. 216-224.

Smith, J. Maynard: The Evolution of Sex. Cambridge Univ. Press, Cambridge, 1978.

Stakem, Patrick H.: Life Versus Computer Capacity. Byte, vol. 4, February 1979, p. 58.

Taneja, Vidya S.; and Walsh, Peter J.: Growth Considerations of a Self-Replicating Field Dependent on External Factors. IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-11, Aug. 1981, pp. 558-563.

Thatcher, James: The Construction of a Self-Describing Turing Machine. Mathematical Theory of Automata. Polytechnic Press, Brooklyn, N.Y., 1963, pp. 165-171.

Thatcher, James: Universality in the von Neumann Cellular Model. In Essays on Cellular Automata, A. W. Burks, ed., Univ. of Illinois Press, Urbana, Illinois, 1970, pp. 132-186.

Trivers, R. L.: The Evolution of Reciprocal Altruism. Quart. Rev. of Biology, vol. 46, 1971, pp. 35-57.

Turing, Alan: On Computable Numbers with an Application to the Entscheidungsproblem. Proc. London Math. Soc., vol. 42, 1936, pp. 230-265.

U.S. Bureau of the Census: Statistical Abstract of the United States, 98th Annual Ed., Washington, D.C., 1977.

U.S. Bureau of the Census: Statistical Abstract of the United States, 99th Annual Ed., Washington, D.C., 1978.

U.S. Bureau of the Mines: Metals, Minerals, and Fuels. In Minerals Yearbook 1976, vol. 1, GPO, Washington, D.C., 1978.

Valdes, Francisco; and Freitas, Robert A., Jr.: Comparison of Reproducing and Nonreproducing Starprobe Strategies for Galactic Exploration. J. British Interplanetary Soc., vol. 33, November 1980, pp. 402-408.

Vondrak, Richard R.: Creation of an Artificial Atmosphere on the Moon. In Advances in Engineering Science, vol. 3, NASA CP-2001, 13th Annual Meeting of the Society of Engineering Science, sponsored by JIAFS, Hampton, Virginia, 1-3 November 1976, pp. 1215-1223.

Vondrak, Richard R.: Environmental Impact of Space Manufacturing. In Proceedings of the Third Princeton/AIAA Conference on Space Manufacturing Facilities, Princeton, N.J., 9-12 May 1977, Paper no. 77-539.

Von Neumann, John: The General and Logical Theory of Automata. In Cerebral Mechanisms in Behavior – The Hixon Symposium, L. A. Jeffress, ed., John Wiley & Sons, N.Y., 1951, pp. 1-31. Also in Collected Works, vol. 5, pp. 288-328.

Von Neumann, John: Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. In Automata Studies, C. E. Shannon, J. McCarthy, eds., Princeton Univ. Press, Princeton, N.J., 1956, pp. 43-98. Also in Collected Works, vol. 5, pp. 329-378.

Von Neumann, John: Theory of Self-Reproducing Automata. Edited and completed by A. W. Burks. University of Illinois Press, Urbana, Illinois, 1966.

Von Tiesenhausen, Georg; and Darbro, Wesley A.: Self-Replicating Systems – A Systems Engineering Approach. NASA TM-78304, July 1980.

Waldron, R. D.; Erstfeld, T. E.; and Criswell, D. R.: The Role of Chemical Engineering in Space Manufacturing. Chemical Engineering, vol. 86, 12 February 1979, pp. 80-94.

Wesley, J. P.: Ecophysics: The Application of Physics to Biology. Charles C. Thomas, Publ., Springfield, Illinois, 1974.

Williams, R. J.; McKay, D. S.; Giles, D.; and Bunch, T. E.: Mining and Beneficiation of Lunar Ores. In Space Resources and Space Settlements, J. Billingham, W. Gilbreath, B. O'Leary, and B. Gossett, eds., NASA SP-428, 1979, pp. 275-288.

Zachary, William B.: A Feasibility Study on the Fabrication of Integrated Circuits and Other Electronic Components in Space Utilizing Lunar Materials. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

APPENDIX 5A

FIRST ATTEMPT TO DEFINE A SELF-REPLICATING SYSTEM
(A personal note contributed by W. E. Bradley, June 1980)

At a recent meeting a member of the NASA Advisory Council expressed excitement at the positive conclusions reached by the June 1979 Woods Hole Symposium concerning self-replicating mechanical systems. He said that he could not understand why a subject of such interest and importance to the exploration and utilization of space should be approached so timidly. Earnestly, he added:

> "After all, a lathe can produce a lathe, properly operated; nowadays numerically controlled lathes are available; so why not program one to reproduce itself?"

My reaction was the following:

- A lathe cannot produce another lathe without many added subsystems (e.g., driving motor, tool grinder, tool bit production, etc.).
- Some contemplation of the self-replicating system problems at the practical engineering level has been undertaken by a few individuals in the past few months. This work is incomplete as yet, but is aimed at practical, demonstrable systems with only a few critical parts supplied from outside the system, including energy and raw materials for device fabrication. Energy and raw materials appear here in the role of "nutrient," the supply necessarily increasing as the system grows.
- The self-replicating system is indeed of great interest on fundamental grounds.
- The subject is appropriate to and important to NASA.

The work of the past few months (prior to the present study) relevant to self-replicating systems (SRS) is incomplete but has brought to light some principles and ideas of interest.

5A.1 Preliminary Investigation of the Self-Replicating Machine Shop

The town of Muncy is located in a somewhat remote part of central Pennsylvania. It is remarkable because of a nearly self-sufficient machine manufacturing capability in the Sprout-Waldron Company (now a division of another corporation, and therefore subject to change without notice). This company for many years has manufactured agricultural and food-processing equipment as well as heavy machinery for the paper industry, especially pulp grinders. I became acquainted with them while searching for machines able to produce dense pellets for use as solid fuel from agricultural cellulosic wastes.

In the course of my visit, I was shown an excellent machine shop, a foundry, a woodworking shop, and a factory assembly space in which their machines were put together, painted and tested. They also had complete drafting and design engineering facilities. Of special interest was their toolmaking and repair shop, with which all of the milling machines, lathes, jig borers, punch presses, and so forth were kept in fine working order.

This complex, with the possible exception of the foundry, seemed to be a system which, with human assistance, could duplicate itself. In retrospect, it seems worthwhile to explore the possibility that the human operators might be replaced by general purpose automata, manufactured almost completely by the complex itself. The result would then be a major component of a self-replicating system. To complete the system would require manufacture of a prime power source which could be expanded as the complex grows, manufacture of a shelter system (sheds with roofs, walls, windows, and doors) similarly expandable, and possibly a casting and/or forging subsystem, and electronic and computer components of the automata. The foundry with its requirement for refractory furnace linings and high temperatures is a special problem and in some versions of the system may be bypassed.

Present machine shops. Each machine in a machine shop has a functional domain or "scope," assuming unlimited operator attention and guidance. Thus, a lathe (with no attachments) is able to produce objects with cylindrical symmetry having axial length and maximum diameter determined by the "bed length" and the "swing" of the machine. It can also make threads (helical structures), and, to a limited extent, can also make straight-line cuts or grooves which are more properly the work of a milling machine. Lathes can drill holes most readily on the axis of a workpiece of cylindrical symmetry and can achieve a high degree of accuracy of concentricity for this one type of drilling. Most drilling, however, is best accomplished on a jig borer.

The second major machine type in a shop is some form of drill press, or, better, a jig borer. The workpiece is held

firmly in an accurately translatable and rotatable fixture, remaining stationary while holes are drilled by a drill or boring tool held in a chuck rotating about the principal axis of the machine. Such a device can produce clusters of accurately located holes with parallel axes.

The third important shop component is the milling machine. The workpiece is clamped firmly to an accurately controlled table. The workpiece moves continuously, slowly, during operations while the rotating milling cutter shaves or saws the surface being worked. The milling machine is usually used to make rectilinear cuts to form accurately related plane surfaces or grooves.

A well-equipped machine shop usually also includes a power hacksaw, a powerful press with forming dies for forming sheet metal and for punching holes with "punch and die" sets, a bending brake, tool grinders, and possibly a surface grinder to be used like a milling machine to produce flat surfaces.

Self-replicating shop and universal machines. Each machine or subsystem of such a shop can be separated into parts from which it can be reassembled. Each machine therefore has a "parts list," and each part either can or cannot be fabricated by the set of machines and subsystems comprising the shop. The criterion for replication thus may be stated as follows:

> If all parts of all machines and subsystems can be fabricated within the shop, then if properly operated the entire shop can be replicated.

"Proper operation" in this context includes supplying raw materials, energy, and manipulatory instructions or actions necessary to carry out the large number of machine operations, parts storage, and parts assembly required. Human labor is now used for these functions.

It is not necessary that the shop be able to produce anything except a replica of itself which is in turn capable of producing another. Therefore, some simplifications appear possible, such as standardization and limitation of scope where feasible. For example, a universal machine can be imagined with a wider cross feed table than a conventional lathe and with a standardized vise and tool holder so that it can be used for milling. All three dimensions of translation and one axis of rotation could be provided on the table. The head stock could be arranged to hold workpieces, milling cutters or drills. Hardened tools for the necessary cutting operations could be fabricated by the machine from carbon steel in the annealed condition, then tempered, drawn, and sharpened by a separate simpler machine including a small furnace and a tool grinding wheel equipped with tool-holder and feeds. By careful standardization of parts, tools, and fixtures, it is conceivable that such a "one-machine shop" could succeed in reproducing itself.

"Factons." After a shop had been tested with human operators and proven capable of self-replication, it would be possible to explore the replacement of the human operators by mobile computer-controlled manipulators, or "factons." Hopefully, all of the "numerical control" features could be contained in these general-purpose programmable devices which could handle the machines like a human operator. The factons would transfer work from operation to operation, adjust the machine, perform each operation, then transfer the work to a parts storage array. Finally, the parts would be assembled by the factons and the entire shop set up in a selected location and floor-plan. The facton itself has a parts list, most designed to be manufacturable by the shop. Here it is practically inevitable that computer chips plus enormous memories will be needed which would fall outside the scope of the shop thus far envisioned. In other words most, but not all, of facton components could be fabricated by them in the shop. Still, given these extra components provided from outside, the factons could probably fully assemble themselves. The shop itself would require some exogenous elements, as noted above. Prime power, shaft power transmission such as belting or electric motors, abrasives, furnace heating arrangements for tool heat treatment, raw material such as basic feedstock including steel rods, strips, and plates are among the most obvious.

Using the same facton design, it should be possible to implement extensions of the shop, including an optical shop, a pneumatic and/or hydraulic equipment manufacturing shop, and ultimately even an integrated circuit shop. Note, however, that only the original shop with its factons and their programs would have to possess the capability for self-replication.

Computer components, probably provided from outside the system, might be furnished in an unprogrammed condition. Thus, factons would program the tapes, discs, or read-only memories by replication (and verification) of their existing programs. This procedure allows for the possibilities of "heritable" changes of program embodying "devolution" (simplification) or "evolution" (capability augmentation) by orderly program amendment.

5A.2 Program Extension Beyond Self-Replication

The "scope" of a self-replicating shop is much larger than is required for self-replication. Apparently the ability to replicate utilizes only a vanishingly small fraction of total capabilities (to produce various sizes and shapes of parts and to assemble them into machines and structures). The essential characteristic for self-replication is that the scope must be adequate to produce every part of every machine in the shop by means of a feasible program. This "closure condition" can be satisfied using only a small part of the shop's full capabilities.

A generic self-replicating shop can therefore, by means of a simple addition to its program, manufacture other machines and structures and, by means of them, interact

with its environment. For example, it can construct and operate foraging systems to procure fuel or materials, waste disposal systems, or transporters to carry replica shops to other locations.

Obviously, self-replication of such an extended system requires replication of the program-memory. This memory can be partitioned into two parts: (1) The self-replication process memory, and (2) the external process (manufacturing) memory. The distinction between these two memories is that the first is required to reproduce the basic unit (shop machines plus factons) while the second memory contains the program to produce process equipment not essential to the self-replicating nucleus.

At this point it is clear that the effect of a self-replicating system on its environment may take many forms dependent on the external process program. Using such a program, the scope of the system can be extended by construction of machines and structures capable of producing complex subsystems including mineral processing plants, solar energy power supplies, etc.

All of these extended self-replicating systems would embody the same basic nucleus of machines, factons and self-replication programming. They would differ only by addition of the external process program segment peculiar to each type.

Reliability and redundancy. Reliability is a primary concern, especially in the case of self-replicating processes. Two ideas are most important here.

First, the self-replicating program accuracy can be verified by comparison with other replicas of the same program. If a discrepancy is found between two self-replicating programs, a third or fourth replica can be consulted and the error pinpointed and corrected. The test of correctness is the ability to self-replicate.

Second, machines tend to wear, and ultimately to fail, from excessive use. On the other hand, if the system can replicate itself it can make spare parts and install them itself. A special program segment, the "maintenance program," should be devised to check machine wear and perform repairs as needed. This segment would be part of the self-replication program, although another somewhat similar maintenance program should probably be used to care for machines and structures of the external process. This external maintenance program would be specialized for each extended system and is properly part of the second memory.

Speciation. Any self-replicating system is actually another species of SRS, the species being dependent upon the contents of the second memory.

A group of interacting extended self-replicating systems may form a still larger self-reproducing system with yet more complex capabilities. It is not immediately apparent what factors limit the possibilities of such systems. Separable subsystems manufactured by a self-replicating shop may be machines of considerable complexity, themselves incapable of self-replication. Their supply is therefore dependent on the self-replication shop and its program.

It is interesting to note that a facton equipped with an aberrant program may function like a virus, visiting a self-replicating shop and using its machines for reproduction of its own type without constructing any other machine. It could then replicate its program for installation in the new "virus facton" and reproduce this way, using materials and energy from a host self-replication shop. This possibility opens up a large field of problems related to the security of self-replication systems from facton defect or infection.

APPENDIX 5B

LMF POSITIONAL TRANSPONDER SYSTEM

According to the baseline mission for a growing, self-replicating Lunar Manufacturing Facility (LMF) presented in section 5.3.4, a 100-ton seed is dropped to the lunar surface and thereupon unpacks itself, sets up the initial factory complex, and then proceeds to produce more of itself (or any other desired output). Clearly, the level of automation and machine intelligence required lies beyond current state-of-the-art, though not beyond the projected state-of-the-art two or three decades hence. Because of the already challenging design problem, it is highly desirable to keep all seed systems as simple as possible in both structure and function. This should help reduce the risk of partial or total system failure and make closure less difficult to achieve at all levels.

One of the more complicated pieces of hardware from the AI standpoint is the "camera eyes" and pattern recognition routines (visual sensing) that may be needed. Although it is possible that standardized robot camera eyes may be developed, it is more likely that each particular application will demand its own unique set of requirements, thus greatly reducing or eliminating any gains in simplicity of camera design. The pragmatic industrial approach (Kincaid et al., 1980) and design philosophy in these cases, especially in the area of computer vision, is to: (1) simplify, (2) use unconventional solutions, and (3) "cheat" (i.e., solve another problem). It may be that the best way to handle the problem of computer vision is to find a way to largely avoid it altogether.

When the seed unpacks itself it opens into a rather wild environment full of hills, bumps, ledges, crevasses, boulders, craters, and rocks. Surface navigation by mobile robots will be a serious challenge to AI technology. How will a machine know where it is, what the terrain ahead may be like, or how to get home? Laser tracking is one possibility, but probably too complicated when out of line of sight. Pattern recognition of geological and geographical landmarks is another possibility, but there are at least three serious deficiencies associated with this solution. First, the pattern recognition routines must be extremely sophisticated and the sensor very high in resolution and in the ranges of illumination that may be accommodated. Second, to recall how to get home after a lengthy perambulation across the lunar surface may require vast amounts of onboard computer memory. Every turn, every detour, every move the robot makes must be recorded, analyzed for spatial displacement geometry, and the present-position pointer augmented against the stored features maps and correlated with the geographic images received through the vision sensors to plot the shortest route home to avoid the inefficiency of retracing the original physical path. Third, since exploration, development, and construction operations are always in progress around the site, each robot would need a memory capacity sufficient to recall in detail all changes in the landscape between the last series of explorations and the present one – the view is always changing. It may not be practical to design this much AI into each mobile robot, nor to require the central computer to exercise full teleoperator control of a large fleet of nonautonomous mobile robots.

5B.1 The Transponder Network

One way to achieve accurate positioning of all mobile robots while retaining their navigational autonomy is to employ a transponder system operating in the gigahertz frequency range. Much like the LORAN and NAVSTAR systems on Earth, these radar beacons would permit the accurate determination of position by simple triangulation for mobile robot devices located anywhere in the vicinity of the seed. A frequency of perhaps 30 GHz, easily within the range of current technology, would be required for 1-cm positioning accuracy. The transponder system could be orbital-based, but for the present design a ground-based system has been assumed with at most a single satellite for purposes of initial calibration.

When the seed unpacks, its first task is to unfurl the "home base" transponder. Power consumption has not been examined in detail but should not exceed 100 W, the amount supplied by a 1 m^2 solar panel. The next step is to establish an accurate navigational baseline between the home transponder and a reference transponder some distance away, perhaps using a relatively simple nonlaser surveyor's transit. A second baseline is similarly established in some other direction, and the whole system then calibrated and synchronized to coherence. Thus deployed, a local radio navigation grid exists which can fix the position of any appropriately equipped receiver to within 1-cm accuracy, horizontally or vertically, anywhere near the seed.

Since the transponder operates on line-of-sight, each transmitter must be placed a certain distance above the ground in order to "see" the entire area for which it is responsible. The general horizon distance formula is

$X = (h^2 + 2hR)^{1/2}$, where X is the distance to the horizon, R is lunar radius, and h is height of the observer/transmitter above ground. Horizon distances for the Moon are given in table 5.8, neglecting surface irregularities.

TABLE 5.8.– HORIZON DISTANCES FOR THE MOON

Observer height h, m	Horizon distance X, m
1.0	1.9
2.0	2.6
3.0	3.2
4.0	3.7
5.0	4.2
10.0	5.9
15.0	7.2
20.0	8.3

As the original facility grows the transponder network also must be expanded. At the very minimum, a mobile robot should remain in communication with at least three noncollinear beacons to accurately fix its location. (The problems of feature shadowing and unit downtime may require the use of four or five stations. The exact number and layout can only be determined after the specific landing site has been selected and mapped from orbit. One possible deployment geometry is a grid of equilateral triangles with sides roughly equal to the desired horizon distance, with transmitters at the vertices. For example, the triangle pattern edges should be roughly 2.6 km if 2-m high antennas are used. This ensures that the range circle of any mobile robot receiver always will encompass at least three transponder units, thus permitting high-accuracy triangulation. (See fig. 5.33.) Depending on the maximum size of the mature LMF and the maximum feasible height for transponder antennae, the number of transmitters necessary to support the growing seed may range from the tens up into the thousands.

In any case, the main seed computer may be presumed to carry lunar topographical maps of the landing locale, assembled prior to landing and accurate to 1-m resolution, in hard memory. This knowledge, plus the accurate positional information provided by the transponder network, should help to eliminate surprises at the expanding LMF site and lessen the need for a highly sophisticated "intelligent" vision-based surface navigation capability.

5B.2 References

Kincaid, William et al.: Summer Study Background Briefing on Computer Vision, Fault-Tolerant Systems, Large Space Structures and Antennas. Lockheed Missiles & Space Company, 7 July 1980.

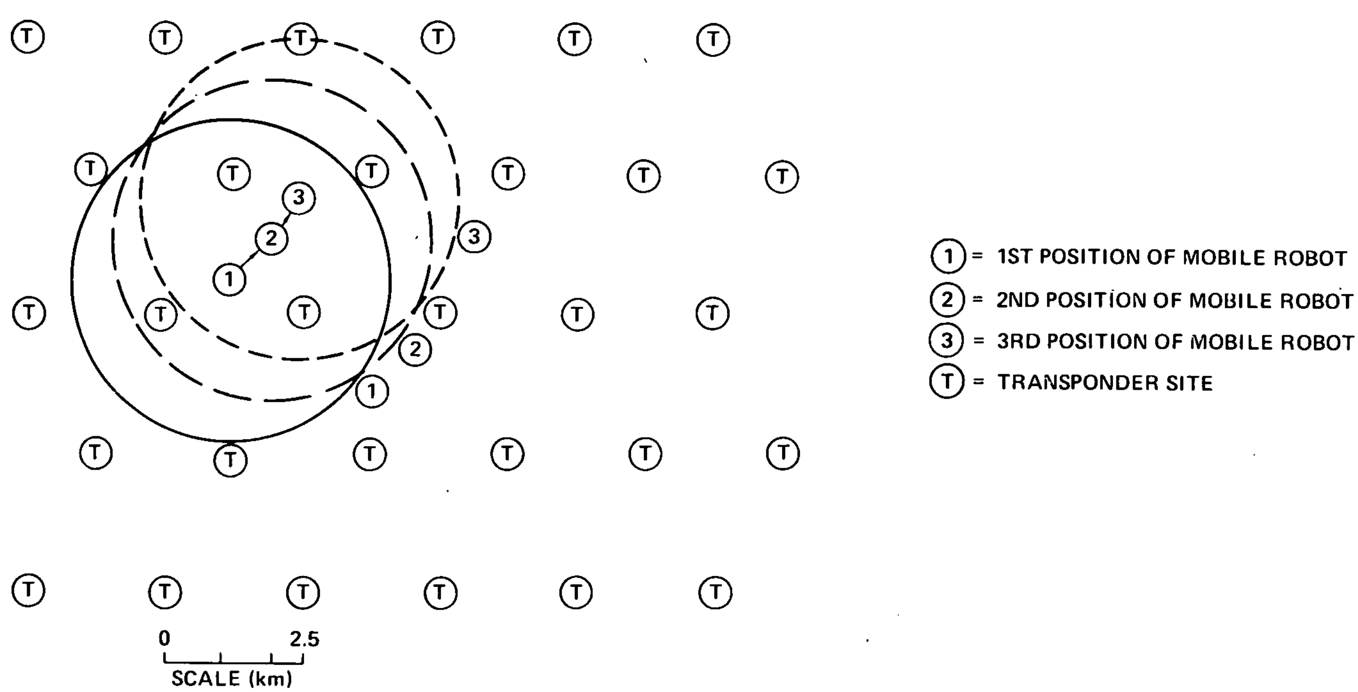


Figure 5.33.– Range circles for mobile robots using LMF transponder network for navigation.

APPENDIX 5C

LMF PAVING ROBOT SUBSYSTEM

The Platform of the Lunar Manufacturing Facility (LMF) described in section 5.3.4 serves as the physical foundation for both the original deployed Seed and the growing and mature LMF manufacturing complexes. According to Nichols (1976), "pavement is a surfacing for traveled areas, which is intended to provide a long-lasting, smooth, clean, supporting surface; to spread loads sufficiently so that base material can support them; and to protect the base against damage by traffic...." These factors are almost as important on the Moon as in terrestrial applications – a simple graded surface would require frequent maintenance, lack cleanliness, and provide no firm foundation base to anchor SRS factory machines. A small crew of platform-building or paving robots is probably necessary for any fully automated lunar factory.

5C.1 Basic LMF Platform Design

The best material for construction of the platform ideally should be plentiful, easy to work, and most suitable for the job in terms of structural strength. Native lunar basalt appears to satisfy all three requirements adequately (Rowley and Neudecker, 1980).

Green (1980a, unpublished Summer Study document) has discussed the properties of lunar basalt at length. Raw lunar soil may be fused at about 1550 K, then allowed to cool and solidify into a very hard, exceptionally strong material. If cooling is virtually immediate – minutes or tens of minutes – the liquid basalt is quickly quenched and becomes a polymeric glassy substance. The material is very strong but also moderately brittle, permitting cracks to propagate rather easily. Using this option, it is necessary to divide the platform into small square-meter-size slabs to help isolate fracture failures and to permit relatively easy maintenance and repair. If the liquid basalt is permitted to cool more slowly – allowing perhaps several hours for the melt to pass from full liquidity at 1570 K to hard solid below about 1370 K – the material anneals into a crystalline form. This method of platform construction takes much longer and requires more energy, but would produce a far less brittle foundation. Such a basalt crystal platform could be prepared as one continuous surface, whereas the glassy basalt platform must be made in slab-sized sections.

Green has also pointed out that Moon soil has characteristics necessary to make an excellent basalt casting due to the uncontaminated, unweathered nature of the lunar material and an extraordinarily low viscosity which is necessary for superior basalt castings. Dunning (1980, unpublished Summer Study document) considered the mechanical properties of cast basalt and found them comparable to those of cast iron and many fine steels, and superior to aluminum, brass, bronze, and copper both in compression and shear strengths. Compression strength is important in many construction applications, and shear strength is a necessary requirement for all foundation materials (U.S. Department of the Interior, 1952). A list of the properties of cast basalt is collected and modified from Anderson (1977), Baumeister and Marks (1967), and several other sources in table 5.9.

Having chosen the foundation material, the team next considered the physical configuration. According to Nichols, concrete pavements for highways are generally about 15-25-cm thick, 30 cm and higher for airport runways. Adjusting for the 0.17-g lunar gravity and the attendant reduced forces to be sustained, the equivalent load bearing strength on the Moon would require a thickness of perhaps 2.6-4.3 cm for highways. Both highways and airport runways encounter heavier use than the LMF platform is expected to receive in normal use, so a choice near the lower end of this range appears justified especially since basalt appears to be stronger than concrete in compression and shear (Baumeister and Marks, 1967; Zwikker, 1954). Consequently, a thickness of 3 cm (Green, 1980b, private communication) was tentatively selected. The square-meter size of individual slabs represents a compromise between limiting possible structural damage caused by fracture propagation and the minimum reasonable size from a practical construction standpoint.

Individual slabs comprising the platform should be formed with a 5-cm margin around the edge (slab separation 0.1 m). Rather than a second sintering pass by the paving robots, slabs are placed close enough so that overheating beyond the nominal square-meter target area for a brief period during each production cycle is sufficient to sinter neighboring blocks. (Some backfilling may be required as about 1-cm horizontal shrinkage is anticipated upon cooling.) A simple diagram of the slab pattern is shown in figure 5.34. Calculations suggest that the baseline design for paving robots should permit each device to prepare about six slabs per day in continuous operation.

TABLE 5.9.– PROPERTIES OF CAST BASALT

Physical properties	Average numerical value, MKS units
Density of magma @ 1473 K	2600-2700 kg/m^3
Density of solid	2900-2960 kg/m^3
Hygroscopicity	0.1%
Tensile strength	3.5×10^7 N/m^2
Compressive strength	5.4×10^8 N/m^2
Bending strength	4.5×10^7 N/m^2
Modulus of elasticity (Young's modulus)	1.1×10^{11} N/m^2
Moh's hardness	8.5
Grinding hardness	2.2×10^5 m^2/m^3
Specific heat	840 J/kg K
Melting point	1400-1600 K
Heat of fusion	4.2×10^5 J/kg (±30%)
Thermal conductivity	0.8 W/m K
Linear thermal expansion coefficient	
273-373 K	7.7×10^{-6} m/m K
273-473 K	8.6×10^{-6} m/m K
Thermal shock resistance	150 K
Surface resistivity	1.0×10^{10} ohm-m
Internal resistivity	1.0×10^9 ohm-m
Basalt magma viscosity	10^2-10^5 N-sec/m^2
Magma surface tension	0.27-0.35 N/m
Velocity of sound, in melt @ 1500 K	2300 m/sec (compression wave)
Velocity of sound, solid @ 1000 K	5700 m/sec (compression wave)
Resistivity of melt @ 1500 K	1.0×10^{-4} ohm-m
Thermal conductivity,	
melt @ 1500 K	0.4-1.3 W/m K
solid @ STP	1.7-2.5 W/m K
Magnetic susceptibility	0.1-4.0$\times 10^{-8}$ V/kg
Crystal growth rate	0.02-6$\times 10^{-9}$ m/sec
Shear strength	$\sim 10^8$ N/m^2

5C.2 Power Requirements for Paving Robots

To obtain a baseline design for LMF paving robots a rough estimate of the power required to fuse the basalt slabs required (in a reasonable amount of time) must be made. For this crude model, basalt platform slabs were taken as square plates with horizontal dimension x and vertical dimension y, with a sintering margin of width s ($2s$ between slabs). A platform of radius R must be constructed within a time τ, so a total of $\pi R^2/(x + s)^2$ slabs must be produced in τ sec, a rate of $t^{-1} = \pi R^2/\tau(x + s)^2$ slab/sec.

The total input power to each square meter of lunar regolith for slab production is given by:

$$P = P_h + P_m + P_r + P_c$$

where P is total power required, P_h is the power needed to heat the basalt material to its melting point, P_m is the power necessary to melt the slab at the melting point, P_r is the rate at which energy is lost due to radiation from the top surface of the slab, and P_c is the rate of energy loss by conduction into the lunar subsurface (modified from Davies and Simpson, 1979). Radiation losses through the thin slab side walls are ignored.

To a first approximation it is sufficient to simply calculate the total energy which must be supplied and divide this by the length of time spent on each slab, hence:

$$P_h = H_s(T_m - T_L)x^2 y\rho/t$$

$$P_m = H_f x^2 y \rho/t$$

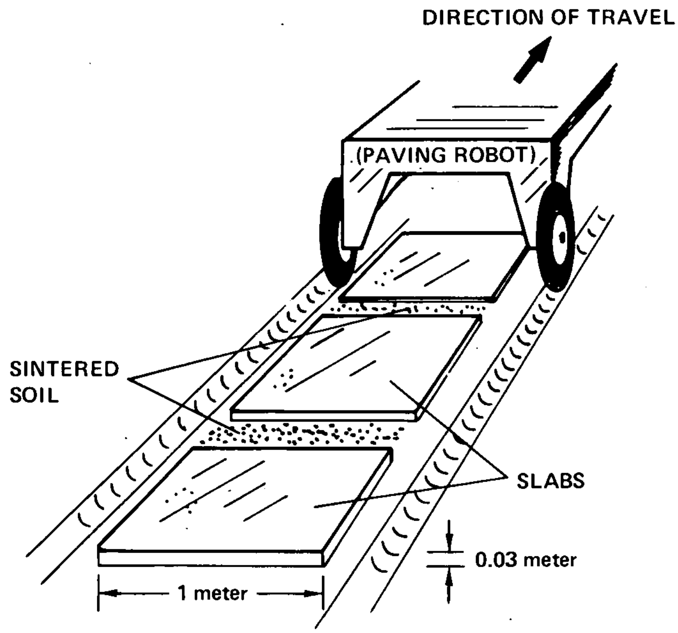


Figure 5.34.– Slab pattern of LMF cast basalt platform.

where H_s and H_f are the specific heat and heat of fusion of lunar regolith, respectively, T_m is the melting point of lunar basalt, T_L is the mean daylight temperature of the lunar surface under direct sunlight at the LMF site, and ρ is the mean density of lunar basalt.

Assuming that heating time is long compared to melting time so that the latter may be neglected, the mean radiative power loss through the exposed face of the slab is given by:

$$P_r \sim \left(\epsilon_L \sigma x^2 \int_0^t T^4 \, dt'\right) / t \, ,$$

where ϵ_L is the emissivity of lunar regolith, σ is the Stephan-Boltzmann constant, and T is temperature at elapsed time t'. If heat is applied such that temperature rises at a linear rate, then:

$$P_r \sim (1/5)\epsilon_L \sigma x^2 (T_m - T_L)^4$$

$$P_c \sim (1/2)(T_m + T_L) C x^2 / \lambda$$

where C is thermal conductivity of lunar soil and λ is the depth at which regolith temperature returns approximately to T_L.

Taking the parameters as listed in table 5.10 as typical, then for a team of five paving robots each capable of processing two slabs at once:

$$t = 10\tau(x + s)^2 / \pi R^2 = 30{,}600 \text{ sec}$$

and

$$P = 20{,}530 \text{ W}$$

TABLE 5.10.– TYPICAL VALUES FOR LMF PAVING ROBOT PARAMETERS

$R = 60$ m	$y = 0.03$ m
$\tau = 1$ yr $= 3.14 \times 10^7$ sec	$T_m = 1573$ K
$x = 1$ m	$T_L = 503$ K
$s = 0.05$ m	$\rho = 2700$ kg/m^3 (for melt)
$H_s = 840$ J/kg	$H_f = 4 \times 10^5$ J/kg
$\epsilon_L = 0.80$ (typical for silica brick and fire brick)	$\sigma = 5.67 \times 10^{-8}$ W/m^2-K^4
	$C = 1$ W/m K
	$\lambda \sim 0.2$ m

5C.3 Paving Robot Design

For the given platform layout there are many possible different modes of operation for paving robots. For instance, each robot might scoop out a hole of the appropriate dimensions, "ingest" the soil and melt it in an internal furnace, then drain the basalt magma back into the hole, neatly filling the depression. Alternative heating techniques may be readily imagined – resistance heating, controlled oxyhydrogen combustion torch with hydrogen recovery, arc furnaces (molten basalt is surprisingly electrically conductive), or induction/dielectric heating using vertical-parallel plates, finger electrodes or "stray field heating" (Cable, 1954; Curtis, 1950; Davies and Simpson, 1979). However, from a pragmatic standpoint, direct solar energy is preferred both for practical convenience and to reduce total external demand on the main LMF power grid.

The solar option for paving robots also has many degrees of design freedom, but for illustrative purposes a comparatively simple model was selected. The basic paving power module consists of a large, spherical polished aluminum mirror, constructed with easily manufactured small planar segments and affixed to a single-axis equatorial-drive turntable with a 90° sweep. This large dish is mounted on the north side of paving robots working in the lunar northern hemisphere. The robots travel east-west to maintain near-constant directional orientation at all times (except when beginning or completing a row of slabs). A planar rectangular mirror is mounted low in front of the dish, leaning forward at about 45° to direct the focus of the solar rays downward onto the carefully graded lunar surface. This second mirror may require three degrees of freedom for tracking and to permit it to project a proper square beam. Assuming accurate dish and plate mirror servo gearing, mirror positions are at all times accurately known. If the position of the robot vehicle is precisely fixed by the transponder network (see app. 5B), and an updated monthly lunar solar ephemeris is provided each robot by the seed central computer when work begins each lunar dawn, then the entire mirror pointing task can be fully automated and sun-tracking sensor apparatus eliminated. The basic optical geometry is shown in figure 5.35.

Main dish size is given by:

$$D = 2(P/\pi k^2 \alpha I \cos \delta)^{1/2}$$

where D is mirror diameter, k is the reflectivity of either of the two polished mirror surfaces (which may range up to 0.86 for aluminized glass, Weast, 1969), α is the coefficient of absorption of solar radiation for lunar basalt (taken as 0.93 for lunar albedo of 7%), I is solar insolation (1400 W/m^2), and δ is the angle between the mirror pointing axis and the Sun. In a worst case of $\delta = 20°$ error, $D = 5.4$ m.

The planar mirror is roughly rectangular, long end pointing downward, of approximate dimensions 2m × 4m. The heat absorbed by this mirror is at most $P(1 - k)/8 + I$ or 1710 W/m^2, corresponding to a blackbody radiation temperature of 417 K which seems manageable. Mirrors should require resurfacing only rarely, since oxidation and meteorite pitting are not expected to be major problems.

The tentative design for the LMF paving robot is shown in figure 5.36. Each machine has a pair of dish and rectangular mirrors. Two small navigational receivers are at either end of the flatbed, permitting the onboard computer to calculate its rotational orientation with respect to the transponder network as well as its position, and a two-axis

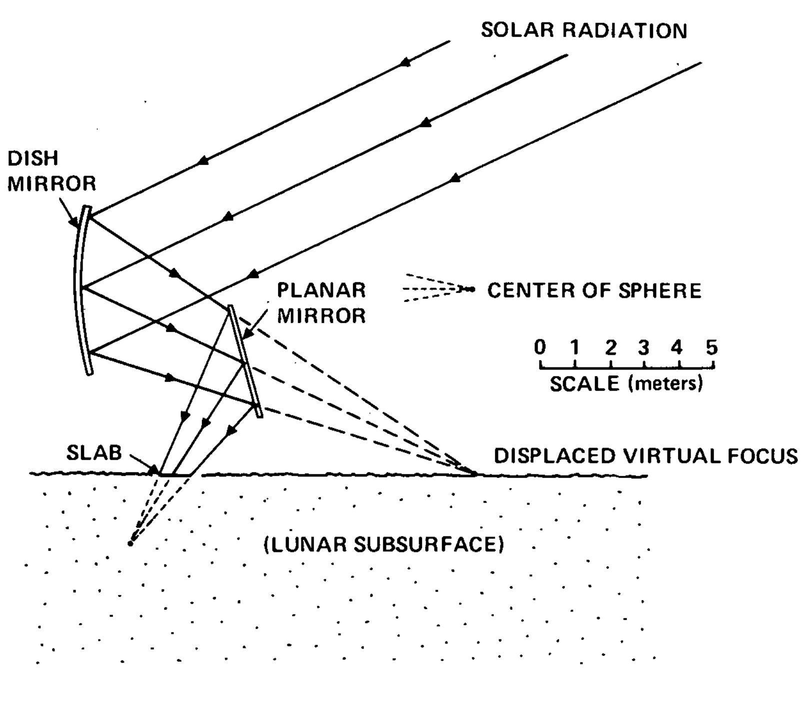


Figure 5.35.– LMF paving robot optical geometry.

level sensor measures tipping angle. Simple retractable IR sensors extend down near the slab working area to monitor energy flux and temperature, and a steerable low-resolution camera with two degrees of freedom (vertical and horizontal rotation) is installed between the two main dish mirrors to check slab placement as construction proceeds. Tires are made of soft woven basalt fibers (see sec. 4.2.2), and the vehicle is driven by four low-power electric motors fore and aft geared to steerable front and rear wheel pairs. Energy requirements for mobility and onboard computing are expected to be modest, so a few square meters of exterior solar cell paneling augmented by a rechargeable fuel cell should suffice.

5C6.4 Mass and Information Estimates

A 5.4 m spherical dish made of aluminum 1-cm thick will have a mass of about 620 kg, or 1240 kg for a pair. Similarly, the total for both planar mirrors is 440 kg. Assuming 2×10^{-2} kg computer/kg serviced (Freitas, 1980), each robot computer is about 50 kg. Camera, sensors, and navigational equipment add another estimated 50 kg. Solar panels and fuel cells may total 100 kg. The aluminum vehicle frame should be able to support its own weight (1700 kg) on Earth, so in low lunar gravity only 280 kg are required to obtain equivalent support. Each tire and drive assembly is about 40 kg, a total of 240 kg for all six wheels.

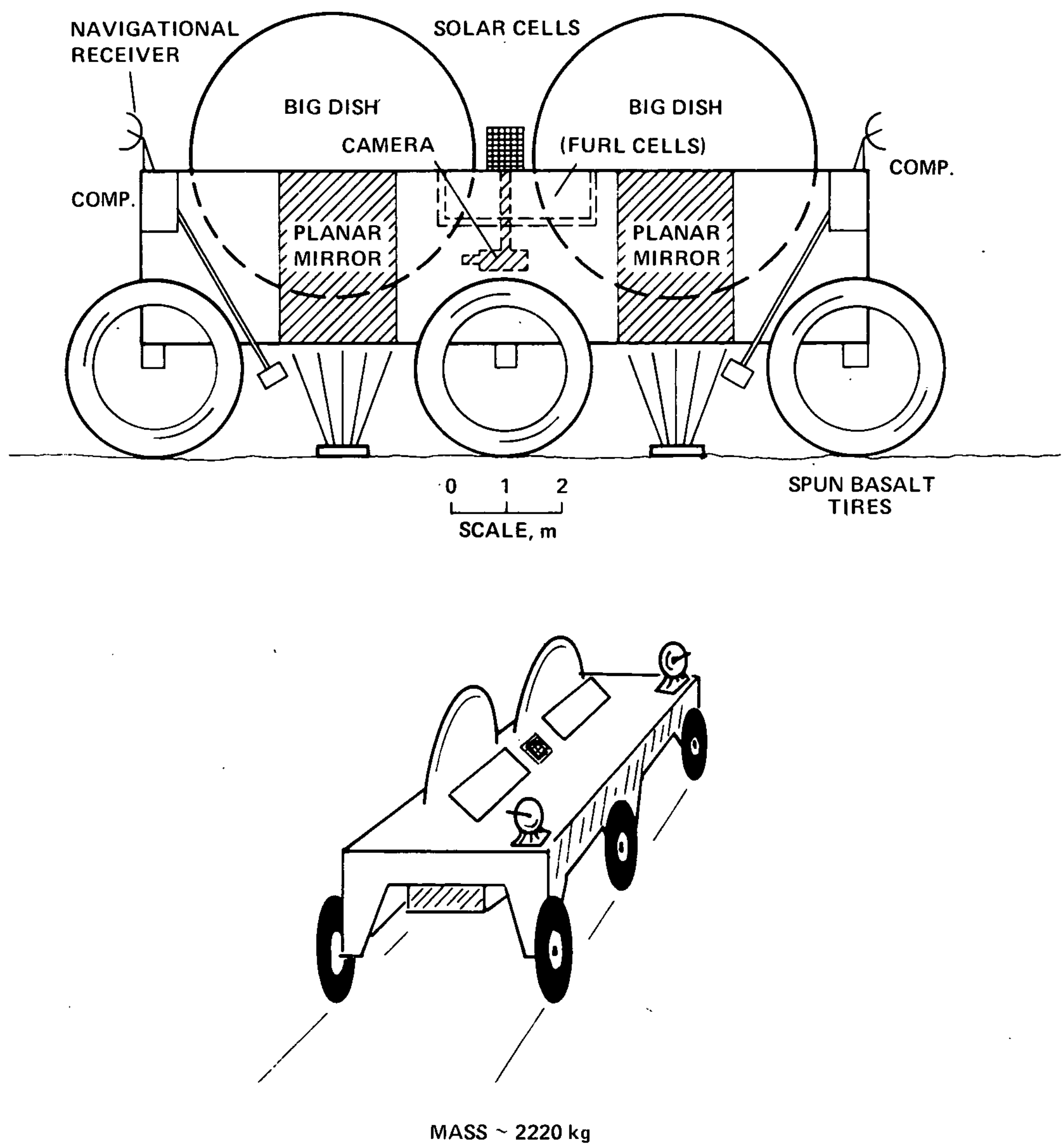


Figure 5.36.– Tentative LMF paving robot design.

Hence, the mass of each paving robot is about 2400 kg. The fleet of five included with the original seed totals 12,000 kg.

Paving robot computers must serve a number of functions, including autopilot, dish mirror guidance and control, planar mirror guidance and control, executive operating program execution, operational "timesheet" memory for the run in progress, traffic pattern coordination with other robots, neighbor machine avoidance, self-diagnostic routines for simple malfunctions, pattern recognition for slab working area imaging, sensor control and data processing, energy system maintenance, lunar solar ephemeris memory and calculation of solar pointing angles, navigation and drive wheel control, and various routines for recognition and verification of task completion. The computation capacity needed to handle these functions probably is in the range 10^6-10^7 bits (about 64 K-512 K bytes). The information necessary to completely describe the machine for purposes of replication is probably about an order of magnitude greater, roughly 10^7-10^8 bits.

5C.5 References

Anderson, Alfred T., Jr.: Basalt. In Encyclopedia of Science and Technology, vol. 2, McGraw-Hill, New York, 1977, pp. 110-110D.

Baumeister, Theodore; and Marks, Lionel S., eds.: Standard Handbook for Mechanical Engineering. 7th Ed. McGraw-Hill, New York, 1967.

Cable, J. Wesley: Induction and Dielectric Heating. Reinhold Publishing Corp., New York, 1954.

Curtis, Frank W.: High-Frequency Induction Heating. 2nd ed., McGraw-Hill Book Company, Inc., New York, 1950.

Davies, John; and Simpson, Peter: Induction Heating Handbook. McGraw-Hill Book Company, Ltd., London, 1979.

Freitas, Robert A., Jr.: A Self-Reproducing Interstellar Probe. J. of the British Interplanetary Soc., vol. 33, July 1980, pp. 251-264.

Nichols, Herbert L., Jr.: Moving the Earth: The Workbook of Excavation. 3rd ed. North Castle Books, Greenwich, Conn., 1976.

Rowley, J. C.; and Neudecker, J. W.: Melted In-Place Lunar Soil for Construction of Primary Lunar Surface Structures. In Extraterrestrial Materials Processing and Construction, David R. Criswell, ed. Final Report, NSR 09-051-001 Mod. No. 24, Lunar and Planetary Institute, Houston, Texas, 31 January 1980, pp. 215-219. (Also published as NASA CR-158870.)

U.S. Department of the Interior: Earth Manual: A Manual on the Use of Earth Materials for Foundation and Construction Purposes. Reprinted by Bureau of Reclamation, with revisions, February 1952.

Weast, Robert C.: Handbook of Chemistry and Physics. 49th ed. CRC Company, Cleveland, Ohio, 1968.

Zwikker, C.: Physical Properties of Solid Materials. Interscience, New York, 1954.

APPENDIX 5D

LMF MINING ROBOTS

5D.1 Mining Robot Functions

The requirements of seed mining robots which are components of the proposed growing, replicating Lunar Manufacturing Facility (LMF) include six basic functions:

- Strip mining – Mining robots must be able to strip mine the lunar surface without the need for overburden removal down to a depth of at least 2-3 m.
- Hauling – Having "mined" a certain amount of unbeneficiated lunar regolith, mining robots must haul their loads back to the central LMF complex for further processing. It is possible that greater efficiency might be achieved by separating the hauling and excavating functions. Also, Carrier (1979) has pointed out that it may be more efficient to beneficiate raw materials at the pit site so that only useful soil components need be transported some distance to the LMF. This would reduce the mass of mining robots required, but would increase the mass of equipment located a distance from the base site.
- Landfilling – On the return leg of each journey from the strip mining pit to the growing seed factory, each robot carries a load of unused slag or waste materials back to the pit where it is packed in as landfill. The pit might perhaps be excavated in a spiral pattern, with the fill site lagging the dig site by a gradually increasing amount.
- Grading – Mining robots must be sufficiently general-purpose to be capable of rough leveling of hilly terrain and then precision centimeter-level grading preparatory to paving robot activities.
- Cellar digging – It is conceivable that the LMF computer at the center of the circular factory complex, and perhaps certain other LMF components as well, will need to be buried under a few meters of lunar topsoil for reasons of temperature control, radiation shielding, and so forth.
- Towing – Miners are the mobile workhorses of the growing LMF complex beyond the confines of the factory platform. When mining robots or other machines break down somewhere outside, miners must go to them and tow them back if they are immobilized. For example, if a robot has become trapped because of pit wall collapse, from a fall in loose lunar soil, or has become jammed into the surface or under fallen rocks, mining robots are ideal rescuers because they are also the LMF excavation machines and are smart and mobile enough to handle such tasks with ease.

Many additional mining-related functions conceivably could be performed by a robot system capable of the most general classes of excavation and mining activites. Indeed, such capacity might be absolutely essential if seed packages are dispatched to other planets than the Moon (e.g., Mars, Titan, Mercury, or Earth). These added functions include drilling, tunneling, blasting, and many others. But the basic six capabilities described above appear both necessary and sufficient for system survival and growth on the lunar surface.

5D.2 Design Alternatives

There exists a bewildering variety of mining and excavation machine technologies from which to draw in conceiving an autonomous vehicle (Nichols, 1976). The final design is a variant of the system devised by Carrier (1979) during a 1978 NASA-sponsored study on extraterrestrial materials processing and construction (Criswell, 1978).

In Carrier's system, strip mining proceeds in an annular sector Ω radians wide as shown in figure 5.37. The total system is designed for gradual expansion, based on Earth supply or lunar colony supply, over a 30-year period. In the first few years of operation, all stripping and hauling to the central processing plant is performed by front-end loaders (also called, variously, the "shovel dozer," "dozer shovel," "tractor loader," "end loader," "front loader," or "loader"). These machines are used on Earth for digging, loading, rough grading, and limited hauling. In the lunar case, according to Carrier, the loader should be used at the outset for long hauling as the easiest way to start ore flowing into the central plant. After a few years the loaders may be augmented by a system of haulers, essentially large-volume ore trucks carrying lunar topsoil back to the central plant. This permits the loaders to strip-mine full time.

While useful as a starting point in the present study, the Carrier system cannot perform all required LMF functions. Figure 5.38 shows the basic design for the LMF mining robot. This machine is a modified loader with a rollback bucket; has a dozer blade formed on the lower face of

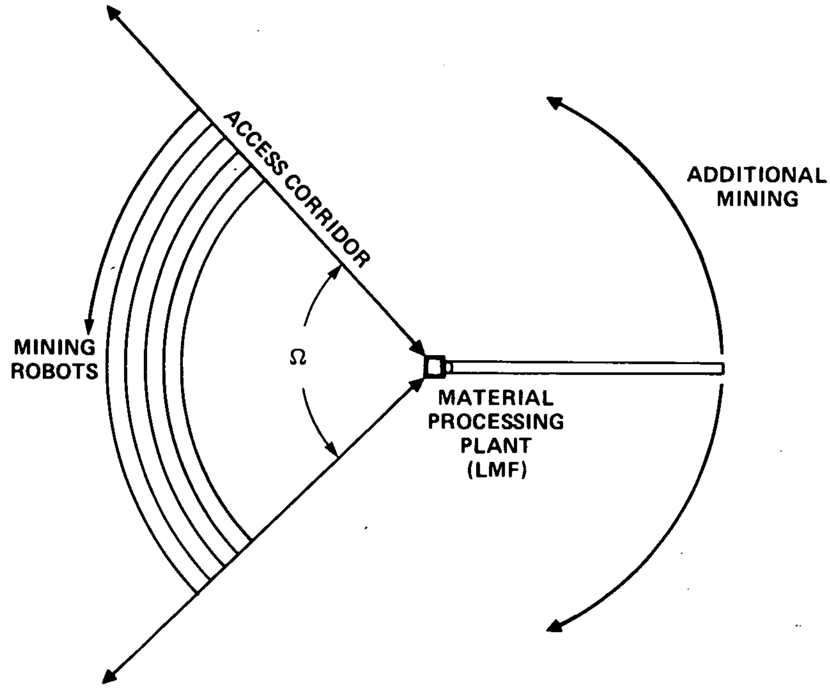


Figure 5.37.– Lunar surface strip mining.

the loader bucket, reinforced so that the bucket can be placed in a locked, elevated position and the robot driven as a dozer; and has three attachments aft which are removed during normal work, including a precision grading blade with surface contour sensors, a simple tow bar, and a somewhat more versatile towing platform.

A loader equipped in this fashion should be able to perform all six basic LMF functions enumerated above. According to Nichols, in a pinch the mining robots should also be able to act as a primitive crane, as a more versatile variable blade pitch bulldozer, as a "reach down" dozer able to cut below the depth accessible to most dozers, and as a backdragger to smooth loose dirt. Finally, it should also be possible for two loaders to join face to face to lift large boulders which neither could conveniently lift alone.

5D.3 Mining Robot Design Specifics

The team considered various specific aspects of LMF mining robot design, including machine mass, power consumption, sensor configuration, and computational and information requirements. The results and conclusions are presented below.

Robot mass and power estimates. According to Carrier (1979), haulers may be much less massive on the Moon than on Earth since the lower gravity enables the same physical structure to carry more payload mass because the force per unit mass is less. In loaders, the vehicle mass is used as a counterbalance to prevent the machine from tipping over when fully loaded, so the mass relations for these machines change little from Earth in the lunar environment. Usual terrestrial practice is to multiply the bucket load mass by a factor of 2.0 to determine a safe tipping mass (the mass of the vehicle used as a counterweight). However, lunar equipment might incorporate automatic sensing systems to prevent tipping over so a safety factor of 1.2 should be sufficient (Carrier, 1979).

If the hauling mass per trip for all mining robots is M_h, m is the rate at which lunar materials must be mined to support the LMF replication schedule, and t is the time required for a robot to complete one cycle of operation (scoop up soil, deliver to LMF, return to pit), then $M_h = mt$. Using a factor of 1.2, the mass of mining robots is approximately $M = 1.2\,M_h = 1.2\,mt$.

Conservatively estimating an average of 40 km travel distance per round trip to the LMF per robot (from a 20 km radius annular pit surrounding the growing seed), an average transport speed of 10 km/hr, and a typical duty cycle of 50% for actual mining work (to leave time for repairs and nonmining labors such as grading, towing, or cellaring), then the mean cycle time

$$t = (40\text{ km})(3600\text{ sec/hr})/(50\%)(10\text{ km/hr}) = 28{,}800\text{ sec}$$

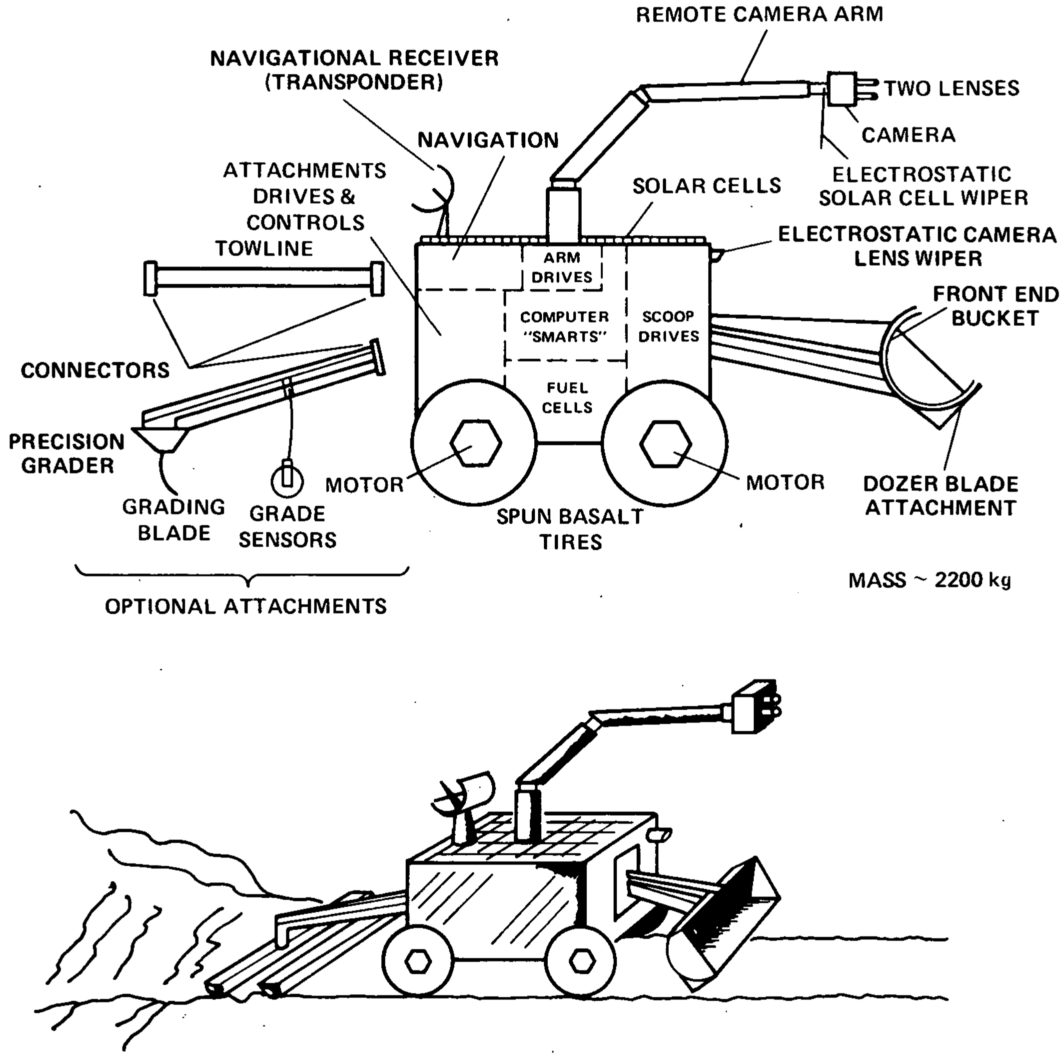


Figure 5.38.– LMF mining robot design.

The annual lunar soil hauling requirement is approximately 4×10^6 kg (see app. 5E) to replicate a new 100-ton seed each working year, so,

$$m = (4 \times 10^6 \text{ kg})/(3.14 \times 10^7 \text{ sec}) = 0.127 \text{ kg/sec}$$

Hence, mining robot mass is

$$M = (1.2)(0.127 \text{ kg/sec})(28{,}800 \text{ sec}) = 4400 \text{ kg}$$

(Approximately 4400 kg/1.2 = 3700 kg of lunar material are transported each cycle.) Note that M is the total mass of robots required, not necessarily the mass per robot. In fact, it is essential that the seed carry at least two such machines so that strip mining can proceed almost continuously given a 50% duty cycle and so that a "spare" is always available in emergency situations. Assuming linear downscaling the mass of each robot is 2200 kg.

In Carrier's strip-mining system the machines require an average of 0.3 W/kg. Mostly this is due to the hauling function, the most energy-intensive operation performed. Hence each mining robot requires about 660 W which may be drawn from 4 m^2 of photovoltaic solar cell panels mounted on every available surface. A fuel cell module (Fickett, 1978) is included in the robot design, for buffer storage and peak load coverage when power consumption may rise as high as 10 kW (as during rescue operations). This module may be recharged at any time from the LMF power grid, but this should not be necessary as the robots should be fully self-sufficient in this regard. Finally, an electrostatic lunar dust wiper is provided to maintain solar cells and camera lenses at maximum efficiency.

Sensor configuration. Sensing equipment on board includes the usual navigational receiver which ties into the high-accuracy transponder network; a two-axis level sensor so the robot knows its tipping angle with respect to the local gravity field; a detachable grading sensor which rolls along the ground just in front of the precision grading blade and provides immediate real-time feedback to permit exact control of grading angle, pitch, and slew.

The most complex sensor system is the remote camera arm. (See discussion of state-of-the-art techniques by Agin, 1979.) The camera is binocular to allow ranging and depth perception, and to provide a spare in case one camera "eye" fails. This is mounted on a long robot arm which can be directed to observe any part of itself or to survey the landscape during roving activity. The camera arm will need at least seven degrees of freedom – rotation of the arm shaft, flexure of the two intermediate joints, bending at the wrist, camera rotation, lens rotation for focus, and telephoto capability for close scrutiny of interesting features in the environment.

The mining robot camera arm is absolutely essential if the vehicle is to function in the versatile manner envisioned for it. It is not enough simply to know position in space, because the environment in which the system must operate is highly complex. It might be possible for the seed computer to give the robot a "road map" to 1 m accuracy, but this would not allow for proper navigation once the miners begin to physically alter their surroundings by digging, hauling, dozing, etc. Also, there may be objects smaller than 1 m that could cause major difficulties such as crevasses and boulders. Hence, it seems necessary to give the mining robots a true generalized "intelligent" roving capability.

Automation and AI requirements. The camera arm will require some high-level AI that lies beyond state-of-the-art. The onboard computer must keep track of the position of the moving arm in order to know where the camera is at all times. There must be routines for avoiding obstacles – for instance, the system should avoid hitting the camera with the loading bucket. Complex pattern recognition routines must be available to permit image focusing, telephoto operation, interpretation of shadows and shapes, differentiation between protrusions and depressions in the surface, and intelligent evaluation of potential risks and hazards of various pathways and courses of action. The onboard computer must have an accurate representation of its own structure stored in memory, so that the camera may quickly be directed to any desired location to inspect for damage, perform troubleshooting functions, or monitor tasks in progress. Finally, the computer must have diagnostic routines for the camera system, in case something simple goes wrong that can easily be corrected *in situ* without calling for outside assistance.

According to Carrier (1979) the automatic haulers can easily be designed to operate in an automatic mode, requiring only occasional reprogramming but substantially more advanced AI pattern recognition systems. (In 1980 a child's toy was marketed which can be programmed to follow simple paths (Ciarcia, 1981; "Toy Robots," 1980).) Carrier suggests that since there are so many variables associated with excavation "it is doubtful that the front-end loader could operate automatically," though the team disputes this conclusion. In addition to sophisticated pattern recognition and vision systems (Williams et al., 1979), the robot miners need a "bulldozer operator" expert system of the kind under development at SRI for other applications (Hart, 1975, and personal communication, 1980). Such an expert system would embody the knowledge and skills of a human excavator and could substitute for human control in most circumstances. In addition, expert systems might be executed remotely by a process called "autonomous teleoperation." In this mode of operation, mining robots can be remote-controlled via transponder network links by the master LMF computer, thus reducing onboard computer complexity.

Additionally, the onboard computer must handle such comparatively mundane chores as clocking, operating drive trains on the wheels, turning controls, blade angle control and configuration, task completion testing and verification, guidance and navigation, and internal diagnostics. An executive program is also required, capable of accepting new orders from the central LMF computer (e.g., "rescue machine X at position Y") and semiautonomously calculating how best to execute them (Sacerdoti, 1980).

Computation and information requirements. A first-cut estimate of the computational capacity required on board reveals that three major computer subsystems are involved: (1) robot camera arm (seven degrees of freedom, binocular vision, rangefinding, sophisticated AI such as pattern recognition and inference); (2) excavator expert system (controls physical operations, understands a world model, has expectations about outcomes, and can troubleshoot simple problems); and (3) high-level executive system (reprogrammability, interpretation, and "common sense" reasoning). Each of these subsystems represents a different problem and must be separately analyzed.

The robot system with mobile camera studied by Agin (1979) engaged in very primitive pattern recognition. This included insertion of bolts into holes, positioning a movable table relative to a fixed camera, velocity tracking (a Unimate PUMA arm, camera in hand, follows an object moving past on a conveyor belt), spot welding on a moving assembly line, and following a curved path in three dimensions at constant velocity (simulating industrial activities such as gluing, sealing, and seam-following). Again's visual recognition routines ran on a PDP-11/40 minicomputer, a 28K application, and the PUMA robot arm was controlled by the usual LSA-11 microcomputer which has a 16K capacity using 32-bit words. The visual system for the proposed

mining robot will be at least 1-2 orders of magnitude more complicated than Agin's system, so we would estimate a control requirement of 10^6-10^7 bytes, or about 10^7-10^8 bits of computer capacity.

The SRI expert system "PROSPECTOR" runs on a DEC-10 computer with a 150K operating program and a 1M database, a total of about 3.2×10^7 bits (Hart, personal communication, 1980). PROSPECTOR "knows" about 1000 different factors related to prospecting. It is difficult to imagine a general excavation expert system requiring more than ten times this, or 10,000 factors, to achieve adequate autonomous operation with troubleshooting capability – the PROSPECTOR expert has generated some impressively accurate results in searches for ore-bearing bodies. If the "EXCAVATOR" expert system is thus about one order of magnitude larger than PROSPECTOR, the basic computational requirement is 10M or 3.2×10^8 bits.

Mining robot executive computer requirements are more difficult to estimate, as there are few previous directly applicable models. A simple passenger aircraft autopilot probably will run on a 32K microprocessor, and a "smart rover" vision-equipped wheeled mobile robot with a 6-degree-of-freedom arm developed in the 1970s at JPL used state-of-the-art microprocessors. Remarks by Sacerdoti (1979, 1980) on the subject of autonomous planning and execution in robotics suggest that the system required for robot miners is perhaps 1 to 2 orders of magnitude beyond current technology; thus the executive system may require a memory capacity of about 1 to 10M, or 3 to 30×10^7 bits.

Summing the requirements for the three major computer subsystems gives an "information bandwidth budget" of 3.6-7.2×10^8 bits, centering on about 500 Mb. The information necessary to completely describe the system for purposes of self-replication is probably on the order of 10^9 bits.

5D.4 LMF Approach and Access Geometry

In the baseline LMF scenario, mining robots must assume all hauling duties beyond the factory platform. Thus, it becomes necessary to specify how these mobile machines, normally bearing loads of strip-mined soil to be processed, will approach the factory and deposit their cargoes at an appropriate input location. A related query is how and where robots will accept waste products for transport to the pit for use as landfill. These questions are of some importance, because as the seed expands to full maturity it may become physically more difficult to exchange raw materials and wastes with interior LMF processing systems unless the access geometry has been designed to accommodate growth.

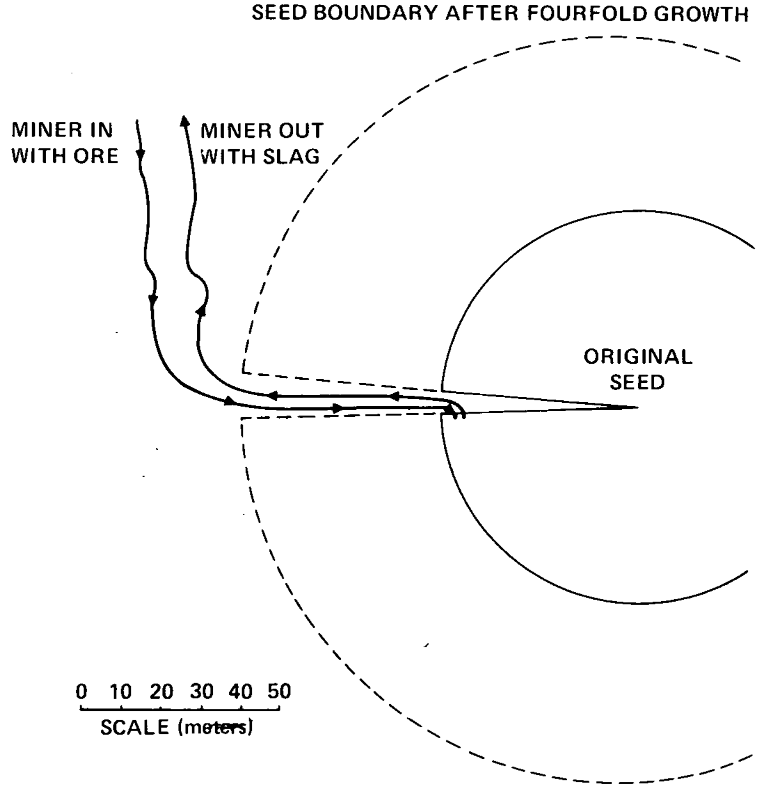
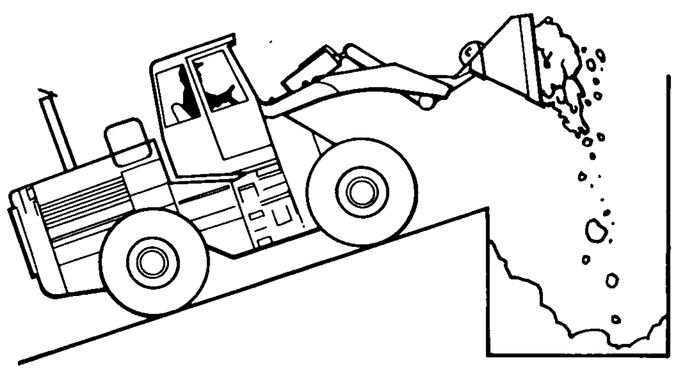


Figure 5.39.– LMF constant-angle wedge corridor access route.

The solution adopted by the team is to earmark a constant-angle wedge corridor for permanent use as a mining robot access road. A 5° angle provides a corridor width of 5 m at the perimeter of the initial 60-m radius seed – comfortably enough room for a mining robot to enter, drop off its cargo, pick up a load of waste materials, and then withdraw. The area of the constant-angle corridor increases as R^2. This is the same dependence on radius exhibited by the area of the growing "seed," hence LMF mass, raw materials requirements, and waste production will increase at the same rate as the access corridor which supports interior factory systems. In other words, the expanding corridor prevents internal LMF systems from becoming "landlocked" as seed mass and radius grow exponentially in time.

The wedge corridor geometry is shown in figure 5.39. Note that as the LMF grows larger, mining robots (or any other external transport vehicle) must traverse ever greater distances, on average, to reach the entry corridor. For this reason a minimum of two such corridors should be provided, with the factory organized as two identical halves as suggested in figure 5.19. Further studies will be required to determine the optimum access and LMF configuration geometry from the standpoint of scheduling, efficiency, and access time.

Mining robots deliver raw materials to an input hopper located in the chemical processing sector, as shown in figure 5.40. Outshipments of waste materials are delivered to them in similar fashion. These hoppers serve as materials depots, able to help sustain LMF operations during periods when the supply of lunar topsoil is interrupted for any reason. Since each of the two initial seed robots makes one round trip about every eight hours, a hopper intended to serve as a one-week buffer must have a capacity of 42 mining robot loads or 76,900 kg of lunar regolith. A roughly cubical hopper constructed of 1 cm sheet aluminum and able to contain the weekly input volume of 42.7 m^3 has a mass of 1650 kg.



COURTESY OF INTERNATIONAL HARVESTER COMPANY

Figure 5.40.– Raw material delivery to input hopper.

5D.5 References

Agin, Gerald J.: Real Time Control of a Robot with a Mobile Camera. SRI International Technical Note 179, February 1979.

Carrier, W. D., III: Excavation Costs for Lunar Materials. Fourth Princeton/AIAA Conference on Space Manufacturing Facilities, Princeton, New Jersey, 14 May 1979, AIAA Paper 79-1376.

Ciarcia, Steve: A Computer-Controlled Tank. Byte, vol. 6, February 1981, pp. 44–66.

Criswell, David R.: Extraterrestrial Materials Processing and Construction. NASA CR-158870, 1978.

Fickett, Arnold P.: Fuel-Cell Power Plants. Scientific American, vol. 239, December 1978, pp. 70-76.

Hart, Peter E.: Progress on a Computer-Based Consultant. SRI International, Menlo Park, January 1975.

Nichols, Herbert L., Jr.: Moving the Earth: The Workbook of Excavation. 3rd ed. North Castle Books, Greenwich, Conn., 1976.

Sacerdoti, Earl D.: Problem Solving Tactics. SRI International Technical Note 189, July 1979.

Sacerdoti, Earl D.: Plan Generation and Execution for Robotics. SRI International Technical Note 209, 15 April 1980.

Toy Robots Gaining Intelligence. Byte, vol. 5, October 1980, p. 186. See also "Inside Big Trak," Robotics Age, vol. 2, Spring 1980, pp. 38-39.

Williams, D. S.; Wilf, J. M.; Cunningham, R. T.; and Eskenazi, R.: Robotic Vision. Astronautics and Aeronautics, vol. 17, May 1979, pp. 36–41.

APPENDIX 5E

LMF CHEMICAL PROCESSING SECTOR

Mining robots deliver raw lunar soil strip-mined from the pit to large input hoppers along the edge of the entry corridors into the chemical processing sector. The primary responsibility of the materials-processing subsystems is to accept lunar regolith, extract from it the necessary elemental and chemical substances required for system growth, replication, and production, and then return any wastes, unused materials, or slag to an output hopper to be transported back to the surrounding annular pit by mining robots for use as landfill.

It is possible to achieve qualitative materials closure (see sec. 5.3.6) – complete material self-sufficiency within the Lunar Manufacturing Facility (LMF) – by making certain that chemical processing machines are able to produce all of the 84 elements commonly used in industry in the United States and the global economy (Freitas, 1980). However, such a complete processing capability implies unacceptably long replication times T (on the order of 100-1000 years), because many of the elements are so rare in the lunar or asteroidal substrate that a vast quantity of raw soil must be processed to obtain even small amounts of them. By eliminating the need for many of these exotic elements in the SRS design, replication times can be cut by as much as three orders of magnitude with current or foreseeable materials processing technologies.

Hence, it is desirable to determine the minimum number of elements and process chemicals and to fix the lowest extraction ratio R (kg input material/kg useful output material, see sec. 5.3.6) which can still maintain closure of the system, thus minimizing the replication time T.

5E.1 Minimum LMF Requirements: Elements and Process Chemicals

The elemental and chemical requirements of the expanding LMF fall into a fairly small number of broad categories summarized in table 5.11. Note that these are the minimum (or very nearly so) requirements for LMF qualitative materials closure – an "adult" LMF entering production phase may need additional chemical processing capabilities which may be programmed into the factory's operational software. Table 5.11, however, lists only those minimum requirements necessary to achieve closure for a seed during the growth phase.

TABLE 5.11.– MINIMUM SEED ELEMENT AND PROCESS CHEMICAL REQUIREMENTS

I. Structural metals, alloys, hard parts, tubing, containers, etc. – Fe, Al, Mg, Ti, Mn, Cr, C, Si, Ca

II. Building materials, insulation, fabrics, glass plate, ceramics, crucibles, furnace linings, chemistry glassware, high-temperature refractories, etc. – lunar soil as found (basalt when fused), anorthite ($CaAl_2Si_2O_8$), silica (SiO_2), alumina (Al_2O_3), magnesia (MgO), feldspar

III. High purity electronics-grade materials for the manufacture of solar cells, computer chips, etc. – Si, O_2, Al, P, B

IV. Magnetic materials – Fe

V. Fluorine chemistry containers – Fe, C, F_2

VI. Process chemicals for bulk manufacturing, high-purity electronics chemical production – H_2O, HF, N_2, H_3PO_4, HNO_3, SiH_4, CF_4 (Freon for microelectronic "dry etching" processes), NaOH, Cl_2, H_2SO_4, $CaCl_2$, Na_2CO_3, NH_3

VII. Process minerals, inputs to chemical processing sectors – olivines, pyroxenes, feldspars, spinels, ilmenite, apatite, anorthite, tincalconite (anhydrous borax).

Total of 18 elements, 12 minerals/mineral types, and 11 additional process chemicals.)

It will be argued that a chemical processing system capable of producing each of the above from raw lunar soil has achieved full self-sufficiency, or materials "closure."

Demonstration of materials closure plausibility. The components in table 5.11 were obtained first by taking a very basic list of necessary elements (the first four categories) for the entire LMF and adding to these any additional substances necessary to chemically produce the original items. This resulted in an increase in the number of items, therefore, all newly added items themselves had then to be similarly checked to ensure that each of *them* could be produced from the materials already at hand. This procedure was iterated until closure apparently was achieved.

The list includes reagents necessary for the production of microelectronic circuitry (Oldham, 1977), even though "wet chemistry" may not be necessary for this application in space manufacturing (Zachary, 1981). The team is unaware of any significant omissions in table 5.11, which demonstrates essential qualitative closure.

5E.2 Derivation of Minimum Requirements: Qualitative Materials Closure

The lunar substrate from which the required substances are extracted or manufactured has a mean global mineral content as shown in table 5.12. Source minerals for boron do not appear in this list, nor do the sources for volatiles implanted by the solar wind. A summary of all elements found to date in the lunar regolith samples returned by Apollo and Luna missions may be found in table 4.1.

To plausibly demonstrate materials closure, it must be shown that every item on the requirements list can be derived from other items on the list and that all elements are derived from those found in the lunar regolith. To fully and rigorously demonstrate closure, a detailed element-by-element breakdown of the entire factory would be required, giving the mass of each element or process chemical required followed by a convincing demonstration that such quantities could indeed be produced using only the amounts of other substances known to be available and an input of lunar material. This latter set of conditions is called quantitative closure.

Preparation of process minerals. A comparison of the list of process chemicals in category VI in table 5.11 with the minerals found in lunar soil (table 5.12) suggests that it may be possible to use raw lunar soil as input to the materials processing extraction machines if these minerals require no beneficiation. In the event such beneficiation is needed to obtain the specific minerals in separated form, the electrophoretic separation technique described in section 4.2.2 may be used. This method involves placing finely divided powdered lunar dust in aqueous (or slag, or other solvent) suspension which has a solvent pH tuned to match the isoelectric potential of the desired mineral species. A cross voltage is applied and all minerals but the one desired migrate away, leaving behind a purified residue – in the present case, anorthite and the category VII (table 5.11) process minerals may be recovered. Preliminary testing of the electrophoretic separation concept with simulated lunar soil has been successful (Dunning and Snyder, 1981).

In addition, the electrophoretic technique may prove invaluable in separating out "trace minerals" from lunar soil, in particular apatite and possible differentiated boron-containing minerals which may exist in the lunar regolith.

Separation of iron. The magnetic properties of lunar soil are due almost entirely to the presence of metallic iron, which occurs in lunar soil as a free element in the amount of 0.5% by weight, roughly 5% of the total iron content of the lunar regolith. Since it is magnetic, metallic iron may be

TABLE 5.12.– MINERALS TYPICALLY FOUND IN LUNAR REGOLITH
(from Williams and Judwick, 1980)

Major	Minor
Olivine $(Mg,Fe)_2SiO_4$	Spinels $(Fe,Mg,Al,Cr,Ti)O_4$
Pyroxene $(Ca,Mg,Fe)SiO_3$	Armalcolite (Fe_2TiO_5)
Plagioclase feldspars $(Ca,Na)Al_2Si_2O_8$	Silica (quartz, tridymite, cristobalite) SiO_2
	Iron Fe (variable amounts of Ni and Co)
	Troilite FeS
	Ilmenite $FeTiO_3$
Trace	
Phosphates	*Oxides*
Apatite[a] $Ca_5(PO_4)_3(F,Cl)_3$	Rutile TiO_2
Whitlockite[a] $Ca_9(Mg,Fe)(PO_4)_7(F,Cl)$	Corundum (?) Al_2O_3
	Hematite (?) Fe_2O_3
	Magnetite Fe_3O_4
Zr mineral	Goethite (?) FeO(OH)
Zircon[a] $ZrSiO_4$	
Baddeleyite ZrO_4	*Metals*
	Copper (?) Cu
Silicates	Brass (?)
Pyroxferroite $(Fe,Mg,Ca)SiO_3$	Tin (?) Sn
Amphibole $(Ca,Mg,Fe)(Si,Al)_8O_{22}F$	*Zr-rich mineral*
Garnet (?)	Zirkilite or zirconolite[a] $CuZrTi_2O_7$
Tranquilletyite[a] $Fe_8Zr_2Ti_3Si_3O_4$	*Meteoritic minerals*
	Schreibernite $(Fe,Ni)_3P$
Sulfides	Cohenite $(Fe,Ni,Co),O$
Mackinawite $(Fe,Ni)_9S_8$	Niningerite (Mg,Fe,Mn)S
Pentlandite $(Fe,Ni)_9S_8$	Lawrencite (?) $(Fe,Ni)Cl_2$
Cubanite $CuFe_2S_3$	
Chalcopyrite $CuFeS_2$	
Sphalerite (Zn,Fe)S	

[a]These minerals are known to exhibit complex substitutions, particularly of elements as Y, Nb, Hf, U, and the rare earth elements that are concentrated in these minerals.

separated from the raw lunar substrate by straightforward electromagnetic techniques directly as the raw input material leaves the input hopper. This Fe will be fairly pure, containing only about 5% nickel and 0.2% cobalt (Phinney et al., 1977).

Structural metals and metal oxides. Of all the chemical materials processing options studied to date, the hydrofluoric (HF) acid leach technique appears to have the best potential for minimum operating mass, ease of element separations to high purity, and favorable energy and heat-rejection requirements (Arnold et al., 1981; Waldron et al., 1979). HF acid leach (Waldon et al., 1979), shown in figure 5.41 in flowsheet form, uses low-temperature hydrochemical steps to separate the silica content of the lunar raw material from metallic oxides in minerals by converting them to fluorides and fluorosilicates. The silica is then vaporized as SiF_4, leaving Ca, Al, Fe, Mg, and Ti fluoro salts to be separated by a variety of solution, precipitation, ion exchange, and electrolytic steps. These are then reduced to the pure metallic form with sodium metal, which is recycled. (HF is added as a major process chemical.)

Sodium for the reduction of metals and silicon may be obtained by a modified Castner cell process, which involves the electrolysis of molten NaOH to produce Na, O_2, and H_2. Iron electrodes can be used in this application. (NaOH must also be added to the process chemicals list.)

Metal oxides and silicon dioxide can be obtained, where needed as ceramics, refractories, or for glasses, by hydrolysis of the fluoride or fluorosilicate with H_2O steam (for the metal oxides), with NH_3 (for silicon dioxide), or by ion exchange methods. (Water and ammonia are thus added to the list of process chemicals.) Electronics-grade silicon may be prepared through zone-refining and other techniques with up to nine-9s purity, although these processes have not been thoroughly investigated in the present study.

In a discussion of the HF acid leach technique, Criswell (1978) points out that the process with its various options is adaptable to several of the potential lunar minerals or concentrates including feldspars, pyroxenes, olivines, and even nonsilicates such as ilmenite and spinels. Beneficiation of these minerals (the major constituents of lunar soil) seems unnecessary since the appropriate separations are performed later on the fluorides and fluorosilicates. However, if necessary, this beneficiation can be accomplished using the electrophoretic method described above.

In addition to Fe, Al, Mg, Ti, Ca, Si, and O_2, it is possible that the HF acid leach process may be used to prepare Cr and Mn. These two elements are present in pyroxene (up to 0.5% MnO, up to 1.25% Cr_2O_3), olivine and spinel (which contain Cr). CrF_2 is slightly soluble in water; MnF_2 is soluble, so the techniques described above should still be applicable although the details of this extension have not been extensively studied.

One final problem unique to the HF process is the question of containers. Process vessels and tubing normally employed in terrestrial industry are attacked by hydrofluoric acid. One solution is to use special carbon steel alloys for this purpose – these are customarily employed for storage of fluorine gas because a protective layer of iron fluoride forms which greatly impedes further chemical attack. A second alternative is to use hydrocarbon-based waxes, paraffins or plastics which are not attacked by HF, applied as a thin layer to the insides of pipes and containers. Yet a third option is to develop new structures perhaps based on sulfur and phosphorus (Allcock, 1974) and other inorganic polymers (Lee, 1979) which could be in reasonably plentiful supply in the lunar factory.

Extraction of volatiles. Lunar soil heated to 1300 K releases 0.1% by weight of the following trapped volatiles: CO, CO_2, N_2, H_2, H_2O, SO_2, H_2S, CH_4, and inert gases (He, Ar, Ne, Kr, Xe). As much as 0.5-1.5% by weight may be released upon heating to 1700 K (Phinney et al., 1977). CO may be reduced to carbon by methanation followed by decomposition of the CH_4 species over a refractory catalyst (such as MgO) to C and H_2. CO_2 may be reduced to CO by making use of the reversible reaction:

$$2CO \underset{1300\text{ K}}{\overset{800\text{ K}}{\rightleftharpoons}} CO_2 + C$$

That is, CO_2 passed over elemental C above 1300 K reduces to CO, which can then be methanated and further reduced to C over hot refractory. N_2, H_2, H_2O, and SO_2 are desirable process chemicals. H_2S may be burned in O_2 to yield SO_2 and water. A sharply limited supply of O_2 results in steam and sulfur vapor. If SO_2 and H_2S are mixed at room temperature, they react to form water and elemental sulfur. Finally, oxygen bubbled through an aqueous solution of H_2S produces a precipitate of elemental sulfur.

Inert gases are useful in lasers and for providing a nonreactive atmosphere, and may be separated by fractional condensation using cold traps at various temperatures.

Boron production. Historically on Earth the most important source of boron has been borax or tincal ($Na_2B_4O_7 \cdot 10H_2O$), though today the more common source is kernite or rasorite ($Na_2B_4O_7 \cdot 4H_2O$). Other boron minerals include colemanite ($Ca_2B_6O_{11} \cdot 5H_2O$), ulexite ($NaCaB_5O_9 \cdot 8H_2O$), priceite ($Ca_4B_{10}O_{19} \cdot 7H_2O$), boracite ($Mg_3B_7O_{13}Cl$) in salt beds, and sassolite (H_3BO_3).

Boron minerals on the Moon are likely associated with phosphorus-bearing apatite species (Dunning, personal communication, 1980), although it is possible that local concentrations of the most common anhydrous boron mineral,

tincalconite ($Na_2B_4O_7$), may be found in the vicinity of ancient lunar volcanic vents. In either case it should be possible to isolate the boron species using a combination of chemical and electrophoretic techniques. However, the details of this process cannot be specified until available boron resources on the Moon are more precisely characterized.

Terrestrial boron-containing minerals are either calcium or sodium borates. A calcium borate may be converted to a sodium borate by treatment with Na_2CO_3, yielding borax and $CaCO_3$ which precipitates out of solution. (Calcium carbonate may be recycled by roasting to obtain CaO and CO_2, from the latter of which elemental carbon can be recovered.) Sodium borates are reduced to boric oxide in two steps:

$$Na_2B_4O_7 + H_2SO_4 + 5H_2O \rightarrow 4H_3BO_3 + Na_2SO_4$$

$$2H_3BO_3 \xrightarrow[\text{(heat)}]{} B_2O_3 + 3H_2O$$

The sodium and sulfur may be recycled by the following steps:

$$Na_2SO_4 + CaCl_2 \rightarrow CaSO_4 + 2NaCl$$

$$CaSO_4 + C \rightarrow SO_2 + CaO + CO$$

$$2NaCl \xrightarrow[\text{(electrolysis)}]{} 2Na + Cl_2$$

(Sulfuric acid and calcium chloride are added to the list of process chemicals.)

Low-purity boron is prepared by reduction of B_2O_3 with Mg, followed by vigorous washing with sodium alkali and HF. The impurities are a mixture of oxides and borides. Almost pure boron (up to 99.9999% is available commercially by this method) for electronics applications may be prepared by vapor phase reduction of BCl_3 (or BBr_3) with hydrogen on electrically heated filaments. BCl_3 is prepared by heating B and Cl_2 directly at 800 to 1100 K. Possible filament materials have not been investigated, but the mass requirement is probably less than 1 kg. Chlorine is added to the process chemicals list, since F_2 cannot be substituted for Cl_2 for vapor phase purification.

Phosphorus and halogens. More than 200 minerals containing up to 5% phosphorus by weight are known on Earth, but the two main species available on the Moon are fluorapatite, $Ca_5(PO_4)_3F$ and chlorapatite, $Ca_5(PO_4)_3Cl$. The other lunar phosphorus-bearing mineral, whitlockite, is generally given as $Ca_3(PO_4)_2$, but often is found associated with Mg, Fe, F, and Cl. Fluorapatite is by far the most abundant and is also the major source of fluorine on the lunar surface. (Amphibole has a trace of fluorine, but this small amount is probably not worth the trouble to extract.) Chlorapatite, very rare by comparison, is the major source of chlorine on the Moon, except for lawrencite (a nickel-iron chloride believed derived from meteorites). Whitlockite is also very rare.

Apatite is separated from lunar soil by the electrophoretic process described above. The calcium phosphate is then reduced to P_2O_5 by heating with silica (available from the HF leach stage) yielding pure phosphorus when treated with carbon:

$$2Ca_3(PO_4)_2 + 6SiO_2 + 10C \xrightarrow[\text{(electric furnace)}]{} P_4 + 6CaSiO_3 + 10CO$$

Alternatively, calcium phosphate dissolved in sulfuric acid gives phosphoric acid plus insoluble calcium sulfate (which may be recycled, see below). The acid is then reduced with carbon to obtain elemental phosphorus.

The sulfuric acid technique appears best for halogen extraction. When acted upon by sulfuric acid, a natural mixture of fluorapatite and chlorapatite undergoes the following net reaction:

$$3Ca_3(PO_4)_2 \cdot Ca(F,Cl)_2 + H_2SO_4 \rightarrow H_3PO_4 + HF + HCl + CaSO_4$$

This results in a solution of the three acids. If heated to above 390 K (but below 486 K), the HF and HCl boil off leaving pure orthophosphoric acid behind. The evaporate is condensed, then separated into HF and HCl by either of two methods. First, the acid solution is desiccated in vapor form over anhydrous $CaCl_2$, then cooled to 273 K. HF condenses and is removed in liquid form, leaving HCl gas to be electrolyzed to obtain H_2 and Cl_2. Or, second, after desiccation with $CaCl_2$ the HF/HCl solution is electrolyzed with the release of H_2 at one electrode and a mixture of F_2 and Cl_2 at the other. This mixture is cooled to 240 K which liquefies the Cl_2 (to be drained off) leaving F_2 gas, which may be combined directly with the liberated H_2 to make HF. This entire problem may also be circumvented if fluorapatite and chlorapatite can be separated using electrophoretic beneficiation.

To recover sulfur, a valuable volatile, from the above process, the calcium sulfate is recycled by roasting according to:

$$CaSO_4 + C \xrightarrow[\text{(heat)}]{} SO_2 + CO + CaO$$

Supporting reagents. Reagents necessary to ensure closure of the LMF chemical processing sector include sodium hydroxide, silane, sulfuric acid, nitric acid, freon, ammonia, calcium chloride and sodium carbonate. The derivation of each is briefly reviewed below.

Lunar pyroxene contains up to 0.2% and lunar plagioclase up to 1.5% Na_2O (Williams and Jadwick, 1980). Specific pyroxene minerals containing Na are acmite or aegirite, $Na_2O \cdot Fe_2O_3 \cdot 4SiO_2$ and jadeite, $Na_2O \cdot Al_2O_3 \cdot 4SiO_2$. Among plagioclase feldspars are anorthoclase, albite, and andesine, $Na_2O \cdot Al_2O_3 \cdot 6SiO_2$. After these minerals are obtained by electrophoresis, roasting causes the Na_2O component to sublime above 1200 K. By 1800 K as much as 70% of the available Na_2O may have evaporated, leaving behind a still solid residue of iron, silicon, and aluminum oxides (Williams and Jadwick, 1980). The liberated sodium oxide is dissolved in water to give NaOH. The small amount of Na produced during boron reduction may be added directly to the HF leach system as metal, or hydrated to form NaOH, as required.

Silane for microelectronic wafer fabrication may be prepared in either of two ways. First, elemental silicon may be heated in the absence of air with magnesium to form the silicide, which is then hydrolyzed with sulfuric acid to silanes and $MgSO_4$ (which can be recycled for sulfur much like calcium sulfate). This hydrolysis gives about 25% yield of silicon hydrides, comprised of 40% SiH_4, 30% Si_2H_6, 15% Si_3H_8, 10% Si_4H_{10}, and 5% of Si_5H_{12} and Si_6H_{14}. These may be separated by fractional distillation; or, if cooled to below 258 K, all species liquefy except SiH_4, which remains a gas and can be removed. A second process suggested by Criswell (1980a) involves hydrolysis of the Mg_2Si with HCl, with the magnesium chloride hydrolyzed by steam to recover the HCl.

Sulfuric acid is relatively simple to prepare, provided a suitable catalyst is available. In the two-step contact process, SO_2 is burned in oxygen and in the presence of catalyst to the trioxide, which is then dissolved in water to yield the acid. The usual catalyst was, traditionally, finely powdered platinum, and more recently vanadium pentoxide. If possible, the use of these substances should be avoided as Pt and V are rare in the lunar regolith. Fortunately, practically all refractory substances have some degree of catalytic activity in the contact process, provided they are immune to impurities. Alternative and plentiful viable catalyst agents include pumice ($SiO_2 \cdot Al_2O_3$), porcelain or powdered ceramic, and ferric oxide (Fe_2O_3), all of which are active and readily available in the LMF.

Nitric acid is more difficult to prepare, primarily because of the difficulty of "fixing" nitrogen chemically. The two most common commercial processes for acid production involve the use either of existing nitrate stocks or of platinum (for the catalytic oxidation of ammonia), neither of which is feasible at the LMF. A third method, not feasible commercially because of its low energetic efficiency, is the electric arc technique first discovered by Priestley in 1772. Elemental nitrogen and oxygen are passed through a spark discharge, producing nitric oxide with a yield of 2.5% under ideal conditions. After rapid quenching of the reaction mixture, the NO reacts rapidly below 873 K in an excess of O_2 to form NO_2, which makes nitric acid upon contact with water. Biological nitrogen fixation using *Rhizobium* and *Azotobacter* microorganisms is an interesting alternative and should be investigated further.

Freon (CF_4) is prepared by fluorination of methane with elemental fluorine. The resulting mixture of CF_4 and HF is separated by dissolution in water. There are two potentially feasible methods for producing ammonia. First is the standard Haber process, in which elemental nitrogen and hydrogen are combined directly at 800 K in the presence of iron and aluminum oxide catalysts. In the second process, magnesium is ignited at 600 K in a nitrogen atmosphere to form the nitride, which is then hydrolyzed to yield ammonia and magnesium hydroxide. Water and MgO are recycled by roasting the hydroxide.

Only very limited amounts of $CaCl_2$ are needed, so direct combination of the elements (both of which are already available) is the preferred production pathway. Sodium carbonate for boron production is obtained by bubbling CO_2 gas through an aqueous solution of NaOH, then gently heating to recover the solute.

5E.3 Quantitative LMF Materials Closure

The arguments presented in section 5E.2 demonstrate that a surprisingly simple system involving 18 elements and perhaps two dozen mineral species and process chemicals can probably achieve virtually 100% materials processing closure. Reagents necessary for electronics parts fabrication were included so that the lunar SRS has the materials needed to replicate its own computer and robot equipment. While the above is probably not the minimum size chemical processing plant that can retain closure, it is certainly one example of such a system. Other possibilities should be pursued in future research. Of course, once a growing seed reaches full adult size, it can install a whole new series of production equipment (say, for the recovery of platinum group metals) making possible a new range of capabilities that were unnecessary during the early growth/replication phases.

Quantitatively, in order to rigorously demonstrate complete materials closure it would be necessary to work through every chemical process described above, calculate the exact materials mass for every structure, robot, and other LMF device on an element-by-element basis, then verify that enough of each could be produced by the system. Such a detailed computation clearly lies beyond the scope of the present study. However, the team has attempted to estimate some of the most critical throughputs and analyze their anticipated effects upon total system closure. In this context, "closure" is a relationship between a given machine design and a particular substrate from which the machine's chemical elemental constituents are to be drawn. Hence, the numerical calculation of closure requires a knowledge of the precise composition both of

the intended base substrate to be utilized as well as of the products which the SRS must manufacture from that substrate.

Following a method suggested by the work of Freitas (1980), the "extraction ratio" R (see sec. 5.3.6) is defined as the total mass of raw substrate material which must be processed (input stream) to obtain a unit mass of useful system output having the desired mass fractions of each required element (output stream). Consider the significance of the extraction ratio to the problem of materials closure. An $R = 1$ means that 1 kg of lunar regolith contains exactly the mass of all necessary LMF elements to manufacture a kilogram of desired output product. $R = 10$, on the other hand, means that 10 kg of lunar regolith must be processed to extract all of the elements required to make 1 kg of final product (see sec. 5.3.6).

For the purposes of the present study the team chose a trial value of $R = 40$ kg/kg. This choice is based on information from previous studies which suggests that 40 represents a good intermediate value between low closure and high complexity SRS materials designs.

On the one hand, for $R < 10$, the available mass fractions of certain critical but relatively rare elements such as H, C, B, and Cl fall too low to remain credible for a system requiring 100% closure. The missing materials must be imported as "vitamins" or the entire SRS must be redesigned to eliminate chemical processing and electronics using these elements. Examples of low closure models include the lunar processing factory designs proposed by Ho and Sobon (1979), $R = 1.7$; O'Neill (1976), $R = 1.7$; Phinney et al. (1977), $R = 1.2$; and Waldron et al (1979), $R = 1.1$. These systems are capable of extracting only half a dozen of the most abundant lunar elements and are not expected to achieve more than 60-90% materials closure.

On the other hand, for $R > 100$ the problem lies not in extracting rare elements but in processing them fast enough to meet a $T = 1$ year replication time deadline. For instance, Freitas (1980) gives an example of a high complexity system which could extract 84 elements from asteroidal material. For $R = 26,800$ the replication time is 500 years. It appears that $10 < R < 100$ is a plausible condition for 100% closure and 1-year replication in SRS. The maximum recoverable mass from lunar soil for each element assuming $R = 40$ is estimated in table 5.13. The question remains whether or not these quantities are adequate to achieve quantitative materials closure.

Certainly 100% closure exists for the six primary structural elements Al, Ca, Fe, Mg, O, and Si. Even if the entire 100-ton seed were comprised entirely of any one of these there is enough available of each. A similar argument may be made for Ti, since 80 tons in theory can be extracted. Steels and other alloys typically have 1% Mn, 0.2% Cr, and 0.1% C or less, which limits the total steel mass to 400 tons, 4000 tons, and 400 tons, respectively. Hence, alloy production will not be materials-limited by these three elements.

Carbon is also used in the boron and phosphorus production cycles. The mass of boron is so low that the carbon requirement is negligible in terms of mass. In the phosphorus cycle, 10 atoms of C are needed to cycle 4 atoms of P. Phosphorus is required as a dopant in silicon microelectronic chip manufacture and in phosphoric acid which is used as a photolithography process chemical and which also appears during the halogen recovery cycle. At most, 40 kg of phosphorus are required, necessitating a carbon budget of 100 kg. This leaves more than 200 kg of carbon to account for losses and special uses such as CO_2 gas lasers.

Boron is used solely as a microelectronic silicon dopant; 4 kg of B can produce perhaps 10^3-10^4 kg of chips, more than enough for the 100-ton seed. A few kilograms of phosphorus (though high purity is required) will suffice for the same purpose, and the use of P as a process chemical should be more hydrogen-limited than phosphorus-limited because of the relative abundance of P in the lunar regolith.

According to calculations by Waldron et al. (1979), about 63 metric tons of H_2, F_2, and Na, half of which is F_2, are needed for an HF acid leach extraction facility having a total mass of about 823 tons. According to Criswell (personal communication, 1980) this model may

TABLE 5.13.– MAXIMUM MASS OF CHEMICAL ELEMENTS EXTRACTABLE FROM LUNAR SOIL, per year, for a 100-ton Seed with Extraction Ratio $R = 40$

Element	Typical global lunar abundance, kg element/kg soil	Maximum extractable mass, kg $(R_{ch})(A)(100\text{ tons})$
Fe	0.10	400,000
Al	.07	280,000
Mg	.05	200,000
Ti	.02	80,000
Mn	.001	4,000
Cr	.002	8,000
C	.0001	400
Si	.2	800,000
O	.4	1,600,000
P	.0005	2,000
B	.000001	4
F	.0001	400
N	.0001	400
H	.00005	200
Na	.003	12,000
Cl	.00002	80
Ca	.07	280,000
S	.001	4,000
(Inert gases) ?	.00001	40

scale almost linearly down as low as 1 ton. The equivalents for a scaled-down 2.5-ton HF leach system are 90 kg F_2 and 100 kg of H_2 and Na. Sodium is about an order of magnitude more abundant than required, and fluorine does not appear to be a limiting factor even if recovery losses and spillages permit only 50% utilization of available stock. The supply of hydrogen, however, is crucial in achieving quantitative materials closure (see below). The 2.5-ton plant described above can output about 91 tons/year, which should be adequate to replicate a 100-ton seed once per year.

The primary use of nitrogen is in making NH_3 for the recovery of silica and as N_2 and HNO_3 for the production of microelectronic chips. The 400 kg N_2 given in table 5.13 is sufficient to prepare a maximum of 490 kg NH_3 or 1800 kg HNO_3. (These applications would require a maximum of 86 kg and 29 kg of H, respectively, hence are not seriously hydrogen-limited.) The amount of nitric acid seems more than sufficient, and the NH_3 can produce 100 to 1000 kg of silica, which should be adequate with recycling and provided losses can be held to a minimum.

Chlorine appears in the boron- and phosphorus-production cycles – in the former it is consumed and must be recycled; in the latter it is incorporated in a deliquescent compound and should not incur serious losses or require chemical recycling. The preparation of 1 mole of boron requires recycling 0.25 mole of Cl, hence (0.5)(4 kg)(35.45/10.8) = 6.6 kg of chlorine are needed to produce 4 kg of boron. As for the phosphorus cycle, 80 kg Cl produces 125 kg of deliquescent $CaCl_2$ which is capable of absorbing roughly its own weight in water. This should be sufficient with recycling (by simple heating) no more often than once a month on a $T = 1$ year schedule.

Sulfur is used primarily in the casting subsystem in the fabrication sector (about 600 kg required) and in the manufacture of sulfuric acid. This product is mass-limited about equally by the amounts of S and H available. The 4000 kg of sulfur can be used to prepare 12,000 kg H_2SO_4, and the 200 kg of hydrogen can make up 9800 kg of the acid. Since hydrogen also has many other uses, available S will be underutilized and perhaps 1 or 2 tons of H_2SO_4 reasonably can be produced. Is this enough? The main uses of sulfuric acid are in the recovery processes for B, P, F, and Cl, and in the preparation of silanes. The ratio of B:H_2SO_4 is about 4:1 moles, so to extract 4 kg B requires 9.1 kg acid. For phosphorus extraction, P:H_2SO_4 :: 2:3 moles, so (3/2)(98.1/31)(40 kg) = 190 kg H_2SO_4. For fluorine extraction, F:H_2SO_4 :: 2:1 moles, which requires (1/2)(98.1/19)(200 kg) = 516 kg acid. For chlorine extraction, Cl:H_2SO_4 :: 2:1 moles, which requires (1/2)(98.1/35.45)(80 kg) = 110 kg H_2SO_4. The quantity of silane needed for microelectronics processing is expected to be minimal, so it appears that adequate supplies of sulfuric acid can be made available with reasonable loss factors to sustain the growth of a fully autonomous LMF on a sulfur budget of about 1500 kg.

The only critical element appears to be hydrogen. This criticality is not especially peculiar to the present design, but rather stems from the relative scarcity of the element in lunar materials and the many chemical processing applications to which it may be put. Any hydrogen-chemistry-based materials processing system will encounter similar difficulties. The 200 kg of available hydrogen could make the maximum quantities of H-bearing compounds listed in table 5.14, although the available hydrogen must be spread among these applications as required with lower masses in

TABLE 5.14.– HYDROGEN-LIMITED MATERIALS PROCESSING REAGENTS

LMF reagent	Estimated LMF requirements, kg	Maximum, limited only by hydrogen available, kg	Fraction of available hydrogen required
NH_3	300	1,100	0.273
H_2SO_4	1,000	9.800	.102
SiH_4	100	1,600	.0625
HNO_3	600	12,600	.0476
HF	190	4,000	.0475
H_3PO_4	100	6,500	.0154
NaOH	100	8,000	.0125
H_2O	700	1,800	.3895
(5% losses)	---	---	.05
			1.000

each case. Expected requirements of hydrogen-bearing reagents are listed in table 5.14. Although these calculations are highly sensitive to the assumptions employed, closure may be achieved if an allowance of 5% for spillage and other losses is adequate. Obviously a major leak could seriously jeopardize a hydrogen-based LMF system.

If hydrogen supply remains a critical problem it may become necessary to: (1) redesign the processing system for greater hydrogen frugality, (2) select a slightly higher extraction ratio R to permit recovery of a greater mass of H, (3) locate and "mine" particular lunar soils extra-rich in H, such as the suggested use of ilmenite as a hydrogen "ore" (Green, personal communication, 1980), (4) accept a replication time longer than 1 year, or (5) relax the 100% closure requirement and permit resupply of small amounts of hydrogen "vitamin" from Earth.

5E.4 Sector Mass and Power Estimates

The overall functional layout of the LMF chemical processing sector is illustrated in figure 5.16. The operations flowsheet shows that there are 13 components within the sector: (1) input hopper, (2) electrophoretic separators, (3) P/F/Cl extractors, (4) boron extractors, (5) sodium extractors, (6) volatiles extractors, (7) HF acid leach system, (8) freon producer, (9) ammonia producer, (10) silane producer, (11) nitric acid producer, (12) sulfuric acid producer, and (13) output hopper.

Mass and power consumption for LMF materials processing may be estimated by comparison with other automated chemical processing designs that have been considered, and which are summarized in table 5.15. For $R = 40$, a 100-ton/year (self-replicating) output demands a 4000-ton/year raw materials input, or 0.13 kg/sec. Taking the range of values given in table 5.15, Sector mass should lie within 18,200 to 78,000 kg. Similarly, the estimated power requirements range from 455 kW up to 10.9 MW, although in this case the lower values seem more appropriate. Dry thermal chemical processing techniques are associated with very high energy requirements, whereas lower values are found in wet chemistry processes – of which the HF acid leach selected for the present design is an example.

5E.5 Information and Control Estimates

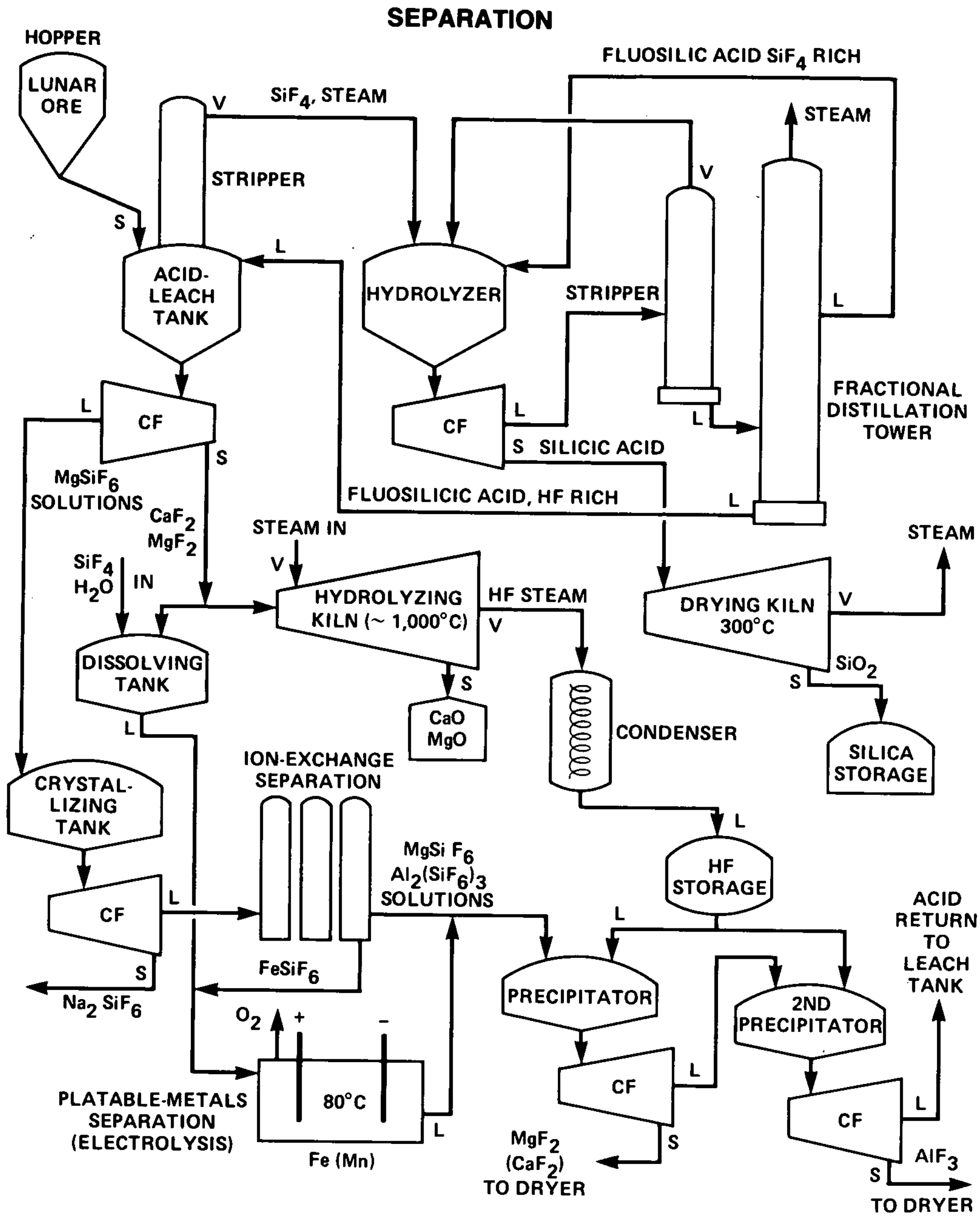
Probably the most complex of the 13 sector components which appear in figure 5.41 is the HF acid leach system. From figure 5.41 this appears to consist of 34 component subsystems such as "precipitator," "dissolving tank," "fractional distillation tower," "centrifuge/filter," "Castner cell," etc. Each subsystem performs a single well defined task. In addition, there are 111 nodes (each denoting a point of connection of a pipe or supply line to another pipe or to a subsystem) each requiring at least one valve and valve control mechanism. At each valve there must be a number of sensors indicating valve position (open, closed, fractionally open), valve malfunction and cause (if simple), and volume or velocity of flow of material through the valves. Interface with actuators and reportage to the subsystem subcomputer are additional requirements.

Assuming each valve can be automated with a 1K computer allocation, and each subsystem can be automated with a 10K memory allocation, then the total computer capability required for continuous leach system operation is $(1)(111) + (10)(34) = 451K$ which is 7.2×10^6 bits using 16-bit words. This should be sufficient to handle normal system operations and troubleshooting, although actual repair must be done by mobile repair robots. Also, any catastrophic malfunctions such as pipe ruptures, jammed fixtures, leaks, heating element burnouts or explosions must be diagnosed and corrected by the mobile repair robots.

The chemical processing sector looks not to be a place where complicated new automation techniques will be

TABLE 5.15.– COMPARISON OF CHEMICAL PROCESSING PLANT MASSES AND POWER REQUIREMENTS FROM PREVIOUS RELATED STUDIES

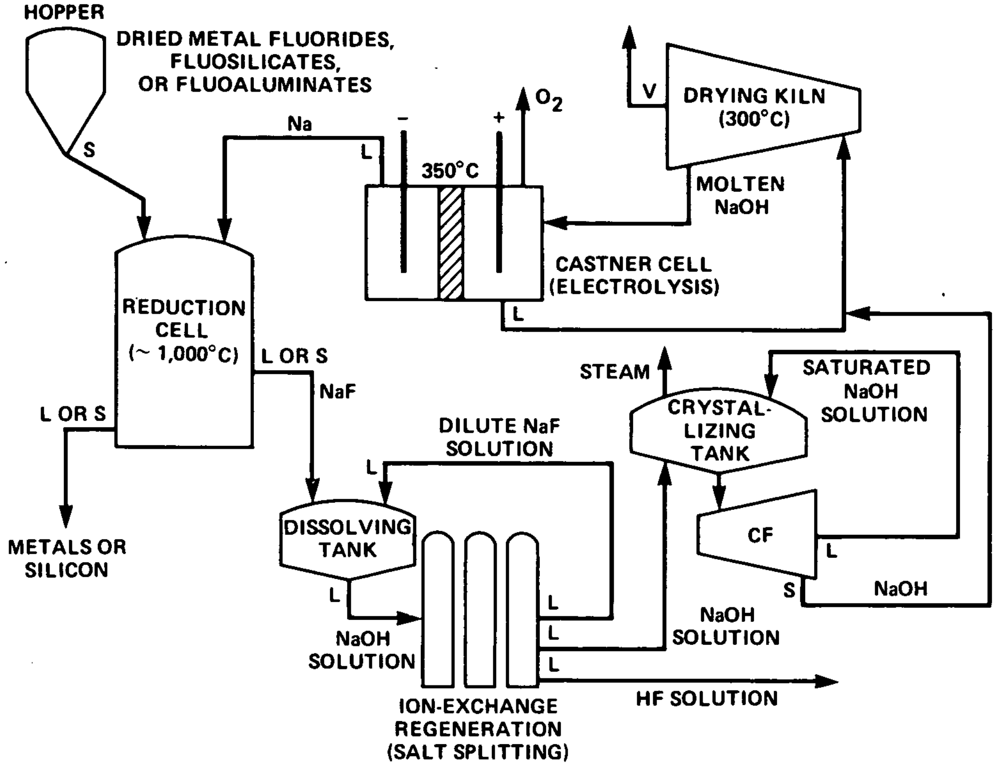
Source	kg plant / kg/sec input	kg chemicals / kg/sec input	kg chemicals / kg plant	Power, W / kg plant
Johnson & Holbrow (1977) Al-processing plant	3.1×10^5	2.6×10^4	8.5×10^{-2}	25
Phinney et al. (1977) carbo/silico-thermic plant	3.3×10^5	---	---	140
Waldron, Erstfeld, & Criswell (1979) HF acid-leach metal extraction plant	$3\text{-}6 \times 10^5$	3.5×10^5	4.1×10^{-1}	37
O'Neill, Driggers, & O'Leary (1980) reference design figure	3.6×10^5	---	---	---
O'Neill (1976) carbothermic Space Manufacturing Center	5.2×10^5	---	---	100
Criswell (1980)	3.6×10^5	---	---	100
Vajk et al. (1979) Space Manufacturing support requirements	1.4×10^5	---	---	---



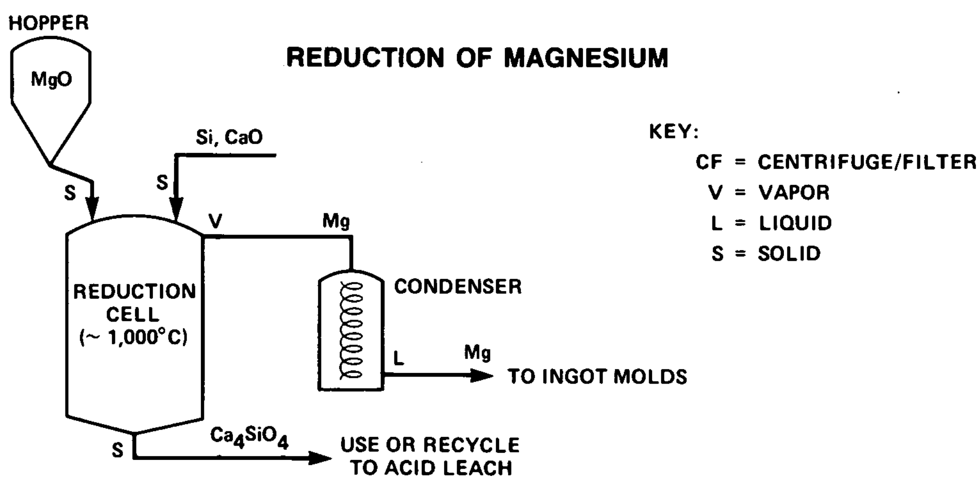
1. $xMO \cdot SiO_2 + (4 + 2x)\ HF = xMF_2 + SiF_4\ (aq) + (2 + x)\ H_2O$
1′. $xMO \cdot SiO_2 + (5 + 2x)\ HF = xMF_2 + HSiF_1\ (aq) + (2 + x)\ H_2O$
2. $SiF_4\ (aq) + nH_2O = SiF_4\ (v) + nH_2O\ (v)$
2′. $HSiF_5\ (aq) + nH_2O = SiF_4\ (v) + HF\ (aq) + nH_2O\ (v)$
3. $(1 - y)\ [SiF_4\ (v) + 4H_2O = Si\ (OH)_4 + 4HF]$
3a. $(1 - y)\ [SiF_4\ (v) + 2H_2O = SiO_2 + 4HF]$
4. $(1 - y' - z)\ [xMF_2 + xH_2O = xMO + 2xHF]$
5. $y\ [SiF_4 + 4Na = Si + 4NaF]$
6. $y'\ [xMF_2 + 2xNa = xM + 2xNaF]$

Figure 5.41.– Flowsheet and process equations for the HF acid-leach process.

REDUCTION OF METALS OTHER THAN MAGNESIUM



REDUCTION OF MAGNESIUM



7. $z\,[xMF_2 + xSiF_4\,(aq) = xMSiF_6\,(aq)]$
8. $z\,[xMSiF_4\,(aq) + xH_2O + \text{electrical energy} = (x/2)O_2$
   $+ xM + xH_2SiF_6]$

8a. $z\,[xMSiF_6\,(aq) + M'SO_3R^{*} = xM'SiF_6\,(aq) + xMSO_3R^{*}]$

9. $m\,NaF + mR^{*}OH = mNaOH + mR^{*}F$

9a. $m\,NaF + (m/2)\,Ca\,(OH)_2 = mNaOH + (m/w)\,CaF_2$

10. $m\,NaOH + \text{electrical energy} = mNa + (m/2)O_2 + (m/2)H_2O$
11. $(1 - y)\,[Si\,(OH)_4 = SiO_2 + 2H_2O]$

R^{*} = ion-exchange resin
$m = 4y + 2xy'$

Figure 5.41.– Concluded.

required (Ayres, 1952; Foster, 1963; Kallen, 1961; Luke, 1972). Component processes are all state-of-the-art. There is no need for visual processing during normal operations, and procedures are standardized so no expert systems or judgmental algorithms are required beyond the simple integration of well defined sensor data. All operations will probably be hard-automated, and materials will be moved about almost entirely in sealed vessels. If there is need for additional transport within the system a materials transit network may be erected using metal or basalt tracks, electric motors and small carrier vehicles.

If each of the 13 sector components is as complex as the HF acid leach system (certainly a gross overestimate), then the total computer control capability required is about 6 megabytes or 9.4×10^7 bits using 16-bit words. The information needed to describe the sector sufficiently for purposes of self-replication must also be estimated. Assuming that each HF leach subsystem requires 2×10^6 bits for complete description (about a 200-page printed book, or 80,000 English words), and that each valve requires about half as much (say, 100 book pages), then the total for the HF leach system is 1.8×10^8 bits. Again conservatively multiplying by 13, the total information to describe the sector components is 2.3×10^9 bits. If sector equipment is distributed across a floor space of 5000 m^2, then to store a map with 1-cm placement resolution requires a memory capacity of 8×10^8 bits assuming one 16-bit word to describe the nominal status of each 1 cm^2 of platform space. Note that large empty areas convey useful information and must be mapped, since they may be used for traffic routes, repair routes, temporary warehousing, etc. The total information for sector replication is thus about 3.1×10^9 bits. The information control budget is 9.4×10^7 bits.

5E.6 References

Allcock, Harry R.: Inorganic Polymers. Scientific American, vol. 230, March 1974, pp. 66-74.

Arnold, J. R.; Criswell, D. R.; and Waldron, R. D.: Progress Report on Experimental Program to Develop HF Acid Leach Process for Refining Lunar Soils. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

Ayres, Eugene: An Automatic Chemical Plant. Scientific American, vol. 187, September 1952, pp. 82-88.

Criswell, David R.: Extraterrestrial Materials Processing and Construction. NASA CR-158870, vol. 1, 1978, and vol. 2, 1980.

Dunning, Jeremy D.; and Snyder, Robert S.: Electrophoretic Separation of Lunar Soils in a Space Manufacturing Facility. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

Foster, David: Modern Automation. Sir Isaac Pitman and Sons, Ltd., London, 1963.

Freitas, Robert A., Jr.: A Self-Reproducing Interstellar Probe. J. British Interplanet. Soc., vol. 33, July 1980, pp. 251-264.

Ho, Darwin; and Sobon, Leon E.: Extraterrestrial Fiberglass Production Using Solar Energy. In Space Resources and Space Settlements, John Billingham, William Gilbreath, Brian O'Leary, and B. Gossett, eds. NASA SP-428, 1979, pp. 225-232.

Johnson, Richard D.; and Holbrow, Charles, eds.: Space Settlements: A Design Study, NASA SP-413, 1977. 185 pp.

Kallen, Howard P., ed.: Handbook of Instrumentation and Controls. McGraw-Hill Book Co., New York, 1961.

Lee, Stuart M.: Lunar Building Materials – Some Considerations on the Use of Inorganic Polymers. In Space Resources and Space Settlements, John Billingham, William Gilbreath, Brian O'Leary, and B. Gossett, eds. NASA SP-428, 1979, pp. 233-241.

Luke, Hugh D.: Automation for Productivity. John Wiley and Sons, New York, 1972.

Oldham, William G.: The Fabrication of Microelectronic Circuits. Scientific American, vol. 237, September 1977, pp. 110-128.

O'Neill, Gerard K.: Engineering a Space Manufacturing Center. Astronautics and Aeronautics, vol. 14, October 1976, pp. 20-28, 36.

O'Neill, Gerard K.; Driggers, G.; and O'Leary, B.: New Routes to Manufacturing in Space. Astronautics and Aeronautics, vol. 18, October 1980, pp. 46-51.

Phinney, W. C.; Criswell, D. R.; Drexler, E.; and Garmirian, J.: Lunar Resources and Their Utilization. In Space-Based Manufacturing from Nonterrestrial Materials, Gerard K. O'Neill, ed., AIAA, New York, 1977, pp. 97-123. (Progress in Astronautics and Aeronautics, vol. 57.)

Vajk, J. P.; Engel, J. H.; and Shettler, J. A.: Habitat and Logistic Support Requirements for the Initiation of a Space Manufacturing Enterprise. In Space Resources and Space Settlements, John Billingham, William Gilbreath, Brian O'Leary, and B. Gossett, eds., NASA SP-428, 1979, pp. 61-83.

Waldron, R. D.; Erstfeld, T. E.; and Criswell, D. R.: The Role of Chemical Engineering in Space Manufacturing. Chemical Engineering, vol. 86, 12 February 1979, pp. 80-94.

Williams, Richard J.; and Jadwick, James J.: Handbook of Lunar Materials, NASA RP-1057, 1980.

Zachary, Wm. B.: A Feasibility Study on the Fabrication of Integrated Circuits and Other Electronic Components in Space Utilizing Lunar Materials. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

APPENDIX 5F

LMF PARTS FABRICATION SECTOR

There are two distinct classes of fabrication production machines in any general-product self-replicating system – parts or "bulk" fabrication and electronics or microcircuit fabrication. Appendix 5F is concerned exclusively with LMF substystems required for bulk manufacturing. Microelectronics production in space manufacturing facilities is considered in section 4.4.3 and is the subject of Zachary (1981); estimated mass of this component of the original LMF seed is 7000 kg, with a power draw of perhaps 20 kW to operate the necessary machinery (Meylink, personal communication, 1980).

5F.1 Overall Design Philosophy

The plausibility of both qualitative and quantitative materials closure has already been argued in appendix 5E. A similar line of reasoning is presented here in favor of a very simple parts fabrication system, to be automated and deployed in a self-replicating lunar manufacturing facility. To rigorously demonstrate parts closure it would be necessary to compile a comprehensive listing of every type and size of part, and the number required of each, comprising the LMF seed. This list would be a total inventory of every distinct part which would result if factory machines were all torn down to their most basic components – screws, nuts, washers, rods, springs, etc. To show 100% closure, it would then be necessary to demonstrate the ability of the proposed automated parts fabrication sector to produce every part listed, and in the quantities specified, within a replication time of $T = 1$ year, starting from raw elemental or alloy feedstocks provided from the chemical processing sectors.

Unfortunately, such a detailed breakdown and analysis probably would require tens of thousands of man-hours even for the simplest of systems. Not only is the seed not a simple system, but the present baseline design is not conveniently amenable to this sort of detailed analysis. Thus, a completely rigorous demonstration of parts closure is beyond the scope of the present study.

However, it is possible to advance a plausibility argument based upon a generalized parts list common to many complicated machines now in use in various terrestrial applications (Spotts, 1968; von Tiesenhausen, unpublished Summer Study document, 1980). Although machines designed for construction and use in space may employ radically different components than their terrestrial counterparts, to a first approximation it may be assumed that they will be comprised generally of the same kinds of parts found in commonplace machines on Earth such as bolt, nut, screw, rivet, pulley, wheel, clutch, shaft, crank, rod, beam, wire, plate, disk, bushing, cable, wedge, key, spring, gasket, seal, pipe, tube, and hose. If this is valid, then a showing that all parts classes in the general parts list can be manufactured by the proposed automated fabrication system may serve as a valuable plausibility argument in favor of parts closure for that system.

The achievement of a sound design which incorporates the advantages of maximum economy in manufacture and functional requirements of a part is dependent upon the designer's ability to apply certain basic rules (Yankee, 1979). There are four recognized rules, equally applicable to terrestrial factories and lunar replicating machine systems, as follows:

(1) Design all functional and physical characteristics for greatest simplicity. As a general principle, service life of a part is greatly increased when design of that part is both simple and sturdy ("robust"). Performance is more predictable and costs (money, build time, repair time) are lower for simpler parts.

(2) Design for the most economical production method. The particular production design selected should, if possible, be optimized for the part or set of parts the system must produce. The production of scrap (input/output ratio) is one valuable index by which optimality may be compared. This factor is relatively simple to evaluate where only one part is manufactured. In multipart production lines the problem is far more complicated, since each of the many parts may be expected to have dissimilar optima. Consequently, only the production of the entire system can be truly optimum.

(3) Design for a minimum number of machining operations. All types of costs are lower when fewer operations are required to produce a part according to specifications. The greatest savings result when the number of separate processing operations necessary to complete a part is reduced. Multiple operations which can be combined into fewer operations, or functionally similar parts requiring fewer production steps, should be changed in a design. "Needless fancy or nonfunctional configurations requiring extra operations and material" should be omitted from the design (Yankee, 1979).

(4) Specify finish and accuracy no greater than are actually needed. If a part will adequately serve its intended purpose at some lower level of accuracy of machining than is technologically possible, then cheaper, simpler production processes may be used which make closure easier to attain. The specification of needlessly close tolerances and an unreasonable degree of surface finish invariably results in a low part production rate, extra operations, high tooling costs, and high rejection rates and scrap losses (Yankee, 1979).

5F.2 Selection of Basic Production Processes

A wide variety of fabrication processes is available using current technology, each of which is optimum for the production of one or more classes of parts or in certain specialized applications (see table 4.17). From inspection of table 4.10 it is reasonable to conclude that there are perhaps only 300 fundamentally distinct fabrication techniques in widespread use today. Ultimately, the LMF factory in production phase may be called upon to perform many if not all of these functions. However, most may be unnecessary for initial system growth or replication. Indeed, optimum seed design should permit maturation to adulthood in the minimum time with the fewest parts using the fewest machine operations possible.

The team concluded that four basic processes – plaster casting, vapor deposition, extrusion, and laser machining – are probably sufficiently versatile to permit self-replication and growth. These four techniques can be used to fabricate most parts to very high accuracy. Plaster casting was selected because it is the simplest casting technique for producing convoluted parts as well as flat-surface parts, to an acceptable level of accuracy. (A number of alternatives have already been reviewed in app. 4B.) The laser machining tool can then cut, weld, smooth, and polish cast parts to finer finishes as required. Vapor deposition is the least complicated, most versatile method of producing metal film sheets to be used as the manufacturing substrate for microelectronics components, mirrors or solar cells, or to be sliced into narrow strips by the laser for use as wire. The extruder is used to produce thread fibers of insulating material, presumably spun basalt drawn from a lunar soil melt as described in section 4.2.2.

5F.3 Casting Robot

The casting robot is the heart of the proposed automated fabrication system. It is responsible for producing all shaped parts or molds from raw uncut elemental materials. The moldmaking materials it works with are of two kinds. First, the casting robot receives thermosetting refractory cement with which to prepare (a) molds to make iron alloy parts, (b) molds to make iron molds to cast basalt parts (but not aluminum parts, as molten aluminum tends to combine with ferrous metal), and (c) individual refractory parts. Second, the robot receives hydrosetting plaster of Paris with which to prepare (a) molds to cast aluminum parts and (b) substrates for the vacuum deposition of aluminum in sheets. According to Ansley (1968), small castings using nonferrous metals (aluminum, magnesium, or copper alloys) may be produced using plaster molds with a surface finish as fine as 2-3 μm and an accuracy of ±0.1 mm over small dimensions and ±0.02 mm/cm across larger surfaces (a drift of 2 mm over a 1 m^2 area).

Traditionally, the plaster casting technique requires a split metal pattern in the shape of the object to be cast. This pattern is used to make a hollow mold into which molten metal is poured, eventually solidifying to make the desired part. Alternatively, patterns may be manually carved directly into the soft, setting plaster, after which metal again is poured to obtain the desired casting.

The casting robot should have maximum versatility. It will have access to a template library located within its reach, containing samples of each small or medium-sized part of which the LMF is comprised. If the SRS seed is designed with proper redundancy, it will use the fewest number of different kinds of parts and there will be large numbers of each kind of part. Assuming that on average there are 1000 pieces of each type of part in the original LMF architecture, then the total template library has a mass of only 100 tons/1000 = 100 kg and there are perhaps a thousand different *kinds* of parts (see below).

In addition, the casting robot is equipped with shaping and carving tools which can create any desired shape in the slowly hardening plaster. (Pure gypsum plaster hardens in 6-8 min after water is added, but this setting time may be extended up to 1-2 hr by adding lime, CaO, to the emulsion. Setting time is also temperature-dependent.) The shaping tools may represent perhaps 100 specific shapes and sizes and should also include at least a dozen "universal" carving instruments.

To make a given part, the robot searches its template library to see if it has a convenient pattern already in stock. If so, it uses the pattern to form the mold; if not, it uses its many tools to carve out a mold of the appropriate size and shape. Plaster of Paris is a hydraulic cement – it sets with the addition of water. Refractory cement is thermosetting and has to be heated to 1300-1400 K in a kiln to set the mold.

Water used to make the plaster molds cannot remain liquid in the lunar vacuum. Thus, the casting robot plaster system must be pressurized, probably with nitrogen gas to permit the pouring of molten aluminum. The triple point of water (the bottom end of its liquid phase) occurs at 608 Pa, but a 1.3×10^4 Pa atmosphere (16 kg N_2 to fill a 100 m^3 working volume) prevents water from boiling off up to about 323 K.

Mass requirements for plaster molding are estimated by assuming that 10% of the volume of each mold contains a

useful part (10% mold volume utilization). If the mean density of LMF parts (mostly aluminum) is taken as 3000 kg/m^3, and the entire plaster mass is recycled once a day, then for a 100-ton seed the robot must have 2600 kg (0.91 m^3) of plaster compound (gypsum, or calcium sulfate) on hand. To hydrate (set) this much plaster requires 483 kg of water, an amount of precious hydrogen already allowed for in LMF materials estimates presented in appendix 5E. Availability of sulfur is not a concern, since 2600 kg of plaster requires only 475 kg of S. Terrestrial plasters commonly have a small amount of strengthener added, but in the lunar application this substance should be designed to be recyclable or must be eliminated altogether.

Plaster casting is not the only way to make parts in a growing, self-replicating factory, but it is definitely one of the easiest both conceptually and in common industrial practice. Plaster methods are especially well suited for producing parts with hard-to-machine surfaces such as irregularly shaped exterior surfaces and in applications where a superior as-cast surface is important (Yankee, 1979). Plaster molded products commonly include aluminum match plates, cores and coreboxes, miscellaneous parts for aircraft structures and engines, plumbing and automotive parts, household appliances, hand tools, toys, and ornaments. The technique is good for manufacturing parts requiring high dimensional accuracy with intricate details and thin walls ($\geqslant$0.5 mm). Castings of less than 0.45 kg and as massive as 11,350 kg have been made on Earth. Commercially, when compared to aluminum die casting, plaster mold casting is considered economical if 1000 parts or less are produced, although production runs up to 2000 parts may also be considered economical if the parts are especially complex.

Refractories. Refractories are materials which remain useful at very high temperatures, usually 1500-2300 K. They are employed primarily in kilns, blast furnaces, and related applications. In the lunar SRS refractories are needed as linings for drying kilns, roasting ovens, in the production of iron molds (to cast basalt parts) and iron parts, and also as material for special individual parts such as nozzles and tools which must operate at very high temperatures.

Refractories are usually, but not always, pure or mixtures of pure metal oxides. Tables in Campbell and Sherwood (1967) list the most important simple and complex refractory substances which LMF designers might choose. There are a few basic considerations, such as vapor pressure. For instance, although magnesia melts at 3070 K and has a useful operating temperature to about 2700 K in oxidizing atmospheres, it cannot be used in a vacuum at temperatures above about 1900 K because of volatization (Johnson, 1950). Similarly, zinc oxide volatizes above 2000 K and tin oxide sublimes excessively at 1780 K even in an atmosphere.

Refractory bodies are fabricated from pure oxides by powder pressing, ramming, extruding, or slip casting. The last of these is the simplest, but requires a very fine powder. This powder is normally prepared by ball milling. Steel mills and balls are used, and the iron is later separated by chemical means. For simplicity in LMF design, the iron alloy powder inevitably mixed with the milled product can be removed by magnetic separation.

High-alumina cements and refractories may be the best option for lunar manufacturing applications. Alumina is a major product of the HF acid leach system in the chemical processing sector, and is capable of producing castable mortars and cements with high utility up to 2100 K (Kaiser, 1962; Robson, 1962). It will permit casting iron alloys, basalts, and low melting point metals such as Al and Mg. Unfortunately, it will not be possible to cast titanium alloys in this fashion, since in the liquid state Ti metal is very reactive and reduces all known refractories.

Alumina can be slip-cast from water suspensions. The oxide powder is first ball-milled as described above to 0.5-1.0 μm, then deflocculated by the addition of either acid (HCl) or base (NaOH), and finally the refractory body is developed by absorbing the liquid in a porous mold (plaster of Paris may be used with a base deflocculant). Gravity and hydrodynamic pressure of the flowing liquid produce a well compacted body of the suspended particles (Campbell and Sherwood, 1967). A fairly comprehensive review of alumina and alumina ceramics may be found in Gitzen (1966).

Metal alloys. A number of different metal alloys will be required for casting various parts and molds. Different alloys of iron may be chosen for the steel balls for ball milling, the basalt casting molds, and the individual part that might be comprised of steel or iron. Various aluminum alloys may be selected for parts, whereas pure metal is required for vapor deposition processes. Castable basalt may require fluxing but otherwise is a fairly straightforward melt.

Metallurgical duties are performed at the input terminus of the fabrication sector. Mobile chemical processing sector robot carriers dump measured quantities of metals and other substances into cold fabrication sector input hoppers (made of cast basalt and perhaps stored under a thin oxygen atmosphere to preclude vacuum welding). Mixing is accomplished by physical agitation, after which the contents are fed into a solar furnace to be melted. If net solar efficiencies are roughly the same as for the 5 kg capacity induction furnace (output 30 kg/hr) described in the MIT space manufacturing study (Miller and Smith, 1979), then about 30 kW of power are required which may be drawn most efficiently from a large collector dish roughly 6 m diam. There are at least three hopper/furnace subsystems required – a minimum of one each for iron, basalt, and aluminum alloys. Possibly another would be needed for

magnesium alloys, and several more to forestall contamination between disparate batches, but three is the absolute minimum requirement.

Parts manufacturing. The construction of a machine system as complex as a lunar SRS will require a great many individual parts which vary widely in mass, shape, function, and mode of assembly. If a complete parts list were available for the seed, then the manufacturing steps for each could be explicitly specified, precise throughput rates and materials requirements given, and closure demonstrated rigorously. Unfortunately, no such list is yet available so the team was forced to resort to the notion of the "typical part" to gain some insight into the performance which may be required of the casting robot.

Modern aircraft have about 10^5 parts and weigh up to about 100 tons, for an average of 1 kg/part (Grant, 1978). The average automobile has 3000–4500 parts, depending on its size and make, so the typical part weights perhaps 0.5 kg (Souza, personal communication, 1980). A study performed for General Motors concluded that 90% of all automotive parts weigh 2 kg or less (Spalding, personal communication, 1980). A design study by the British Interplanetary Society of a very advanced extrasolar space probe assumed a figure of 9 kg per typical part (Grant, 1978). Conservatively estimating that the typical LMF part is only 0.1 kg, then a 100-ton seed is comprised of roughly a million parts.

If most components may be made of aluminum or magnesium then the density of the typical part may be taken as about 3000 kg/m^3, so the characteristic size of the typical part is $(0.1/3000)^{1/3}$ = 3.2 cm. This result is consistent with Souza's (personal communication, 1980) suggestion that the average automobile part could be characterized as "roughly cylindrical in shape, an inch in length and half an inch in diameter." The casting robot must be able to cast all 10^6 parts within a replication time T = 1 year. If the casting bay is only 1 m^2 in horizontal extent, and only 10% of that area is available for useful molding, then each casting cycle can prepare molds for 0.1 m^2 of parts. The characteristic area of the typical part is $(0.1/3000)^{2/3}$ = 0.001 m^2, and dividing this into the available area gives 100 parts/casting cycle as the typical production rate for the robot. To produce 10^6 parts/year the casting robot must achieve a throughput rate or 10,000 cycles/year, or about 52 min/cycle. This in turn implies that the system must be able to carve or mold at an average rate of 30 sec/part. Since most parts should be simple in form or will have patterns available, this figure appears feasible. After the casting robot makes molds for the parts, the molds are filled with molten aluminum alloy. The metal hardens, the mold is broken, and the pieces are recycled back into plaster of Paris; the aluminum parts formed in the mold are conveyed to the laser machining and finishing station.

Very thin sheets of aluminum also are required in various applications, among them solar cell manufacture, production of microelectronic components, and solar furnace mirror surfaces. Extrusion, rolling, and direct casting were considered and rejected on grounds of lack of versatility and complexity. Vapor deposition, currently used in industry to apply coatings to surfaces and to prepare thin sheets of aluminum and other substances, was tentatively selected both because of its tremendous versatility (any curved surface may be coated) and because it is state-of-the-art technology. The major problems with the process in terrestrial applications are maintenance of the vacuum and high energy consumption, neither of which are factors on the lunar surface or in an orbital environment.

Plaster molds to be surfaced are passed to a laser honing station where they are finished to any desired accuracy, after which they move to the vapor deposition station and are coated with appropriate metals or nonmetals to the requisite thickness. The process is expected to proceed much as described by Miller and Smith (1979). The plaster mold is then removed and recycled, and the fabricated aluminum sheet is passed on to the electronic fabrication system or is sliced into wires by a fine cutting laser (Miller and Smith, 1979).

Mass throughput rates for this system appear adequate. Assuming that 10^4 m^2 of solar cells are needed for the original seed (Freitas, 1980) and that the casting bay is about 1 m^2 in area, then for T = 1 year the required deposition rate to produce 0.3 mm thick aluminum sheet is r_d = (10^4 m^2 solar cells/year)(3×10^{-4} m thick/sheet) (1 sheet/m^2)(1 year/5.23×10^5 min)(10^6 μm/m) = 5.7 μm/mm. State-of-the-art deposition rates attained for aluminum commercially are about 50 μm/min (Miller and Smith, 1979), nearly an order of magnitude higher than required. (The above throughput rate would also be equivalent to 1 m/sec of 0.3 mm aluminum wire production if cutting and wrapping can keep pace with deposition.) Cycling time is about 52 min/sheet. Following Johnson and Holbrow (1977), a heat of vaporization of 10^7 J/kg for 10^4 solar cells each made of 0.3 mm Al of density 3000 kg/m^3 requires a continuous power draw of only 2.9 kW, which can be supplied by a small solar collector mirror 2 m diam.

A small number of LMF parts are expected to be made of cast basalt – fused as-found lunar soil perhaps with fluxing agent additives. Most parts will probably be aluminum because Al is an easily worked metal with high strength, low density (hence supporting structures need not be large), and relatively low melting point (hence is easily cast). The major advantages of basalt are its easy availability, its tolerance of machining, good compressive strength, and high density in some uses. Anticipated applications include machine support bases, furnace support walls, robot manipulator tools (to avoid vacuum welding), and other special parts where weight is not a problem. Because

plaster fuses at 1720 K – very near the melting point of basalt – and loses its water of crystallization around 475 K, it cannot be used to make basalt castings. Iron molds cast from refractory templates are required; they may be reused or recycled as necessary.

Another principal application for basalt is as an insulating fiber. Spun basalt threads can be used to wrap electrical conductors to provide insulation, woven to produce "mineral fabrics" as filler to strengthen cements, shock-absorbing resilient packing material, filters and strainers for materials processing, or as thermal insulation or to prevent cold welding of metals (Green, unpublished Summer Study document, 1980). The technology for producing spun basalt products (Kopecky and Voldan, 1965; Subramanian and Kuang-Huah, 1979), basalt wool, and drawn basalt fibers (Subramanian et al., 1975) is well established commercially and customarily involves extrusion or simple mechanical pulling from a melt (see sec. 4.2.2).

Ho and Sobon (1979) have suggested a design for a fiberglass production plant for the lunar surface using a solar furnace and materials obtained from lunar soil (anorthite, silica, alumina, magnesia, and lime). The entire production facility has a mass of 111 metric tons and a power consumption of 1.88 MW, and produces 9100 metric tons of spun fiberglass per year. Assuming linear scaling, the production for the replicating LMF of even as much as 10 tons of fiberglass thread would require a production plant of mass 122 kg and a power consumption of 2.1 kW (a 2-m solar collector dish).

A small number of LMF parts will also be made of iron (from refractory molds) and refractory cements (carved directly from ceramic clay by the casting robot) in order to take advantage of the special properties of these substances. The total mass of such items is expected to be relatively low. Used refractory molds may be fed to the ball mill and recycled if necessary.

5F.4 Laser Machining and Finishing

The plaster casting parts manufacturing technique was chosen in part because of its ability to produce ready to use "as-cast" components. Thus, it is expected that the majority of parts will require little reworking, machining, or finishing. A small fraction, perhaps 10%, of all lunar SRS parts may require more extensive machining. A laser machining system was selected for this function in the LMF. The characteristic circumference of the typical part is $3.14(0.1/3000)^{1/3}$ or about 10 cm. If surface articulations cause an increase by a factor of ten in the total average path length that must be machined, then the mean operating speed of the laser system must be (10^6 parts/year) (10% machinables)(0.1 m/part)(10 m path/m circum.) (1 year/8722 hr) = 11.5 m/hr. Table 5.16 compares the performances of several different types of lasers, and table 5.17 gives specific performance parameters for high-power gas lasers used in industry for welding (butt, lap, corner, and edge) and for cutting. Inspection of these values suggests that a 5-10-kW continuous-wave (CW) carbon dioxide laser should be able to weld and cut "typical parts" with characteristic dimensions up to 3 cm at the required throughput rate.

TABLE 5.16.– CHARACTERISTICS AND PERFORMANCE OF VARIOUS LASERS COMMONLY USED FOR WELDING (Acharekar, 1974)

Laser	Operation	Pulse length, msec	Pulse energy, J	Peak power, W	Maximum weld thickness[a]		Speed of welding	
					in.	mm	in./min	mm/sec
Ruby	Pulsed	3-10	20-50	1-5k	0.005 to .020	0.13 to .50	3.0	1.2
Nd:glass	Pulsed	3-10	20-50	1-5k	.005 to .020	.13 to .50	1.5	0.63
Nd:yag	Pulsed	3-10	10-100	1-10k	.005 to .025	.13 to .60	5.0	2.1
CO_2	Pulsed	5-20	0.1-10	1-5k	.005	.13	3.0	1.2
Nd:yag	CW			1000	.150	3.81	30.0	12.7
CO_2	CW			1000	.025	.60	30.0	12.7
Gas dynamic	CW			20 k	.750	19.0	50.0	21.2

[a]Maximum thickness given here is for Type 304 stainless steel.

TABLE 5.17.– TYPICAL PERFORMANCE DATA FOR CO_2 WELDING/ CUTTING LASERS

Demonstration butt welds on tanker construction steels (Nagler, 1976)					
Thickness		Laser power, kW	Weld speed		Comment
in.	mm		in./min	mm/sec	
0.375	9.5	10.8	50	21.2	Single pass
0.375	9.5	10.8	45	19.0	Single pass
0.5	12.7	12.0	27	11.4	Single pass
0.5	12.7	12.0	30	12.7	Single pass
0.625	15.9	12.0	24	10.2	Single pass
0.75	19.1	12.0	45	19.0	Dual pass
1.0	25.4	12.0	30	12.7	Dual pass
1.0	25.4	12.0	30	12.7	Dual pass
1.125	28.6	12.8	27	11.4	Dual pass
0.375–0.5	9.5–12.7	11.0	90	38.1	Tee joint
0.375–0.5	9.5–12.7	7.5	65	27.5	Tee joint
1.0	25.4	12.0	27	11.4	Dual pass[a]
1.0	25.4	12.0	25	10.6	Dual pass[a]

[a]0.001-in. (0.03 mm) aluminum foil preplaced at weld interface.

Material	Thickness		Weld type	Laser power, kW	Weld speed		Number of pieces
	in.	mm			in./min	mm/sec	
HY-130 steel	0.25	6.4	Butt	5.5	50	21.2	3
HY-180 steel	0.062	1.6	Butt	5.5	160	67.7	2
HY-180 steel	0.062	1.6	Lap	5.5	140	59.2	1

Typical cutting and drilling rates for a 1-kW CO_2 laser (Yankee, 1979)				
Metal thickness, in.	Cutting rates (in./min)			
	Stainless steel	Aluminum	Galvanized steel	Titanium
0.020	750	800	250	---
0.032	650	---	---	---
0.040	550	350	100	250
0.062	450	200	50	150
0.080	325	100	---	100
0.125	200	---	---	---

Drilling rate: Less than 1 msec is required to drill each of these holes:		
Material	Thickness	Hole diameter
Tungsten	0.020 in. (0.51 mm)	0.020 in. (0.51 mm)
Ceramic	0.101 in. (2.57 mm)	0.050 in. (1.27 mm)
Brass	0.010 in. (0.25 mm)	0.250 in. (6.35 mm)

Laser cutting speeds typically are as much as 30 times faster than friction sawing (Yankee, 1979). Cutting accuracy is about 0.01 mm/cm under closely controlled conditions. All metals – including high-strength, exotic, and refractory alloys such as Inconel and titanium, as well as aluminum, stainless steel, and brass – and nonmetals such as diamond, ceramics, and plastics may be vaporized by laser beams. Hence, parts of these materials may be easily machined. Burr-free laser holes may be drilled as small as 10-100 μm. Lasers can also be used for pattern cutting, gyro balancing, insulation stripping, surface hardening, trimming, photoetching, measurement of range and size to 1 μm accuracy or better, scribing 5-10 μm lines on microelectronic wafers, flaw detection, marking or engraving parts, and impurity removal (e.g., carbon streaks in diamond). Laser beam machining is "especially adaptable and principally used for relatively small materials processing applications such as cutting, trimming, scribing, piercing, drilling, or other delicate material removal operations similar to milling or shaping" (Yankee, 1979).

Dunning (unpublished Summer Study document, 1980) has suggested a variety of space and lunar applications for laser machining, including flash trimming of cast basalt parts; engraving bar codes on parts to enable quick and accurate recognition by robot vision systems; drilling holes in workpieces an inch thick or less; internal welding of cast basalt joints, pipe, and structural members; impurity removal from lunar-produced semiconductor chips; cutting operations on gossamer structures (Brereton, 1979) in orbit; and case hardening of cast basalt or metal parts. Dunning has also suggested two potential major problems associated with the use of lasers in the context of a self-replicating, growing lunar manufacturing facility: (1) the need for gas jets, and (2) the requirements of closure.

In normal industrial usage, vaporized workpiece material is carried away by a gas jet, usually oxygen (Yankee, 1979). The gas serves three functions: (1) to oxidize the hot working surface, decreasing reflectivity, (2) to form a molten oxide (i.e., the metal "burns") which releases a large fraction of the useful cutting energy, and (3) to remove slag and hot plasma from the path of the beam. There is no problem maintaining a moderate-pressure O_2 atmosphere around the laser work area, as the beam penetrates air easily. In this case the usual gas jet can still be used. Or, the laser could be placed outside the pressurized working area, shooting its beam through a transparent window. If pressurization must be avoided, laser machining can be done entirely in vacuum and the ionized plasma wastes removed by a magnetic coil following the cut or weld like an ion "vacuum cleaner." However, it is estimated that up to 80% of the laser cutting energy comes from the exothermic oxidation reaction, so in this latter case laser energies would have to be on the order of five times the value for the equivalent O_2-atmosphere machining.

The problem of closure is even more critical in a replicating autonomous remote factory. The materials closure problem is solved in large measure by resorting to CO_2 gas laser technology. This gas is available in limited quantities on the Moon, whereas materials for solid state lasers such as yttrium, ruby, garnet or neodymium are generally very rare (although Dunning has suggested that spinel, which is plentiful on the Moon, might be substituted for garnet). Quantitative materials closure may be argued as follows. A typical CO_2 laser uses three gases for high-power operation – carbon dioxide to lase, nitrogen to sustain the reaction, and helium for cooling because of its excellent heat conducting properties. Since oxygen is plentiful, the three limiting elements are C, N, and He. From appendix 5E, the LMF in one year can produce 400 kg C, 400 kg N_2, and about 40 kg inert gases (at least 90% of which is He). This is sufficient to make 747 m^3 (33,300 moles) of CO_2, 320 m^3 (14,300 moles) of N_2, and 224 m^3 (10,000 moles) of He, at STP. Even if the laser machining device requires several hundred moles of these gases (a few thousand liters at STP), still only a few percent of available LMF stocks of these elements need be diverted for this purpose, a negligible resource drain.

The problems of parts and assembly closure cannot be answered satisfactorily at the present time. However, it is often asserted that machining the laser end mirrors to high accuracy may be a major roadblock to automated manufacture of lasing devices. Nazemetz (personal communication, 1980) has pointed out that a laser is accurate enough to surface a rough-hewn mirror to the accuracy required for its own construction. In a pinch, concave mirrors could be hewn from solid metal or basalt blanks simply by sweeping the laser beam radially across the disks, applying higher power nearer the center so more material volatizes there, thus creating a perfect spherical or parabolic surface gradient. There appear to be no major unresolvable difficulties associated with the use of lasers in an autonomous lunar manufacturing facility.

After parts leave the laser machining station they may require some slight further treatment such as annealing or coating to prevent cold weld, though this latter function may be unnecessary if laser welding takes place in an oxygen atmosphere (a thin layer of metal oxide prevents the vacuum-welding effect). Once fabrication is completed each part may have one of three possible destinations: (1) assembly sector, where the part is given to a mobile robot for transport to wherever it is needed, (2) parts warehouse (which serves as a buffer supply of extra parts in the event of supply slowdowns or interruptions), where the part is taken to storage by a mobile robot, or (3) fabrication sector, when more fabrication must be performed upon an already manufactured "part" (e.g., solar cell aluminum sheets), where a mobile robot carries the part to wherever

it is needed in the fabrication sector. A general flowchart of the entire automated parts fabrication process appears in figure 5.17.

5F.5 Parts Fabrication: State-of-the-Art

In the operation of any general-purpose fabrication machine (mill, lathe, laser machining system, casting robot, there are seven distinct functions which must be performed either manually or automatically, according to Cook (1975):

(1) Move the proper workpiece to the machine,
(2) Load the workpiece onto the machine and affix it rigidly and accurately,
(3) Select the proper tool and insert it into the machine,
(4) Establish and set machine operating speeds and other conditions of operation,
(5) Control machine motion, enabling the tool to execute the desired function,
(6) Sequence different tools, conditions, and motions until all operations possible on that machine are complete, and
(7) Unload the part from the machine.

Traditionally all seven operations were performed by the human operator. The development of numerical-control (N/C) machining relieved human operators of the need to manually perform step (5), and automatic tool-changing systems supplanted step (3). Although most modern computer-controlled machining systems have "a finite number of tool-storage locations – 24, 48, or 60 tools, for example – the number that could be built into a system runs into the thousands" (Gettleman, 1979). If the seed is comprised of about 1000 different kinds of parts, each requiring a template pattern for the casting robot, Gettleman's estimate for N/C machine tooling makes plausible the satisfaction of this requirement by extensions of current technology. Adaptive control of N/C machine tools, with sensors that measure workpiece and tool dimensions, tool application forces, vibration and sound, temperatures, and feed rates to optimize production have already been developed (Nitzan and Rosen, 1976) but will require further improvements to achieve the kind of generalized capability required for a lunar SRS.

The next logical developmental step is the design of a completely computer-managed integrated parts manufacturing system. Cook (1975) describes such a system developed and built by Sunstrand Corporation. One version in operation at the Ingersoll-Rand Company is used primarily for fabricating hoists and winches, while another at the Caterpillar Tractor Company is used for making heavy transmission casing parts (Barash, 1976). As of 1975 there were about ten similar systems in operation in the U.S., Japan, Germany, and the U.S.S.R. (Barash, 1975).

The Ingersoll-Rand system consists of six N/C tools – two 5-axis milling machines, two 4-axis milling machines, and two 4-axis drills – arranged around a looped transfer system as shown in figure 5.42. Machining operations include milling, turning, boring, tapping, and drilling, all under the control of an IBM 360/30 central computer. At any given time about 200 tools are in automatic tool-changing carousels, available for selection by the computer, although about 500 are generally available in the system. The computer can simultaneously direct the fabrication of as many as 16 different kinds of parts of totally different design which are either being machined, waiting in queue to be machined, or are in the transfer loop. The entire system is capable of manufacturing about 500 completely different parts. During each 12-hr shift the system is run by three human operators and one supervisor. It is calculated that to achieve the same output using manual labor would require about 30 machines and 30 operators. Finally, the circular pallets used to present parts to each control station have maximum dimensions which fit inside a 1-m cube, exactly the scale discussed earlier in connection with the casting robot.

Another major advance is the variable-mission manufacturing system developed by Cincinnati Milacron Inc. This system not only has the general character of computer-managed parts manufacture seen in other systems but also provides for the processing of low-volume parts at higher rates than those which can be achieved with more conventional N/C machines. For instance, an ingenious five-axis "manufacturing center" automatically changes clusters of tools mounted on a single head so that a number of operations can be performed simultaneously. By means of a novel scheme of handling workpieces from above, the Cincinnati Milacron system provides efficient management of coolants and chips, together with easy access for inspection and servicing (Cook, 1975).

The Japanese have been most aggressive in pursuing the "total automation" concept. During 1973 through 1976 their Ministry of International Trade and Industry (MITI) supported a survey and design study entitled "Methodology for Unmanned Manufacturing" (MUM) which forecast some rather ambitious goals. The MUM factory was to be operated by a 10-man crew, 24 hr/day, and replace a conventional factory of about 750 workers. The factory will be capable of turning out about 2000 different parts at the rate of 30 different parts (in batches of about 1-25) per day, which will be inspected and assembled to produce about 50 different complex machine components such as spindle and turret heads, gear boxes, etc. Machining cells, based on the principle of group technology, will be controlled by a hierarchy of minicomputers and microcomputers, and will receive workpieces via an automated transfer system. Each machine cell will be equipped with inspection and diagnostic systems to monitor such useful

parameters as tool wear, product quality, and the conditions of machine operation. Assembly cells, much like the machining cells, will be equipped with multiple manipulators fashioned after present industrial robots, together with an automated transfer system for movement of assemblies (Nitzan and Rosen, 1976). One ultimate program goal, explicitly stated, was to design a system "capable of self-diagnosis and self-reproduction ... [and] capable of expansion" (Honda, 1974).

Following this initial study, MITI in 1977 initiated a 7-year national R&D program at a funding level of 12 billion yen (about $57 million) to develop, establish, and promote technologies necessary for the design and operation of a "flexible manufacturing system complex," a prototype

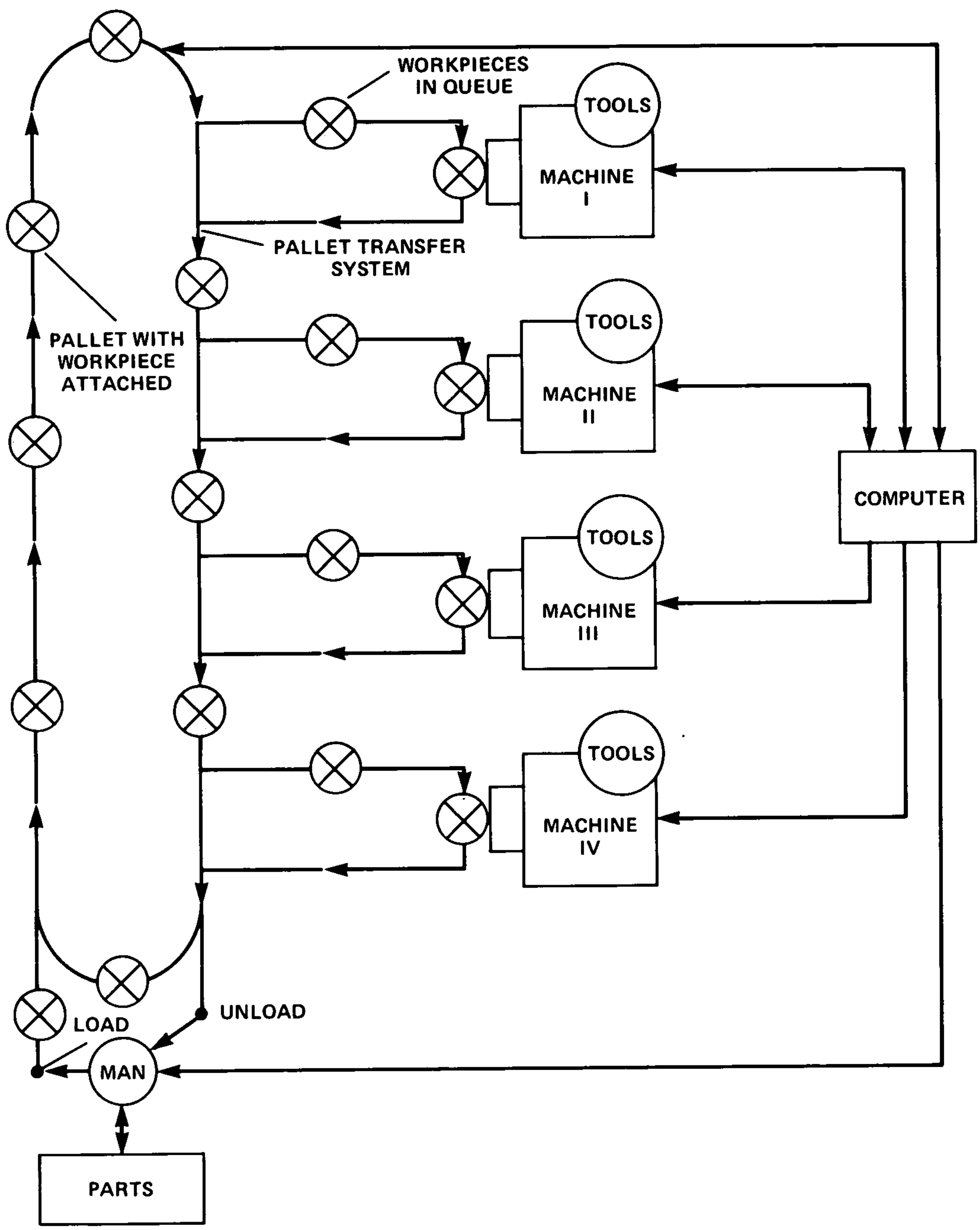


Figure 5.42.– Computer-managed parts manufacturing.

"unmanned" factory to be built sometime in the mid-1980s (Ohmi et al., 1978). The technologies currently receiving emphasis include:

- Optimum design and integrated control of manufacturing systems including blank fabrication, machining and assembly,
- Flexible machining for mechanical parts and components,
- Enlargement of the flexibility of blank fabrication,
- Enlargement of the applicable area of automatic assembly and automatic transfer,
- Application of high-power (20 kW) CO_2 lasers to metalworking,
- Automatic diagnosis of manufacturing facilities to detect malfunctions, and
- Planning and production management to optimize system operation.

MUM presently is being pursued vigorously by three government research institutes and 20 private companies, and is being managed by the Agency of Industrial Science and Technology of MITI (Honda et al., 1979).

The original forecast was that MUM technology would go into operation sometime during the 1980s. At a conference in Tokyo in September of last year, Fujitsu FANUC Ltd., a leading international manufacturer of numerical control (N/C) machining equipment, announced its plans to open a historic robot-making factory near Lake Yamanaka in Yamanashi Prefecture in late November. At the plant, then still under construction, industrial robots controlled by minicomputers would produce other industrial robots without major human intervention save minor machine operation and administrative tasks. The plant is the first "unmanned" factory in the world machinery industry, producing robots and other equipment worth about $70 million in the first year of operation with only 100 supervisory personnel. In 5 years the plant is expected to expand, perhaps with some of the robots it itself manufactures, to a $300 million annual output with a workforce of only 200 people, less than a tenth the number required in ordinary machine factories of equivalent output. The mainstay products are to be various kinds of industrial robots and electronic machines. A spokesman said that FANUC's fully automated system is suitable not only for mass production of a single product line but also for limited production of divergent products (IAF Conference, 1980).

An automated plant in which robots make robots is a giant first step toward the goal of a practical self-reproducing machine system. When a factory such as the FANUC plant can make all of the machines and components of which it itself is comprised, its output can be specified to be itself and thus it can self-replicate. It appears likely that the automation technology required for LMF fabrication and assembly operations could become available within the next 10-20 years, given adequate funding and manpower support targeted specifically to the development of such a system.

5F.6 Automation of Specific LMF Systems

It is useful at this point to consider the automation potential of specific LMF systems. Most critical are the casting robot and the laser machining system, but several other subsystems will also require automation.

Casting Robot Automation

There are two potential precursor technologies to the general-purpose casting robot described in section 5F.3, in addition to established robotics devices such as the Unimate 4000 that produces lost wax ceramic molds for use in investment casting (Moegling, 1980). One of these lines of development has been in the field of precision machining, the other in the area of art and sculpturing.

Engraving and tracer milling are well established machining techniques. These machines use high-speed spindles mounted on pantograph mechanisms guided by master patterns which permit the cutting tools to be guided from an original which may be larger or smaller than the workpiece. The original pattern may be wood, plastic, or metal; the operator follows it with a guide and the machine faithfully reproduces each motion – but enlarges or reduces it as desired (Ansley, 1968).

Modern machines work in three dimensions and can be used for very intricate carving in metal from arbitrary solid originals. A contour milling machine developed by Gorton Machine Corporation uses numerical control to replace entirely the master pattern and the human operator (Ansley, 1968). A skilled technician can preprogram the complete machining cycle for any given part. The Lockheed CAD/CAM system (see below) permits still more sophisticated computerized design and parts fabrication. It seems but a few conceptually simple steps from this level of technology to that required for a "universal" contour-carving device like the casting robot. Such a system will require a vision system, excellent tactile sensing, an automatic tool-changing and pattern-changing capability, and development of an automatic feedstock handling system for metals, gases, and refractories.

Another possible precursor technology to the casting robot may be found in the area of artistic sculpting, otherwise known as "three-dimensional portraiture." An excellent summary of 19th-century attempts to construct machines able to automatically size and shape a human head for personalized sculptures has been written by Bogart (1979). In the last 10 years two very different descendants of the 19th-century efforts to produce sculpted likenesses (thus bypassing the creative artist) have been spawned. The

first of these is modern holography techniques, which permit the generation of 3-D images using laser beams and, more recently, white light sources.

The second technology, often called "solid photography," requires that the human model pose in front of eight cameras shooting simultaneously from different angles. Linear patterns of light are projected onto the subject's face and all three-dimensional information is coded by the cameras. The coded films are then read by an optical scanner which converts the code into digital information which is processed by a computer to produce an accurate surface map of the person or object. This map is then translated into a series of cutting instructions which are passed to two cutting instruments.

In the system operated by Dynell Electronics Corporation of Melville, New York, instructions are first passed to a "coarse replicator" which rough-hews the shape of the human head in paralene wax (high melting point) in 90° sections. After about 30 min, the rudimentary carving is completed and is passed to the "fine-cut replicator" which is also computer-controlled. This time, instead of a single rotating bit, the tooling consists of 20 rotating blades that finish the work to a very high accuracy in about 40 min of work. Human hands are used only for touch-up of very fine details or for imparting skin-like smoothnesses; witnesses to the procedure are impressed with the results – excellent representations of eyebrows, locks of hair, creases, even moles (Field, 1977). Clearly, the Dynell automated sculpting system is not too distant from the casting robot, conceptually or technologically. If treated as a serious item for further development, it is likely that casting robot technology could be ready in a decade or less starting from the current state-of-the-art.

Laser Machining System Automation

Nonlaser spot welding has been a standard automated industrial technique for many years. Welding robots at Chrysler's Hamtramck assembly plant put uniform spot welds on parts assemblies with positional accuracy exceeding 1.3 mm. Typical operation includes a sequence of 24 welds on four automobile assemblies at once (Tanner, 1979). One of the largest and most fully automated welding lines in the world operates at Volvo's Torslanda plant in Gothenburg, Sweden. The new welding line consists of 27 Unimate robots which replace 67 workers with 7. The installation is fully automated, including loading and unloading stations, intermediate assembly of all automobile body parts, lining, and clamping preparatory to welding. The line does a total of 754 spot welds per assembly, and each Unimate is directed by 2-8K programmable controller computers (Mullins, 1977). Kawasaki Unimate robots have been applied to arc welding of motorcycle frames and automobile rear axle housings (Seko and Toda, 1974). Accuracy in arc welding is more difficult to achieve than in spot welding, but apparently much progress has been made in this area.

Nonlaser machining is also highly automated. The generalized machining center can perform a number of functions in typical operation including milling, drilling, boring, facing, spotting, counterboring, threading, and tapping, all in a single workpiece setup and on many different surfaces of the workpiece (Gettleman, 1979). A numerical-control machine operated by the Giddings and Lewis Machine Tool Company has an automatic tool changer with 40 tools. It machines all sides of a workpiece with one setup. (Setup time is usually 50-90% of total machining time, and a typical part might normally require a dozen setups or more, so this is a substantial savings.) A machined block requiring 174 separate operations can be completed automatically in 43 min; the former method required 4 machines with 3 operators and took 96 min to finish the part. Piggott (personal communication, 1980) estimates that a "typical part" weighing 0.1 kg will require about 20 machining operations. If 10% of all LMF parts must be closely machined after casting, a *single* Giddings N/C robot could perform all 2,000,000 necessary machining operations in just 0.94 year. Since several such robots could be available in the early LMF, this item is noncritical.

A more sophisticated methodology (Luke, 1972) is used in the Lockheed CAD/CAM system. In this system, the user designs a part of arbitrary shape in three dimensions on an interactive computer-driven TV console. This description is processed to yield a series of machine operations and is then passed to a set of 40 sophisticated N/C machines which make the part "from scratch" out of feedstock supplied at one end. On the average, parts are machined correctly five out of every six tries.

If all LMF parts had already been designed and placed in memory, a shop in space using the Lockheed system could manufacture each of the 1000 different SRS parts. With the addition of pattern recognition software capable of recognizing any part presented to a camera eye, in any physical condition (e.g., rotated, broken, partly melted, partly obscured) (Perkins, 1977), and a simple goal-setting command hierarchy, the Lockheed system might be able to recognize and repair damaged parts presented to it randomly.

The purpose of describing the above nonlaser welding and machining systems is to suggest that laser machining should be equally automatable because the laser may be viewed as another modality for delivering heat or cutting action to a workpiece. Any nonlaser automated welding/machining technology in principle may be modified to accept a laser as its active machining element.

Lasers already have found many automated applications in industry. Computer-driven lasers presently perform automated wire-to-terminal welding on relay plates for electronic switching circuits (Bolin, 1976). There are

automated laser welding lines for manufacturing metal-enclosed gasprotected contacts for telephone switchgear (Schwartz, 1979). A computer-controlled laser welding system at Ford Motor Company allows welding parameters for a number of different automobile underbody designs to be stored in the central memory and retrieved as required for seam welding body-pans (Chang, personal communication, 1978). In the garment industry, the cutting of patterns from single-ply or multilayer stacks of fabrics is easily fully automated and rates of up to 61 m/min have been achieved (Luke, 1972; Yankee, 1979). Flash trimming of carbon resistors has been successfully automated. Automated marking and engraving (with alphanumeric characters) is another application of computer-guided lasers (Yankee, 1979). Numerous other laser applications have already been put into operation (see sec. 5F.4) but are not yet automated. Lasers for many automobile body assembly tasks are impractical today because the component metal pieces to be welded, which are stamped metal sheet, are too inaccurate to permit a close enough fit for laser welding to be feasible – though, according to Schwartz (1979), "this situation may change gradually in the future."

Lunar seed lasers should be able to operate at many different power settings, preferably spanning a broad continuum. Precision machining of liquid- and air-tight valves, laser mirror surfaces, and various other small intricate parts will demand the closest scrutiny of the rate at which energy is delivered to the workpiece. Lasers may also be used for super-accurate ranging and sizing measurements, which require an ultralow power capability as well as sophisticated optics, timing, and data processing systems. Automation of the LMF Laser Machining System will require close computer/mechanical control to perform each of the seven basic machining steps described earlier in section 5F.5.

Some consideration should also be given to the architecture of beam delivery to the workpiece. Laser power may be transmitted directly, in which case the entire laser assembly must be swiveled as various operations are performed. One alternative is to use a system of lightweight movable mirrors to angle laser energy in the desired direction to impact the workpiece. Reflectivities up to 0.86 for aluminum on glass would give an absorbed power density of 14 to 140 W/cm^2 for a 1-10% efficient 10 kW laser beam with a 1 cm^2 cross section. This heating may be reduced by at least an order of magnitude by "jiggling" the mirrors along their plane to spread the beam impact spot over a wider area while maintaining precise directional control. Another possible solution is to locate a high power laser in some central location and convey the beam to its destination via large fiber-optic light pipes. There are possible materials closure problems with fiber-optics, and absorbed energy may damage or destroy the glass, but this alternative offers many interesting opportunities and cannot be logically ruled out.

The team recognizes that lasers may not be the optimum technology for an autonomous replicating lunar facility. Their inclusion in the present design is intended as a heuristic device to illustrate, not unequivocally select, a particular option. For example, industrial experts in manufacturing technologies are split over whether lasers or electron beams are generally superior or more versatile, e.g., Schwartz (1979) favors lasers and Yankee (1979) favors e-beams. The MIT study group selected electron-beam cutting over lasers because "lasers are less efficient and require more maintenance and repair than EB guns" (Miller and Smith, 1979), a conclusion not adequately documented in their final report.

Nor is it absolutely clear that conventional machine tools such as mills, lathes, or drills are unsuitable for use in space. The problem most often cited in this context is that the tool bit and workpiece may vacuum weld during machining. However, cold welding is known to occur only between identical metals or between those with very similar crystallographic characteristics (such as aluminum and magnesium). Steel, for instance, will not vacuum weld to aluminum. Neither will any metal part cold weld to cast basalt.

Further, ceramic cutting tools have recently been developed which have increased the cutting speeds of mills and lathes dramatically. When tungsten carbides were introduced in 1929, cutting speeds quadrupled to 100 to 200 m/min. Since the 1950s, ceramic and other cemented oxide (alumina) and refractory tool materials such as nitrides and borides have been successfully employed in achieving cutting rates of 300 m/min and higher (Ansley, 1968). Ceramic tools will not cold weld to anything.

A more critical problem would seem to be the seizing of internal machine components, rather than vacuum welding between tool and workpiece. This difficulty could perhaps be surmounted by bathing enclosed machinery in lubricants, a light oxygen atmosphere trapped by airtight seals, or by using basalts or ceramics to construct or merely protectively coat internal machine moving parts.

Automation of Other Systems

The remaining subsystems within the parts fabrication sector must also be automated for full LMF autonomous operation. These subsystems include:

(1) Kilns and metallurgical furnaces: The extraterrestrial fiberglass production system using solar energy, designed by Ho and Sobon (1979), is designed to be automated. This system includes melting and drawing operations. According to the authors, "the systems will be automated, but minimum manpower will be required for maintenance. For the lunar plant, maintenance will be required at the beginning of each lunar day to begin the drawing process."

(2) Basalt threads: The system of Ho and Sobon will be automated. Also, a series of eleven specific steps which a

manufacturing robot such as a Unimate must perform in order to completely automate the thread-drawing procedure is given in appendix 4D.

(3) Wire wrapping: An automatic insulation wire-wrapping machine has been described in some detail by Miller and Smith (1979).

(4) Sheet metal and cutting operations: Miller and Smith (1979) discuss in some detail aluminum ribbon and sheet operations. Vacuum vapor deposition as a fabrication technique is also described in Johnson and Holbrow (1977). These will be at least partially automated.

(5) Refractory and cement production: Ansley (1968) has described a concrete batching plant equipped with electronic controls permitting the selection of some 1500 different formulas and which give twice the output of manually operated plants. Batches are prepared by inserting a punched card into a reader to specify the formula to be used, and the system does the rest automatically if adequate materials have been supplied.

(6) Ball mills and magnetic purification: These are standard automated technologies, assumed available in space processing models provided by O'Neill (1976), Phinney et al. (1977), and others.

5F.7 Sector Mass and Power Estimates

In lieu of a complicated breakdown of fabricator sector component subsystems with detailed analysis of each, table 5.18 illustrates a more practical approach. This information was assembled from various sources and gives typical masses and power requirements for parts fabrication facilities in previous studies.

The nominal annual output of the original lunar seed is 100 tons/year. Using the most extreme machine productivity values given in table 5.18, fabrication sector mass may range from 137 kg up to 20,400 kg. A similar comparison with the power requirements values gives a range of 0.3-345 kW for sector energy consumption. The upper ranges of these estimates are probably most appropriate in the replicating lunar factory application.

5F.8 Information and Control Estimates

Even in the absence of a detailed analysis of the necessary control operations, it is obvious that the complete description of all parts will dominate computer memory requirements. Since each typical part has a characteristic

TABLE 5.18.– COMPARISON OF FABRICATION PLANT MASSES AND POWER REQUIREMENTS FROM PREVIOUS RELATED STUDIES

Source	Plant mass, kg/kg sec output	Power requirement, W/kg plant
Johnson and Holbrow (1977) annealing, trimming, pressing plate silica glass plant	8.3×10^5	2.2
Ho and Sobon (1979) fiberglass threads/rods plant	3.8×10^5	16.9
Johnson and Holbrow (1977) bulk processing and heavy industry estimate for human workers	4.3×10^4	2
O'Neill, Driggers, and O'Leary (1980) estimated range for machine shop bulk fabrication systems	3.6×10^5 - 3.6×10^6	---
Miller and Smith (1979) MIT Study on Space Manufacturing Facility	6.4×10^6	12
Vajk et al. (1979)	3.6×10^5	---

surface area of 10^{-3} m^2, then if the surface of each is mapped to 1 mm^2 resolution per pixel, each part will require 1000 pixels for complete coverage. Each pixel must identify three position coordinates, materials used, machining operations to be performed, etc. If 100 bits/pixel is adequate, then roughly 10^5 bits/part are required in memory for a total of 10^{11} bits of storage for all 1,000,000 parts in the original lunar seed. This crude estimate is intended as a combined total for description and operation of the system.

Subsystem control hardware is likely to use vastly less computer capacity than this. The entire Sundstrand integrated parts manufacturing line is managed by an IBM 360/30 central computer with microcomputers driving each robot station. While some tricks might be employed to reduced redundancy (such as "chunking" large similar areas), more convoluted surfaces will require extra description. It is likely that the main driver will be the requirements for parts description.

5F.9 References

Acharekar, M. A.: Lasers Finding Growing Use in Industrial Welding. Welding Engineering, December 1974, pp. 9-11.

Ansley, Arthur C.: Manufacturing Methods and Processes. Chilton Book Company, Philadelphia, 1968. Revised and enlarged edition.

Barash, M. M.: Optional Planning of Computerized Manufacturing Systems. Proc. 3rd NSF/RANN Grantee's Conference on Production Research and Industrial Automation, Cleveland, October 1975.

Barash, M. M.: Integrated Machining Systems with Workpiece Handling. In Proc. 6th Intl. Symp. Industrial Robots and Industrial Robot Technology, University of Nottingham, March 1976, International Fluidics Services, Kempston, 1976, pp. H4/29-36.

Bogart, Michele: In Art the Ends Don't Always Justify Means. Smithsonian, vol. 10, June 1979, pp. 104-108, 110-111.

Bolin, S. R.: Bright Spot for Pulsed Lasers. Welding Design and Fabrication, August 1976, pp. 74-77.

Brereton, Roy G., ed.: Discussion Meeting on Gossamer Spacecraft (Ultralightweight Spacecraft): Final Report, JPL Publication 80-26, Jet Propulsion Laboratory, Pasadena, California, 15 May 1980. 174 pp. NASA CR-163275.

Campbell, Ivor E.; and Sherwood, Edwin M., eds.: High-Temperature Materials and Technology. Wiley, New York, 1967.

Cook, Nathan H.: Computer-Managed Parts Manufacture. Scientific American, vol. 232, February 1975, pp. 22-29.

Field, Roger: Computerized Cameras, Knives Sculpt Quickly. Science Digest, vol. 81, April 1977, pp. 77-78.

Freitas, Robert A., Jr.: Baseline Reference Design for a Growing, Self-Replicable Lunar Manufacturing Facility. Summer Study Document, 8 July 1980. 56 pp.

Gettleman, K. M.: Fundamentals of NC/CAM. Modern Machine Shop 1979 NC/CAM Guidebook, 1979.

Gitzen, W. H.: Alumina Ceramics. U.S. Air Force Rept., AFML-TR-66-1621, 1966.

Grant, T. J.: The Need for On-Board Repair. In Project Daedalus – The Final Report on the BIS Starship Study, Anthony R. Martin, ed., British Interplanetary Society, 1978 (Journal of the British Interplanetary Society, Supplement, 1978), pp. S172-S179.

Ho, Darwin; and Sobon, Leon E.: Extraterrestrial Fiberglass Production Using Solar Energy. In Space Resources and Space Settlements, John Billingham, William Gilbreath, Brian O'Leary, and B. Gossett, eds. NASA SP-428, 1979, pp. 225-232.

Honda, Fujio, ed.: Methodology for Unmanned Metal Working Factory. Project Committee of Unmanned Manufacturing System Design, Bulletin of Mechanical Engineering Laboratory No. 13, Tokyo, 1974.

Honda, F.; Kimura, M.; Ozaki, S.; Tamagori, K.; Ohmi, T.; and Yoshikawa, H.: Flexible Manufacturing System Complex Provided with Laser – A National R&D Program of Japan. Proc. IFAC Symposium on Information-Control Problems in Manufacturing Technology, September 1979, pp. 7-11.

IAF Conference: Robots To Produce Robots At Fujitsu. From an English-language newspaper handout at the Conference, Tokyo, 21-28 September 1980. See also The Washington Post, 13 October 1980, p. A-38.

Johnson, P. D.: Behavior of Refractory Oxides and Metals, Alone and in Combination, *in Vacuo* at High Temperatures. J. Am. Ceram. Soc., vol. 33, no. 5, 1950, pp. 168-171.

Johnson, Richard D.; and Holbrow, Charles, eds.: Space Settlements: A Design Study, NASA SP-413, 1977.

Kaiser Aluminum Company: New Kaiser Super Refractories, May 1962.

Kopecky, Lubomir; and Voland, Jan: The Cast Basalt Industry. In Geological Problems in Lunar Research, Harold Whipple, ed., Annals of the New York Academy of Science, vol. 123, July 16, 1965, pp. 1086-1105.

Luke, Hugh D.: Automation for Productivity. John Wiley and Sons, New York, 1972.

Miller, Rene H.; and Smith, David B. S.: Extraterrestrial Processing and Manufacturing of Large Space Systems, vols. 1-3, NASA CR-161293, September 1979.

Moegling, Frank: Robot-Controlled Mold-Making System. Robotics Today, Fall 1980, pp. 30-31.

Mullins, Peter J.: Robot Welding of Car Bodies. Automotive Industries, 1 February 1977. Reprinted in Industrial Robots: Volume 2 – Applications, William R. Tanner, ed., Society of Manufacturing Engineers, Dearborn, Michigan, 1979, pp. 151-152.

Nagler, H.: Feasibility, Applicability, and Cost Effectiveness of LBW of Navy Ships, Structural Components and Assemblies. Contr. N00600-76-C-1370, vols. 1-2, 22 December 1976.

Nitzan, David; and Rosen, Charles A.: Programmable Industrial Automation. IEEE Trans. Comp., vol. C-25, December 1976, pp. 1259-1270.

Ohmi, T.; Kanai, M.; and Honda, F.: Research and Development of a Flexible Manufacturing System Complex Provided with Laser – A Japanese National Project. Paper presented at the Workshop on Flexible Manufacturing Systems, 11-14 September 1978, Peoria, Illinois.

O'Neill, Gerard K.: Engineering a Space Manufacturing Center. Astronautics and Aeronautics, vol. 14, October 1976, pp. 20-28, 36.

O'Neill, Gerard K.; Driggers, Gerald; and O'Leary, Brian: New Routes to Manufacturing in Space. Astronautics and Aeronautics, vol. 18, October 1980, pp. 46-51.

Perkins, W. A.: Model-Based Vision System for Scenes Containing Multiple Parts. General Motors Research Laboratories Publication, GMR-2386, June 1977.

Phinney, W. C.; Criswell, D. R.; Drexler, E.; and Garmirian, J.: Lunar Resources and Their Utilization. In Space-Based Manufacturing from Extraterrestrial Materials, Gerard K. O'Neill, ed., AIAA Progress in Astronautics and Aeronautics Series, vol. 57, AIAA, New York, 1977, pp. 97–123.

Robson, T. D.: High-Alumina Cements and Concretes. John Wiley and Sons, Inc., New York, 1962.

Schwartz, M. M.: Metals Joining Manual. McGraw-Hill Book Co., New York, 1979.

Seko, K.; and Toda, H.: Development and Application Report in the Arc Welding and Assembly Operation by the High-Performance Robot. In Proc. 4th Intl. Symp. Industrial Robots, Tokyo, Japan, November 1974, Japan Industrial Robot Assn., Tokyo, 1974, pp. 487-596.

Spotts, M. F.: Design Engineering Projects. Prentice-Hall, Inc., New Jersey, 1968.

Subramanian, R. V.; Austin, H. F.; Raff, R. A. V.; Sheldon, G.; Dailey, R. T.; and Wullenwaber, D.: Use of Basalt Rock for the Production of Mineral Fiber. Pacific Northwest Commission Annual Report, Contract #NR-3001, College of Engineering, Washington State University, Pullman, Washington, June 1975. 79 pp.

Subramanian, R. V.; and Kuang-Huah, Shu: Interfacial Bonding in Basalt Fiber-Polymer Composites. 34th Annual Technical Conference, Reinforced Plastics/Composites Institute, Section 17C, 1979, pp. 1-10.

Tanner, W. R., ed.: Industrial Robots: Volume 2 – Applications, Society of Manufacturing Engineers, Dearborn, Michigan, 1979. 287 pp.

Vajk, J. P.; Engel, J. H.; and Shettler, J. A.: Habitat and Logistic Support Requirements for the Initiation of a Space Manufacturing Enterprise. In Space Resources and Space Settlements, John Billingham, William Gilbreath, Brian O'Leary, and B. Gossett, eds., NASA SP-428, 1979, pp. 61–83.

Yankee, Herbert W.: Manufacturing Processes. Prentice-Hall, Englewood Cliffs, New Jersey, 1979. 765 pp.

Zachary, Wm. B.: A Feasibility Study on the Fabrication of Integrated Circuits and Other Electronic Components in Space Utilizing Lunar Materials. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

APPENDIX 5G

LMF ASSEMBLY SECTOR

5G.1 Assembly Sector Components and Technology Assessment

After raw lunar soil has been processed by the chemical processing sector into metallic and nonmetallic elements, and the parts fabrication sector has used these substances to manufacture all parts needed for LMF construction activities (growth, replication, or production), it is the job of the assembly sector to accept individual completed parts and fit them together to make working machines and automated subsystems themselves capable of adding to the rate of construction activities. A number of basic functions are required to perform sophisticated assembly operations. These are outlined in the assembly sector operations flowchart in figure 5.18. Each functional subsystem is discussed briefly below.

Parts Input

Parts produced by the fabrication sector are delivered either to inventory or directly to the assembly sector via mobile Automated Transport Vehicle (ATV) which runs on wheels or guide tracks. Parts are also retrieved from inventory by the ATVs. All retrieved or delivered parts are placed in segregated bins as input to the automated assembly system.

Parts Recognition/Transport/Presentation (RTP) System

The Recognition/Transport/Presentation (RTP) system is responsible for selecting the correct parts from the input bins, transporting them to within the reach of assembly robots, and presenting them in a fashion most convenient for use by the assembly robots. This will require a manipulator arm, vision sensing, probably tactile sensing, and advanced "bin-picking" software.

Early research concentrated on the identification and handling of simple blocks. For instance, at Hitachi Central Research Laboratory prismatic blocks moving on a conveyor belt were viewed, one at a time, with a television camera and their position and orientation determined by special software. Each block was then tracked, picked up with a suction-cup end-effector, and stacked in orderly fashion under the control of a minicomputer (Yoda et al., 1970). In another early experiment performed at Stanford University, a TV camera with color filters and a manipulator arm was developed that could look at the four multicolored blocks of an "instant Insanity" puzzle, compute the correct solution to the puzzle, and then physically stack the blocks to demonstrate the solution (Feldman et al., 1974).

At the University of Nottingham, the identity, position, and orientation of flat workpieces were determined one at a time as they passed under a down-looking TV camera mounted in a vertical turret much like microscope lens objectives. A manipulator then rotated into a position coaxial with the workpiece and acquired it (Heginbotham et al., 1972). More recently, software developed by General Motors Laboratories can identify overlapping parts laid out on a flat surface. The computer analyzes each part, calculates geometric properties, then creates line drawing models of each object in the scene and memorizes them. Subsequently, objects coming down the conveyor belt which resemble any of the memorized parts in shape – even if only small sections of a part can be seen or the lighting is poor – will be identified correctly by the system (Perkins, 1977).

In a recent series of experiments performed at SRI International, workpieces transported by an overhead conveyor were visually tracked. The SRI Vision Module TV camera views a free-swinging hanging casting through a mirror fixed on a table at 45°. An LSI-11 microprocessor servos the table in the x-y plane to track the swinging part. If a part is swinging over a 20 cm arc at about 0.5 Hz, the tracking accuracy is better than 1 cm continuously (Nitzan, 1979; Nitzan et al., 1979; Rosen, 1979). A moderate research and development program could produce an arm capable of tracking and grabbing a swinging part.

At Osaka University a machine vision system consisting of a television camera coupled to a minicomputer can recognize a variety of industrial parts (such as gasoline engine components) by comparing visual input from unknown parts with stored descriptions of known parts. The system can be quickly trained to recognize arbitrary new objects, with the software generating new internal parts models automatically using cues provided by the operator. The present system can recognize 20-30 complex engine parts as fast as 30 sec/part, and new objects can be

learned in 7 min (Yachida and Tsuji, 1975). Another system developed at SRI International can determine the identity, position, and orientation of workpieces placed randomly on a table or moving conveyor belt by electro-optical vision sensing, then direct a Unimate industrial robot arm to pick up the workpiece and deliver it to the desired destination (Agin and Duda, 1975).

Contact sensing may also be used in parts recognition. Takeda (1974) built a touch sensing device consisting of two parallel fingers each with an 8 × 10 needle array free to move in and out normal to the fingers and a potentiometer to measure the distance between the fingers. As the fingers close, the needles contact an object's surface contour in a sequence that describes the shape of the object. Software was developed to recognize simple objects such as a cone.

Of direct relevance to the lunar self-replicating factory RTP system is the "bin-picking" research conducted at SRI International. This involves the recognition and removal of parts from bins where they are stored by a robot manipulator under computer control. Three classes of "bins" may be distinguished: (1) workpieces highly organized spatially and separated, (2) workpieces partially organized spatially and unseparated, and (3) workpieces in completely random spatial organization. Simple machine vision techniques appear adequate for bin picking of the first kind, essentially state-of-the-art. Semiorganized parts bins (second class) can be handled by state-of-the-art techniques, except that picking must be separated into two stages. First, a few parts are removed from the bin and placed separately on a vision table. Second, standard identification and manipulation techniques are employed to pick up and deliver each part to the proper destination. Parts bins of the third class, jumbled or random pieces, require "a high level of picture processing and interpretive capability" (Rosen, 1979). The vision system has to cope with poor contrast, partial views of parts, an infinite number of stable states, variable incident and reflected lighting, shadows, geometric transformations of the image due to variable distance from camera lens to part, etc., a formidable problem in scene analysis. Some innovations have been made at General Motors in this area (Perkins, 1977), but researchers believe that progress using this technique alone will be slow, and that practical implementation will require considerably faster and less expensive computational facilities than are presently available (Rosen, 1979).

At SRI an end-effector with four electromagnets and a contact sensor has been built to pick up four separate castings from the top of a jumbled pile of castings in a bin. A Unimate transports the four castings to a backlighted table and separates them. Then a vision subsystem determines stable states, position, and orientation, permitting the Unimate gripper to pick up each casting individually and transfer it to its proper destination (Nitzan et al., 1979).

Although clearly more work needs to be done, a great deal of progress already has been made. It is possible to imagine a 5-10 year R&D effort which could produce the kind of RTP system required for the LMF assembly sector. Considerably more effort will be required to achieve the level of sophistication implied by Marvin Minsky's reaction to a discussion of current bin-picking and conveyor belt picking technology: "On this question of the variety of parts on assembly lines, it seems to me that assembly lines are silly and when we have good hand-eye robots, they will usually throw the part across the factory to the machine who wants it and that machine will catch it" (Rosen, 1979). The RTP system for the self-replicating LMF does not require this extreme level of robot agility.

Parts Assembly Robots

Once the correct parts have been identified, acquired, and properly presented, assembly robots must put them together. These assemblies – electric motors, gearboxes, etc. – are not yet working machines but rather only major working components of such machines. Thus it may be said that assembly robots assemble simple parts into much more complex "parts."

There has been a certain amount of basic research on aspects of programmable assembly. At MIT in 1972 a program called COPY could look at a simple structure built of children's building blocks, then use a manipulator to physically build a mirror image of the structure to prove its "understanding" of the block shapes and orientations. It would do this by withdrawing the blocks it needed from a collection of objects in its field of view, randomly spread out on a table (Winston, 1972). In Japan, a Hitachi robot called HIVIP could perform a similar task by looking at a simple engineering drawing of the structure rather than at the physical structure itself (Ejiri et al., 1971). In Edinburgh the FREDDY robot system could be presented with a heap of parts comprising a simple but disassembled model. Using its TV cameras and a manipulator, the system sorted the pieces, identified them correctly, then assembled the model. Assembly was by force and touch feedback, using a vise to hold partial assemblies, and parts recognition was accomplished by training (Ambler et al., 1975).

Research has also begun on the problems involved in fitting parts together or "parts mating." For instance, Inoue (1971) programmed a manipulator to insert a peg into a hole using force sensing at the manipulator joints. A more sophisticated version was later built by Goto at Hitachi Central Research laboratory. This version consisted of a compliant wrist with strain gauge sensors to control the insertion of a 1.2-cm polished cylinder into a vertical hole with a 7 to 20 μm clearance in less than 3 sec (Goto et al., 1974).

Besides fitting, assembly operations also include fastening. The most common methods include spot welding, riveting, arc welding, bolting, nailing, stapling, and gluing, all of which have been automated to some degree.

Numerical-control (N/C) riveting machines have replaced human riveters in the production of jetliner wings at Boeing Aerospace (Heppenheimer, 1977). At Westinghouse Electric Corporation a four-joint programmable manipulator under minicomputer control performs arc welding along curved trajectories (Abraham and Shum, 1975). According to information gleaned from Ansley (1968) and Clarke (1968), the Gemini spacecraft required 0.15 m/kg of seam welds and 6.9 spot welds/kg. Thus, for a 100-ton LMF seed equal to the Gemini capsule in its welding requirements, 15,000 m of seam welding would be required. This should take about a month of continuous work for a dedicated 5-10 kW laser welder (see appendix 5F). Another alternative is to make positive use of vacuum welding. Surfaces of parts to be fastened would be cleaned, then pressed gently together, causing a cold weld if they are made of the same or similar metallic material. Cast basalt end-effectors will probably be required for handling in this case.

At a high level of sophistication, assembly of certain well-defined machines from basic parts has been studied. Abraham and Beres (1976) at Westinghouse have described a product line analysis in which assembly line automation sequences were considered for constructing ten candidate assemblies, including a continuous operation relay (300 assembly steps), low voltage bushings (5 parts), W-2 low voltage switches (35 parts), fuse assembly (16 steps), and a small motor rotor assembly (16 steps). The tasks and implementation list for a sample motor rotor assembly is shown in table 5.19. This research has evolved into the Westinghouse APAS System, which uses state-of-the-art industrial robots and can automatically assemble complete electric motors of eight different classes representing 450 different motor styles discovered in a broad survey of all motors (van Cleave, 1977).

Other major industry and laboratory accomplishments include the following:

- Typewriter assemblies – At IBM Research Laboratories a program has been under way to use a multi-degree-of-freedom manipulator with a computer-controlled system for assembling small but complex parts. A high-level programming language for mechanical assembly was developed and used to acquire and assemble irregular typewriter parts (Will and Grossman, 1975).

- Water pump assembly – At Stanford University a manipulator called the "Stanford Arm" was programmed to assemble a water pump consisting of a total of 9 parts (base, gasket, top, and six screws). Joint forces were determined indirectly from measurements of drive motor currents. The software compensated for gravity and inertial forces, and included force feedback to locate holes for inserting two pins used to align the gasket (Bolles and Paul, 1973).

TABLE 5.19.– ASSEMBLY TASKS FOR A ONE-ROBOT CONFIGURATION, TO ASSEMBLE SMALL MOTOR ROTORS

Sequential tasks	Task implementation methods
1. Heat core in oven	New vertical in-line oven
2. Place shaft in hot core	Dedicated assembly unit
3. Quench cool	Water spray
4. Transfer subassembly to in-line conveyor	Pick and place device #1
5. Stake shaft	Automatic stake machine
6. Test subassembly	Automatic test device
7. (Optional – remove reject subassembly) 8. Retrieve switch from vision table 9. Place switch on shaft 10. Retrieve top sleeve 11. Place top sleeve on shaft	Computer-controlled robot #1
12. Press top sleeve and switch 13. Assemble bottom sleeve and press 14. Assemble rubber washers	Dedicated assembly units
15. Transfer subassembly to conveyor	Pick and place device #2
16. Assemble nylon washers	Dedicated assembly unit

- Compressor cover assembly – An assembly station using computer vision, various other sensors, and a robot arm with a force-controlled gripper and an x-y table has been developed to place and fasten the cover on an air compressor assembly (see fig. 5.43). There are 10 parts in the assembly operation, although one "part" is a preassembled compressor housing (McGhie and Hill, 1978).

- Motor and gearbox assemblies – Kawasaki Laboratories has demonstrated that complex motor and gearbox assemblies can be put together with precision

feedback sensors and appropriate manipulator grippers and fixtures. Kawasaki uses vibratory motion to jiggle parts with suitable bevels and tapers into place during assembly which automatically compensates for minor misalignments or tolerance variations (Thompson, 1978).

Automobile alternator assembly – A programmable robot assembly station built at the Charles Stark Draper Laboratory can assemble a commercial automobile alternator which consists of 17 individual parts, in a total of 162 sec using 6 tools (Nevins and Whitney, 1978). Simple changes such as using multiple head screwdrivers and assembling several units at once should bring the assembly time down to 60 sec/unit (Thompson, 1978). Figure 5.44 shows the functional components and flow pattern of the Draper machine. The Japanese have made similar advances. In fact, one such robot has been successfully assembling automotive alternators on a production basis in a standard factory environment for more than 3 years (Thompson, 1978).

- Gasoline engine assembly – Kawasaki's most impressive undertaking is the development of a pilot line for the automated assembly of small gasoline engines (Seko and Toda, 1974). Under control of one minicomputer, the assembly proceeds sequentially through five work stations, each including two small Kawasaki Unimates, a table, special jigs and tools, parts feeders, and special end-effectors. Controlled by the minicomputer but working independently, each robot performs a sequence of previously taught assembly operations including parts acquisition, parts mating, and, if necessary, parts fastening operations. No sensors were used for manipulative control and, consequently, there is heavy reliance on expensive jigging for orientation of workpieces. By the mid-1970s, the system was slow and not cost effective, but significant improvements were already being planned (Nitzan and Rosen, 1976).
- Expert system assembler – Some work has been done by Hart (1975) in developing a computer-based consultant able to "talk someone through" the assembly

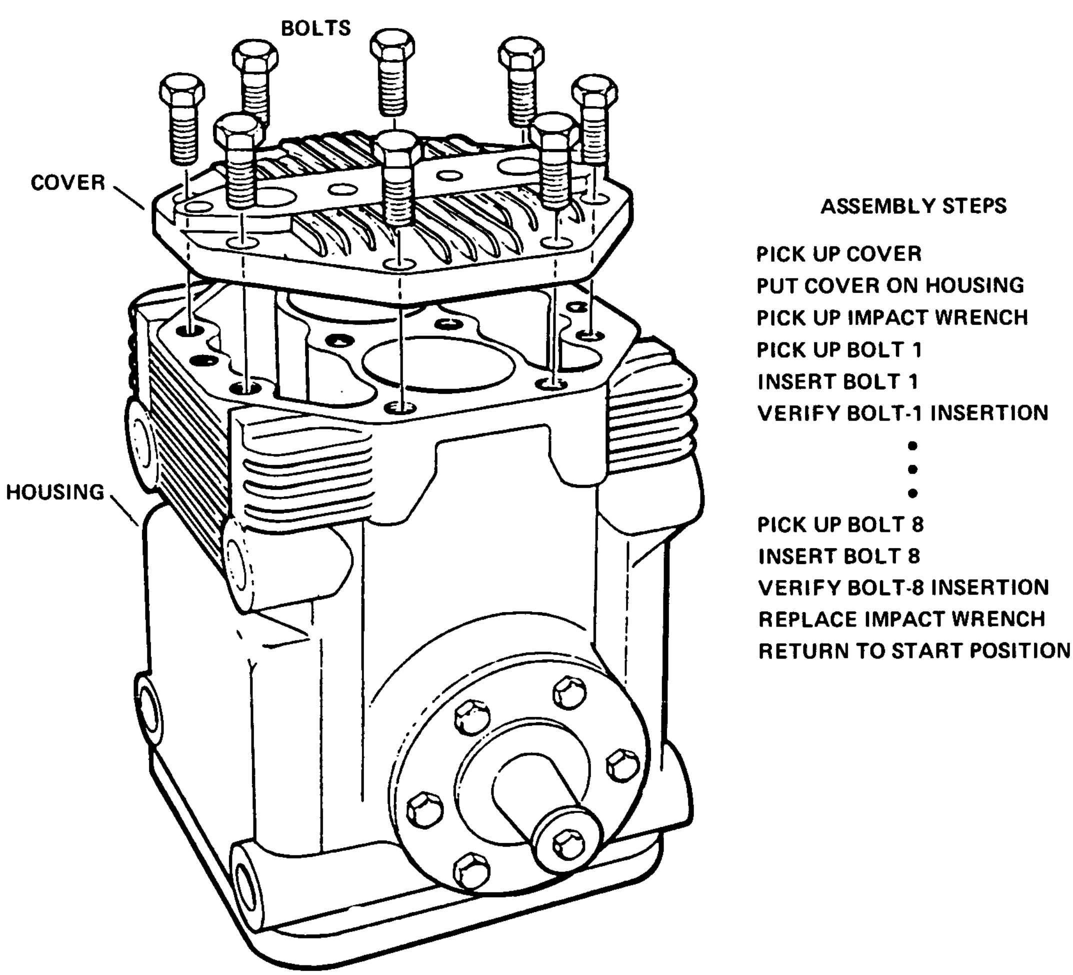


Figure 5.43.– Exploded view of SRI compressor cover assembly. (Rosen et al., 1978.)

of a complicated air-compressor assembly. In principle, the same kind of system could be used to "talk a robot," such as a repair robot with many different functions or a rescue robot, through the same assembly steps.

Clearly, a great deal of progress has been made, but much more remains to be made in all areas before an LMF-capable universal assembly system could be designed. Nitzan, private communication, 1980) estimates such a system might become available commercially by the end of the present century at the present rate of development. The amazing progress of the Japanese in developing "unmanned manufacturing" systems confirms this estimate, and suggests that by the end of the present decade a serious effort to design a universal assembly system of the type required for the lunar SRS might be successful.

If the original LMF seed has about 10^6 parts which must be assembled within a replication time T = 1 year, then parts must be assembled at an average rate of 31 sec/part. If subassembly assembly is included with successive ranks of ten (i.e., 10 parts make a subassembly, then 10 subassemblies make a more complex subassembly, etc.), then 1.111111×10^6 assembly operations are required which is only 28 sec/part. This is about typical for assembly operations requiring 100% verification at each step, using state-of-the-art techniques. The Draper robot described earlier assembles 17 parts in 162 sec, or 9.5 sec/part, and the improvement to 60 sec for the whole alternator assembly task would decrease this to 3.5 sec/part, an order of magnitude less than the mean continuous rate required for successful LMF operation.

Assembly Inspection Robots

After parts have been assembled by assembly robots with 100% verification at each step, the final assembly must be inspected as a final check to ensure it has been correctly built from the correct parts. According to Rosen (1979),

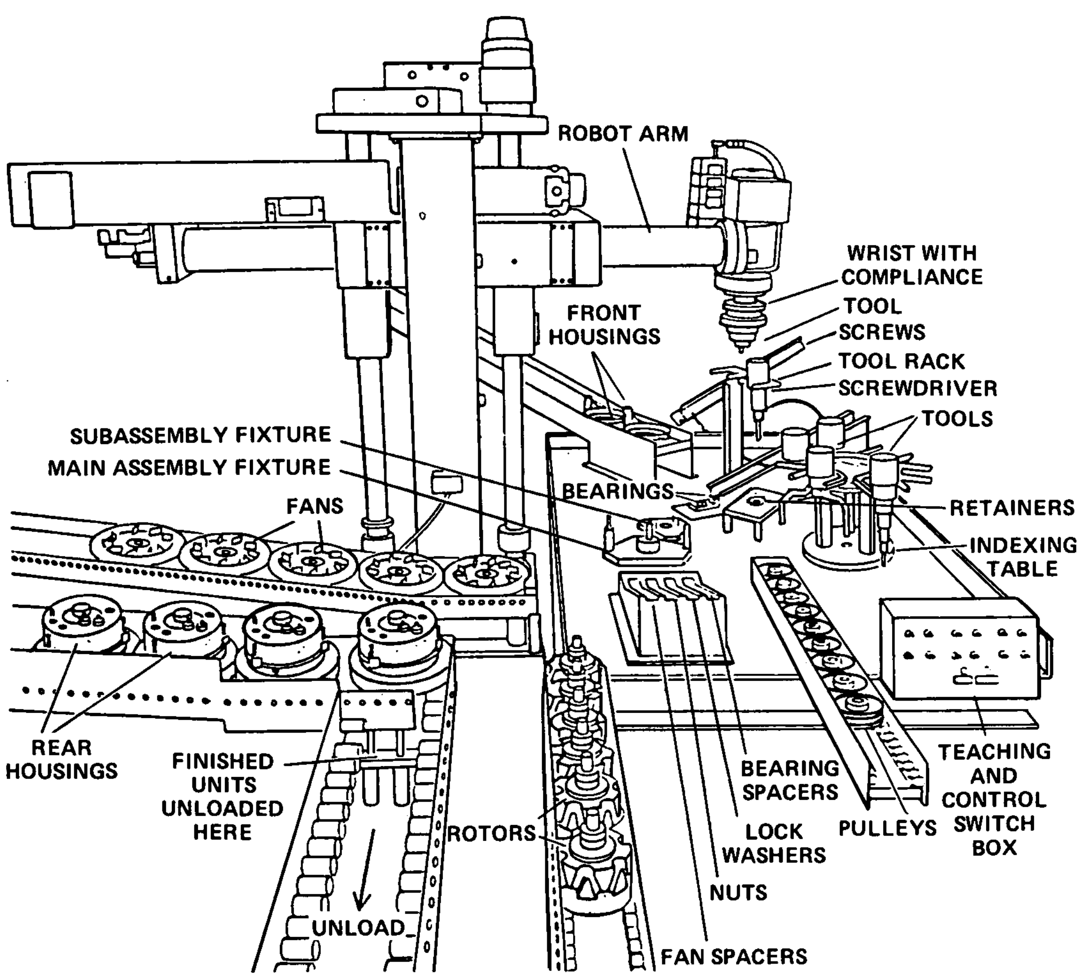


Figure 5.44.– Functional components of the Draper automobile alternator assembly robot. (Nevins and Whitney, 1978.)

machine vision for inspection may be divided into two broad classes: (1) inspection requiring highly quantitative measurement, and (2) inspection that is primarily qualitative but frequently includes semiquantitative measures.

In the quantitative inspection class, machine vision may be used to inspect stationary and moving objects for proper size, angles, perforations, etc. Also, tool wear measurements may be made. The qualitative inspection class includes label reading, sorting based on shape, integrity, and completeness of the workpiece (burrs, broken parts, screws loose or missing, pits, cracks, warping, printed circuit miswiring), cosmetic, and surface finishes. Each type of defect demands the development of specialized software which makes use of a library of subroutines, each affecting the extraction and measurement of a key feature. In due course, this library will be large and be able to accommodate many common defects found in practice. Simple vision routines utilizing two-dimensional binary information can handle a large class of defects. However, three-dimensional information, including color and gray-scale, will ultimately be important for more difficult cases (Rosen, 1979).

With the SRI-developed vision module, a number of inspection tasks have been directed by computer. For example, washing machine water pumps were inspected to verify that the handle of each pump was present and to determine in which of two possible positions it was. A group of electrical lamp bases was inspected to verify that each base had two contact grommets and that these were properly located on the base. Round and rectangular electrical conduit boxes were inspected as they passed on a moving conveyor, the camera looking for defects such as missing knockouts, missing tabs, and box deformation (Nitzan, 1979).

An inspection system developed by Auto-Place, Inc. is called Opto-Sense. In one version, a robot brings the workpiece into the field of vision. Coherent laser light is programmed by reflection off small adjustable mirrors to pass through a series of holes and slots in the part. If all "good part" conditions are met, the laser light is received by the detector and the part is passed. In addition to looking at the presence or absence of holes and object shape, the laser system can also check for hole size and location, burrs or flash on parts, and many other conditions (Kirsch, 1976). Range-imaging by lasers is well suited for the task of inspecting the completeness of subassemblies (Nitzan et al., 1977).

An inspection system designed for an autonomous lunar factory would need an internal laser source, a three-dimensional scanning pattern, at least two detectors for simple triangulation/ranging, a vision system for assembly recognition and position/orientation determination, and a large library of parts and assemblies specifications so that the inspection system can determine how far the object under scrutiny deviates from nominal and a valid accept/reject/repair decision may be made.

Electronics Assembly Robots

Electronics components, including resistors, capacitors, inductors, discrete semiconductor components (diodes, thyristors), and microelectronic "chips" (microprocessors, RAMs, ROMs, CCDs) are produced by the Electronics Fabrication System in the fabrication sector. Aluminum wire, spun basalt insulation, and aluminum base plates are provided from the bulk or parts fabrication system described in appendix 5F. After these parts are properly presented to the electronics assembly robots, these robots must assemble the components into major working electronics systems such as power supplies, camera systems, mini/microcomputer CPUs, computer I/O units, bulk memory devices, solar cell panels, etc. Electronics assembly appears to require a technology considerably beyond the state-of-the-art.

Present techniques for automated electronics assembly extend mainly to automatic circuit board handling. For instance, Zagar Inc. uses an automatic PCB drilling machine, and Digital Systems Inc. has an N/C automatic drilling machine with four speeds for drilling four stacks of boards simultaneously (Ansley, 1968). A circuit-board assembly line at Motorola allows automatic insertion of discrete components into circuit boards – the plug-in modular 25-machine conveyor line applied 30,000 electrical connections per hour to printed circuit modules used in Motorola Quasar television sets (Luke, 1972). Using four specialized assembly machines developed for Zenith, a single operator can apply more than half a million electrical contacts to more than 25,000 PCBs in one 8-hr shift (Luke, 1972).

Probably one of the most advanced electronics assembly systems currently available is the Olivetti/OSAI SIGMA-series robots (Thompson, 1978). The minicomputer-controlled SIGMA/MTG two-arm model has eight degrees of freedom (total) and a positioning accuracy of 0.15 mm. In PCB assembly, boards are selected individually from a feeding device by a robot hand, then positioned in a holding fixture. This method frees both hands to begin loading integrated circuit (IC) chips into the boards. The robot hands can wiggle the ICs to make them fit if necessary. ICs are given a cursory inspection before insertion, and bad ones are rejected. Assembly rates of 12,500 IC/hr are normally achieved (50 IC/PCB and 250 PCB/hr) for each robot arm pair, 2-3 per human operator. The two arms are programmed to operate asynchronously and have built-in collision avoidance sensors. In other operations, different SIGMA-model robots assemble typewriter parts such as ribbon cartridges, typewriter key cap assemblies, and mechanical key linkages.

The SIGHT-1 computer vision system developed by General Motors' Delco Electronics Division locates and calculates the position of transistor chips during processing for use in car and truck high-energy ignition systems. It also checks each chip for structural integrity and rejects all defectives (Shapiro, 1978). The simple program logic for the IC chip inspection is shown in figure 5.45.

A most serious gap in current technology is in the area of inspection. There are few if any systems for automatic circuit verification – at present, inspection is limited to external integrity and structural irregularities or requires a human presence. At present, neither IC nor PCB performance checking is sufficiently autonomous for purposes of SRS.

Bin Packing for Warehouse Shipment

Bin packing (or crate loading for shipment) is a straightforward problem in robotics provided the parts and crate presentation difficulties have already been solved. SRI International has done a lot of work in this area. For example, using feedback from a proximity sensor and a triaxial force sensor in its "hand," a Unimate robot was able to pick up individual preassembled water pumps from approximately known positions and pack them neatly in a tote-box. In another experiment boxes were placed randomly on a moving conveyor belt; the SRI vision system determined the position and orientation of each box, and permitted a Unimate robot arm to pack castings into each box regardless of how fast the conveyor was moving (Rosen et al., 1978). At Hitachi Central Research Laboratory, Goto (1972) built a robot "hand" with two fingers, each with 14 outer contact sensors and four inner pressure-sensitive conductive rubber sensors that are able to pick up blocks located randomly on a table and pack them tightly onto a pallet.

A related and interesting accomplishment is the stenciling of moving boxes. In an experiment at SRI International, boxes were placed randomly on a moving conveyor and their position and orientation determined by a vision system. The visual information was used by a Unimate robot to place a stencil on the upper right corner of each box, spray the stencil with ink, then remove the stencil, thus leaving a permanent marking on each box (Rosen et al., 1976). An immediate extension of this technique would be to use the vision module to recognize a particular kind of

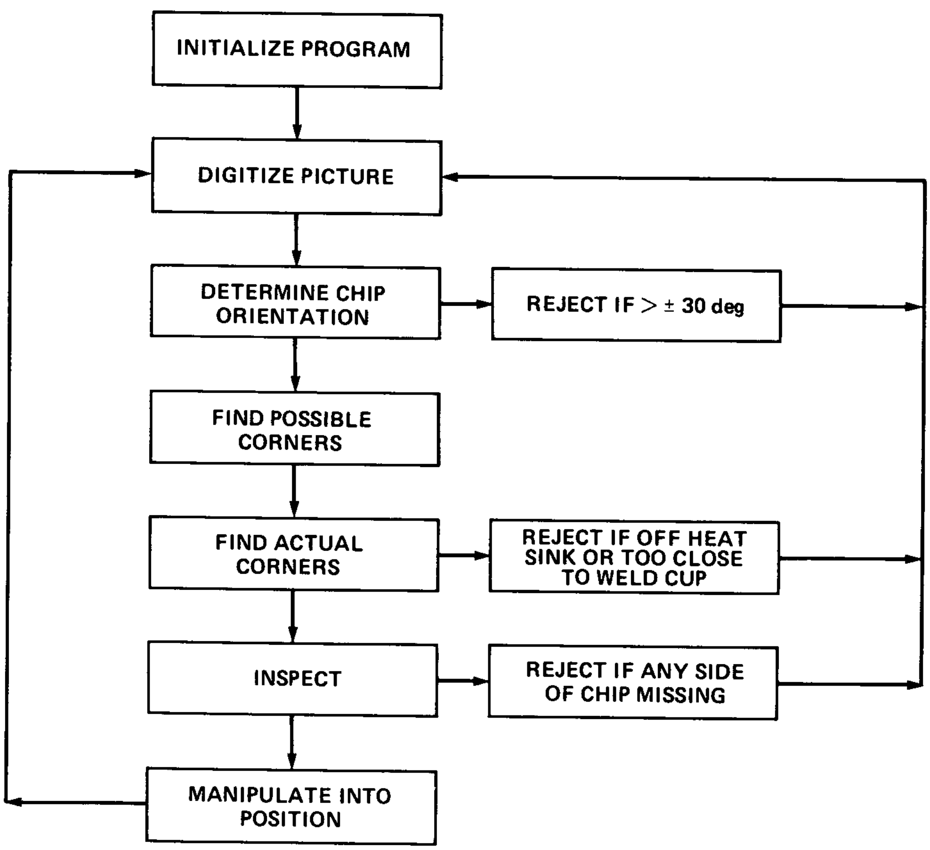


Figure 5.45.– Program logic for the GM/Delco IC "chip" inspection system.

box coming down the conveyor line, and then choose one of many possible stencils which was the "name" of that kind of box. Then the stenciling could be further extended to objects in the boxes, say, parts, in which case the end result would be a robot capable of marking individual objects with something akin to a "universal product code" that warehouse or assembly robots could readily identify and recognize.

Automated Transport Vehicles

Automated Transport Vehicles (ATVs), or "parts carts," are responsible for physically moving parts and subassemblies between sectors, between robot assembly stations, and in and out of warehouses in various locations throughout the LMF. Mobile carriers of the sophistication required for the lunar seed do not exist, but should be capable of development within a decade given the present strong interest in developing totally automated factories on Earth.

Luke (1972) describes a tow-cart system designed by SI Handling Systems, Inc., for use in manufacturing plants. These "switch-carts" serve as mobile workbenches for assembly, testing and inspection, and for carrying finished products to storage, shipping areas, or to other work areas. Carts can be unloaded manually or automatically, or loaded, then "reprogrammed" for other destinations. However, these carts are passive machines – they cannot load or unload themselves and they have no feedback to monitor their own condition (have they just tipped over, lost their load, had a load shift dangerously, etc.?) They have no means of remote communication with a centralized source of control, and all destination programming is performed manually. The ideal system would include vision and touch sensors, a loading/unloading crane, vestibular or "balance" sensors, an onboard microcomputer controller, and a radio link to the outside. This link could be used by the ATV to periodically report its status, location, and any malfunctions, and it could be used by the central factory computer to inform the ATV of traffic conditions ahead, new routes, and derailed or damaged machines ahead to avoid or to assist.

A major step forward was the now legendary "Shakey" robot, an SRI project during 1968-1972 (Raphael et al., 1971). Shakey was, in essence, a prototype mobile robot cart equipped with a TV camera, rangefinder, and radio link to a central computer. The system could be given, and would successfully execute, such simple tasks as finding a box of a certain size, shape, and color, and pushing it to a designated position. The robot could form and execute simple plans for navigating rooms, doorways, and floors littered with the large blocks. Shakey was programmed to recover from certain unforeseen circumstances, cope with obstacles, store (learn) generalized versions of plans it produced for later use, and to execute preliminary actions and pursuance of principal goals. (In one instance, Shakey figured out that by moving a ramp a few feet it could climb up onto a platform where the box it needed to move was resting.) The robot also carried out a number of manipulative functions in cooperation with a Unimate robot arm – Shakey had no manipulators of its own.

Work of a similar nature is now in progress in French laboratories. For example, the mobile robot HILARE is a modular, triangular, and computer-controlled mobile cart equipped with three wheels (two of them motor-driven), an onboard microcomputer, a sophisticated sensor bank (vision, infrared, ultrasonic sonar/proximity, and telemetry laser), and in the future a manipulator arm will be added (Prajoux, 1980). HILARE's control systems include "expert modules" for object identification, navigation, exploration, itinerary planning, and sensory planning.

The Japanese have also made significant progress in this area. One design is an amazing driverless "intelligent car" that can drive on normal roads at speeds up to 30 km/hr, automatically avoiding stationary obstacles or stopping if necessary (Tsugawa et al., 1979). Other Japanese mobile robot systems under development can find pathways around people walking in a hallway (Tsukiyama and Shirai, 1979), and can compute the relative velocities and distances of cars in real time to permit a robot car to be able to operate successfully in normal traffic (Sato, 1979).

Automated Warehouse Robots

Workpieces and other objects delivered to LMF warehouse facilities for storage must be automatically stowed away properly, and later expeditiously retrieved, by the warehouse robots. Numerous advanced and successful automated warehouse systems have already been installed in various commercial operations. A typical system in use at Rohr Corporation efficiently utilizes space and employs computer-controlled stacker cranes to store and retrieve standardized pallets (Anderson, 1972). The computer keeps records on the entire inventory present at any given time as well as the status of all parts ingoing and outgoing.

Similar techniques were used in the semiautomated "pigeonhole" storage systems for sheet metal and electric motors (in the 3/4 to 30 hp range) first operated by Reliance Steel and Aluminum Company decades ago. Each compartment contained one motor or up to 2250 kg of flat precut aluminum, magnesium, or high-finish stainless or galvanized steel stored on pallets. Retrieval time was about 1 min for the motors and about 6 min for the entire contents of a sheet metal compartment (Foster, 1963; Luke, 1972).

The technology in this area appears not to be especially difficult, although a "custom" system obviously must be designed for the peculiarities of lunar operations.

Mobile Assembly and Repair Robots

A Mobile Assembly and Repair Robot (MARR) must take complex preassembled parts (motors, cameras, microcomputers, robot arms, pumps) and perhaps a limited number of simple parts (bolts, washers, gears, wires, or springs) and assemble complete working LMF machines (mining robots, materials processing machines, warehouse robots, new MARRs). A MARR requires mobility, because it easily permits complex assembly of large interconnected systems and allows finished machines to be assembled *in situ* wherever needed in any LMF sector (Hollis, 1977). A MARR needs full mobility independent of specialized tracks or roadways, a wide range of sophisticated sensors (including stereo vision, IR and UV, radar and microwave, and various contact, contour, and texture sensing capabilities) mounted on flexible booms perhaps 4 m long. MARRs also require at least one "cherry picker" crane, a minimum of two heavy-duty manipulator arms, two light-duty manipulator arms with precision end-effectors, and a wide selection of tools (e.g., screwdrivers, rivet guns, shears, soldering gun, and wrenches). A radio link and onboard computer-controller are also essential.

MARRs have an omnibus mission illustrated by the diversity of the following partial list of tasks:

- Receive assembled subassemblies via automated transport vehicles
- Assemble subassemblies into working LMF machines *in situ* during growth phase(s)
- 100% verification of each final assembly step, with functional checkout as well as structural verification
- Debugging, dry-running, final checkout, and certification of operational readiness of each final assembly
- Repair by diagnostics, followed by staged diasssembly if necessary to locate and correct the fault (Cliff, 1981; see appendix 5H)
- Assemble new LMF seeds during replication phase(s)
- Assemble useful products during production phase(s)

According to van Cleave (1977), when General Motors began to consider the design of automated assembly systens for automobiles "the assembly of vehicles was rejected as being too complex for the time being so studies are confined to subassemblies." This area is identified as a major potential technology driver – insufficient research has been conducted on the development of systems for complete automated final assembly of working machines from subassemblies in an industrial production setting.

For instance, at General Motors Research Laboratories the most progress made to date is an experimental system to mount wheels on automobiles (Olsztyn, 1973). The location of the studs on the hubs and the stud holes on the wheels were determined using a TV camera coupled to a computer, and then a special manipulator mounted the wheel on the hub and engaged the studs in the appropriate holes. According to Rosen and Nitzan (1977), "although this experiment demonstrated the feasibility of a useful task, further development is needed to make this system cost-effective." The prospects for semiautonomous assembly robots have recently been favorably reviewed by Leonard (1980).

In Japan, much recent work has dealt with the design and construction of robot "hands" of very high dexterity of the sort which might be needed for fine precision work during delicate final assembly and other related tasks. Takese (1979) has developed a two-arm manipulator able to do tasks requiring cooperation between the arms – such as turning a crank, boring a hole with a carpenter's brace and bit, sawing wood, driving nails with a hammer, and several other chores. Okada (1979), also of the Electrotechnical Laboratory in Tokyo, has devised a three-fingered robot hand of incredible dexterity. Each finger has three joints. The hand of Okada's robot can tighten nuts on a threaded shaft, shift a cylindrical bar from side to side while holding it vertically, slowly twirl a small baton, and rotate a ball while holding it. Further research will extend into more complex movements such as tying a knot, fastening buttons, and using chopsticks.

Although some of the needed technologies for final assembly are slowly becoming available, many are not. Further, no attempt has yet been made to produce a final assembly robot, let alone a truly universal final assembly robot such as the MARRs required for the LMF. Such is a leap beyond even the ambitious Japanese MUM program mentioned in appendix 5F – even MUM envisions a minimum continuing human presence within the factory.

Conceptually, final assembly seems not intractable – a typical machine can be broken down into perhaps a few dozen basic subassemblies. But little research has been done so potential difficulties remain largely unknown. Major problem areas may include verification and debugging, subassembly presentation and recognition, actual subassembly interconnection or complex surfaces mating, and heavy lifting; today flexible robot arms capable of lifting much more than their own weight quickly, accurately, and dexterously do not exist.

The MARR system is a major R&D area which must be explored further before LMF design or deployment may practically be attempted.

5G.2 Assembly and LMF Computer Control

As with other sectors, LMF assembly is controlled by a computer which directs the entire factory. The assembly sector minicomputer, on the other hand, directs the many microcomputers which control its various assembly robots,

transport robots, and warehouse robots. The entire manufacturing system is thus controlled by a hierarchy of distributed computers, and can simultaneously manufacture subsets of groups of different products after fast, simple retraining exercises either programmed by an "intelligent" central computer or remotely by human beings. Plant layout and production scheduling are optimized to permit maximum machine utilization and speed of manufacturing, and to minimize energy consumption, inventories, and wastage (Merchant, 1975).

Merchant (1973) suggests that a fully automatic factory capable of producing and assembling machined parts will consist of modular manufacturing subsystems, each controlled by a hierarchy of micro- and minicomputers interfaced with a larger central computer. The modular subsystems must perform seven specific manufacturing functions:

(1) *Product design* by an advanced "expert system" software package or by humans, remotely or interactively, using a computer design system that stores data on models, computes optimal designs for different options, displays results for approval, and allows efficient process iteration.

(2) *Production planning,* an optimized plan for the manufacturing processes generated by a computer on the basis of product-design outputs, scheduling, and line-balance algorithms, and varying conditions of ore-feedstock deliveries, available robot resources, product mix, and priorities. Planning includes routing, timing, work stations, and operating steps and conditions.

(3) *Parts forming* at work stations, each controlled by a small computer able to load and unload workpieces, make parts and employ adaptive control (in-process operation sensing and corrective feedback), and incorporate diagnostic devices such as tool-wear and tool-breakage sensors.

(4) *Materials handling* by different computer-controlled devices such as lifts, warehouse stacking cranes, carts, conveyors, and industrial robots with or without sensors that handle (store, retrieve, find, acquire, transport, load, unload) parts, tools, fixtures, and other materials throughout the factory.

(5) *Assembly of parts* and subassemblies at computer-controlled work stations, each of which may include a table, jigs, industrial robots with or without sensors, and other devices.

(6) *Inspection of parts,* subassemblies, and assemblies by computer-controlled sensor systems during and at the end of the manufacturing process.

(7) *Organization of production information,* a large overseeing computer system that stores, processes, and interprets all manufacturing data including orders; inventories of materials, tools, parts, and products; manufacturing planning and monitoring; plant maintenance; and other factory activities (Nitzan and Rosen, 1976).

Such a completely computer-integrated factory does not yet exist, though various major components of this kind of system have been constructed and are in use in industry in the United States, Europe, and Japan. The most ambitious plan to reach Merchant's level of full automation is the Japanese MUM program which aims at "unmanned manufacturing" (computer-controlled operations, man-controlled maintenance) in the 1980–1985 time frame and "complete automatic manufacturing" (computer-controlled operations and maintenance) by 2000–2005 (Honda, 1974).

According to advanced planning notes, the most advanced and expensive MUM system would be "metabolic," "capable of being expanded," and "capable of self-diagnosis and self-reproduction.... With a built-in microcomputer, it is a self-diagnosis and self-reproduction system which can inspect functional deteriorations or abnormal conditions and exchange machine elements for identical ones. It is a hierarchy-information system with built-in microcomputer, middle computer, and central control computer. It can alleviate the burden on the central computer, and is capable of rapid disposal in case the computer fails. It is also capable of expansion" (Honda, 1974). Plans to open an automated robot-making factory at Fujitsu in accordance with the MUM philosophy are proceeding smoothly (see appendix 5F).

5G.3 Sector Mass and Power Estimates

A set of mass and power estimates for assembly systems was obtained from several sources and is displayed in table 5.20. Taking the extremes in each range, and given the known required throughput rate to replicate the original LMF seed in 1 year, we find that mass of assembly sector machinery lies between 83–1100 kg and the power consumption between 0.083–19 kW. If the warehouse robots and their fixed plant have a mass of about 1% of the stored goods (parts for an entire 100-ton seed) and a power requirement of about 10 W/kg, their mass is about 1 ton and their power draw about 10 kW.

The automated transport vehicles may have to carry the entire seed mass as often as ten times during the course of a year's growth, replication, or production. This is a hauling rate of 3.2×10^{-2} kg/sec or 0.32 parts/sec. If the average trip for an ATV is 100 m (initial seed diam), with a mean velocity of 1 km/hr (taking account of downtime for repairs, reprogramming, on- and off-loading, rescues, etc.), then the ATV trip time is 360 sec (6 min) and the average load is 11.5 kg/trip or 115 "typical parts"/trip. While a properly designed hauler should be capable of bearing at least its own weight in freight, ATVs require special equipment for manipulation rather than hauling. A conservative estimate for the ATV fleet is 100–1000 kg. If a typical vehicle power consumption is 20 (J/m)/kg (Freitas, 1980), the power requirement for the fleet is 0.56 to 5.6 kW total.

As for MARRs, the "warden" robots in the Project Daedalus BIS starship study (Martin, 1978) served a similar function and were allocated to the main vessel in the amount of 10^{-7} robots/kg-year serviced. To service a

TABLE 5.20.– MASS AND POWER ESTIMATES FOR ASSEMBLY SYSTEMS FROM VARIOUS SOURCES

Source	Plant mass, kg/kg per sec output	Plant power, W/kg plant
Johnson and Holbrow (1977) – Bulk processing and heavy industry estimate for human workers	4.3×10^4	2
Criswell (1980) – for "Cold Macro Assembly"	3.6×10^5	1
PUMA (1980) arm and controller computer, assuming 88 kg mass, 1500 W power, speed 1 part/30 sec assembly, part mass 0.1 kg/part	2.6×10^4	17

100-ton LMF Seed for a century would require one "warden" of mass 1 ton and a power draw of 10 W/kg. Conservatively assigning one MARR each to chemical processing sector, parts and electronics fabrication sectors, and assembly sector results in a total mass of 4 tons and draws 40 kW of power for the fleet of four MARRs. The main seed computer has a mass of 2200 kg, with 22.2×10^{-2} kg computer/kg serviced as in Martin (1978). With 17 W/kg as for the PUMA robot arm controller computer (Spalding, personal communication, 1980), seed computer power requirements are 37 kW.

5G.4 Information and Control Estimates

The team assumed that the assembly of a typical part may be described by 10^4 bits (about one page of printed text), an extremely conservative estimate judging from the instructions printed in Ford Truck (1960) and Chilton (1971), and especially if the seed has only 1000 different kinds of parts. Thus (10^4 bits/part)(10^6 parts/seed) = 10^{10} bits to permit the assembly sector to assemble the entire initial seed. To operate the sector may require an order less capacity than that needed for complete self-description, about 10^9 bits. Applying similar calculations to other sector subsystems gives the estimates tabulated in table 5.1 – ATVs lie between mining and paving robots in complexity, and warehoused parts, each labeled by 100 bits, require a total of 10^8 bits for identification, and perhaps an order of magnitude less for the computer controller that operates the warehouse and its robots.

5G.5 References

Abraham, Richard G.; and Beres, James F.: Cost-Effective Programmable Assembly Systems. Paper presented at the 1st North American Industrial Robot Conference, 26-28 October 1976. Reprinted in William R. Tanner, ed., Industrial Robots, Volume 2: Applications, Society of Manufacturing Engineers, Dearborn, Michigan, 1979, pp. 213-236.

Abraham, Richard G.; and Shum, L. Y.: Robot Arc Welder with Contouring Teach Mode. In Proc. 5th International Symposium on Industrial Robots, IIT Research Institute, Chicago, Illinois, September 1975, Society of Manufacturing Engineers, Dearborn, Mich., 1975, pp. 239-258.

Agin, Gerald J.; and Duda, Richard O.: SRI Vision Research for Advanced Industrial Automation. In 2nd USA-Japan Computer Conference, Session 5-4-5, 1975, pp. 113-117. Proceedings, Aug. 26-28, 1975. American Federation of Information Processing Societies, Montvale, N.J., 1975.

Ambler, A. P.; Barrow, H. G.; Brown, C. M.; Bonstall, R. M.; Popplestone, R. J.: A Versatile System for Computer-Controlled Assembly. Artificial Intelligence, vol. 6, Summer 1975, pp. 129-156.

Anderson, R. H.: Programmable Automation: The Future of Computers in Manufacturing. Datamation, vol. 18, December 1972, pp. 46-52.

Ansley, Arthur C.: Manufacturing Methods and Processes. Chilton Book Company, Philadelphia, 1968. Revised and enlarged edition.

Bolles, R. C.; and Paul, R.: The Use of Sensory Feedback in a Programmable Assembly System. Computer Science Department, Stanford Univ., Stanford, California, October 1973. Stan-CS-73-396, AD-772064.

Chilton's Auto Repair Manual, 1964-1971: Chilton Book Company, Philadelphia, 1971.

Clarke, Arthur C.: The Promise of Space. Harper and Row Publ., New York, 1968.

Cliff, Rodger A.: An Hierarchical System Architecture for Automated Design, Fabrication, and Repair. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

Criswell, David R.: Extraterrestrial Materials Processing and Construction. NSR 09-051-001 Mod. 24, Final Report, 31 January 1980.

Ejiri, M. *et al.*: An Intelligent Robot with Cognition and Decision-Making Ability. Proc. 2nd Intl. Joint Conf. on Artificial Intelligence, London, Sept. 1971, British Computer Society, London, 1971, pp. 350-358.

Feldman, J. *et al.*: The Use of Vision and Manipulation to Solve the Instant Insanity Puzzle. Proc. 2nd Intl. Joint Conf. on Artificial Intelligence, London, Sept. 1971. British Computer Society, London, 1971, pp. 359-364.

Ford Truck Shop Manual: Ford Motor Company, 1960.

Foster, David B.: Modern Automation. Pitman, London, 1963.

Freitas, Robert A., Jr.: A Self-Reproducing Interstellar Probe. J. British Interplanet. Soc., vol. 33, July 1980, pp. 251-264.

Goto, T.: Compact Packaging by Robot with Tactile Sensors. Proc. 2nd Intl. Symp. Industrial Robots, Chicago, 1972. IIT Research Institute, Chicago, Illinois, 1972.

Goto, T.; Inoyama, T.; and Takeyasu, K.: Precise Insert Operation by Tactile Controlled Robot HI-T-HAND Expert-2. Proc. 4th Intl. Symp. Industrial Robots, Tokyo, November 1974, pp. 209-218.

Hart, Peter E.: Progress on a Computer Based Consultant. SRI International Publication 711, January 1975.

Heginbotham, W. B.; Kitchin, P. W.; and Pugh, A.: Visual Feedback Applied to Programmable Assembly Machines. Proc. 2nd Int. Symp. Industrial Robots, Chicago, 1972. IIT Research Institute, Chicago, Illinois, 1972, pp. 77-88.

Heppenheimer, T. A.: Colonies in Space. Stackpole Books, PA, 1977.

Hollis, Ralph: NEWT: A Mobile, Cognitive Robot. Byte, vol. 2, June 1977, pp. 30-45.

Honda, Fujio, ed.: Methodology for Unmanned Metal Working Factory. Project Committee of Unmanned Manufacturing System Design, Bulletin of Mechanical Engineering Laboratory No. 13, Tokyo, 1974.

Inoue, H.: Computer Controlled Bilateral Manipulator. Bull. Japanese Soc. Mech. Eng., vol. 14, March 1971, pp. 199-207.

Johnson, Richard D.; and Holbrow, Charles, eds.: Space Settlements: A Design Study, NASA SP-413, 1977. 185 pp.

Kirsch, Jerry: Progression of Intelligence in Limited Sequence Robots. Paper presented at the 1st North American Industrial Robot Conference, 26-28 October 1976.

Leonard, Raymond S.: Automated Construction of Photovoltaic Power Plants. Paper prepared as in-house document, Bechtel National, Inc., San Francisco, California, 1980. 36 pp.

Luke, Hugh D.: Automation for Productivity. Wiley, New York, 1972.

Martin, A. R., ed.: Project Daedalus – The Final Report on the BIS Starship Study. British Interplanetary Soc., 1978. (J. British Interplanetary Soc., Supplement 1978.)

McGhie, Dennis; and Hill, John W.: Vision-Controlled Subassembly Station. Paper delivered at Robots III Conference, Chicago, Illinois, 7-9 November 1978.

Merchant, M. Eugene: The Future of CAM Systems. In National Computer Conference and Exposition, Anaheim, 1975. AFIPS Conference Proceedings, vol. 44, 1975, pp. 793-799.

Merchant, M. E.: The Future of Batch Manufacture. Phil. Trans. Royal Soc. London, vol. 275A, 1973, pp. 357-372.

Nevins, James L.; and Whitney, Daniel E.: Computer-Controlled Assembly. Scientific American, vol. 238, February 1978, pp. 62-74.

Nitzan, David: Robotic Automation at SRI. Proc. of IEEE Midwest Conference. MIDCON/79, Chicago, Illinois, 6-8 November 1979. Western Periodicals, Hollywood, California, 1979, Paper 5.1.

Nitzan, D.; Brain, A. E.; and Duda, R. O.: The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis. Proc. IEEE, vol. 65, February 1977, pp. 206-220.

Nitzan, David; and Rosen, Charles A.: Programmable Industrial Automation. IEEE Trans. Computers, vol. C-25, December 1976, pp. 1259-1270.

Nitzan, D.; Rosen, C.; Agin, G.; Bolles, R.; Gleason, G.; Hill, J.; McGhie, D.; Prajoux, R.; Park, W.; and Sword, A.: Machine Intelligence Research Applied to Industrial Automation. 9th Report, SRI International, August 1979.

Okada, Tokuji: A Versatile End-Effector with Flexible Fingers. Robotics Age, vol. 1, Winter 1979, pp. 31, 3-39.

Olsztyn, J. T., *et al.*: An Application of Computer Vision to a Simulated Assembly Task. Proc. 1st International Joint Conference on Pattern Recognition, Washington, D.C., 1973, pp. 505-513.

Perkins, W. A.: Multilevel Vision Recognition System. 3rd International Joint Conference on Pattern Recognition, Coronado, Calif., 8-11, 1976. New York, IEEE, 1976, pp. 739-744.

Perkins, W. A.: Model-Based Vision System for Scenes Containing Multiple Parts. General Motors Research Laboratories Publication GMR-2386, June 1977a.

Perkins, W. A.: A Model-Based Vision System for Industrial Parts. General Motors Research Laboratories Publication GMR-2410, June 1977b.

Prajoux, Roland: Robotics Research in France. Robotics Age, vol. 2, Spring 1980, pp. 16-26.

Raphael, B. *et al.*: Research and Applications – Artificial Intelligence. NASA CR-131991, 1971.

Rosen, Charles A.: Machine Vision and Robotics: Industrial Requirements. In Computer Vision and Sensor-Based Robots, George G. Dodd, Lothar Rossol, eds., Plenum Publ. Co., 1979, pp. 3-20, 20-22 (discussion).

Rosen, Charles A.; and Nitzan, David: Use of Sensors in Programmable Automation, vol. 10, December 1977, pp. 12-13.

Rosen, C. A.; Agin, C.; Andeen, G.; and Berger, J.: Machine Intelligence Research Applied to Industrial Automation. 6th Report, SRI International, November 1976. PB-289827/8, NSF/RA-761655.

Rosen, C.; Nitzan, D.; Agin, G.; Bavarsky, A.; Gleason, G.; Hill, J.; McGhie, D.; and Park, W.: Machine Intelligence Research Applied to Industrial Automation. 8th Report, SRI International, August 1978.

Sato, T.: Automotive Stereo Vision Using Deconvolution Technique. From 6th Intl. Joint Conf. on Artificial Intelligence, Tokyo, Japan, 1979. Stanford Univ., Computer Science Dept., Stanford, Calif., 1979.

Seko, K.; and Toda, H.: Development and Application Report in the Arc Welding and Assembly Operation by the High Performance Robot. In Proc. 4th Intl. Symp. Industrial Robots, Tokyo, Japan, November 1974, pp. 487-596. Japan Industrial Robot Assn., Tokyo, 1974.

Shapiro, Sydney F.: Digital Technology Enables Robots to See. Computer Design, January 1978. Reprinted in William R. Tanner, ed., Industrial Robots, Volume 2: Applications, Society of Manufacturing Engineers, Dearborn, Michigan, 1979, pp. 271-276.

Takeda, S.: Study of Artificial Tactile Sensors for Shape Recognition – Algorithm for Tactile Data Input. Proc. 4th Intl. Symp. Industrial Robots, Tokyo, Japan, November 1974, Japan Industrial Robot Assn., Tokyo, 1974, pp. 199-208.

Takese, Kunikatsu: Force Control of a Multi-Jointed Robot Arm. Robotics Age, vol. 1, Winter 1979, pp. 30, 32-36.

Thompson, Terrence: I See, Said the Robot. Assembly Engineering, Hitchcock Publ. Co., 1978. Reprinted in William R. Tanner, ed., Industrial Robots, Volume 2: Applications, Society of Manufacturing Engineers, Dearborn, Michigan, 1979, pp. 265-270.

Tsugawa, S. *et al.*: An Automobile with Artificial Intelligence. Paper delivered at 6th Intl. Joint Conf. on Artificial Intelligence, Tokyo, Japan, 1979. International Joint Conference on Artificial Intelligence, 1979, pp. 893-895.

Tsukiyama, T.; and Shirai, Y.: Detection of the Movements of Men for Autonomous Vehicles. From 6th Intl. Joint Conf. on Artificial Intelligence, Tokyo, 1979. International Joint Conference on Artificial Intelligence, 1979.

Van Cleave, David A.: One Big Step for Assembly in the Sky. Iron Age, vol. 28, November 1977. Reprinted in William R. Tanner, ed., Industrial Robots, Volume 2: Applications, Society of Manufacturing Engineers, Dearborn, Michigan, 1979, pp. 209-212.

Will, P. M.; and Grossman, D. D.: An Experimental System for Computer Controlled Mechanical Assembly. IEEE Transactions on Computers, vol. C-24, September 1975, pp. 879-888.

Winston, P. H.: The MIT Robot. In Machine Intelligence, B. Meltzer, D. Michie, eds., vol. 7, Edinburgh Univ. Press, 1972.

Yachida, M.; and Tsuji, S.: A Machine Vision for Complex Industrial Parts with Learning Capacity. Proc. 4th Intl. Joint Conference on Artificial Intelligence, Tbilisi, Sept. 1975. Artificial Intelligence Laboratory, Publications Dept., MIT, Cambridge, Mass., 1975.

Yoda, H.; Ikeda, S.; Ejiri, M.: New Attempt of Selecting Objects Using a Hand-Eye System. Hitachi Review, vol. 22, no. 9, 1973, pp. 362-365.

APPENDIX 5H

HIERARCHICAL SYSTEM ARCHITECTURE FOR AUTOMATED DESIGN, FABRICATION, AND REPAIR

The desire to create the self-replicating telefactor (Bekey and Naugle, 1980; Heer, unpublished draft notes of the Pajaro Dunes Goal Setting Workshop, 1980) leads to a number of interesting systems design problems. Early theoretical work by von Neumann (1966) showed that self-replicating machines could in principle be built. Laing (1975, 1977) has further elaborated this theme in a novel approach to the problem. Practical considerations in the creation of self-replicating machines have been treated by von Tiesenhausen and Darbro (1980). Freitas (1980) and Valdes and Freitas (1980) have dealt with the application of self-replicating machines to the exploration of deep space.

This appendix presents an architecture for a system which can perform automated design, fabrication, and repair of complex systems. This methodology should be a useful component of any self-replicating system.

5H.1 System Level Architecture

This section describes the architecture of a hierarchical fabrication system which starts with raw materials and outputs finished products. At the system level, each layer or "rank" of the hierarchy looks just like any other rank; however, the internal details of the various ranks may be entirely different. The present approach was inspired by Miller (1978). Figure 5.46 shows the basic system architecture which consists, from left to right, of rank upon rank of fabricators. Adjacent ranks of fabricators are separated by a transportation and communication subsystem. Ultimately, final products issue from the system at the extreme right.

For generality, the transportation and communication subsystems for each rank are shown to be disjoint. Indeed, it is evident that the subsystem T_1 which handles raw materials, such as ores, will no doubt differ in detail from the

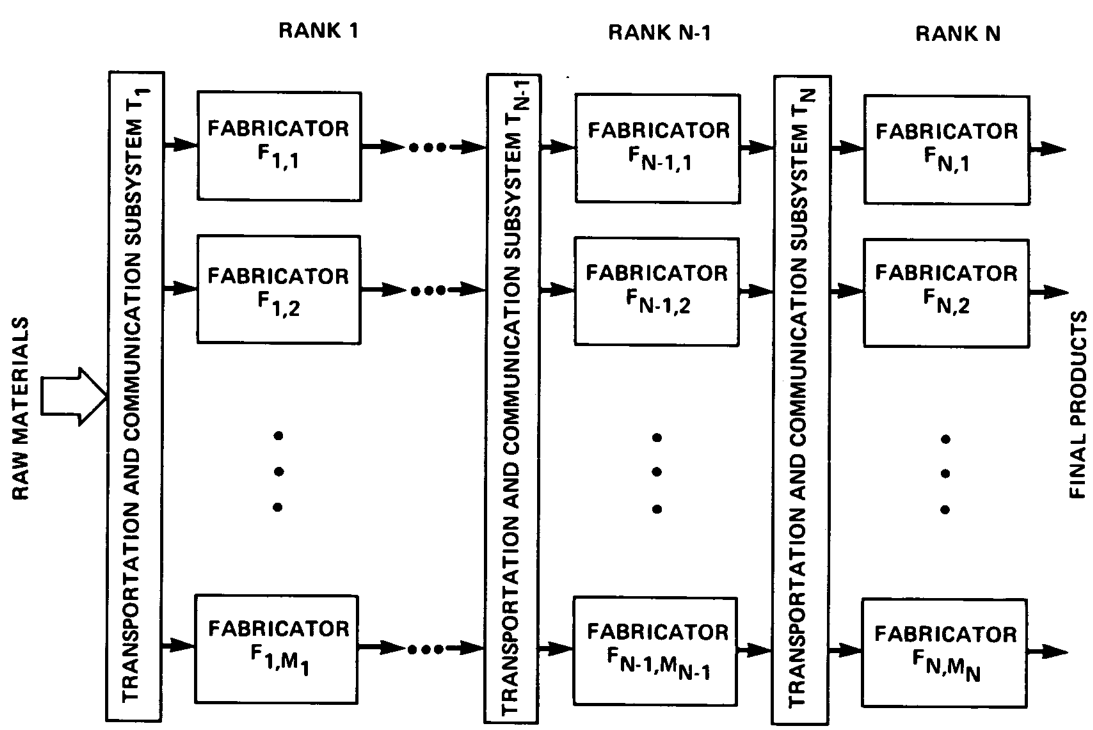


Figure 5.46.– Basic hierarchical system architecture.

subsystem T_L which handles electronic parts, such as electronic circuits. Furthermore, they will both differ from the system T_N which handles major subassemblies, such as complete power plants or complete computer systems. However, at the system level, each transportation and communication subsystem performs the same function: T_L handles the transfer of information and material between the fabricators at rank L and those at rank L+1 as shown in figure 5.47.

Although the internal details of the transportation and communication subsystem, T_L, need not concern us here, we shall consider how they appear logically to their adjacent ranks of fabricators. To the fabricators they look like a random-access, nonblocking switching network for information (e.g., our telephone system), and like a network of roads and delivery trucks for products. In other words, each message or product is dispatched from a fabricator with a unique address (e.g., telephone number or street address) attached to it. It is the function of T_L to see that the messages and products reach their specified destinations.

In figure 5.47, it will be seen that a fabricator at rank L+1 can request (via the status links) information on the types of product and their availability from each of the fabricators at rank L. Then the fabricator at rank L+1 transmits orders to the rank L fabricators for those products it requires as inputs to its process. The fabricators at

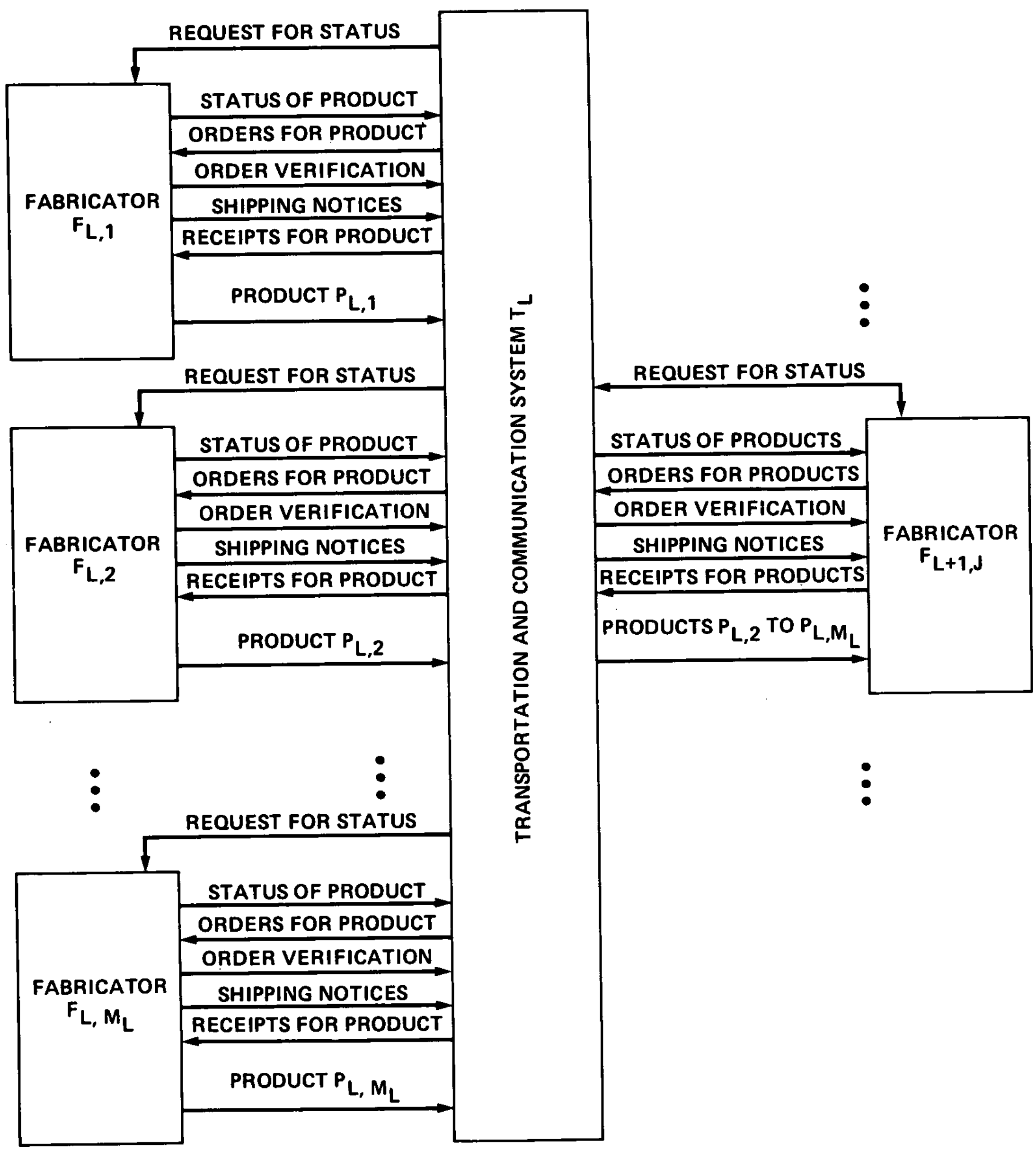


Figure 5.47.– Interrank interface.

rank L respond with order verifications which include expected time of shipment (to aid scheduling at rank L+1). Physical transportation of product through T_L is expected to be slow compared to information transfer; therefore, provision has been made for the transmission of shipping notices. Although this is logically redundant it can aid error recovery if the physical transport of materials goes awry. When the product arrives at its rank L+1 destination a receipt is returned to the rank L shipper to complete the transaction.

It may appear that the complexity of this interrank interface is not absolutely necessary. While strictly speaking this may be so, it is intended to aid in error recovery and to facilitate adaptive multipurpose behavior throughout the entire system. Indeed the multiple hierarchical feedback loops are borrowed both from living organisms and from human industrial economies (Miller, 1978).

Although some modest amount of diversity is expected in the transportation and communication subsystems, the fabricators are expected to be extremely diverse. They will range from ore smelters, to rolling mills, to high precision lithography for integrated circuits, to final assembly of complex products. The next section of this appendix will show, however, that at the system level all fabricators are essentially similar.

5H.2 Fabricator Morphology

Although great diversity of fabricators is envisioned, relatively few subsystem level primitives are required regardless of the rank L of the fabricator and these primitives are common to all ranks L.

A basic (Morph I) fabricator node appears as figure 5.48. It consists of six subdivisions which correspond to the production oriented parts of a business. The arrows are intended to denote material flow. Associated with each arrow, but not shown, are bi-directional information exchanges analogous to those associated with interfabricator communication in figure 5.47. The intrafabricator communications can in many instances be simpler because of tighter coupling between the communicating entities. Feedback and ability to manage error recovery should be preserved, however.

Material enters a rank L+1 fabricator node through a receiver which places orders for those rank L products which have been requested by the kitter. The receiver handles the shipping protocol for interface with transportation and communication subsystem T_L.

The kitter is driven by the parts list for whatever product the fabricator is to produce. The kitter orders the parts (through the receiver) and provides complete kits to the producer.

The producer converts the kitted parts with which it is provided into the output product of the fabricator node. This conversion could be a chemical process such as ore smelting or a mechanical process such as milling or PCB assembly.

After the product is produced by the producer, it is tested by the tester. In the basic Morph I, fabricator node product which does not pass the test is simply expelled as waste. Product which passes the test is sent to the stocker which accumulates it for shipment.

The shipper handles the protocol with the rank L+1 transportation and communication subsystem. It transmits status information, receives orders, and ships product.

Communication between the transportation and communication subsystems of adjacent ranks can be provided by degenerate fabricator nodes as shown in figure 5.49. Such a fabricator node might consist, as a minimum, of only a receiver and a shipper. This is sufficient to interface the protocols of the two different transportation and communication systems. In the simplest case it only provides a buffer. In a slightly more complex case, perhaps a repackaging or aggregation of product is performed. Kitters and/or stockers could be added, to perform these functions. Note that a Morph IIB fabricator node provides a material flow from higher to lower ranks. This capability is useful in dealing with recyclable scrap.

Figure 5.50 shows how a fabricator node could deal with product which fails its test, but which can profitably be recycled at some lower rank. The additional stocker and shipper can send the failed product, through a series of Morph IIB nodes, to an appropriate rank for recycling. This could be applied, for instance, to a milled part which is out of tolerance and can be recycled at less cost than producing an equivalent amount of material from raw ore.

If, on the other hand, the production process is reversible (such as putting a number of printed circuit cards into a card cage) then a more advantageous approach is shown (fig. 5.51). The disassembler performs the inverse of the

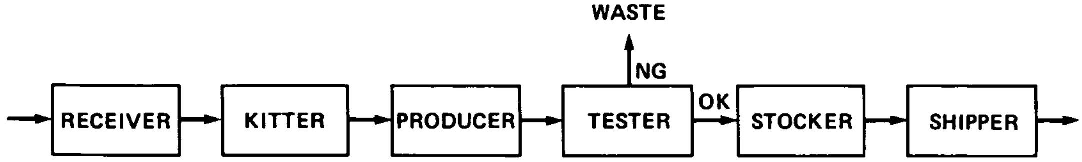


Figure 5.48.– Morph I fabricator node.

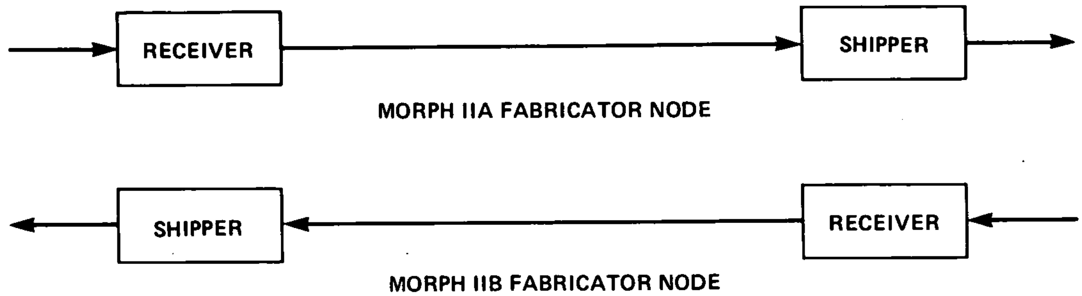


Figure 5.49.– Morph II fabricator node.

production process. The subassemblies are then returned whence they came to be re-tested as subassemblies. Also shown in the lower right hand corner of the figure are a receiver and stocker for returned (potentially faulty) product from the next higher rank.

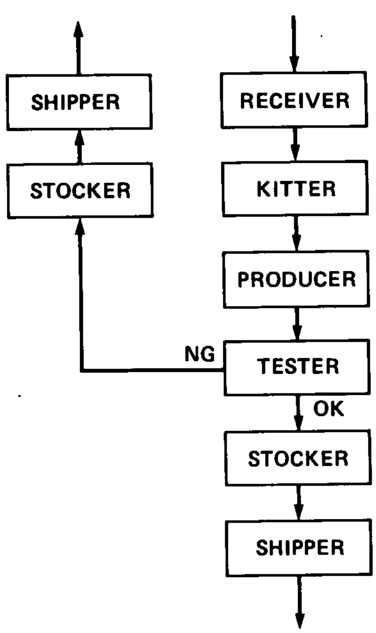


Figure 5.50.– Morph III fabricator node.

One could conceive of doing incoming inspection at the receiver; however, this does not seem to be the best approach. It seems more effective to associate production and test together in the same node (and thus the same rank) since they use much the same information. Furthermore, it seems wasteful to test product both after production and after shipment. This is especially true because the test capability would have to be duplicated at several receivers. Accordingly, for this architecture, test is uniquely associated with production, and suspect product is sent back to its producer for re-test.

We now consider a further embellishment of the fabricator node. Rather than entirely disassembling a faulty product, it may be profitable to rework it. Rework is in general much harder than simple disassembly because it can entail fairly complex diagnosis and repair. We deal here with a simple form of rework, namely, part exchange, as shown in figure 5.52.

In a Morph IV fabricator node, product which fails test (again a cage full of printed circuit cards is a good example) has its parts replaced one at a time. After each replacement, the product is tested again. If it passes it goes to the stocker and the faulty part is recycled or becomes waste. If the product fails again, a different part is exchanged, and so on until the culprit is identified. The decision of which part to exchange can be made on the basis of diagnostic tests, or it can be made at random.

5H.3 Automated Repair

The above has described a system level architecture of an automated fabrication facility. That architecture incorporated a test function after each production function as a way to catch errors as close to the source as possible and to prevent wasted effort. The architecture included paths for recycling product which failed its test. This also was to promote economy. A side effect of this architecture (which was designed only for efficient fabrication) is that it can also perform automated repair. Depending on circumstances this repair can be effected by selective disassembly or selective rework.

First consider selective disassembly for a suspect computer system that needs repair. For the purposes of example it is assumed that the final product is a computer system consisting of four racks of card cages and that each card cage contains 32 printed circuit boards. The PCBs are populated by 11 integrated circuit (IC) types. In order to simplify the discussion, inter-rack cabling, the card cages themselves (with back planes), and power supplies are not considered. Clearly, these can be accommodated in the same way as the components explicitly considered. It is further assumed that plugging ICs into PCBs, plugging PCBs

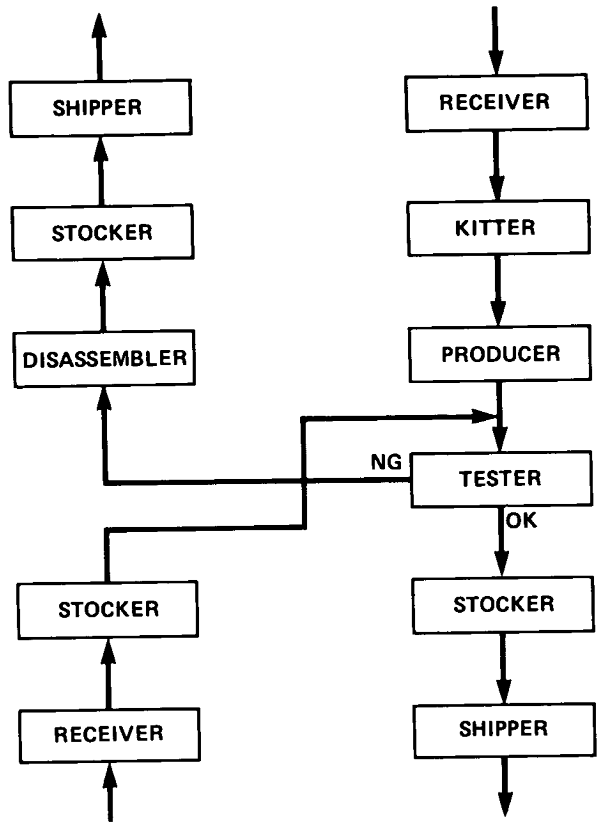


Figure 5.51.– Morph IV fabricator node.

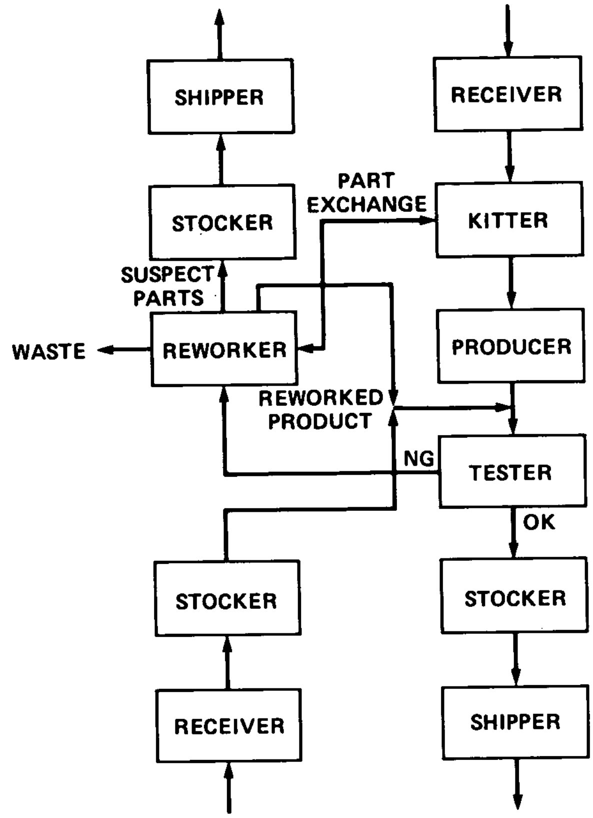


Figure 5.52.– Morph V fabricator node.

into card cages, placing card cages into racks (with interconnection), and cabling the racks together are reversible processes. Furthermore, since the Morph IV fabricator node is the simplest which can accept suspect product from a higher rank source, it is assumed that Morph IV fabricator nodes are used to perform the aforementioned assembly operations. Since IC manufacture is inherently nonreversible (one does not get useful subassemblies by disassembling a finished chip), ICs are assumed produced by a somewhat degenerate Morph IV fabricator node which simply scraps product which fails test rather than disassembling it. Finally, it is also assumed that the entire production facility is idle, there being no current need for additional finished computers.

The repair process begins when the suspect computer is presented to the fabricator node which produced it. This is node $F_{5,1}$ of figure 5.53. In this figure only those parts of the fabrication network which actually participate in the repair are shown. At node $F_{5,1}$ the computer is tested and determined to indeed be faulty. It is then disassembled into four component racks, each of which is sent to the node which produced it. At these nodes the individual racks are tested and in the example the third one is found to be faulty. Then this rack is disassembled into card cages, etc.

The selective disassembly proceeds until finally at node $F_{1,3}$ a faulty IC is found. With say, 100 ICs per card, 32 cards per cage, 10 cages per rack, and 4 racks per computer, this process has in rather straightforward manner isolated the *one* out of 128,000 ICs that was faulty. Node $F_{1,3}$ provides another IC of that type and scraps the faulty one.

Meanwhile, node $F_{5,1}$, having tested a computer which failed, has placed orders with nodes $F_{4,1}$ through $F_{4,4}$ for a set of racks from which to fabricate a replacement computer. Nodes $F_{4,1}$, $F_{4,2}$, and $F_{4,4}$ return their racks to node $F_{5,1}$. Node $F_{4,3}$, however, places orders for a set of card cages from which to fabricate a replacement for the third rack. Cages 1, and 3 through 10, are returned to node $F_{4,3}$ after testing, while cage 2 is disassembled. The PCBs from cage 2 are pulled out and sent to their respective fabrication nodes. Boards 1 through 12 and 14 through 32 are returned to node $F_{3,2}$. Meanwhile, board 13 is disassembled and the ICs are tested at nodes $F_{1,1}$ through $F_{1,11}$. (It is assumed here that of the roughly 100 ICs per card there are only eleven different types.)

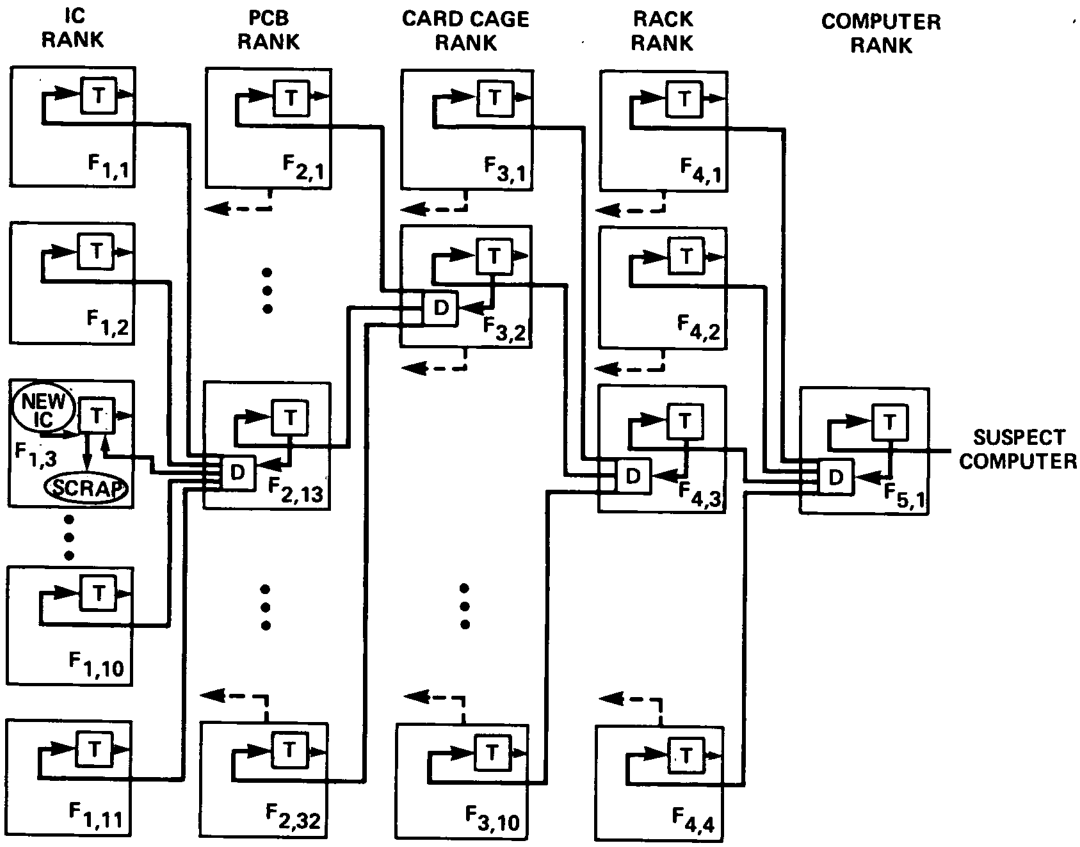


Figure 5.53.– Selective disassembly of failed system.

Now final reassembly can commence as shown in figure 5.54. The heavy lines in the figure trace the path of the replacement IC back into a repaired computer. When the replacement IC reaches node $F_{2,13}$, PCB 13 is reassembled. When the PCB rack reaches node $F_{3,2}$, card cage 2 is reassembled. When this card cage reaches node $F_{4,3}$ rack 3 is reassembled. And, finally, when rack 3 reaches node $F_{5,1}$ the original computer reappears with the one faulty IC replaced.

If, on the other hand, the production facility had been in use when the suspect computer was presented to node $F_{5,1}$, then the original computer would not re-emerge at node $F_{5,1}$. Instead, one more new computer would be produced. The subassemblies obtained by the selective disassembly of the failed computer would be incorporated into many different new computers. However, the failed computer plus one new IC will have resulted in the construction of one new computer to replace the failed one.

In either case, the replacement of the faulty computer is automatic – a consequence of the fabrication system architecture. No additions to the architecture were required to obtain the repair function. Similar arguments hold true if Morph V fabricator nodes are used, but selective rework, rather than selective disassembly, occurs.

5H.4 Automated Design

The system architecture described above readily lends itself to top-down modular design techniques. Using this discipline, at each level of detail, a designer (usually a human) receives a specification for the product he is to design. He then consults a catalog of available lower level products and selects those to be incorporated into his design. If he needs a lower level product which is not available, he generates specifications for that lower level product. The designer also generates assembly instructions for his product and a parts list. The assembly instructions go to the production department and the parts list to the procurement department. The original specification is used by the test department to verify that the product is what was originally requested.

Figure 5.55 shows how such a design function is added to a fabricator node. The designer in this node functions in a capacity analogous to that of a human designer. Automated design is a function which requires a fairly intelligent machine. Indeed, this is a topic of active current interest in the machine intelligence community.

Although the figure does not show the details of the interface, it is intended that the designer of rank L query

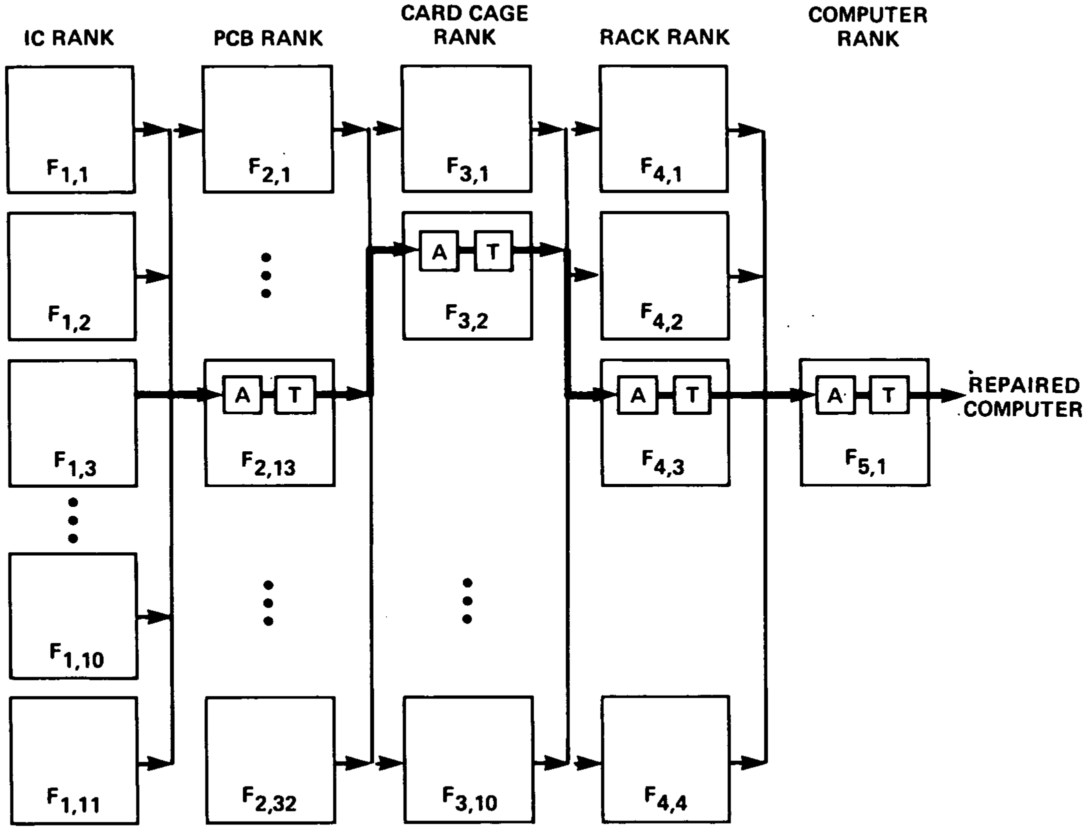


Figure 5.54.– Reassembly of repaired system.

the nodes of rank L-1 to ascertain what rank L-1 products are available, in much the same way as the kitter does. If the designer finds it needs a product which is not available, it chooses a free node at rank L-1 and sends to it the specification for the required product. If there are no free nodes, then the system as configured is not large enough to produce the desired final product. As the system architecture presently stands, outside intervention is required at this point.

5H.5 Utility of Node Morph Architecture

The node morphs enumerated in this paper are not meant to be exhaustive. Compound node morphs may be made by combining elements of two or more morphs. Degenerate morphs from which some functions have been deleted can also be useful. Those morphs enumerated are the ones found useful in the exposition of the system architecture. Additional stocker functions may be desirable in an actual physical system. Potentially useful locations are between the kitter and producer and between the producer and the tester. Also, multiple stockers may be required following a disassembler to handle the number and diversity of components. Redundant transportation and communication subsystems at each rank would make the overall system more robust, as would redundant fabricator nodes. This could be readily incorporated into a physical system.

This appendix has described the architecture of an automated system which has the following interesting properties:

(1) If it is presented with a final product specification (within its capabilities) it will do the detailed design (all the way down to raw materials if necessary) and then manufacture that product.

(2) If the system is presented with a faulty final product, it will repair it.

Interesting extensions of this architecture would be the ability to add additional fabricator nodes when required and the ability to add entire additional ranks when needed. This is presently under consideration (Cliff, 1981).

The similarities between the system described herein and the industrial complex of a developed nation are fairly obvious and indeed intentional. There are some significant differences, however. The automated system presented here is much more regular: The interrank transportation and communication systems are disjoint, one from another,

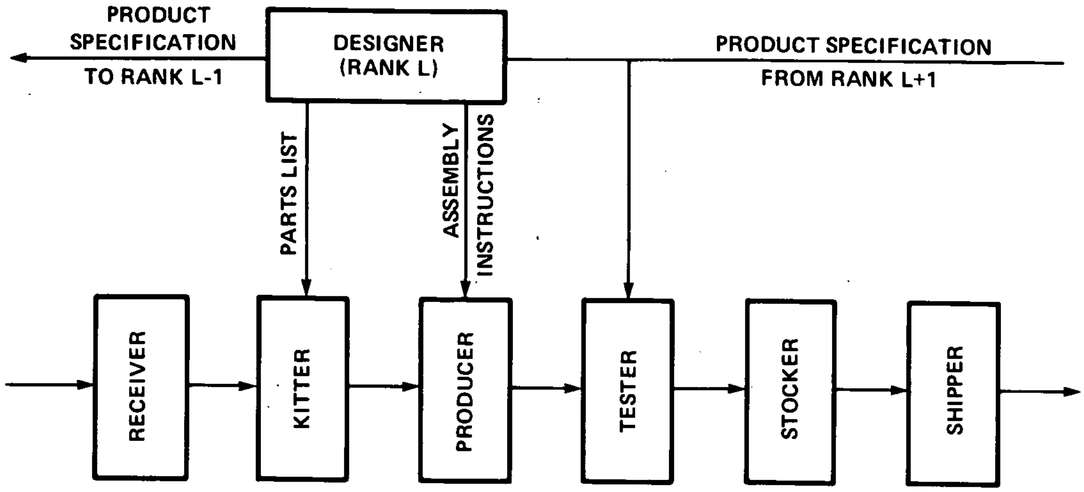


Figure 5.55.– Morph VI fabricator node.

and any one fabricator node makes only one product. The result is that the graph of the system is a lattice, rather than a random network. Furthermore, one could expand the number of nodes in each rank in such a way that the graph of the system becomes a tree. This should facilitate mathematical analysis of the system.

The idea (borrowed from top-down structured programming) is to produce a system which is at once extremely large and complex, but is still both understandable by humans and rigorously analyzable mathematically. This will be especially useful when growth by addition of nodes and ranks is studied.

5H.6 References

Bekey, Ivan; and Naugle, John E.: Just Over the Horizon in Space. Astronautics and Aeronautics, vol. 18, May 1980, pp. 64-76.

Cliff, Rodger, A.: An Hierarchical System Architecture for Automated Design, Fabrication, and Repair. Paper presented at the 5th Princeton/AIAA/SSI Conference on Space Manufacturing, 18-21 May 1981, Princeton, NJ.

Freitas, Robert A., Jr.: A Self-Reproducing Interstellar Probe. J. Brit. Interplan. Society, vol. 33, July 1980, pp. 251-264.

Laing, Richard: Some Alternative Reproductive Strategies in Artificial Molecular Machines. J. Theoret. Biol., vol. 54, Oct. 1975, pp. 63-84.

Laing, Richard: Automation Models of Reproduction by Self-Inspection. J. Theoret. Biol., vol. 66, June 7, 1977, pp. 437-456.

Miller, James G.: Living Systems. McGraw-Hill Book Company, N.Y., 1978.

Valdes, Francisco; and Freitas, Robert A., Jr.: Comparison of Reproducing and Nonreproducing Starprobe Strategies for Galactic Exploration. J. of the Brit. Interplan. Soc. vol. 33, November 1980, pp. 402-408.

Von Neumann, John: Theory of Self-Reproducing Automata, A. W. Burks, ed., Univ. of Illinois Press, Urbana, Illinois, 1966.

Von Tiesenhausen, Georg; and Darbro, Wesley A.: Self-Replicating Systems – A Systems Engineering Approach. NASA TM-78304, July 1980.

APPENDIX 5I

LMF SOLAR CANOPY POWER SUPPLY

The solar canopy provides electrical power for the entire lunar factory complex described in section 5.3.4. The canopy consists of many sections of automated (active or passive) solar energy collection devices. Mobile robots begin erecting the canopy after a useful fraction of the LMF base platform has been laid down and the central computer system installed in a depression near the hub. The canopy is just a simple framework of lightweight vertical metal/basalt struts snapped into universal connectors bolted into the heavy basalt foundation of the LMF. Horizontal wires or thin crossbeams support the solar panel mechanisms. The solar canopy is designed to be broken into relatively small sections for ease of assembly, installation, maintenance, and repair.

5I.1 LMF and Solar Canopy Geometry

One of the major constraints on LMF shape is the necessity for solar energy collection. The LMF may be visualized geometrically as a very broad, squat cylinder with some net density d_L (kg/m^3), mass (exclusive of platform) M, radius R, and height H. All factory energy is gathered using a "rooftop" surface area approximately the same size as the underlying foundation platform, so the fundamental constraint on factory size may be expressed by the condition $M/\pi R^2 = d_L H \leqslant MP_s/P$, where P_s is the usable energy delivered to the LMF by its solar collectors (roughly 150 W/m^2 for high quality photovoltaic devices at 45° angle of incidence) and P is the total power required by the initial lunar seed (about 1.7 MW; see sec. 5.3.4-5).

For a factory mass $M = 10^5$ kg, $R \geqslant 60$ m, the figure used elsewhere in this report. Estimates from O'Neill et al. (1980) that solar power systems (SPS) in the 100 kW range can be assembled for 8 kg/kW suggest a total mass for canopy collector panels (1.7 MW) of 13,600 kg, although this figure was derived from space-based SEPS and SPS design studies. A mass of 22,000 kg was adopted for the canopy, which includes transformers, diodes, cabling, and other necessary support devices. Since $d_L H = 8.8$ kg/m^2, the LMF in fact will be quite "roomy" inside – a "typical" population of 1-ton factory machines would be separated by an average distance of $2[10^3/(\pi d_L H)]^{1/2} = 12$ m.

Another major factor in determining basic factory configuration is the degree of isolation desired from the external surroundings. There appear to be few compelling reasons for solid massive walls enclosing a fully automated lunar manufacturing facility. Inclement weather, cleanliness, provision of human-habitable volume, protection from the dangers of seismic activity, noise/pollution abatement, and theft prevention are the usual reasons for heavy walls on Earth, yet these factors should have little if any impact upon factory construction in space or on the Moon. Further, rigid solid walls hinder growth and might delay reconfiguration as the LMF expands in size. The cleanliness problem in an open factory is expected to be minor, as mobile robots are designed either for external or internal operation but not for both (though in special circumstances MARR machines can be towed to external sites by mining robots).

The simplest solar canopy configuration is a web-like metal structure overlaid with flat solar panel assemblies. These cells are suspended from a series of crossbeams spaced at regular intervals along chords of the circular LMF. These crossbeams may be as thin as wires if adequately supported by strategically placed vertical columns. Calculations of stress reveal that a 1 mm radius aluminum rod (typically 10^8 N/m^2 tensile or compressive strength) should be strong enough to support a 22-ton canopy structure with a loading safety factor of about 5. Support posts are 1-cm diam aluminum/basalt columns placed at intervals of 10 m across the factory floor and anchored with universal connectors and several braces and struts for stabilization. These posts have an overload factor of more than 100, hence should be able to sustain low-speed accidental impacts by mobile robots without buckling. The total mass for the entire framework is well under 1 ton.

Ideally, all lunar operations should be conducted continuously with only scheduled maintenance shutdowns. However, continuous operation is possible only if continuous power is also available. A number of options for power storage during the lunar night have been considered in the context of a lunar base (Criswell, 1979; Vajk et al., 1979) in the literature. Possibilities include nuclear plants, volume heat capacity storage, chemical storage (batteries, fuel cells, exothermic reactants), capacitor banks, gravitational energy storage, pressurized gas, flywheels, and SPS transmission from orbit to lunar surface collection stations. These many promising alternatives, however, were not explored in depth. Without such an option, the baseline LMF must be placed on standby during the lunar night with one working year requiring two calendar years.

5I.2 Solar Canopy vs Lunar Igloo Designs

In the solar canopy LMF design the entire automated factory complex is erected on a fused basalt platform resting on the lunar surface. Above the factory floor is a relatively flimsy framework of solar energy collectors which provide system power.

The "lunar igloo" is an alternative in which geodesic domes of 120 m diam are constructed over each seed factory. Additional factory growth is accommodated by adjacent domes of similar size built with a network of connecting tunnels. Each dome is covered with at least 2-5 m of lunar topsoil which may be sufficient to permit the retention of an internal 0.3-atm oxygen atmosphere. This configuration might be handy in preventing accidental vacuum welding and could simplify servicing and troubleshooting by humans during system failures. Light could be admitted to the underground LMF via a converging reflective geometry (Hyson, personal communication, 1980).

Since these models represent fundamentally different design concepts (see fig. 5.56), the team compared the two directly on a number of significant factors enumerated in table 5.21. The conclusion was that the canopy model is possibly superior in the present fully automated self-replicating LMF application, but that the igloo model is not precluded in other scenarios.

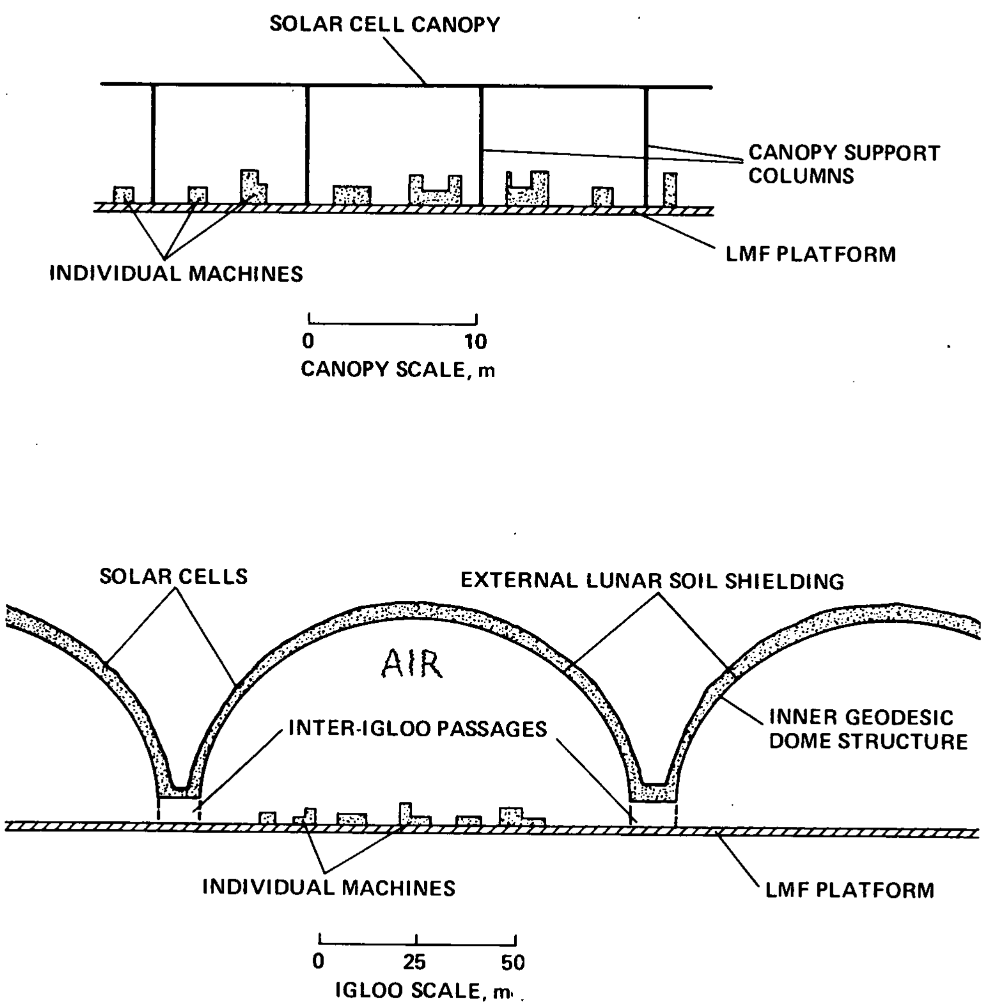


Figure 5.56.– Schematic of Solar Canopy and Lunar Igloo models of self-replicating or growing LMF.

TABLE 5.21.– COMPARISON OF IMPORTANT FACTORS FOR SOLAR CANOPY AND LUNAR IGLOO MODELS OF SELF-REPLICATING OR GROWING LMF

Some important factors	Solar canopy	Lunar igloo
1. Maintain useful atmosphere?	no	yes
2. Maintain useful vacuum?	yes	yes
3. Prevent solar cell degradation?	no	no
4. Prevent external optics degradation?	no	no
5. Prevent internal optics degradation?	no	yes
6. System temperatures easily controlled?	no	yes
7. Low mass foundation structure?	yes	yes
8. Low mass total structure?	yes	no
9. System construction mechanically easy?	yes	less easy
10. Easy maintenance of system integrity?	yes	less easy
11. Internal lighting easily available?	yes	no
12. Human repairman accessible?	yes	yes
13. Human repairman habitable?	no	yes
14. Easy horizontal mass flow?	yes	yes
15. Simplicity of overall system design?	yes	less simple
16. Easy to expand LMF system size/mass?	yes	no
17. Waste heat easily rejected?	yes	no
18. Terrestrial manufacturing processes easily transferred?	less easy	yes

5I.3 References

Criswell, David R.: The Initial Lunar Supply Base. In Space Resources and Space Settlements, John Billingham, William Gilbreath, Brian O'Leary, and Beulah Gossett, eds., NASA SP-428, 1979, pp. 207-224.

O'Neill, G. K.; Driggers, G.; and O'Leary, B.: New Routes to Manufacturing in Space. Astronautics and Aeronautics, vol. 18, October 1980, pp. 46-51.

Vajk, J. P.; Engel, J. H.; and Shettler, J. A.: Habitat and Logistic Support Requirements for the Initiation of a Space Manufacturing Enterprise. In Space Resources and Space Settlements, John Billingham, William Gilbreath, Brian O'Leary, and Beulah Gossett, eds., NASA SP-428, 1979, pp. 61-83.

APPENDIX 5J

COMPLEXITY AND LEVEL OF DETAIL IN ROBOT PROGRAMMING

Programming one robot manipulator to assemble another robot manipulator is not a trivial task. The finest level of detail involves controlling the individual steps of the various stepping motors which in turn control the independent degrees of freedom of the manipulator.

In order to gain some insight into robot programming, the Replicating Systems Concepts Team visited Dr. Charles H. Spalding at the research laboratories of Unimation, Inc., in Mountain View, California. Dr. Spalding demonstrated the operation of the PUMA 500 robot manipulator for the team. This manipulator system consists of a five-degree-of-freedom electrically servocontrolled arm combined with an electronics package containing a DEC LSI-11 control computer, individual microcomputer systems for each degree of freedom, and drivers for the servo motors.

In a system such as the PUMA with separate microcomputers for each degree of freedom the individual microcomputers must receive commands specifying either the required rate of motion for their respective degrees of freedom, or the desired position of that degree of freedom, or both. In the PUMA, the individual microcomputers are controlled by a larger, more powerful microcomputer (a DEC LSI-11, a member of the PDP-11 family). The LSI-11 can direct the end effector of the robot manipulator to trace out a number of different predetermined paths in three-dimensional space. In the present configuration (depending on the complexity of the selected paths), on the order of 1000 programmed motion steps can be accommodated. The PUMA robot has about 500 distinguishable "parts," about 50 in the wrist assembly alone.

The next order of sophistication in robot control is at the level of elementary assembly operations. The command "put a washer on the bolt" requires the performance of subtasks such as:

(1) Move the end effector to the washer supply.
(2) Grasp a washer.
(3) Move the end effector to the end of the bolt.
(4) Orient the washer so it is perpendicular to the bolt.
(5) Translate the washer so the axis of the bolt passes through the center of the hole in the washer.
(6) Translate the washer along the axis of the bolt.
(7) Release the washer.
(8) Retract the end effector.

Still more sophisticated operations include the joining of subassemblies. To join two subassemblies each one must be brought into the proper relative position and several washers, nuts, connectors, etc., must be installed. It is not clear, without further study, just how much of this hierarchy of operations could be controlled by the LSI-11 that has become an industry standard. However, the team has no doubt that a suitably powerful computer can be constructed in a module not exceeding 1 m^3 in volume, which would also serve as a base for an advanced robot manipulator. Spaulding estimated that 5 years of adequate funding and manpower support could probably produce a robot manipulator system capable of assembling a duplicate of itself from prefabricated parts.

The team discovered no fundamental difficulties with the software, although the programming task will be extremely challenging. In current industrial robotics applications, each manipulator has a very limited number of tasks to perform. To use one manipulator to perform as many tasks, each of the complexity required by the SRS demonstration, the state-of-the-art in robot programming should be advanced considerably.

A top-level description of the steps required to produce a robot manipulator system complete with control computer and required electronics support (see fig. 5.29) might include the following sequence:

(1) Assemble base frame and bolt it to floor.
(2) Install card cages in frame.
(3) Install cables between card cages.
(4) Insert printed circuit cards into card cages.
(5) Assemble manipulator waist joint support to base frame.
(6) Install waist joint.
(7) Install manipulator trunk (vertical member).
(8) Install shoulder joint.
(9) Install upper arm.
(10) Install elbow joint.
(11) Install forearm.
(12) Install wrist joint.
(13) Install end effector.
(14) Install television camera mast.
(15) Install television camera.
(16) Connect television camera to electronics in base.
(17) Connect manipulator to electronics in base.
(18) Connect AC power and turn on computer.
(19) Transfer construction software.
(20) Boot up the new computer.

Having been replicated as thus detailed, the new robot is on its own.

APPENDIX 5K

ISSUES AND CONCEPTS FOR FURTHER CONSIDERATION

During the present study the Replicating Systems Concepts Team considered numerous concepts relating to the problems of self-replicating systems (SRS). The following is a partial list of various notions, ideas, suggestions, and research directions which came to the team's attention but which could not be adequately explored in the time available.

5K.1 Definitions

(1) Reproduction – What is a good, precise definition of "self-reproduction" or "self-replication"? What exactly is a "self-replicating system"? Does replication include any assembly of a physical copy of self? A copy of patterns? Is full assembly from molecular or atomic primitives required? Shall minimal reproduction be defined in terms of basic functions, bits of information processed, or some other measure? Is there some irreducible minimum necessary for "reproduction"? Most regard simple autocatalysis or Ashby's falling dominoes as not representative of "true" replication. However, perhaps a New Guinea islander would regard the cafeteria tray line (with seemingly equal justification) as "not real" when the source of human reproduction – viewing our environment as "too well-ordered to be believable."

(2) Growth – Exactly what is the distinction between growth and reproduction? What is the difference between these concepts and the notion of "self-organization"? What about "self-assembly"? These are common terms in regular use, and need to be more precisely characterized.

(3) Repair – What is the difference between self-repair and self-reproduction? Ordinarily replication involves duplication of the whole system, whereas repair involves replacement of only some subset of it. But at what point does "repair" become "reproduction"? Is machine self-repair or self-reproduction more difficult from a technical standpoint, and why? (Self-repair may require an analytical intelligence, whereas much of reproduction can be accomplished by "rote.")

(4) Telefactor, teleoperator, intelligent tools, autonomous, etc. – precise definitions are needed. Is there a clear dividing line between biological reproductive systems and advanced self-replicating robot systems?

5K.2 Evolutionary Development

(1) Which theoretical models would be easiest to cast into physical engineering form: the von Neumann kinematic model, the Laing self-inspection approach, the Thatcher methodology, or some other alternative? Under what conditions would each be desirable from a pragmatic engineering standpoint? The Laing approach, for instance, may prove superior to the von Neumann kinematic model in the case of extremely large, complex self-reproducing systems where the universe of components is so vast that self-inspection becomes essential to maintain order or where rapid evolution is desired.

(2) Specific "unit growth" and "unit replication" models of SRS were considered in detail during the present study. Under what conditions is one or the other optimum? Are there any fundamental differences between the two in terms of performance, stability, reliability, or other relevant factors? What might SRS emphasizing "unit evolution" or "unit repair" be like?

(3) Can SRS be designed to have few or no precision parts? Can milling and turning operations be eliminated? What substitutes might be found for the usual precision components such as ball bearings, tool bits, metering instruments, micron-feature computer chips, etc.? It is possible to imagine Stirling engines, solar mirrors, electromagnets, and mechanical gear-trains using only native lunar basalt, iron, and gases with no chemical processing – but are complete (but simple) SRS possible using just two or three non-chemically recovered elements/minerals? Could SRS be patterned after terrestrial biological protein synthesis, in which the factory is made up of perhaps two dozen fundamental "building blocks" (similar in function to amino acids) assembled in virtually limitless combinations?

(4) To what extent is intelligence a prerequisite for reproduction? (Amoebas appear to replicate well with almost no intelligence at all.) Does increasing intelligence make more efficient the processes of biological, and potentially machine, replication? Is there a law of diminishing returns, or does more intelligence always produce a superior reproductive entity?

(5) What forms of machine intelligence might possibly be required for a fully autonomous SRS, that are not now being adequately pursued by artificial intelligence researchers? A few possibilities include learning, memory structure, advanced task planning, adaptivity, association, creativity, intuition and "hunch" formation, hypothesis generation, self-awareness, survival motives, sophisticated database reasoning, symbolic meaning of knowledge, autonomous problem solving, and insight. Similarly, the state-of-the-art in robotics and automation from the viewpoint of SRS development needs to be examined.

(6) What is the least complex biological self-replicating system? How does it work? Can similar processes and analogies be drawn upon for use in the development of self-replicating machine technology? What is the minimum critical mass for a stable ecosystem? For a machine economy with closure?

(7) What is the possibility of semisentient workpieces? This concept is sometimes referred to as "distributed robotics." Perhaps each workpiece in an assembly process could be imbued with some small measure of machine intelligence using advanced microelectronic circuitry. Parts could then assist in their own assembly and subsequent installation and maintenance.

(8) Can computers be programmed to write their own self-assembly software? Perhaps an "artificial intelligence expert system" is required?

(9) What can be said about the possibility of machine "natural" or "participatory" evolution? How fast might machines "evolve" under intelligent direction? Is there any role for the concept of "sex" in machine replicating systems?

(10) Competing machines of different types, loyalties, or functions may interact destructively. For example, machines could disassemble others and cannibalize the parts. This might be viewed as adaptive or aggressive, if the disabled machine is willing; ecological if the stricken device is already dysfunctional and of no further use, etc. Or, competing machines could inject neighbors with senility software to accelerate deterioration as a prelude to subsequent cannibalism; "Frankenstein programs" in which the infected machine returns to its point of origin and adversely affects its creators; "hidden defect programs" which cause output of defective product so that the affected machine will be retired early; or "virus programs" which cause the host machine to begin producing output as directed by the invader to the exclusion of all else.

5K.3 Cost Effectiveness

(1) What are the proper tradeoffs among production, growth, and reproduction? Should these proceed serially or simultaneously? Should the LMF be permitted to grow indefinitely, or should useful production be siphoned off from the start? How big is big enough? What are the tradeoffs between "litter size" and number of generations in terms of efficiency and cost effectiveness? How long a replication time or doubling time is economically acceptable and feasible? Are there "diseconomies of scale" that make a small seed factory difficult to achieve? Should whole systems, or just their components, be replicated? At what point should factory components specialize in particular functions? Should these components be permitted to replicate at different rates within the expanding factory complex? What is the optimum mix of special-purpose and general-purpose robots? What are the other relevant factors involved?

(2) How and under exactly what conditions can a replicating system "exponentiate"? What should be exponentiated – economic value, number of items, quality, or complexity of product? What are the fundamental limitations and most significant factors? What are the important considerations of reliability, mean lifespan, replication time, unit and system costs? How does component reliability relate to replicating system lifespan? Multiple redundancy increases the mean time to failure but concurrently increases system complexity, which might lead to higher costs and added difficulty in overall design and coordination. How can error propagation in SRS be quantified and analyzed mathematically? Should evolutionary biological notions such as "mutation" and "survival of the fittest" be made a part of SRS designs?

(3) How can closure be defined, studied, and achieved? What are the different aspects of closure? How can closure be demonstrated? Is less than full closure acceptable in some applications? What might be the principles of "closure engineering"? To what extent should/can/must reproducing machines be energetically and materially self-sufficient? How many "vitamin parts" can be imported from Earth and still retain economic viability? Can artificial deposits of special materials be created on other worlds for the convenience of SRS machines (e.g., crash a comet into the Moon)?

(4) What sorts of useful output might self-replicating robot systems produce? Would there be an emphasis on services or products? Would terrestrial or extraterrestrial consumption dominate?

5K.4 Man and Machine

(1) What is the most appropriate mix of manned and automated functions in complex, self-replicating machine systems? Does this optimum mix vary from mission to mission, or can certain general categories be established?

For manned functions, what is the most efficient mix of physical and mental labors?

(2) What is the cost tradeoff between man and machine? Is, say, a fully automated lunar factory cheaper to design, deploy, and operate than one which is fully manned, or remotely teleoperated? Is a lunar base populated by humans cheaper than a "colony" of replicating machines? Is the oft-heard assertion that "in a factory with automation, productivity is inversely proportional to the number of human workers involved" true? What should be the ratio of biomass/machine mass in SRS factories?

(3) Is it possible that very highly advanced machines could evolve to the point where humans could no longer understand what their machines were doing? Would "their" interests begin to diverge from ours? Would they replace us in the biosphere, or create their own and not displace us? Would they keep us happy, feeding us the information we request while spending most of their time on higher-order operations "beyond our understanding"?

CHAPTER 6

TECHNOLOGY ASSESSMENT OF ADVANCED AUTOMATION FOR SPACE MISSIONS

A principal goal of the summer study was to identify advanced automation technology needs for mission capabilities representative of desired NASA programs in the 2000-2010 time period. Six general classes of technology requirements derived during the mission definition phase of the study were identified as having maximum importance and urgency, including autonomous "world model" based information systems, learning and hypothesis formation, natural language and other man-machine communication, space manufacturing, teleoperators and robot systems, and computer science and technology.

The general classes of requirements were individually assessed by attempting to answer the following sequence of questions in each case:

(1) What is the current state of the relevant technology?

(2) What are the specific technological goals to be achieved?

(3) What developments are needed to achieve these goals?

After the mission definition phase was completed, summer study personnel were reorganized into formal technology assessment teams with assignments based on interest and expertise. The results of this activity are summarized below.

6.1 Autonomous World Model Based Information Systems

The first assessment team considered the technology necessary to autonomously map, manage, and re-instruct a world-model-based information system, a part of which is operating in space. This problem encompasses technology needs for a wide range of complex, computerized data systems that will be available twenty or thirty years hence. The concept of a world model aboard a satellite operating without human intervention appears useful for a variety of satellite missions, but is specifically required for the terrestrial applications IESIS (Intelligent Earth-Sensing Information System, see chapter 2) and Titan exploration (see chapter 3) missions defined during the summer study.

The world model in space serves as a template by which to process sensor data into compact information of specific utility on Earth. It can consist of mapping data and modeling equations to describe, by past experience, the expected features the spacecraft will encounter. The use of the model requires algorithms in conjunction with the spacecraft sensors. A companion central model of higher sophistication will be required to further process, analyze, and disseminate the information and to update the entire world model. In the IESIS this component is on Earth and in the Titan mission it is centered around Titan, but in either case the entire system requires autonomous management.

The following are some very general requirements for an autonomous world model-based information system; however, only the first two are discussed here.

- Development of mission specific philosophy for handling the mission data
- Model of the user and user requirements
- Realistic mission simulation techniques to test mission designs
- Modular satellite components
- Satellite serviceable in space
- Fault-tolerant design
- Autonomous navigation assistance
- Communications network
- Autonomous pointing, navigation, and control
- Standardized software to run and maintain satellites
- Data return

It is obvious that each NASA space mission should have specific information goals and that the data handling required in each must suit those goals. Costly data transmission and storage beyond that strictly required for mission operations should be eliminated. The sensor set adopted for a mission and its use must directly serve mission goals.

The goals of the Titan mission differ widely from those of the intelligent Earth-sensing system. In comparison with Earth, Titan is basically unknown. The space exploration mission goal is generally to explore and to send back as much general information as possible to terrestrial researchers about Titan. The Earth is much better known, so a major IESIS goal is to return very specific information in response to user requests or system demands. (In this latter mission, raw pixel data should be returned only under very restricted circumstances. Users requiring raw data should pay a premium for it and should accept archiving responsibility as well.) Each mission will develop a uniquely relevant data-handling philosophy. This, of course, presupposes that models are available of mission users who are the final recipients of the data.

TABLE 6.1.– SOME SIGNIFICANT LANDMARKS IN WORLD MODEL CONSTRUCTION

Year	Landmarks
1988	Autonomous on ground construction and test of a world model directly from advanced Landsat data
1990	Shuttle demonstrations of intelligent satellite system begin
1990	Primitive world model for Titan mission
1992	Completed user models by opening advanced Landsat ground test to selected user
1994	Autonomous satellite demonstration
1995	Titan intelligent demonstration mission launch
2000	Start of Intelligent Earth Sensing Information System

Table 6.1 lists a few milestones in the production of a completely autonomous and sophisticated satellite world model system. The Titan mission proposed in this report would be scheduled for launch in about 1995 and the Earth-sensing system would go into operation in 2000 AD, although a more primitive version of the world model could be ready by 1990. Since Titan is largely unknown, its world-model system must be capable of constructing a database almost entirely from first-hand on-orbit observations of the planet, hence should most properly be termed a "modeler." The Titan modeler and Earth model initially will be developed autonomously on the ground using incoming imaging data from an advanced Landsat-type satellite using conventional computers, memory, and Space Shuttle demonstrations (Spann, 1980). Test operations will characterize the operation of world model systems, and as testing continues the Earth model portion can be opened to selected users for terrestrial applications purposes. User access will allow development of worthwhile user models for the forthcoming IESIS mission (Rich, 1979). If the world model programs are successful, launch of the Titan modeler could take place in 1995 and initiation of IESIS could begin in the year 2000.

The important features in the operation of the world model arranged from its internal database through its construction, sensing, management, and user interface are:

- Techniques for autonomous management of an Intelligent Satellite System
- Mapping and modeling criteria for creation of a compact world model
- Autonomous mapping from orbital imagery
- Efficient rapid image processing techniques against world models
- Advanced pattern recognition, signature analysis algorithms for multisensory data-knowledge fusion
- Models of the users
- Fast high density computers suitable for space environment.

Autonomous hypothesis formation and natural language interfaces are important additional techniques discussed in detail in the remainder of this report, and a summary of specific recommendations of the remaining sections are in the following categories:

1. Land and ocean models
2. Earth atmosphere modeling
3. Planetary modeling
4. Data storage in space
5. Automatic mapping
6. Image processing via world model
7. Smart sensors
8. Information extraction techniques
9. Active scanning
10. Global management of complex information
11. Systems plan formation and scheduling

6.1.1 Land and Ocean Database

Each world model is specific for a given mission goal. For a land-sensing Earth mission the satellite model may be as simple as a flat map with discrete "niches" specified by type, coordinates, rough boundaries, and nominal sensor and characteristic values. The niche type may be separately catalogued and a file stored of important niche characteristics, sensor combinations useful in determining boundaries between two niches, normal anomalies, and information extraction and sensor-use algorithms. Sensor combinations most useful in determining niche boundaries must be developed. The ground component of the model will be more advanced, combining finer detail, historical data, local names, seasonal and temporal information, and complex modeling equations. Oceanic (and atmospheric) components of the world model will require sophisticated dynamic representation.

The satellite model is the component of the world model used for direct on-board processing. Without the satellite component, it is not possible to accomplish the very large data reduction inherent in model-based systems. The satellite model must be stored so that it is compact, sensor specific, capable of updating, and consistent with its use in image processing and in the particular orbit overpass.

In the image processing on board the satellite, the large number of pixel elements spanning a niche in each sensor is replaced by a small set of niche sensor characteristics such as area, average value, variance, slope, texture, etc. A highly convergent representation of desirable descriptors is required so that these few niche-dependent characteristics can faithfully represent the multitude of pixel points.

Oceanic niches and cells are commonly quite large. The ground model is necessary as a master for updating and against which the various satellite models may be calibrated. The ground model includes the library and archiving functions of the whole system.

Land and ocean database technology requirements are:

- Identification and characterization of important niches on land, ocean, atmosphere, and in boundary regions between
- Optimum niche size for use in space image processing
- Determination of well separated, easily identified niches to serve as geographical footprints
- Compact representation of niche boundaries
- Optimum sensor combinations for each niche
- Optimum sensor combinations for boundaries
- Anomaly specifications for niches
- Convergent set of niche specific characteristics
- Nominal values for niche characteristics in each niche in each sensor and for various sensor combinations such as sensor ratios
- Dynamic models for temporal variations of land, ocean, and atmospheric niches
- Optimum coordinate system for storing world model for computer readout in strips during orbital pass
- Optimum distribution of a complex world model within a multicomponent system
- Advanced data cataloguing
- Models of the users and their requirements.

6.1.2. Earth Atmospheric Modeling

The choice of sensor measurements most appropriate for terrestrial meteorological monitoring will require great advances in our present understanding of the atmosphere. Because of their dynamic and highly interactive character, the boundaries of homogeneous atmospheric three-dimensional niches will be far more difficult to define than surface niches whose features are essentially stationary by comparison. An important stage in the development of the Intelligent Earth-Sensing Information System will thus be the definition of useful niche concepts. Choosing measurements important for monitoring the Earth's atmosphere will also require great advances in present understanding of both lower and upper atmospheric phenomena.

Lower atmosphere. Examples of possible lower atmosphere niches might be regions where (a) certain temperature or pressure regimes such as low-pressure cyclones are operative, (b) there is a concentration of a particular molecular species, or (c) there is a characteristic cloud pattern indicative of an identifiable dynamic process. Such niches will often overlap, being highly interactive and transient. If the concept of a niche is to be efficient its boundaries should be essentially independent of its major properties, although property dependent niches could also be very useful. Lower atmospheric niches are time-varying in size and location, constantly appearing, disappearing, and merging.

The size of each niche will depend partly on the complexity of the atmospheric region and partly on requirements for effective monitoring or modeling. For example, atmospheric niches near the Earth's surface will undoubtedly be smaller than those in the upper atmosphere because of the complexity of surface weather patterns and of the need to have detailed niche descriptions to develop adequate meteorological models.

Properties measured in lower atmospheric niches will include a wide range of parameters – pressure, temperature, humidity, cloud cover, wind speed, rainfall, atmospheric components, etc. Each property will have its measured values processed within the three-dimensional niche in a useful niche-dependent manner. This may be used to extract data showing, for example, the average rainfall in a niche area, its gradient toward niche boundaries, patchiness of the rainfall pattern, and higher-order characteristics. Since the atmospheric niche sizes are large, the savings from averaging three-dimensional data can be huge. To ensure that niche properties such as rainfall are faithfully reproduced over the niche, the number of higher-order characteristics such as Fourier components of the data may be large, perhaps several hundred.

Another alteration of the land sensing concept must be made when comparing incoming observations to a resident world model on board the satellite. To meet Earth's needs, satellite descriptions of local weather should be continually updated together with models of the processes involved so that predictions may be made. The ephemeral and interrelated nature of many of the weather pattern-defining niches will make comparisons of current with previous observations difficult to interpret. The changing values and spatial extent of niches characterizing temperature, moisture, or pressure must be understood within the context of a complete model of weather activity. If local weather models are part of the resident world model, elaborate adaptive modeling must occur on board to correlate the incoming niche observations and to fit them into a model. In the case of lower atmospheric weather, it may be most efficient to transmit complete niche descriptions from every pass of the satellite for on-ground modeling to determine, say, the appearance of storms (high and low pressure areas) using complex pattern recognition algorithms, weather expert systems, and large computer storage.

Niches which are large or do not involve complex interfacing or modeling in conjunction with other niches lend themselves more easily to comparison with world land models. Changes in large-scale gradients and global trends in

temperature, pressure, particulates, cloud cover, and rainfall could be detected by continuous matchings with an atmospheric world model. On a smaller scale the detection of an increase in concentration of a particular molecular species (e.g., an SO_x pollutant) could also be made by simple comparison. Table 6.2 summarizes possible categories of large-scale spatial niches.

Upper atmosphere. Earth's upper atmosphere involves complex chemistry, photochemistry and transport processes. Although significant progress has been made in understanding these processes, there is still a great deal of uncertainty in present knowledge of the stratosphere, mesosphere, and lower thermosphere.

The upper atmosphere covers the range of 15 to 150 km in altitude. Absorption and emission of radiation occur over a wide range of the electromagnetic spectrum at these heights. Satellite systems presently exist or are in the planning stages which perform high resolution passive radiometry measurements in both down-looking and limb-sounding modes of vibrational (IR) and rotational (mm) molecular transitions. Limb-sounding microwave techniques will for the first time allow study on a continuous global scale since spectral lines are observed in emission. Microwave receiver technology is rapidly advancing to submillimeter wavelengths which will enable the measurement of many additional minor atmospheric constituents that play a part in radiative transfer processes. Distribution of such constituents is determined by various chemical and photochemical reactions and by atmospheric motions on both small and large scales.

TABLE 6.2.– POSSIBLE CATEGORIES OF PROPERTIES OF LARGE SCALE SPATIAL NICHES USEFUL FOR EARTH MONITORING

- Humidity profiles
- Precipitation location and rates
- Air pressure profiles and gradients
- Air temperature profiles and gradients
- Clouds – cloud top temperature, thickness, height, extent/location, albedo
- Atmospheric electrical parameters – lightning, magnetospheric electric field
- Atmospheric winds
- Aerosol size and concentration
- Particulate size and concentration
- Oxidant levels
- Molecular species – natural and man-generated

$CFCl_3$	HCl
CF_2Cl_2	HF
CF_3Cl	HNO_3
CH_4	H_2O
$ClONO_2$	HN_3
CO	NO
C_2O	N_2O
CO_2	SO_2
	O_3

- Also: Atmospheric transmittance, solar constant, solar flare activity, solar particle detection, Earth radiation budget

The current research interest in the field of atmospheric studies reflects the present level of understanding of the atmosphere. Of particular importance are measurements improving the knowledge of how man's increasing technological activities may perturb stratospheric processes and affect the maintenance of the stratospheric ultraviolet-shielding ozone layer. These upper atmospheric studies require long-term precision composition and thermal measurements. An understanding of the role of the stratosphere in climatic change and atmospheric evolution is also needed. This includes understanding stratospheric warnings, their impact on chemistry, the spatial distribution of aerosols, and interactions with the troposphere below and the mesosphere above. Measurements of the mesosphere and lower thermosphere are needed to determine composition and variability. Little is known of the basic meteorology in these regions (temperature, pressure, wind variations). Possible variations in O_2 in these levels may affect ozone concentration at lower altitudes.

The long-term goal is the development of an intelligent Earth-sensing information system which can compare synopses of complex numerical models of the upper atmosphere with specific observations which are a subset to the original observations required to design those models. Comparisons could be simply the matching of predicted or acceptable values with observations. The actual models to be flown will have varying degrees of complexity. Most models may just be listings of predicted values derived from complex numerical models. These listings could be compared with observed values (for developmental purposes). Subsequently, measurements might be reduced to those spectral lines which yield the most information with optimum redundancy. For the purpose of testing systems which will be flown on the Titan mission it will be necessary to fly models which are or can be self-modifying to account for any observed discrepancies.

Since the Earth's atmospheric modeling will be done in much greater detail than is necessary for planetary exploration, tests of adaptive radiative transfer and hydrostatic equilibrium modeling should be kept simple. In planetary exploration, relatively crude remote sensing to determine composition, winds, atmospheric structure, cloud cover, and temperature profiles will be needed to obtain a general understanding of the planet's atmosphere. However, it may be valuable to include complex modeling systems to explicate possible organic chemistries.

To reach the required level of understanding of the atmosphere, extensive studies must be undertaken to

develop and validate complex models complete in their inclusion of aerial chemistry, distribution of minor constituents, radiation fields, and large-scale dynamics as a three-dimensional time-dependent problem. When the upper atmosphere is sufficiently understood, appropriate parameters to be monitored and modeled can be determined. Useful techniques for verifying models will involve checking model predictions with the observed distribution and concentration of chemically active species (some of which may also be useful as tracers of atmospheric motions).

Current planning for the versatile microwave limb-sounders seems to be moving in a direction compatible with Earth and planetary sensing requirements. The radiometers will be modularly constructed so that they can be easily exchanged as measurement priorities change and technology advances. Limb-sounder instruments will probably be capable of accommodating several radiometers for simultaneous measurements. Instruments in different spectral ranges will be employed for complementary measurements. The antenna, scanning, data handling, and power supplies should be common to any complement of radiometers used in the system.

The Earth atmospheric modeling technology requirements are:

- Definition of lower and upper atmosphere niches (spatial location or characteristic properties)
- Adaptive modeling of weather
  - complex pattern recognition algorithms
  - weather expert system
- Sensors for measuring lower atmospheric properties
- Determination of set of properties in an atmospheric niche which give consistent boundaries
- An understanding of the atmosphere sufficient to know what parameters need to be monitored
  - development of high resolution satellite microwave techniques for measurement of minor constituents
- Use of microwave limb-sounding techniques for continuous global coverage
- Development of an optimum sensor set for monitoring the upper atmosphere.

6.1.3. Planetary Modeling

For a relatively unknown body, surface and atmospheric modeling must evolve in greater detail during the course of the mission as more information on important planetary characteristics is obtained. A systematic methodology is required for understanding and exploring a new environment using high sensor technology. This methodology must determine what questions should be asked, and in what order, to efficiently and unambiguously model an uncharted atmosphere and planetary surface. A decision must be made early in the planetary mission whether to place emphasis on elaborate remote sensing from orbit, which may ensure survivability but will not allow all of the scientific objectives to be met, or to physically probe the atmosphere, thus exposing a mission component to increased danger but allowing more precise determination of useful atmospheric properties.

The planetary probe must be capable of orbiting, investigating, and landing during a single mission. This is a difficult task to accomplish in one fixed design because of the uncertainties in the nature of the unknown planetary environment. The resulting planetary modeling requirements are:

- Systematic methodology for exploring an initially unknown environment
- Decision ability in the face of lethal dangers
- Modeling ability to establish norms of a planetary surface which allow recognition of interesting sites
- Autonomous creation and updating of planetary models using a variety of complementary sensors
- Adaptive programming of atmospheric modeling to establish atmospheric parameters
- Complex modeling or organic chemistry processes
- Expert systems for spectral line identification of complex and ambiguous species
- Develop general spacecraft capable of adaptation under uncertain atmospheric and surface conditions and which possesses a broad set of sensors
- Exchangeable radiometers, each capable of simultaneous measurements, and with wide spectral range and self-tuning ability
- Mass spectrometers and radio spectrometers based on range of organic compounds considered important or highly probable
- Instruments with interchangeable and reconfigurable basic elements
- Development of space qualified subsystems, instruments with long lifetimes
- Development of smart probe sensors and high speed image processors able to operate in the short period of time available during descent
- Use of sensors which record only significant variations in incoming data
- Simple redundancy so spacecraft will not be overloaded with back-up instrumentation
- Automated failure analysis systems, self-repairing techniques.

These requirements will engage numerous disciplines and thus create challenging instrumentation and design engineering problems for mission planners.

6.1.4 Data Storage

The terrestrial world model will require satellite storage of from 10^{10} to as much as 5×10^{11} bits and perhaps 10^{14} bits on the ground. Forecasts (Whitney, 1976) give estimates of 10^{14} bits of in-space memory and $>10^{16}$ bits of a typical on-ground memory by the year 2000. The data storage should be structured in a manner compatible with build-up of an image and extraction of image processing during orbital overpass. Optical disc, electron beam, and bubble memories are possible candidates in addition to more conventional alterable memories.

The technology requirements include a high density, erasable memory suitable for use in the space environment, optimum memory architecture for readout of the world model during orbit overpass, and error-correcting memory design.

6.1.5 Automatic Mapping

Terrestrial automatic mapping by IESIS can be accomplished using geographical data already obtained from Earth or from satellite data alone. The Defense Mapping Agency has developed digital mapping techniques for regions of the globe (Williams, 1980). By contrast, the mapping of Titan must be accomplished almost exclusively from orbit. In either case, information in the form of niche identification, basic modeling equations, and known planetary parameters will be supplied from Earth both initially and during operations. Automatic mapping from space requires state-of-the-art AI techniques including boundary and shape determination, optimum sensor choice, niche identification, and learning techniques. Full autonomous learning by abduction and inference to build new knowledge is presently beyond the capability of AI (see section 6.2). Though use of such advanced AI techniques would tremendously enhance the utility of a satellite world-model-based information system, they are not considered essential in this application.

Mapping technology ultimately must prove sensor-independent since the map produced should reflect a reality existing in the absence of the sensor data. However, specific sensor combinations will produce a completed map more rapidly and reliably depending upon the niche environment to be mapped. Orbits which repeat over fixed portions of the planet are especially advantageous in assisting automatic mapping and memory structuring.

Technology requirements, summarized briefly, are:

- Rapid autonomous mapping techniques from orbital data
- Optimum sensor combinations for reliable and rapid mapping
- Determination of relative advantage of radar, optical, IR mapping
- Optimum orbit height and orbit type for automatic mapping
- Techniques to rapidly, reliably, and automatically update world model components in satellites and on ground directly from orbital image data
- Digital mapping techniques
- Autonomous hypothesis formation techniques.

6.1.6 Image Processing via World Model

The satellite memory component of the world model is used for image processing. The actual image data from one or several sensors must be cross-correlated with a pass map (retrieved from memory) in strips along the orbit to produce an optimum match of imaged niches with their mapped locations. This process rectifies the sensed image and produces geometrical corrections necessary to adjust the sensed image to the reality reflected in the stored map. This process also will help determine the precise satellite location (Kalush, 1980). Boundaries must then be identified from the actual imagery and compared to the nominal boundaries of the map. The boundary area is an important and simply determined characteristic of the niche. Other characteristics such as anomalies are determined by new boundaries, altered location of boundaries, or changes in the determined sensor readings from their expected values. Temporal, sensorial, and solar corrections must be applied to the sensor readings and defining labels supplied for all niche characteristics for complete referencing purposes (Schlienn, 1979).

The satellite location can be combined with velocity and navigation information from a global positioning system to prepare for the next image in the sequence. This preparation allows minimum processing in the subsequent image rectification and permits determination of the optimum sensor combination for the next imagery. Instructions from Earth ground control or a central satellite autonomous manager must be incorporated into the preparation and image processing procedures.

Very sophisticated computer technology is required aboard the satellite to accomplish the image processing. Such processing is not found on any present-day satellites, and is done on-ground only in very limited form today. Fully parallel processing techniques are anticipated as a possible alternative to serial processing (Gilmore et al., 1979; Matsushima et al., 1979; Schappell, 1980). Optical processing methods should also be investigated since these

techniques are naturally parallel (Vahey, 1979). Technical advances are needed in the following:

- Automatic techniques to rapidly correlate memory stored mapping and modeling information with visual and radar imagery obtained in orbital pass
- Fast image enhancement and threshold techniques
- Rapid cross-correlation techniques
- Rapid boundary-determination techniques
- Rapid Fourier transform techniques
- Algorithms for improved automated data associations
- High density rapid computers for use in space environment
- Parallel processing computer techniques involving large wafers, advanced cooling techniques, advanced interconnection techniques between array elements, more logic functions between elements performed in one clock cycle and advanced direct data output from array to a central controller
- Ability to load and unload imaging data in full, parallel manner at all stages of handling raw data
- Investigation of possible use of optical processing techniques such as holographic process or integrated optics for satellites processing of imagery via world model
- Techniques to rapidly, reliably and automatically update world model in satellite and on ground directly from image data
- Advanced data compression and compaction techniques for data transmission and storage.

6.1.7 Smart Sensors

Complex sensor configurations are required for both IESIS and Titan missions. A high degree of autonomous sensing capability is needed within the sensors themselves (Haye, 1979; Murphy and Jarman, 1979). These sensors must be smart enough to perform automatic calibrations, compensations, and to reconfigure themselves automatically – tasks requiring advanced memory capabilities and operating algorithms (Schappell, 1980). Desirable characteristics of such smart sensors on satellites (Breckenridge and Husson, 1979) are:

- Introduces no anomalies into data
- Performs analytical and statistical calculations
- Performs all operation in simplest form
- Adapts (handles) new data acquisition and processing situations
- Interactive sensor configurations
- Adjusts to different environmental conditions
- Extracts information in a useful form
- Makes decisions.

The use of a world model in conjunction with smart sensors would confer an extraordinary degree of intelligence and initiative to the system. In order to mate sensors most efficiently with the world model, the model should itself possess models of the sensor components. Since the sunlight at Titan is weak and the planet cold, efficient, visible, and infrared sensors are also necessary.

Technology requirements of smart sensors are:

- Advanced efficient solid-state imaging devices and arrays
- Sensor operation at ambient spacecraft temperature
- Electronically tunable optical and IR filters
- Advanced automatic calibration and correction techniques
- Distributed processing sensors
- Rapid, high responsivity detectors in near IR up to 3 μm
- Optimum set of sensors arrays for particular planetary mission
- Sensor models
- Silicon-based sensors with dedicated microprocessors and on-chip processing
- Investigation of piezoelectric technology for surface wave acoustic devices
- Sensor sequence control which can adapt to conditions encountered
- Precision pointing and tracking sensor mounts.

6.1.8 Information Extraction Techniques

Information can be extracted from sensory data originating from an object by recognizing discriminating features of the object. Such features are of three kinds: (1) physical (color), (2) structural (texture and geometrical properties), and (3) mathematical (statistical means, variance, slope, and correlation coefficients).

Humans generally use physical and structural features in pattern recognition because they can easily be discerned by human eyes and other common means. However, human sensory organs are difficult to imitate with artificial devices so these methods are not always effective for machine recognition of objects. But by using carefully designed algorithms, machines can easily extract mathematical features of patterns which humans may have great difficulty in detecting.

The algorithms will often involve a fusion of knowledge across multisensor data. As an example, the radiance observed from an object is a function of its reflectance,

incident illumination, and the media through which it is viewed. The ratio of the radiances at two different wavelengths can be used to separate water and vegetation from clouds, snow, and bare lands (Schappell and Tietz, 1977; Thorley and Robinove, 1979). (The radiance ratio for clouds, snow, and bare land is essentially the same, so these features must be separated on the basis of absolute radiance.) These two sensor procedures can be improved by using data from several sensor bands simultaneously in a multidimensional sensor space. Machines can process such complicated algorithms and "see" clusters in higher dimensions. The higher the multidimensional volume, the more accurate the discrimination between closely related sensorial characteristics.

The intelligent use of a world model requires autonomous, real-time identification of niches (through their features) and determination of characteristics. Real-time pattern recognition and signature analysis also must be accomplished to supply useful information to the user. Algorithms should be developed for identification, pattern recognition, and signature analysis.

Statistical procedures arise naturally in various classification schemes because of the randomness of data generation in various pattern classes. Statistical theory can be used to derive a classification rule which is optimal because it yields the lowest probability of classification error, on average. Various studies have developed decision functions from sets of finite sample patterns of classes. These decision functions partition the measurement space into regions containing clusters of the sample pattern points belonging to one clan. Some clustering transformations have been used in the development of such functions. Once a function has been selected, the main problem is the determination or estimation of its coefficients. For efficient coefficient estimation, time-dependent training samples are needed.

A wide variety of additional algorithmic techniques are needed. For example, texture analysis can be accomplished using gray-tone statistics and the time rate of change of spatial contrast along scan lines to distinguish among wheat, rye, and oats (Haralick et al., 1974; Mitchell et al., 1977). Below is a summary of technology requirements:

- Rapid methods for area centroid and orientation determination
- Rapid partitioning of image features
- Motion and relative motion detection
- Development of wide range of classification algorithms for user-defined applications
- Multispectral signature ratioing analysis and multisensor correlations
- Rapid texture analysis
- Investigation of usefulness of focal plain transformations for satellite use
- Schemes to allow disparate algorithmic techniques to interact to speed recognition process
- Determination of parameters of decision functions for various classification schemes

6.1.9 Active Scanning

The sensors discussed to date have been essentially passive – they do not generate the radiation they detect. For a variety of purposes, some satellite systems will engage in active scanning by highly efficient RADAR or LIDAR, all-weather imagery, night-time imagery, absolute and differential height determination, absolute and differential velocity determination, atmospheric probing, and leading-edge scanning. Of course the mission to Titan, relatively far from the Sun, will not have large amounts of power available for this purpose.

Additional technology requirements include a fast, efficient computer for generating imagery from SAR, the ability to determine height differentials to within several centimeters at boundaries, and the ability to determine differential velocities to within about 1 km/hr at boundaries.

6.1.10 Global Management of Complex Information Systems

Each mission explored by the study group consists of a very large, complex array of equipment and people widely geographically distributed, all of whom must work in a cooperative and coordinated fashion to achieve mission objectives. An important concern thus becomes the overall architecture of such a system, the way decisions are made and communicated, the coordination of tasks within the system, and the flow of information. These types of difficulties are not new in human endeavors and have been addressed within several disciplines which focus on specific aspects of the problem. A brief review of relevant fields resulted in several recommendations for high priority research in systems theory and control, summarized below.

Classical control theory. Systems which evolve according to well behaved physical laws describable in the form of differential equations have long been the domain of classical control theorists. The aerospace industry has been a prime user of this technology in the guidance and control of missiles and in the development of automatic pilots for aircraft. The system is usually modeled as shown in figure 6.1 which envisions an idealization of a physical system subject to stochastic disturbances (typically Gaussian). The system is observed and digressions from the preferred trajectory are noted. A controller working with the idealized model (expressed in the form of differential equations) and a specific objective (such as "hit a target within a given

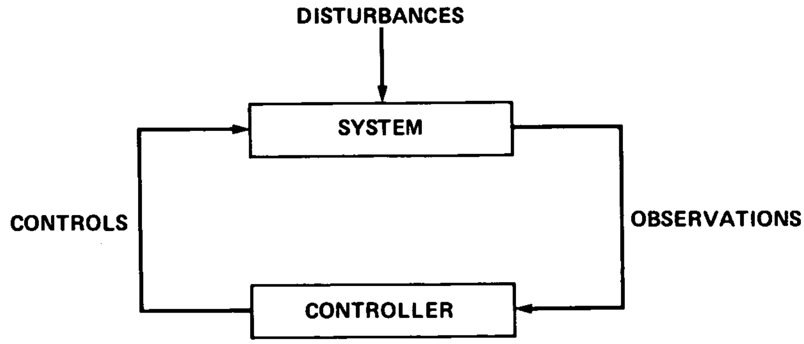


Figure 6.1.– Classical control theory systems model.

tolerance with a limited fuel budget") typically expressed mathematically in a quadratic form computes linear control to correct the system trajectory to meet the stated objective. This type of formulation is known as the LQG (Linear Quadratic Gaussian) formulation and has received wide attention within the control theory community.

Clearly, this theory is applicable to navigation and process control problems but will make only a rather minor contribution to the theory of how systems operate as a whole. This is not considered a critical mission technology since it is a fairly well developed and active field. Further, application depends on the notion of a single centralized controller. This is appropriate for micro control applications but inappropriate for macro control of large decentralized systems.

Game theory. Systems which employ multiple decisionmakers have been addressed by game theorists. Much of this work has been defense-related although economics has also provided an applications base. The basic notion of game theory is that there are an arbitrary number of decisionmakers, each of whom has an individual objective-function which may be (and likely is) in conflict with the objectives of the other decisionmakers. Each decisionmaker attempts to develop strategies which independently maximize the "payoff" to himself. Much work has been done on the "zero sum game" in which one decisionmaker's gain is another's loss.

If one envisions a cooperative, coordinated mission scenario, the current focus of game theory on threat strategies more appropriate to hostile environments seem ill-suited to peaceful space activities. A more appropriate meta-model is required for NASA's applications which reflects the necessity for cooperative coordination among the men and machines of the mission.

Nonclassical information control theory. The decentralized control problem for large-scale systems with a common (or at least coordinated) objective has received increasing attention in recent years. Initial work on "team theory" (Radner, 1962) has centered on a team which is considered to have as its fundamental problem the coordination of decentralized activities utilizing delayed and imperfect information. The meta-model employed appears to be appropriate for the large-scale space missions considered in this report. Team theorists envision an autonomous ensemble of decisionmakers, each of which senses a local environment ("perfect" information) and can communicate in a delayed fashion with other decisionmakers ("imperfect" information). The ensemble, or team, shares a common objective and attempts to communicate as necessary for collective progress toward that objective to be optimized in some sense. This leads to the notion of an information structure among the members of the team.

The team concept has since been adopted within the control theory community and has led to "nonclassical control theory" – control theory which addresses multi-decisionmaker types of problems (Ho, 1980; Sandell et al., 1978). Much of this work is supported heavily by the Department of Defense (DOD) and focuses on problems of little direct relevance to NASA. Vigorous support by NASA of work in nonclassical control theory is recommended to develop more appropriate theories for the types of systems which comprise the space missions of the future. For instance, much of the DOD work addresses guidance and control problems, whereas NASA's prime interest would be more appropriately in information systems control. Supporting disciplines include probability and Markov decision processes. These are areas which are required to advance the state of the art in systems theory and control and to apply it effectively to NASA missions.

Prior work in these fields tends to focus on performance optimality as an objective. While optimality is a laudable goal, it is not clear that this should override other concerns such as stability and performance predictability. The formal tools currently available to evaluate the stability of a large decentralized system are virtually nonexistent. The major recommendation of the study group in this area is that NASA seriously consider the system-wide objectives of its future systems and support a program of basic and applied

research which develops the theoretical basis to achieve these objectives. The major relevant disciplines include, as a minimum set, nonclassical control theory, probability, queueing theory, and Markov decision theory. Technology requirements include the determination of system-wide objectives of missions and the development of theoretical and practical bases to achieve those objectives; development of nonclassical control theory of complex man-machine information systems; probability theory applicable to complex information systems; and Markov decision theory for complex information systems.

6.1.11 Plan Formation and Scheduling

In those cases where robots are called upon to operate in sophisticated task environments, the machine system first performs some computation which can be considered problem-solving, then takes action based upon the problem-solving result which is called the "plan formation" process. The part of the resulting plan which identifies the times at which actions are to occur is called the schedule. Whether the machine system is a relatively small mobile robot as might be used in planetary surface exploration, or a large distributed intelligence such as an Earth-sensing information system, several common features dominate in achieving effective and flexible operation (Sacerdoti, 1979):

- The ability to represent the state of the relevant parts of the world (the "world model")
- The deductive ability to recognize consequences of a particular world state description
- The ability to predict what changes will occur in the world state, possibly due to some action or actions a complex autonomous system itself might perform.

In most realistic environments it is impossible to completely build a detailed plan and execute it in unmodified form to obtain the desired result. Several difficulties preventing such a direct line of attack are:

(1) The external reality may not be known in sufficient detail to accurately predict the outcome of some action. If the action in question is the final one in a plan, then it may not achieve the overall intention of the plan. If it is an earlier action in a several-step plan, then it may not produce a required intermediate state for the overall sequence of actions to achieve the goal of the plan. If the goal is to make an observation to obtain information about the environment, the information obtained may not be adequate.

(2) Even if a perfect, or effectively perfect, model of the external environment is available to the robot, there may still be inaccuracy associated with the robot's control of itself (e.g., mechanical inaccuracy of motion).

(3) Other agents, with goals of their own, may alter the environment in unpredictable ways before the robot can complete the execution of its plan. In such cases some form of overall coordination is necessary. It is not adequate simply to have the main goals of all of the active agents compatible. Even with this precaution, it is still possible to have a contention for resources or intermediate configurations in achieving the common goal. Aside from the problem of avoiding explicit conflict among several active agents there is the inverse problem of achieving efficiency increases by proper cooperative action among the agents.

For these reasons, a robot must continually monitor the results of its actions during plan execution, and modify the plan – in essence, re-plan – during plan execution.

A further complication arises when the plans must meet real-time constraints – that is, definite short-term requirements for actions where failure to meet the timing requirements carries significant undesirable consequences. Two types of real-time constraints, "hard" and "soft," may be distinguished. A "hard real-time constraint" is that the failure to carry out a successful plan that attains the relevant goal within the limits will result in a consequence so undesirable that extreme care must be taken not to overrun the time boundary. An example in the area of large-scale space construction might be the joining of two relatively massive but fragile substructures. Failure to initiate timely deceleration of substructures approaching each other could result in large economic losses. An example of a "soft real-time constraint" is in the maximization of the utilization of a costly resource, such as the observation satellites in an Earth-sensing system where it is important to schedule observations in such a way so as to minimize the number of satellites necessary to provide a given level of observational coverage. In this case, each individual failure to meet the real-time constraints has, in general, only minor consequences, but a continuing high-frequency of failure will result in economic losses through inefficient operation.

Because of the need to re-plan during plan execution, and because of the necessity to meet real-time constraints, it is important that complex autonomous systems have plan formation capabilities well in excess of current state of the art.

Current assessment. A considerable amount of work has been done in AI on problem-solving in general, and on planning and plan execution in particular. In the last 10 years the problem-solving emphasis has shifted away from planning towards the perceptual processes of vision and speech recognition. Table 6.3 lists some techniques for problem-solving and planning, and various representational schemes (NASA SP-387, 1976).

The frame notion of Minsky initially generated much interest and discussion, but little has been accomplished to date in terms of applications. There are attempts from several different perspectives to implement frame-based languages for programming, as for example KRL (Bobrow and Winograd, 1977), FRL (Goldstein and Roberts, 1977),

TABLE 6.3.– FORMS OF REPRESENTATION AND PROBLEM-SOLVING TECHNIQUES USED IN ARTIFICIAL INTELLIGENCE

Representations

- State space (Van de Brug and Minker, 1975)
- First order predicate logic (Nilsson, 1971)
- Semantic nets (Woods, 1975)
- Procedural embedding of knowledge (Hewitt, 1970)
- Frames (Minsky, 1975)
- Production rules (Newell, 1963)

Problem-solving techniques

- Backtrack programming (Goulomb and Baumert, 1965)
- Heuristic tree search (Pohl, 1977)
- GPS means-ends analysis (Newell, 1963; Ernst and Newell, 1969)
- Problem reduction (Amarel, 1968)
- Theorem-proving (Nilsson, 1971)
- Debugging almost correct plans (Sussman, 1975)
- Procedural (Hewitt, 1970)
- STRIPS (Fikes and Nilsson, 1971)
- ABSTRIPS (Sacerdoti, 1974)
- Cooperating knowledge sources (Erman and Lesser, 1975)
- Rule-based systems, expert systems (Shortliffe, 1976)

and MDS (Srinivasan, 1976). These attempts were ambitious, and while all met with serious difficulties, the possibility remains that the problems can be overcome. An ideal frame-based programming language could make it easier to structure knowledge into larger coherent units than would otherwise be practical.

The controversy between procedural and declarative philosophies of embedding knowledge has dwindled. It is now realized that each has its particular function to perform in an overall system, and that neither alone nor in combination are they an adequate underlying basis for AI theories or for sophisticated program organization.

There is a growing trend toward considering the first-order predicate calculus, or minor modifications of it, as the primary mode of representing declarative knowledge in AI systems. The reason is that this calculus has a well defined semantics and other declarative representation schemes tend to be simply different notional systems struggling to capture the same semantic notions as the predicate calculus.

Interest in formal theorem-proving techniques has remained high, and perhaps has even increased slightly, despite the slow progress in increasing the efficiency of mechanical deduction. While theoretical understanding of mechanical theorem-proving is increasing, to date there is little advancement in efficiency beyond that of a decade ago. Theoretical work on model use in theorem-proving has progressed only slightly (Reiter, 1972; Sandford, 1980), and applications methodologies are nonexistent.

Theoretical work has progressed in using first-order Horn logic as a programming language (Kowalski, 1974). Horn logic is a subset of the first-order predicate calculus in which a large number of interesting problems can be expressed. A truly unexpected development is the successful implementation of a workable programming system for Horn logic in which several nontrivial programs have been written (Warren and Pereira, 1977).

Much interest has developed in a rule-based type of knowledge embedding for restricted domains. These systems are commonly called expert systems, and have shown interesting and relatively strong problem-solving behavior. A variety of reasoning task domains have been implemented (Feigenbaum et al., 1971; Shortliffe, 1976), and the rule-based knowledge embedding method is robust in its performance. However, several severe defects of such systems must be addressed before realistic problem domains can be adequately handled. Defects include extremely limited domains of application, the large efforts required to construct the knowledge base, and the inability to access a basic theory and perform an *a priori* analysis. Work is in progress to devise systems avoiding these particular problems (Srinivasan and Sandford, 1980). There are also some relatively minor human interfacing problems with the present systems (see section 6.3).

There is a general increased awareness of the importance of the role of meta-knowledge (knowledge *about* knowledge) in problem-solving and in planning. The important related area of reasoning relative to open world databases is just beginning to be investigated (Reiter, 1980).

The general problem of representing the external world in an appropriate machine representation is a fundamental unsolved problem. While many facts are representable in many ways, no known representation is adequate to handle even such common phenomenon as a glass of water falling to the floor and breaking. It is likely that a fundamental shift in current approaches is required to achieve adequate representations for much of "common" world knowledge. There is little indication at this time what these new approaches should be. However, certain such "common" world knowledge is at least partially tractable with current techniques as, for example, the acquisition and use of knowledge about large-scale spaces (Kuipers, 1977).

Identification of critical research areas. Table 6.4 lists a set of critical research areas in the general AI fields of problem-solving, plan formation, scheduling, and plan

TABLE 6.4.– CRITICAL AI RESEARCH AREAS IN ROBOT PROBLEM-SOLVING AND PLANNING

1. General robot reasoning about actions
2. Combining AI problem-solving and plan formation with operations-research-scheduling techniques and real-time constraints
3. Techniques for classifying problems into categories and selecting the appropriate problem-solving method to apply to it
4. Expert systems
5. Generalized techniques for dynamic accumulation of problem-specific knowledge during a problem-solving attempt
6. Techniques for abstraction, and the use of abstraction for search guidance
7. Methods of combining several representations and search techniques together in a coherent manner
8. Ways to structure systems to have both fundamental theories to allow *a priori* reasoning along with a procedural level of skill to allow efficient real-time response
9. Models and representations of reality

execution. Table 6.5 gives the relevant mission requirements in these areas, the missions to which they might apply, and the identification of which items from table 6.4 are most relevant.

Recommended actions. Traditionally AI has been predominantly a research-oriented activity which implemented systems primarily for experimental purposes. There is a growing awareness among AI researchers that the time has come to produce limited capability but useful working systems.

In like manner, NASA should obtain experience at the earliest possible date with elementary space-robot systems in such areas as fully automatic spacecraft docking and sophisticated Earth-sensing satellites. Theoretical research in AI problem-solving and planning techniques will be an active area for several decades to come. If NASA is to become effective in directing this research toward its own goals, then early experience is necessary with elementary state-of-the-art techniques – although substantial advantages can be obtained even with relatively unsophisticated, near-term AI planning and execution monitoring techniques. Most of the areas listed in table 6.4 will progress both at the theoretical and applications levels without NASA taking action. This theory will generally be supportive of NASA's needs, particularly that done by DOD for space applications. Communication between NASA and DOD is thus important in overall planning for both organizations. While DOD interests in the mission requirements listed in table 6.5 are likely to be restricted to categories

TABLE 6.5.– CORRELATIONS BETWEEN MISSION REQUIREMENTS, MISSIONS AND RESEARCH AREAS FROM TABLE 6.4

Mission requirement, MR	Mission	Relevant areas from table 6.4
1. Automated housekeeping functions for long-duration spacecraft	TM,[a] ES[b]	1,2,4,5,7,8
2. Fully autonomous sequencing of observations, active and passive, from orbit, from landers, and during interplanetary flight, for a variety of sensors	TM,ES	2,3,4,7,8,9
3. Automatic docking, refueling, repair and maintenance of semi-independent probes	TM, ES	1,3,4,5,6,7,8,9
4. Automatic deployment of landers and orbiters from a central orbiter bus or busses	TM	1,4,9
5. Automatic landing capability on a planetary body where the lander is physically designed as a general-purpose lander capable of achieving planet fall on planets with a variety of atmosphere densities, wind velocities, and surface characteristics	TM	1,2,9
6. Automatic sample-taking of atmosphere and soil samples, and automatic low level sequencing of a variety of chemical and physical analysis techniques	TM	1,3,4,5,8

[a]Titan mission.
[b]Earth mission.

MR1, MR2, MR3, and possibly MR4, these cover most of the research areas from table 6.4. If this is indeed the situation with respect to DOD, then NASA can concentrate primarily on implementation projects.

However, certain needs and operating scenarios are peculiar to NASA and are not likely to develop in theory or applications without direct NASA guidance. Two very

specific examples are the development of robot patterning techniques and the development of "show and tell" robot control (see section 6.3). While commercial industrial robots have long employed patterning methods, these methods are used only in a rudimentary form and further applications technology development is needed for them to become useful to NASA. The show and tell mode of robot action has apparently not been identified and investigated to any large degree, and seems to be an area where NASA should take an immediate and large interest both in the theory and the applications aspects.

6.2 Learning and Hypothesis Formation

The Titan exploration mission description, documented in chapter 3, discusses the characteristics of a machine intelligence system possessing autonomous self-learning. This capability, its relation to state-of-the-art AI, and the new research directions it demands are summarized below.

6.2.1 Characteristics

For a machine to learn a previously unknown environment involves both the deployment of knowledge structures correct for known environments and the invention (or discovery) of new knowledge structures. A machine intelligence system which learns could formulate (1) hypotheses which apply existing concepts, laws, theories, generalizations, classification schemes, and principles to the events and processes of the new environment, and (2) hypotheses which state new concepts, laws, and theories whenever the existing ones are inadequate.

Different logical patterns of inference underlie the formation of these types of hypotheses. Analytic inferences support the formation of hypotheses which apply existing concepts, laws, and theories. Inductive and abductive inferences support the invention of hypotheses which state *new* concepts. Analytic, inductive, and abductive inference are mutually and logically distinct – one of them cannot be replaced by some combination of the others (see section 3.3.3 and compare Fann, 1970; Hanson, 1958; Lakatos, 1970a, 1970b, 1976; Peirce, 1960, 1966).

6.2.2 State-of-the-art in AI

State-of-the-art AI lacks adequate and complete treatments of all three inferential classes necessary for the development of machine intelligence systems able to learn in new environments.

Analytic inferences receive the most complete treatment. For instance, rule-based expert systems can apply detailed diagnostic classification schemes to data on events and processes in some given domain and produce appropriate identifications (Buchanan and Lederberg, 1971; Duda et al., 1978; Feigenbaum, 1977; Martin and Fateman, 1971; Pople, 1977; Shortliffe, 1976). An expert system such as PROSPECTOR can identify a restricted range of ore types and map the most likely boundaries of the deposit when given survey data about possible ore sites (Duda et al., 1978). However, these systems consist solely of complicated diagnostic rules describing the phenomena in some domain. They do not include models of the underlying physical processes. In general, state-of-the-art AI treatments of analytic inference fail to link the detailed classification schemes used in these inferences with the fundamental models required to deploy this detailed knowledge with maximal efficiency.

Inductive inferences receive a less complete treatment although some significant advances have been made. For example, Hajek and a group of co-workers at the Czechoslovak Academy of Sciences have, over the past 15 years, developed and implemented systems of mechanized inductive generalization (Hajek and Havranek, 1978). They do not take the approach of "inverse deduction" which has been explored by Morgan (1971, 1973). Instead, the Czech group has developed techniques for moving from data about a restricted number of members of a domain, to observation statement(s) which summarize the main features or trends of these data, to a theoretical statement which asserts that an abstractive feature or mathematical function holds for all members of the domain. (For instance, see table 6.6.) Though they allow a role for what they call "theoretical assumptions," in moving from observation to theoretical statements they have concentrated their work on formulating the rational inductive inference rules for bridging the gap between the two –

TABLE 6.6.– SAMPLE INFERENCE DATA

Rat no.	Weight, g	Weight of rat kidney, mg
1	362	1432
2	373	1601
3	376	1436
4	407	1633
5	411	2262
Observation statement	Therefore, the observed weights of the kidneys have the same order as the weights of the rats with one exception.	
Theoretical statement	Therefore, the weight of a rat's kidney is positively dependent on the weight of the rat.	

though it is not clear that their system captures the full range of influence which fundamental models exercise over inductive inference (see section 3.3.3). An independent research effort in the United States attempts to integrate fundamental models with specific abstractive, or generalizing, techniques (Srinivasan, 1980; Srinivasan and Sandford, 1980). However, unlike the Czech group the American team is still at the stage of theory development – a working system has yet to be implemented in hardware.

Abductive inferences have scarcely been touched by the AI community, but nevertheless some tentative first steps do exist. A few papers on "nonmonotonic logic" were delivered at the First Annual National Conference on Artificial Intelligence at Stanford University in August 1980 (e.g., Balzer, 1980), and much discussion followed. However, this attempt to deal with the invention of new or revised knowledge structures is hampered (and finally undermined) by their lack of a general theory of abductive inference – with one notable exception, the recent work of Frederick Hayes-Roth (1980). Hayes-Roth takes a theory of abductive inference developed by Lakatos (1976) for mathematical discovery and makes two of the low-level members of the family of abductive inferences which Lakatos identifies operational. Still, this work is only a preliminary step toward implemented systems of mechanized abductive inference and, unfortunately, it seems to represent the extent of theory-based AI work on abductive inference to date.

In summary, state-of-the-art AI treatments of analytic and inductive inference provide no fundamental models as a theoretical foundation to support the detailed knowledge structures and inference techniques upon which the treatments are built. Yet these models are an essential and integral element of analytic and inductive inferences. State-of-the-art AI virtually lacks treatments of abductive inference. However, model-based analytic and inductive inference systems and an abductive inference system are all necessary prerequisites for machine learning systems.

There appears to be a growing acceptance within the AI community of the above problems and that overcoming these gaps in current treatments of analytic, inductive, and abductive inference is an important future research direction for the entire field. For example, at the First National Conference on Artificial Intelligence, Peter Hart admitted that the fact that rule-based expert systems lack a fundamental model to ground the detailed rules makes them superficial and inflexible. Charles Rieger at the University of Maryland is beginning to address the question of layering models under rule-based systems. Several recent AI initiatives with respect to inductive and abductive inference have already been noted.

A concerted and serious attack on the problem of developing a theory of abductive inference for machine intelligence could pay enormous dividends. First, machine learning systems cannot possibly possess a full learning capability unless they can perform abductive inferences. Second, a successful mechanization of abductive inference would require the solution of problems which must also be solved for the successful mechanization of analytic and inductive inference. These problems include: (1) how to represent the fundamental models of the processes which underlie the detailed occurrences of domains, (2) how to inferentially relate these to more detailed knowledge structures such as laws, principles, generalizations, and classification schemes, and (3) how to map the representations of a domain occurrence in one "language," say, that of the model, onto its representation in another "language," say, that of a set of diagnostic rules. Since an investigation of abductive inferences seems to hold many keys to solving the problem of machine learning, and since recent developments in AI seem to promise receptivity to such an investigation, the development of a theory of abductive inference for machine intelligence appears to be the preferred research direction for work on machine learning systems.

6.2.3 Two Barriers to Machine Learning

Two points from the above discussion must be emphasized. First, state-of-the-art AI work on hypothesis formation is almost totally devoid of research on abductive inference. However, machine systems must have this capability in order to be true learning systems. Second, current AI work on analytic and inductive inference tends to proceed in the absence of relevant theories, and this seems to be the reason why state-of-the-art AI treatments fail to give fundamental models their proper role in inference systems. However, adequate theories of all three types of inference are a necessary foundation for successful machine learning systems.

Both of these barriers to machine learning – the abductive inference barrier and the theory barrier – must be bridged before machine intelligence systems can be given a full learning capability. The abductive inference barrier has already been fully treated, but some additional discussion of the "theory barrier" is useful here.

Historically, technology has developed in two distinct patterns – empirical and theoretical. Empirical technology is a "black box" approach. Given the problem of producing action A from some set of inputs $(I_1, ..., I_j)$, it leaves the real-world process connecting $(I_1, ..., I_j)$ with A unanalyzed. Because a theoretical model of the process is not available, rules for producing A must be obtained exclusively by empirical discovery. For instance, gunpowder was discovered and utilized by people who did not have a theory of combustion adequate to explain chemical explosive action. Various steelmaking technologies were developed by medieval European and Arabian smiths in the complete

absence of an understanding of how and why their techniques worked. However, given the same problems, theoretical technology utilizes a theoretical model of the real-world process connecting $(I_1, \ldots, I_j)$ with A and derives rules of production for A from the model. Examples of theoretical technology include radar, lasers, the Polaroid Land camera, digital computers, and integrated circuits.

Although these two patterns are distinct, many specific technologies have a "mixed mode" pattern of development. In such cases, a model of the full process connecting $(I_1, \ldots, I_j)$ with A is still not available, but refinements and extensions of the empirically discovered rules of production are based on partial models of the process. That is, some decomposition of the full process into its subprocesses has been made, and models for these subprocesses have been constructed. This is not true theoretical technology, however, because no general model of the full process is available and, consequently, an integrated set of model-derived rules of production is not possible.

Empirical technology, but not theoretical technology, is ultimately self-limiting within any given field of technology. That is, there is a level of technological capability beyond which empirical techniques cannot penetrate. This level is a function of empirical technology's pattern of development, *not* of the world itself. The reason for this self-limiting characteristic is the absence of theoretical models. Empirical methods develop via trial and error, small incremental refinements and extensions of empirically discovered rules of production. Since the rules are not based on a model of real-world processes, however, these modifications cannot be orchestrated and integrated, but are instead *ad hoc* "fixes" that hold only over a limited domain. Once the modified empirically based rules of production reach a sufficient level of complexity, the probability becomes very high that the next *ad hoc* "fix" may undo a previous one. Further development in the particular technological field (development in the sense of increased technical capability) stops at this point.

Theoretical technology need not be self-limiting. Since it is based on a model, the above effect may not be present. Theoretical technology is thus able to push technological development in a given field to the maximum extent consistent with whatever real-world limitations characterize the field.

This discussion sets the stage for a consideration of the type of intelligence capability which can realistically be expected from machine intelligence research. The question of machine intelligence has been replaced by the question of the machine formulation of hypotheses. If we define a scale of hypothesis formulating capability (*HYP*) as $HYP \equiv TH + CRED$, where *TH* is the theoretical content of the hypothesis and *CRED* is the credibility of the hypotheses, then the design goal for advanced forms of machine intelligence is to be as high on this scale as possible.

Either an empirical technology or a theoretical technology pattern can be followed in developing machine intelligence. However, with respect to the *HYP*-level which can be achieved by the two different patterns, the empirical technology approach is ultimately self-limiting at a level of hypothesis formulating capability which is lower than that prerequisite for automated space exploration (see sections 3.2 and 3.3). It is clear that automated space exploration and other applications requiring very advanced machine learning are possible only if the theoretical technology approach to machine intelligence is employed.

Unfortunately, AI is currently taking the empirical technology approach to hypothesis formulation. There is nothing mysterious about the theoretical approach – it may be started by research into the patterns of logical inference by which hypotheses are formulated. Such an approach is limited only by the degree to which hypothesis formulation is logical and inferential. On the condition that it is, then the theoretical technology does not face a real-world barrier to achieving full machine-hypothesis generating capability.

6.2.4 *Initial Directions for NASA*

Several research tasks can be undertaken immediately by NASA which have the potential of contributing to the development of a fully automated hypothesis formulating ability needed for future space missions:

(1) Continue to develop the perspective and theoretical basis for machine intelligence which holds that (a) machine intelligence and especially machine learning rest on a capability for autonomous hypothesis formation, (b) three distinct patterns of inference underlie hypothesis formulation – analytic, inductive, and abductive inference, and (c) solving the problem of mechanizing abductive inference is the key to implementing successful machine learning systems. (This work should focus on abductive inference and begin laying the foundations for a theory of abductive inference in machine intelligence applications.)

(2) Draw upon the emerging theory of abductive inference to establish a terminology for referring to abductive inference and its role in machine intelligence and learning.

(3) Use this terminology to translate the emerging theory of abductive inference into the terminology of state-of-the-art AI; use these translations to connect abductive inference research needs with current AI work that touches on abduction, e.g., nonmonotonic logic; and then discuss these connections within the AI community. (The point of such an exercise is to identify those aspects of current AI work which can contribute to the achievement of mechanized and autonomous abductive inference systems, and to identify a sequence of research steps that the AI community can take towards this goal.)

(4) Research proposals for specific machine intelligence projects should explain how the proposed project contributes to the ultimate goal of autonomous machine intelligence systems which learn by means of analytic, inductive, and abductive inferences. Enough is now known about the terms of this criterion to distinguish between projects which satisfy it and those which do not.

6.3 Natural Language and Other Man-Machine Communication

It is common sense that various specific communication goals are best served by different forms of exchange. This notion is borne out by the tendency in technical fields of human activity to spawn jargon which only slowly (if ever) filters into more widespread usage. In the general area of communication between man and machine, a few tasks are already well handled by available languages. For example, in the area of numerical computations the present formal languages, while not perfect, are highly serviceable.

When one considers the introduction of sophisticated computer systems into environments where it is necessary for them to communicate frequently, competently, and rapidly with people who are not specialists in computer programming, then the need for improvement in man-machine communication capability quickly becomes apparent.

A natural language capability in computers is required primarily in two kinds of circumstances – (1) where the nature of the information to be transferred warrants the flexibility and generality of a natural language, in distinction to a more specialized language, and (2) where, because of the number, nature, or condition of the humans involved, it is impractical to have the humans communicate in a formal language. There are additional considerations of convenience to the user, and attracting users who otherwise might be reluctant to use the available computer facilities.

For directing the global actions of robot devices of all kinds, as well as the interrogation of question-answering systems, the ability to use natural language considerably eases or eliminates the problem of training individuals to use these resources. In particular, the user population of an Earth-sensing information system can be significantly and economically extended through direct communication between users and the system in natural language. Unfortunately, at present the domain is essentially that of a research domain, with relatively few natural language interfaces operating in production environments.

Man-machine information exchanges can be segregated into iconic communication, such as pictures, and symbolic communication, such as formal computer languages and human natural language (see fig. 6.2). These differ significantly in the amount and kind of interpretation required to understand and to react to them. For instance, formal computer languages are largely designed to be understood by machines rather than people. For purposes of further discussion, man-machine communication is subcategorized as follows:

(1) Machine understanding of keyed (typed) natural language

(2) Machine participation in natural language dialogue

(3) Machine recognition/understanding of spoken language

(4) Machine generation of speech

(5) Visual and other communication (includes iconic forms).

6.3.1 Keyed Natural Language and Man-Machine Dialogue

In those instances in which the environment is highly restricted with respect to both the domain of discourse (semantics) and the form of statements which are appropriate (syntax), serviceable interfaces are just becoming possible with state-of-the-art techniques. Primary examples are the LUNAR (Woods et al., 1972) and SOPHIE (Brown and Burton, 1975) systems. However, any significant relaxation of semantic and syntactic constraints produces very difficult problems in AI. Intensive research is presently underway in this area. It seems that the semantic aspects of normal human use of language override a large part of the syntactical aspects. Computer languages traditionally have been almost entirely syntax-oriented, and so the considerable knowledge available concerning them has very little relevance in the natural language domain. Progress in flexible natural language interfaces is likely to be tied to progress in areas such as representation of knowledge and "common sense" reasoning. These lie at the heart of intelligent information processing – full natural language competency at the level of human performance requires a machine with intelligence and world knowledge comparable to that of humans. At this time there is little work in progress on the necessity or appropriateness of specialized hardware for natural language processing.

Accepting the close relationship between human-grade natural language proficiency and general intelligence level, and the improbability of near-term attainment of human-grade general intelligence in machines, it is appropriate to focus instead on achieving usable natural language interfaces at a lower level of machine performance. This leads to an examination of man-machine dialogues in which the goal of the man is to communicate a clear and immediate statement of information, or a request for information or action, to the machine, where the information or request is in a domain for which the machine has a competent model. In this sphere of activity the following component capabilities are thought to be highly desirable, and probably necessary, for efficient and effective communication: domain model, user model (general, idiosyncratic, contextual), dialogue model, explanatory capability, and reasonable default assumptions.

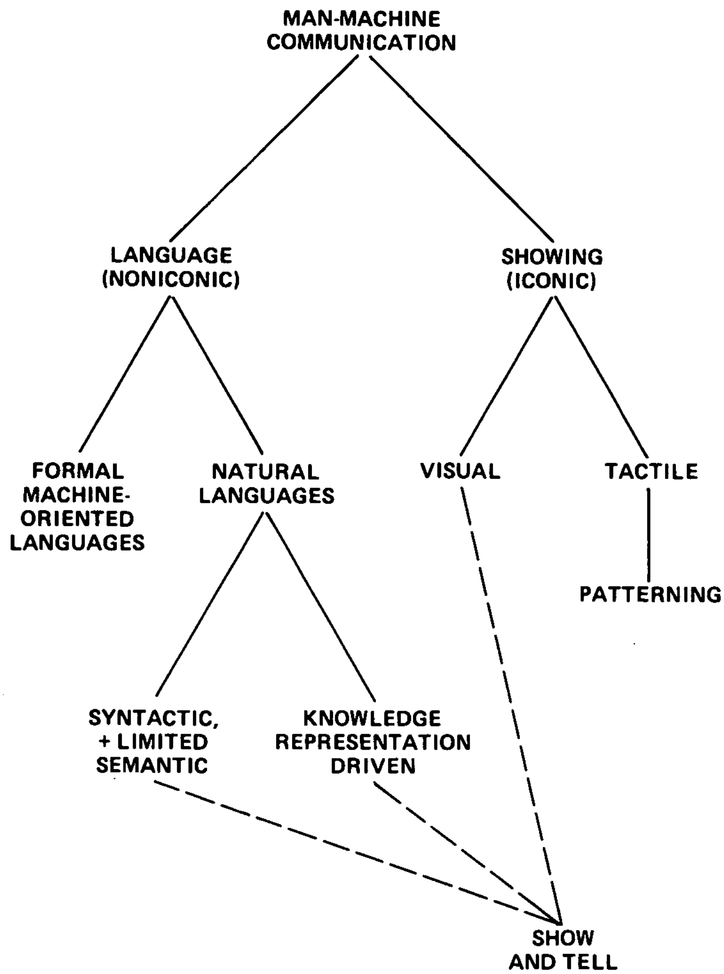


Figure 6.2.– Overview of man-machine communication.

Domain model. The machine must be able to act upon the information it receives, so it is assumed to have competency in some domain, called the object domain. To communicate about this object domain, the machine must have some additional knowledge called the domain model. When the communication environment is such that each linguistic transaction cannot have an immediate conclusion or effect in the object domain, then it is essential that the domain model be used to determine if a particular transaction makes sense. Otherwise the machine will not be able to make any inferences about the information it is being given, and the dialogue may become a monologue on the side of the human. The efficiency of the transfer of information will deteriorate, particularly for naive users. The machine will be accepting information which may prove inadequate later when it is handled in the object domain.

Individual user models. One of the reasons for natural language is to accommodate a wide range of humans in direct and efficient communication. This is best accomplished by taking into account at least some characteristics which are either specific to particular individuals, or specific to classes of individuals. One example is in default values and assumptions. Different users will have different expectations concerning the values of implicit parameters of the conversation, and will have different underlying assumptions. In order not to burden each user with the necessity to make all of these explicit, it is necessary to make these assumptions and defaults a function of the type of user.

Dialogue model. The machine must have a working knowledge of what constitutes an acceptable dialogue. Such things as timing and absolute and relative explicitness are considerations pertinent to all users, and may vary from one person to another. In addition, the machine should avoid long series of questions posed to the human in order to clarify discussion, a particular problem with current expert systems such as MYCIN (Shortliffe, 1976). While part of the solution lies in proper default values and assumptions, there may still be need for the human to supply information in response to a perceived need by the

machine for this information. The best way to guide the human into supplying this information, while avoiding a tedious and long series of direct queries, is a largely unstudied area. Also, since the structure of the knowledge involved in the dialogue differs considerably between the human and machine, it will be necessary to map the initial internal "need to know" requests perceived by the machine into the general flow of the dialogue in a human-oriented way.

6.3.2 Machine Recognition and Understanding of Spoken Language

Recognition and understanding of fluent spoken language add further complexity to that of keyed language/phoneme ambiguity. In noise-free environments where restricted vocabularies are involved, it is possible to achieve relatively high recognition accuracy though at present not in real-time. In more realistic operating scenarios, oral fluency and recognition divorced from semantic understanding is not likely to succeed. The critical need is the coupling of a linguistic understanding system to the spoken natural language recognition process. Thus the progress in speech recognition will depend upon that in keyed natural language understanding.

Early applications have involved single word control directives for machinery that acts upon the physical world, using commands like "stop," "lower," "focus," etc. Some commercial equipment is available for simple sentences, but these require commands to be selected from a small predetermined set and necessitate machine training for each individual user.

6.3.3 Machine Generation of Speech

At the present time mechanical devices can generate artificial-sounding but easily understood (by humans) spoken output. Thus the physical aspects of speech generation are ready for applications, although some additional aesthetics-oriented technology work would be desirable. (The more important aspects of deciding what to say and how to phrase it were covered in the foregoing discussion of keyed natural language.)

6.3.4 Visual and Other Communication

Some motor-oriented transfer of information from humans to machines already has found limited application. Light pens and joysticks are rather common, and some detection of head-eye position has been employed for target acquisition. Graphics input/output is also an active research area, and three-dimensional graphical/pictorial interaction is likely to prove useful.

An interesting alternative approach in communicating information to robot systems is called "show and tell." In this method a human physically manipulates an iconic model of the real environment in which the robot is to act. The robot observes this action, perhaps receiving some simple coordinated information spoken by the human operator as he performs the model actions, then duplicates the actions in the real environment. The distinctions between show and tell and typical teleoperator modes of operation are:

- Show and tell does not assume real-time action of the robot with the human instruction.
- For show and tell, the robot has the time to analyze the overall plan, ask questions and generally form an optimal course of action by communicating with the human.
- The fidelity of the robot actions to the human example can vary in significant useful ways, allowing the robot to optimize the task in a manner alien to human thinking.
- The show and tell task can be constructed piecemeal, thus allowing a task to be described to the machine which requires many simultaneous and coordinated events, or which requires input from teams of human operators which is then chained together into a single more complex task description.

Show and tell permits a high degree of cooperative problem-solving and reasoning about actions between humans and machines. This novel technique probably has an important functional role to play somewhere between autonomous robots and pure teleoperation.

6.3.5 Recommendations

The team recommendations to NASA regarding directed research and development in the field of natural languages and other man-machine dialogue are as follows.

Natural language and man-machine dialogue. Theoretical work in keyed and spoken natural language for managing restricted domain databases will proceed with NASA involvement. The impact of such systems is widely recognized, and much research is in progress. In applications, DOD is already involved in funding research whose results will probably be directly applicable to NASA database interactions in the immediate future. It is recommended that NASA now make plans to initiate implementation of systems using keyed natural language for internal use within NASA. Such implementation not only will provide useful production tools for NASA, but also will generate the in-house experience necessary to provide these techniques to outside users of space-acquired data as in the IESIS mission.

More sophisticated uses of natural language, such as in directing almost autonomous robots in tasks like space

construction or exploration, should be studied by NASA, but immediate application of a pure natural language communication channel does not seem possible at this time or in the near future. The first uses of fluent natural language in controlling robots will probably best be done in a context such as "show and tell."

Spoken language. The development of fluent spoken language recognition is expected to evolve in step with the ability of machines to understand and reason about the object domain. Thus, a NASA orientation toward funding research in this area would be misdirected. There is no obvious pressing need at this time for the Agency to intervene in the development of isolated word recognition control of robots, as this area will develop very rapidly on its own.

Speech generation. Serviceable speech generation is technologically current, for the physical generation, and NASA need not take any particular steps in this area until specific implementation demands it. The more important area of machine decision of what to say is a much more difficult and undeveloped research area, and is essentially the same problem as in keyed input-visual output dialogue systems.

Visual and other communication. The areas of motor and graphic interaction are ready for current implementation. NASA should consider these as tools appropriate both for its own internal use and, as with the keyed natural language, for outside users of NASA-collected data. Show and tell communication would be extremely useful in zero-g robot-assisted construction, and may have application in planetary exploration and space or lunar industrial processes – current research efforts are minimal. Many of the specific capabilities of potential interest to NASA will not be developed if the space agency does not take a direct, active role.

A very rudimentary form of show and tell, called "patterning," should be implemented as soon as possible for all NASA spacecraft with manipulator or other movable components under computer control. In patterning, a prototype or other model of the actual spacecraft is physically articulated in the way the actual spacecraft should behave. The model is connected to a computer through appropriate proprioceptors, and the computer writes a program which can be uploaded to the spacecraft to direct its actions. It should also be required that the model be able to execute the program in order to verify its correctness. Such a capability would greatly extend the flexibility of control of both complex devices in space and exploration craft on planets, and yet are relatively easily implementable with current techniques.

6.4 Space Manufacturing

To achieve the goal of nonterrestrial utilization of materials and factory self-replication and growth, space manufacturing must progress from terrestrial simulation to low Earth orbit (LEO) experimentation with space production techniques, and ultimately to processing lunar materials and other nonterrestrial resources into feedstock for more basic product development. The central focus of this assessment is upon the technologies necessary to acquire a major space manufacturing capability starting with an automated Earth orbiting industrial experimental station established either as an independent satellite or in conjunction with a manned platform such as a manned orbiting facility or "space station."

6.4.1 Earth Orbiting Manufacturing Experiment Station

There are four major components of any production system: (1) extraction and purification of raw materials, (2) forming of product components, (3) product component assembly, and (4) system control. The Earth-orbiting station will conduct experiments to determine the relative merits of alternative methods of implementing these elements in a space manufacturing facility.

Product formation involves two general operations – primary shaping to achieve the approximate shape and size of the component and finishing to meet all surface and dimensional requirements. The most promising primary shaping technologies for space manufacturing are casting and powder-processing techniques. When properly controlled, both methods produce parts ready for use without further processing. Casting techniques appear more versatile in terms of the range of materials (metals, ceramic, metal-ceramic) that can be shaped, but powder processes may outperform casting for metallic components. A determination of the relative utility of these two processes should be one of the primary goals of the space manufacturing experiment station.

The casting process is a fairly labor-intensive activity on Earth and has not been highly automated, with the exception of Strand and other continuous casters. Automated casting facilities do not generally produce a variety of parts configurations; instead, they usually make just a single shape (usually a bloom or billet) which later requires a great deal of expensive and time-consuming processing before it is usable as a machine component.

Many of the finishing operations can be eliminated if the material is cast into (approximately) its final configuration using a specialized mold. The production of these molds has been automated in two instances. In investment casting, the dipping of the wax forms into ceramic slurry has been

accomplished using industrial robots, although actual pattern formation remains largely manual. In permanent die casting, the Nike sports shoe subsidiary in Massachusetts produces tapes for its N/C electrical discharge machining apparatus which drives the tool to form the dies automatically from drawings of the shoe's sole constructed using the plant's CAD/CAM system.

On the whole, however, the formation of patterns and disposable molds (especially green sand casting molds) has remained manual; only equipment for lifting and turning the flasks has come into widespread use. Robots have been employed to unload hot parts from die casting machines as well as place the (hot) castings into trimming dies. Almost all automation of casting has been in high-volume applications where one standard shape is produced ten thousand times per year or more. High-volume production is not likely to be the general mode of space manufacturing, which will probably call for small lot, intermittent production. Methods of performing automatic casting, especially using disposable molds, and doing this efficiently in low- or zero-gravity conditions are required. Elimination of molds using containerless forming techniques should also be investigated and, if successful, will significantly reduce the high capital costs of forming molds and dies. The problems of heat removal and control of the rate of cooling to control grain size in the castings requires both sensor development to sense the internal temperatures and new heat dissipation technologies.

Powder processing has been somewhat automated on Earth, but has not been used extensively due to the tremendous costs associated with purifying and maintaining a non-oxidizing environment for manufacturing. This environment is available in space and on the lunar surface. But, as in the case of casting, powder processing uses dies to form parts. Again, the study of containerless forming techniques may be fruitful, with powder processing alleviating some of the heat dissipation problem, since sintering temperatures are lower than those required for casting. The applicability of powder formation via liquefaction and spraying should be assessed. Grinding and milling must also be examined, since the cold welding phenomenon between similar pure metals may be turned to advantage if it can be used to facilitate coalescence of the metals without sintering or melting. Intensive study of this effect is best performed in space, as pure powders are extremely difficult to prepare and maintain on Earth. Cold welding also has important implications for machining and lubrication.

Machining, or chip formation processes, are the usual finishing operations. These have been extensively automated, but significant problems with heat dissipation and cold welding may be encountered in space if the tools are run in a vacuum. The primary cause of tool wear is the temperature generated at the tool/chip interface. Removal of this heat through the use of cutting fluids will be difficult because all terrestrially used fluids are either petroleum or water-based – two commodities expensive in space and difficult to control in a zero-g environment. Cold welding will decrease chip forming in two ways – first, by the formation of built-up edge on the tool face (although temperature and pressure may still be the determinants of this effect), and second, by the reattachment of the pure metal chips to the cut or uncut surface or machine table by vacuum welding. Use of lasers to finish may eliminate many of these problems and thus may be of tremendous utility, especially if casting or powder techniques can be expected to produce high-tolerance parts. The use of ultra-high speed machining in which most of the heat of cutting is carried away by the molten chip could also be a partial solution, and also the use of ceramic tool bits and cast basalt tables. (See also appendix 5F.)

Assembly requires robotic/teleoperator vision and end-effectors which are smart, self-preserving, and dexterous. Accuracy of placement to 0.001 in. and repeatability to 0.0005 in. is desirable for electronics assembly. Fastening technologies, including nonvolatile adhesives, cold welding, mechanical fasteners, and welding all require special adaptation to the space environment.

Control of a large-scale space manufacturing system demands the use of a distributed, hierarchical, machine-intelligent information system. Material handling tasks require automated, mobile robots/teleoperators. In support of these activities, vision and high-capacity arms, multi-arm coordination, and dexterous end-effectors must be developed. For inventory control, an automated storage and retrieval system well suited to the space environment is needed. The ability to gauge and measure products (quality control) benefits from automated inspection, but a general-purpose machine-intelligent high-resolution vision module is necessary for quality control of complex products.

6.4.2 Materials Processing and Utilization

While it is expected that the orbiting space manufacturing experiment station initially will be supplied with differentiated raw feedstock for further processing, some interesting experiments in systems operations and materials extraction are possible and should be vigorously pursued. One such experiment could be a project to build one reasonably complex machine tool using a minimum of human intervention and equipment. Two logical candidates emerge. The first is a milling, grinding, or melting device that could be used to reduce Shuttle external tanks to feedstock for further parts building or experiments. This project would allow experimentation in material separation and processing using a well-defined and limited input source which can be obtained at relatively low cost whenever the Space Shuttle carries a volume-limited rather than a weight-restricted load. Such a large-scale experiment could be used as an "extra-laboratory" verification of extraction,

manipulation and control mechanizations, as well as providing relatively easy access to pure metal powders for research. A second candidate project is the fabrication and assembly of a beam-builder for use in large structure construction experiments. These two machine tool projects could then be combined to study materials handling and storage problems by having the first project provide feedstock for the second. Additional experimentation on producing feedstock from lunar materials would be a logical outgrowth of this development.

While the space manufacturing experiment station is largely viewed as an experiment station for capital equipment production and as a stepping stone to the establishment of a lunar manufacturing facility, it should be noted that the station can also be used for biological research and the preparation of products such as drugs and medicines for terrestrial consumption. For example, many pharmaceutical components require a zero-g environment for their separation. Additional products for terrestrial consumption would be perfect spheres or flat surfaces made by joining bubbles.

The technology required for permanent facilities to process nonterrestrial materials on the lunar surface or elsewhere lies far beyond currently proposed space materials processing capabilities. Numerous workers have suggested processes such as electrolysis, hydrogen fluoride leaching and carbochlorination (see section 4.2.2), which are adequate for short-term usage but cannot reasonably be expected to meet long-term growth requirements. Processes must be developed which yield a far broader range of elements and materials, including fluorine, phosphates, silica, etc. Volatiles such as water and argon, and desirable rock types such as alkalic basalts and hydrothermally altered basalts, could be acquired as a result of lunar-surface exploration. High-grade metals can probably be retrieved from asteroids in the more distant future.

Sophisticated and highly automated chemical, electrical, and crystallization processing techniques must be developed in order to supply the wide variety of required feedstock and chemicals. Some possible solutions may be generated by studying controlled fractionation and chemical doping of molten lunar materials in order to achieve crystallization of desired phases. Zone refining and zone melting techniques may also be fruitful areas for investigation. New oxygen-based chemical processing methods should also be examined.

6.4.3 Technology Requirements

The control of individual machine tools has continued to advance in feedback and feedforward control modes. The control of a diverse, highly integrated industrial complex requires advances in computer systems. High-speed data access in linked hierarchical computer networks will be needed. These computers will require coordination in real time. For example, the material handling computers must relay messages to the material handling devices telling them which machines need to be emptied or loaded and the material handling devices must know where to place the removed product. Advances in autonomous planning and scheduling in a dynamic environment are required, using new scheduling algorithms and shop floor control techniques. Large database requirements will soon become apparent. Repair robots must have the capability to hypothesize probable causes and sources of malfunction.

The establishment of space or lunar manufacturing facilities require the development of the following technologies:

- Basic research on materials processing in the space environment
- Improvement in primary shaping technologies of casting and powder processing for metals and nonmetals with emphasis on the economic elimination of manual mold production, possibly by the use of containerless forming
- Improvement in heat dissipation abilities in relation to the tool/chip interface in space, and control of cooling rates in castings
- Comprehension of cold-welding as a limiting factor for metal curing and as a joining technique
- Improvement of robot dexterity and sensors (especially vision)
- General and special purpose teleoperator/robot systems for materials handling, inventory control, assembly, inspection, and repair
- Improvement in computer control of large, integrated, dynamic hierarchical systems using sophisticated sensory feedback
- Study and improvement of lasers and electron-beam machining devices
- Embodiment of managerial skills in an autonomous, adaptive-control expert system

6.5 Teleoperators and Robot Systems

A teleoperator is a device that allows action or observation at a distant site by a human operator. Teleoperators represent an interim position between fully manned and autonomous robot operation. Teleoperators have motor functions (commanded by a human) with many possible capabilities, and have sensors (possibly multiple, special-purpose) to supply information. The human being controls and supervises operations through a mechanical or computer interface. As technology advances and new requirements dictate, more and more of the command and control functions will reside in the computer with the man assuming an increasingly supervisory role; as artificial intelligence methods are developed and are applied, the computer eventually may perform "mental" functions of greater

complexity, making the system more autonomous. The following discussion concerns teleoperators and their functions, applications to NASA programs, necessary supporting technologies, and the evolutionary path of robotics.

Teleoperators have been developed to expand man's physical capabilities across great distances and in hostile or inaccessible environments. Typical applications include (1) safe, efficient handling of nuclear or toxic materials, (2) undersea mining and exploration, (3) medical and surgical techniques, and (4) fabrication, assembly, and maintenance on Earth and in space. An artificial limb is considered a teleoperator because it restores lost dexterity to an amputee.

Teleoperators are not new. In 1954 Argonne National Laboratory developed a master/slave hand system with force feedback via cables and pulleys. In 1958 William Bradley (1980) operated an area-of-interest television camera system mounted on a truck to provide a display to the "driver" located 15 km away. In the 1960s General Electric engineers designed "Hardiman," an exoskeletal teleoperator with 15 degrees of freedom and the capability of manipulating 700-kg loads with ease (Corliss and Johnsen, 1968). Research is progressing once again in manipulators, sensors, and master/slave systems. Further technology advances will be made as NASA develops teleoperators for space operations.

6.5.1 Teleoperator Applications

Advanced teleoperators for future space missions present new challenges in the development of spaceborne man-machine systems (Bejczy, 1979; Bradley, 1967; Corliss and Johnsen, 1968). Teleoperators are robotic devices having video or other sensors, manipulator appendages, and some mobility capability, all remotely controlled via a telecommunications channel by a human operator. The man can exercise direct in-the-loop control using a joystick or other analog device, or can choose more indirect means of command such as an AI system in which he shares and trades control with a computer (NASA Advisory Council, 1978). Heer (1979) estimates that flight demonstrations of automated Shuttle manipulators can begin as early as 1982, for automated construction devices in 1986, and for a free-flying automated teleoperator by 1987.

A teleoperator will be on the first operational Space Shuttle flight. The Shuttle has a six-degree-of-freedom general-purpose Remote Manipulator System (RMS) with a 15-m reach (Meade and Nedwich, 1978; Raibert, 1979). The RMS lifts heavy objects in and out of the payload bay and assists in orbital assembly and maintenance (Meade and Nedwich, 1978; Raibert, 1979). An astronaut controls the rate of movement of the RMS using two three-axis hand controllers (Lippay, 1977). One proposed follow-on is installation of a work platform so that the RMS could be used as a "cherry picker," carrying astronauts to nearby work sites.

One RMS will be mounted on the port longeron with provisions for a second RMS mounted on the starboard side. Conceptually the RMS arm is much like a human arm, with yaw and pitch at the shoulder joint, pitch at the elbow, and yaw, pitch and roll at the wrist (Lippay, 1977). The upper arm is 6.37 m long and the lower arm is 7.06 m long, providing a 15-m reach. The RMS can move a 14,000-kg payload at 6 cm/sec with the arm fully extended, or up to 60 cm/sec with no load (Space Shuttle, 1976, 1977, 1978).

Two other distinct classes of teleoperation will be required for complex, large-scale space operations typified by the space manufacturing facility described in chapter 4. The first is a free-flying system which combines the technology of the Manned Maneuvering Unit with the safety and versatility of remote manipulation. The free-flying teleoperator could be used for satellite servicing and for stockpiling and handling materials (Schappell et al., 1979). Both of these operations require autonomous rendezvous, stationkeeping, and attachment or docking capabilities. Satellite servicing requires the design of modular, easily serviceable systems and concurrent development of teleoperator systems.

The Teleoperator Maneuvering System (TMS) is an unmanned free-flying spacegoing system designed to fit in the Shuttle Orbiter, with the capability to boost satellites into higher orbits, service and retrieve spacecraft, support the construction assembly and servicing of large space platforms, capture space debris, and perform numerous other tasks in orbit. TMS has the potential, with developing robotics technology, to greatly extend and enhance man's capabilities in space. As presently defined by NASA, TMS is propelled with hydrazine or cold gas thrusters, controlled by operators at ground stations or in the Orbiter's aft flight deck, and can be placed under automated control using its onboard computational capabilities. TMS eventually will be equipped with antennas, manipulators, video equipment, dexterous servicing mechanisms, a solar power array, and other equipment as needed to position spacecraft, rendezvous with and service satellites, position large platform sections, and act as a "smart" free-flying subsatellite for performing specialized missions. It can perform all known LEO payload retrieval missions within 1 km of the Shuttle, and retrieval at distances of 800–1600 km from the Orbiter could be demonstrated by the mid-1980s (OAST, 1980).

Manufacturing processes and hazardous materials handling may utilize mobile or "walking devices," the second distinct class of teleoperators. The teleoperator would autonomously move to the desired internal or external site and perform either preprogrammed or remotely controlled operations. For manufacturing and repair, such a system could transport an astronaut to the site and the manipulator could be controlled locally for view/clamp/tool operations or as a workbench. Of course, the size and level of teleoperator mobility (free-flying or walking) is dictated by mission needs.

Probably the two classes will be combined into one device in actual practice. Such a combination could be characterized as a remote free-flying teleoperator equipped with a highly specialized manipulator of the general-purpose or spacecraft-services system type. An extension can be envisioned as a teleoperator vehicle combined with single or multiple manipulator arms used to align and attach beams in direct support of space construction activities. Depending on the complexity of the task at hand it may be necessary for humans to be directly in charge in a master-slave relationship and be housed in a life support module on the free-flyer. The next logical development step delegates this function to a manlike robot, thus freeing the system to work autonomously at extended operational ranges without the cumbersome remote or local presence of man.

In the reference Space Manufacturing Facility developed by Miller and Smith (1979), the large number of similar components in the solar-cell factory and the X-ray environment precludes direct human labor. This suggests automated maintenance and repair, so the solar-cell factory was designed for tending by automated and remote devices. A free-flying hybrid teleoperator (FHT) can do on-site repairs at the solar-cell factory. The FHT can be operated either fully automated (tied into an AI-capable computer system or using preprogrammed routines), automatically with human override, or fully remote-controlled by a human operator (teleoperation).

Free-flying teleoperators or robot servicing units will have the capability to autonomously rendezvous, close, and attach to a satellite, first in LEO near the main station and later in GEO (Schappell et al., 1979). In some cases satellite retrieval, rather than servicing, will be desired. This would be a precursor to automated asteroid retrieval missions, requiring completely autonomous systems for navigation, guidance, sensing and analysis, attachment, and mining (Shin and Yerazunis, 1978).

On-board and free-flying teleoperators will be required throughout the postulated mission plan. They will extend man's senses and dexterity to remote locations while the human supervises and controls from a safe, comfortable environment. Teleoperators are a logical step in the evolution to fully automated (robot) systems needed for efficient extraterrestrial exploration and utilization. Previous sections have already discussed the role of man, the role and configuration of such teleoperators, and the role and development required for completely automated, possibly self-replicating, systems.

6.5.2 Teleoperation Sensing Technology

The uniqueness and utility of teleoperators lies not in their mode of locomotion, but rather in the "telepresence" they provide – the ability of the man to directly sense and remotely affect the environment (Minsky, 1980). Sensor and manipulator technology is advancing apace, largely through rapid growth in the fields of industrial robotics and computer science.

Approximately 40% of human sensory input is in the form of vision, so it is appropriate that most work in physical perception relates to visual information processing and remote scene interpretation. Algorithms and specialized sensors developed for satellite on-board pattern recognition and scene analysis can enable the teleoperator to perform many of these functions. Teleoperation has several unique characteristics such as viewing and working in three dimensions under variable conditions of scene illumination, and options of wide or restricted fields of view. Three-dimensional information can be obtained from stereo displays (Chin, 1976; Duda and Hart, 1978; IEEE, 1979), lasers (Shin and Yerazunis, 1978), planar light beams (Baum, 1979), radar and proximity sensors (Schappel et al., 1979), or it may be recovered from two-dimensional pictures (Tenenbaum, 1979).

Besides its use in autonomous tasks, a computer "world model" can be utilized in two ways. First, it can provide the man a computer-generated display from any point in the "world." Theoretically, from an overall view of the entire scene (including the teleoperator itself) the camera eye could zoom down inside a crevice or behind an object. Using data from a scanning laser ranging system, the system described by Shin and Yerazunis (1978) could construct a perspective model of nearby terrain and superimpose the route through the terrain determined by an optimal path-selection algorithm.

Second, using camera location as a reference point and overlaying the "world model" over the camera picture would permit correlation of the world model with the real world, thus enabling the operator to immediately detect anomalies or inaccuracies in the knowledge base. This "knowledge overlay" would allow corrections for sensor errors and keep autonomous manipulator operations properly referenced. Without such a knowledge overlay the man is severely handicapped in acting as supervisor of largely autonomous operations.

Besides vision, a teleoperator should give the human a "feel" for the task. Minsky (1980) notes that no present system has a true sense of feel, and insists that "we must set high objectives for the senses of touch, texture, vibration, and all the other information that informs our own hands." In addition to communicating via sight and touch, an audio interface between man and computer also is feasible (see section 6.5.3). Voice input/output systems are commercially available and in use. Research continues, though, in artificial intelligence and computer science on natural language understanding, faster algorithms, and connected speech processing. However, it should be noted that teleoperators with simple bilateral force reflection can achieve most immediate goals in space. These were demonstrated by Ray Goertz as early as 1955, and can be used now.

6.5.3 *Teleoperation Manipulator Technology*

Much of a teleoperator's capability is sensory; much is associated with manipulation. Although configurational details require further definition of task requirements, overall general-purpose space teleoperator characteristics can be partly inferred. A teleoperator arm must have enough freedom so that the manipulation and arm locomotion systems can position the hand or end-effector at any desired position in the work environment. There must also be a locus of points which all of the teleoperator's hands can reach simultaneously. If such a region does not exist, manipulator cooperation is precluded – cooperation and coordination of multiple manipulator arms and hands give teleoperators (and humans) tremendous potential versatility.

How many manipulator arms might the general-purpose teleoperator have? Despite man's two arms, the teleoperator will probably need three. Most mechanical operations require just two hands – one to grasp the material and the other to perform some task. A third hand would be useful in holding two objects to be joined, or in aiming a television camera (or other appropriate sensor). In many two-handed operations on Earth the human worker moves his head "to get a better look" – the third teleoperator arm would move the man's remote eyes for that purpose. Indeed, the third arm can be used to couple the TV motion to the man's head motion. Bradley (1980) notes that this gives a strong feeling of telepresence. Finally, three fingers probably are sufficient for duplicating most of the functions of the human hand – this is the minimum number necessary for a truly stable and controllable grasp of small objects.

6.5.4 *Robot Systems*

Teleoperators will always be vital to many operations in space because they extend man's senses and motor functions to remote locations. But extraterrestrial exploration and utilization and other advanced systems will require remote autonomous systems – systems with on-board intelligence. These robot systems will evolve along with current AI efforts at representing knowledge functions in a computer. The integration of AI technology with teleoperator/robot systems is a major development task in its own right and should be timed to support space programs that require this capability.

Aspects of artificial intelligence which must be addressed in regard to robot systems include memory organization, knowledge retrieval, search, deduction, induction and hypothesis formation, learning, planning, perception and recognition (Lighthill, 1972; Nilsson, 1974; Sagan, 1980; Winston, 1978). Teleoperation and robotics technology requirements are: time lag compensation methods, sensory scaling, adaptive control methods, touch sensing, hands, hydraulics, actuators that are many times lighter than the masses that they lift, onboard power for autonomous operation (this is a major problem), parallel computers, clamp and hold servoing of arms (extra hands are needed to hold parts while soldering and connecting), homeostasis, survival instincts, world models, laser data links, and laser sensors. Computer science, cybernetics, control theory and industrial process control are all relevant fields in this research. Interactive systems are being developed whereby the computer works, not autonomously, but as a partner or intelligent assistant. Kraiss (1980) discusses the design of systems resulting from cooperation of human and robot systems in four specific areas – computers capable of learning and adapting, computer support in preparation and evaluation of information, computer support in decisionmaking, and computer assistance in problem-solving.

6.5.5 *Telefactor Technology Development Recommendations*

The advantages of the availability of telefactor systems for development of subsequent fully automatic and replicating systems have already been described in this report. However, it is worth noting that: (1) all of the technical information and components to build a telefactor system were available, and the basic subsystems (e.g., master-slave manipulators and head-aimed television systems built and demonstrated) before 1965, and (2) to date, no one has built a complete system (Bradley, 1967).

Construction of a standard telefactor system is long overdue. NASA should include this important step in an early phase of its automation program. Some of its applications to the NASA program are the following:

1. A telefactor system can be used to oversee and operate a materials processing activity to establish requirements for full automation of such activity and also for manned intervention.
2. A telefactor system can provide a built-in maintenance and repair facility in a complex spacecraft.
3. A telefactor system could perform satellite inspection, modification, or other EVA operations from the Shuttle, even with uncooperative objects.
4. All of the actions and observations of a telefactor system can be taped for later playback, permitting retrospective task analysis.
5. Demonstration of the frequently proclaimed versatility and effectiveness of telefactor systems is overdue and much needed.
6. A standard telefactor system can be used as a comparison-piece in the field of robotics. Differences in task performance and in characteristic deficiencies among telefactors versus robots would be of great interest.
7. Since computers can be readily inserted into a standard telefactor system, these could become powerful tools in the development of fully automatic or supervisory control systems.

8. A standard telefactor system would be a convenient starting point for development of rovers for lunar and planetary exploration and prospecting.

9. A standard telefactor system would incidentally be a useful starting point for development of terrestrial remote control or remotely piloted vehicle equipment for use in hostile environments.

To achieve construction of a prototype standard telefactor system with minimum cost, it would appear appropriate to utilize some of the personnel already familiar with this and who have had practical experience in the construction and operation of the major subsystems. A conventional aerospace contractor even if well provided with funding and facilities is likely to misunderstand some of the problem areas discovered and resolved during 1956-1966, thus requiring costly reworking and rediscovery of old techniques. NASA should find means of implementing such an effort with leadership at one of its centers and with interested participation by NASA Headquarters staff. After attainment of a satisfactory prototype of a standard telefactor system, several should be constructed and made available where needed most in the agency's automation program.

6.6 Computer Science and Technology

NASA's role, both now and in the future, is fundamentally one of information acquisition, processing, analysis, and dissemination. This requires a strong institutional expertise in computer science and technology (CS&T). Previous study efforts and reports have made recommendations to integrate current technology more fully into existing NASA programs and to develop NASA excellence in selected relevant fields of computer science. Recent studies have explored the research and development requirements of NASA field centers, and have identified particular R&D goals and objectives relevant to CS&T (ASC, 1980; EER, 1980; Sagan, 1980). In this section, the recommendations are considered from the perspective of the CS&T study team, together with the implications for CS&T of the various missions defined earlier in the report.

Of particular concern in the present technology assessment is the evolving CS&T program required within the space Agency to support a major involvement of automation and machine intelligence capabilities in future NASA missions. The agency presently is not organized to support such a vigorous program in CS&T. Most apparent is the lack of a discipline office at the Headquarters level which supports research and development in computer science and which serves as an Agency advocate for the incorporation of state-of-the-art capabilities into NASA programs.

NASA technical requirements with relevance to CS&T are presented and correlated with specific CS&T disciplines in this section. A general upgrading of computing facilities is recommended. Building an organization to maintain a state-of-the-art capability in the computing and information sciences is perhaps the greatest challenge for the future. The study group is hardly qualified to offer specific organizational recommendations to NASA, but encourages the agency to consider an organizational response and suggests some ideas which may be helpful. Finally, maintenance of a solid computing science institutional capability depends on a vigorous and continuing program of intellectual exchange with peer organizations in academia, industry, and government. A few suggestions are presented as to the possible components of such a program.

6.6.1 NASA Technology Requirements

This report, together with the report of the NASA Study Group on Machine Intelligence and Robotics (Sagan, 1980), has explored the application of advanced automation within NASA. In addition, there are general computer science capabilities required to develop and implement the types of missions described in the present document. These include robotics, smart sensors, mission operations, computer systems, software, data management, database systems, management services, human-machine systems, engineering, and system engineering.

Robotics. The principal requirements associated with robotics which call upon the disciplines of CS&T include visual perception, manipulator control, and autonomous control. This latter category includes problem-solving and plan generation – the ability of a robot device to plan and to pursue its own macroscopic course of action. NASA requirements also argue for a robotic capability to perform intelligent data gathering and in some instances to provide a telepresence capability for a remote human operator.

Smart sensors. Current programs such as NEEDS (see chapter 2) address requirements for smart sensing devices which selectively acquire data and analyze it for information value prior to consuming communications and storage capacity. These requirements include visual perception, image processing, pattern recognition, scene analysis, and information extraction. In addition, the notion of model-based sensing shows promise for intelligent data acquisition. To conserve communications bandwidth, user-oriented data compression techniques are required which can result in a several orders-of-magnitude reduction in the amount of data transmitted.

Mission operations. In the area of mission operations, a rather general symbolic modeling and representation capability is required to do planning, scheduling, sequencing, and monitoring, as well as fault modeling and diagnosis. This draws on problem-solving techniques within artificial

intelligence and can benefit greatly from a hypothesis formation capability. As mission operations are presently conducted, machine intelligence can benefit the coordination of manpower, as well as enhance the mission software development and integration process. As mission operations are envisioned in the future, involving autonomous spacecraft operations and automatic mission control, a strong dependence on CS&T in general and machine intelligence in particular is unavoidable.

Computer systems. In ground-based systems, and especially in spaceborne applications, NASA has a fundamental dependence on computer systems. Requirements include LSI and VLSI circuit design, fabrication and test techniques as well as fault-tolerance, error detection and recovery, component reliability, and space qualification. Beyond the component level, very significant primary and secondary storage requirements emerge. System-level issues become dominant, such as computer architecture (e.g., parallel processors) and system architecture (e.g., computer networks). Many of NASA's systems have severe real-time constraints, and techniques for adequate system control demand attention.

Software. Much of NASA's technology resources are spent on software, yet relatively modest attempts have been made to improve the process of software development, management, and maintenance. Given the exciting prospects for computer-based advanced automation in future missions, a program for more efficient, effective, and timely software development, management, and maintanance is mandatory. Principal software requirements are in the areas of programming languages, the software development environment, software validation, algorithm design, fault tolerance, and error recovery. Automatic programming should also be considered as a vehicle for improving the quality of software and the process of developing it.

Data management. Data management requirements comprise a very large part of the CS&T-related requirements within NASA, and include most of the interfaces to the user community. The public perception of NASA's systems will be derived largely from their ability to use them and to derive benefit from them. Both the NEEDS and ADS projects have realized this, and have diligently considered the end-user interface requirements. Data management requirements include data compression, staging, integration, and dissemination, as well as the implied requirements of data autonomy. Scheduling, performance monitoring, and system control also imply data management requirements, as does sensor management. A fundamental element of a user-oriented system is an extensive directory service as well as a capability to model the user, to know the context of his requests and his level of sophistication. On-line tutorial capabilities are appropriate for a diverse user community, and provide valuable input to the development of a user model. Knowledge-base systems and constructs such as semantic networks can contribute greatly to NASA's data management capability.

Database systems. NASA's current database requirements are not considered to be extraordinary from the CS&T perspective, although future systems supporting a geographically dispersed, technically diverse user community attempting to analyze or correlate sets of data spanning several distinct databases will require a sophisticated capability currently beyond the state-of-the-art. The requirements in this area include "traditional" database systems as well as relational database systems. The capability to satisfy queries which require access to several geographically separate databases is considered fundamental, as is a complete archiving capability.

Management services. NASA has, to a large extent, avoided the application of contemporary CS&T (let alone machine intelligence) to the management of the agency itself and its own programs. Current commercial offerings in management information and word processing systems can substantially enhance the efficiency and effectiveness of NASA management, both at Headquarters and at the field Centers. State-of-the-art capabilities in on-line records management, calendar coordination, and "bulletin boards" can likewise have a significant positive impact. The automated office is a concept evolving from this work which could revolutionize NASA's management techniques. Some obvious requirements in this area are manpower coordination, document preparation, and forms processing (e.g., travel orders and procurement requests). Presently unexplored is the potential application of contemporary machine intelligence techniques such as problem-solving, reasoning, and hypothesis formulation to the management of projects and the exploration of policy alternatives.

Human/machine systems. NASA has extensive requirements relating to human/machine interactions and currently has several efforts exploring the application of machine intelligence to these problems, primarily in the areas of hand-eye coordination and natural language processing. Requirements are primarily in the areas of human/machine control processes and the interface between a human and an "intelligent" computer system. Coordinated work between the computing sciences and cognitive psychology may be required to make substantial progress in this field.

Engineering. NASA is currently applying state-of-the-art technology in the engineering disciplines, particularly in computer-aided design, manufacturing and testing. The requirements of future missions, including mining and

manufacturing in nonterrestrial environments, mandate a continuing vigorous program in this area, embracing robotics technology.

System engineering. There are many component technologies which must come together to build a system. CS&T can aid in the process of engineering systems solutions, rather than component solutions, to systems problems. Formally managing the definition of requirements for a system is one example. Other contributions of CS&T include formalized design methodologies, techniques for performance monitoring and evaluation, and quasi-rigorous approaches to system architecture and control. Requirements in each of these areas pervade NASA programming.

6.6.2 Relevant CS&T Disciplines

For purposes of the present discussion the scope of CS&T is considered to be that classified by the Association for Computing Machinery (Computing Reviews, 1976) in their recently published "Categories of the Computing Sciences." This basic taxonomy was reviewed by the CS&T study team. Those components which appeared to relate most strongly to NASA's anticipated future requirements were scrutinized in more detail. The results of this analysis are summarized briefly below, with the ACM classification number included parenthetically for completeness.

Applications (3.). "Applications" focuses on the uses of computers and the relationships between human cognitive and perceptual processes and computers. NASA-relevant subcategories include:

(3.1) Natural sciences (astronomy, space, earth sciences)
(3.2) Engineering (aeronautical, electronic, mechanical)
(3.4) Humanities (language translation, linguistics)
(3.5) Management (policy analysis, manufacturing, distribution)
(3.6) Artificial intelligence (induction, pattern recognition, problem-solving)
(3.7) Information retrieval (content analysis, file maintenance, searching)
(3.8) Real-time systems (process control, telemetry, spacecraft simulation)

Software (4.). This category includes "the procedures, instructions, techniques, and the data required to apply a computer to a given task." Relevant subcategories include:

(4.2) Programming languages (procedure, problem-oriented)
(4.3) Supervisory systems (multiprogramming, database systems)
(4.4) Utility programs (debugging, program maintenance)
(4.6) Software evaluation, test, and measurements (software modeling, algorithm performance monitoring)

Mathematics of computation (5.). The category of mathematics of computation consists of "the intersection of mathematics and computer science, the category embraces subcategories that cover the mathematical treatment of numbers, mathematical metatheory, symbolic algebraic computation, the study of computational structures (algorithms, data structures) as mathematical objects, and mathematical methods that lend themselves to computer-aided solutions." This entire category was considered relevant to NASA. The subcategories are:

(5.1) Numerical analysis (error analysis, numerical integration)
(5.2) Metatheory (logic, automata, formal languages, analysis of programs)
(5.3) Combinatorial and discrete mathematics (sorting, graph theory)
(5.4) Mathematical programming (linear and nonlinear programming, dynamic programming)
(5.5) Mathematical statistics and probability (regression and correlation analysis, stochastic systems)
(5.6) Information theory (decision feedback, entropy)
(5.7) Symbolic algebraic computation (symbolic differentiators, symbolic interpretors)

Hardware (6.). The hardware category includes all of the physical components of digital computers. The relevant subcategories include:

(6.1) Logical design, switching theory (functional design, switching networks, Boolean algebras)
(6.2) Computer systems (packet switching networks, time-shared hardware, parallel processors)
(6.3) Components and circuits (LSI/VLSI, control and storage units)

Functions (8.). This category deals with major computer functions and techniques. NASA-relevant subcategories include:

(8.1) Simulation and modeling (applications, techniques, theory)
(8.2) Graphics (display processors, image processing, plotting)
(8.3) Operations research/decision tables (PERT, scheduling, search theory)

6.6.3 Correlation of NASA CS&T and Technology Requirements

Thus far, NASA's anticipated CS&T requirements have been presented, together with an outline of relevant CS&T disciplines. In this section, they are correlated through a matrix. Each element of the matrix in table 6.7 can assume one of five values, assigned on a subjective basis by CS&T assessment team members after consultation and thorough consideration.

TABLE 6.7.– NASA COMPUTER SCIENCE AND TECHNOLOGY REQUIREMENTS

NASA – relevant computer science and technology disciplines	NASA technology requirements										
	Robotics	Smart sensors	Mission operations	Computer systems	Soft-ware	Data manage-ment	Data base systems	Manage-ment services	Human/ machine services	Engi-neering	System engi-neering
(3) Applications											
(3.1) Natural Sciences		3	3								0
(3.2) Engineering	1	0	0	1					2	2	0
(3.4) Humanities (natural languages)	2		2						2		0
(3.5) Management								0			0
(3.6) Artificial intelligence	2	3	3		1	2		1	2	1	0
(3.7) Information retrieval						2	2	0	1	0	0
(3.8) Real-time systems	3		3	2		2			1	1	0
(4) Software											
(4.2) Programming languages			0	0	1					0	0
(4.3) Supervisory systems	2		1	2	2	2	2	0	0	1	1
(4.4) Utility programs			1	1	1	1	1				
(4.6) S/W evaluation tests and measurements	1		1		2	0	0			0	
(5) Mathematics of computation											
(5.1) Numerical analysis			0		0					0	
(5.2) Metatheory	0			1	1						
(5.3) Combinatorial and discrete math	0		0	0	0						
(5.4) Mathematical programming	0		0	0	0						0
(5.5) Mathematical statistical problems	0	1	0	0	0					0	1
(5.6) Information theory		1	1	0	0						0
(5.7) Symbolic algebraic computation	0		0		0						0
(6) Hardware											
(6.1) Logic design/switching theory	1	1		1						0	0
(6.2) Computer systems	0	0	0	2						1	1
(6.3) Components and circuits	1	1		1						0	0
(8) Functions											
(8.1) Simulation and modeling	2	2	2	2	2	2	2		2	2	2
(8.2) Graphics			1	0	0				1	1	0
(8.3) O.R./decision tasks	1		1		1	1	1		1	1	0

0 = Monitor technology status and transfer technology into NASA programs.
1 = Support outside research and development.
2 = Perform in-house research, maintain peer level in state-of-the-art community.
3 = Become technology leader.

The five values are 0, 1, 2, 3, and null. A null is used where a given CS&T discipline is not expected to contribute significantly to the satisfaction of a given class of NASA technical requirements. A numerical value indicates that the CS&T discipline is expected to be a significant element in satisfying the NASA requirements in a particular area.

Further resolution is given, addressing the level of NASA commitment required to apply the CS&T disciplines successfully to the NASA requirements. Zero implies that the discipline is receiving adequate and NASA-relevant support from other sources, and the agency need only monitor the status of the technology and apply it to NASA requirements. A value of 1 means that some agency support is required to adopt a technology to applications within NASA, but R&D activities are strictly applied and can be performed through well-defined contract activities. A value of 2 implies that a substantial commitment is required to develop a discipline and apply it to NASA requirements. This commitment will involve both basic and applied research, and will establish the agency as a peer in the community of state-of-the-art researchers in the given discipline. This is a substantial commitment by NASA to a particular discipline, and will require the development of a "critical mass" of capable personnel and a stable funding environment over a period of several years. The final matrix value notation is 3, which is used in those special instances where NASA should become the recognized technology leader in a given CS&T discipline.

The correlation between NASA technology requirements of section 6.6.1 and the CS&T disciplines of section 6.6.2 are shown in matrix form in table 6.7. In general, the table shows that the agency has a wide multidisciplinary dependence on CS&T. This suggests a position of leadership for NASA in the areas of natural sciences and artificial intelligence as applied to mission operations and remote sensing, and real-time systems for robotics and mission operations. It further argues for a substantial commitment to engineering applications such as CAD/CAM technology, natural language processing, artificial intelligence and real-time systems in general, information retrieval, supervisory software, computer systems technology, simulations and modeling.

A cursory and admittedly incomplete review of existing capability within NASA suggests that state-of-the-art technology already is a part of Agency programs in the natural sciences, engineering, and simulation and modeling. Further, some good work is being done in an attempt to bring NASA's capability up to the state-of-the-art in natural language processing, although primarily through contracted research activities. But in order to fully realize the potential of CS&T within the space Agency, it appears that a substantial commitment to research in machine intelligence, real-time systems, information retrieval, supervisory systems, and computer systems is required. In many cases it was concluded that NASA has much of the requisite in-house expertise in isolated individuals and organizations, but that the agency as a whole has been reluctant or disinterested in applying this expertise. An apparent lack of expertise *does* exist in the field of "mathematics of computation" (with a possible exception in the engineering area). This discipline can easily be overlooked as seemingly irrelevant, but in fact is a fundamental theoretical component of a broad-based and effective machine-intelligence institutional capability.

6.6.4 Facilities

To develop an institutional state-of-the-art capability in CS&T as described above will require good people and good facilities. Neither can do the job alone. Unfortunately, competent CS&T research-oriented professionals currently are in very short supply, and those few that exist are being attracted to industry and the universities through incentives of high salaries, outstanding working conditions, and intellectual freedom. None of these are offered by NASA at present, so the agency would probably be frustrated even if it were to attempt to hire the right talent. There is little NASA can do regarding salaries, so its focus in providing competitive incentives must be elsewhere. In this section, several specific recommendations are made with respect to facilities which the CS&T team considers prerequisite to any serious attempt by NASA to develop significant in-house capabilities in CS&T.

Interactive, on-line programming environment. Most programming is currently done within NASA on 10-year-old batch-oriented computer systems, where programmers still manipulate card decks and experience turnaround times measured in hours or even days. In order to attract competent researchers and to provide an environment in which they can labor productively, an absolute prerequisite is a fully interactive, on-line programming environment. For instance, Teitelman (1979) describes a typical state-of-the-art interactive system of the type required. NASA will find that this type of system, when made generally available to its personnel, will yield a very significant increase in programmer productivity. It is expected that this increase will be sizable enough to more than offset the additional cost of the on-line capability.

NASA has historically met its computing requirements through the purchase of computing equipment (e.g., instead of leasing). Due to the intricacies of the government ADP procurement process, 5 years typically will lapse between the conception of a new system and its actual operation – and then that system will remain in operation for 10-15 years, so that a system will be 15-20 years behind the state-of-the-art at its retirement (and 10-15 years behind during the "prime" of its life). NASA may wish to consider as an alternative for its nonmission (and, specifically, R&D) computing requirements the purchase of time-sharing services from a quality commercial vendor, so that it always has access to the best of the commercial offerings at any given time.

Computer communications network. NASA currently lacks any effective mechanism to provide digital communication between its computers in a general way. Several small efforts at individual centers have addressed intercomputer communication, and NASA has actively participated in international negotiations on intercomputer communication protocols, but no agency-wide effort has been made to apply this technology within NASA.

A strong case can be made for NASA to develop such a capability. It facilitates regular communication among geographically dispersed personnel and enables the sharing of both hardware and software resources. This can be particularly important for coordinating joint research among the centers.

One is tempted to envision within NASA a network structured logically as a hierarchy, in which the "standard equipment" (much as is a desk and chair) of an individual is a terminal which gives him direct access to local computing resources. These local computers are then aggregated into local computer networks to provide load balancing and resource sharing for a community of users as well as access to resources outside the local network such as other local networks within NASA, and extending to include the ARPANET and commercial facilities such as TYMNET and TELENET. It is not difficult to consider a NASA network which links all individuals within a Center, Centers to each other and to Headquarters, and provides access to non-NASA resources. There is little doubt that such a system will eventually become a reality. The telephone system currently provides just this type of capability for voice communication. The question becomes more one of *when* than *if* it will happen.

The ARPANET was developed largely as an experiment in this type of technology and has evolved into a primary vehicle for communication among researchers in the artificial intelligence community as well as other CS&T disciplines. As a first step toward integrating this type of capability into NASA systems, the agency should seriously consider negotiating with the Defense Communications Agency (DCA) for agency-wide access to the ARPANET. Not only would such a step provide a communications link with a large part of the research community of CS&T, but it would also provide the opportunity to perform communications experiments within NASA with a minimum investment of agency resources. This experience should equip NASA with the knowledge and expertise it will need to consider the implementation of an in-house networking capability.

Office automation. A significant proportion of NASA's resources are consumed manipulating documentation in many forms, including standard government forms, design documentation for software and hardware systems, intercenter and inter-agency agreements, and scientific papers; yet most of these processes are largely performed through manual means. Given on-line, interactive systems and a good communications capability, it becomes a minor step to provide a word processing capability which enables the author of a document to generate a document, have it reviewed by others, revise the document, and publish it without typing any portion of it more than once, and without standing over a copying machine and circulating review copies.

The current state of the art in "expert systems" is well suited to managing standardized administrative forms. One can easily envision a "Travel Expert System," for example, which knows government travel regulations, per diem rates, etc., and could interactively assist an individual (secretarial or professional) in constructing a set of travel orders. Such a system would also know the approval required and could automatically route the orders to the appropriate signature authorities. Changes in travel regulations would then be integrated into the expert system directly and applied as necessary, avoiding the costly notification process to all concerned individuals, with the assurance that everyone will be using up-to-date information.

Utilizing state-of-the-art office automation technology within NASA to manage documentation, coordinate manpower, and provide communication among personnel could significantly improve the productivity of NASA personnel. Integrating state-of-the-art machine intelligence capabilities into such an office environment could provide untold improvements in the efficiency and effectiveness of the organization, including the potential for significantly enhanced project management and rigorous statistical exploration of policy alternatives and their impacts.

6.6.5 Organizations

NASA is not presently organized to support a vigorous program in CS&T. The most apparent lack is a discipline office at the Headquarters level which supports research and development in computer science and serves as an agency advocate for the incorporation of state-of-the-art capabilities into NASA programs. There presently exist within the space agency many computer scientists capable of pursuing state-of-the-art research and of integrating contemporary technology into NASA programs, but there is no place for them to go for support other than mission-oriented offices whose goals and objectives are not consistent with supporting long-term commitments in CS&T R&D.

In addition to recognizing the requirement for a Headquarters CS&T discipline office, the study team fully supports the recommendation of the NASA Study Group on Machine Intelligence and Robotics that an advisory council composed of industry leaders in CS&T should be formed. This council would assist the agency in developing its computer science programs in order to assure a proper focus and to construct the appropriate relationships with other research organizations.

It is beyond the scope of the present study to recommend how NASA should organize institutionally to develop its CS&T capabilities, but several ideas have surfaced which may be useful to the agency in its consideration of future courses of action. In order to be effective, it would seem appropriate that NASA's CS&T endeavors maintain a multimission focus. A possible starting point may be a nucleus of discipline specialists to develop the program coupled with an agency-wide matrix management strategy to apply contemporary CS&T in mission environments and a vigorous encouragement of the development of CS&T "centers of excellence" at the Centers. Before embarking on any major organizational changes, however, it is useful to perform a systems analysis to fully explore the organizational possibilities and their ramifications. In this regard, the techniques developed by Krone (1980) are highly recommended.

In consideration of the two-fold objective of maintaining state-of-the-art expertise in CS&T and applying this expertise to NASA programs, one is confronted with the dilemma of providing both an effective R&D environment and a line organization capability to apply CS&T to real missions. If one assumes the existence of line organizational entities that now exist within NASA but applied to CS&T endeavors, then a possibility to be considered is the augmentation of the line management positions with staff researchers as illustrated in table 6.8. The positions of "Fellow" are intended to be highly competitive and attractive positions open to employees of NASA or other government agencies, industry, and academia. They might be treated similarly to professorships within universities, where highly talented individuals may be tenured in a position, but many are rotated through positions on temporary assignments of several years' duration.

An intriguing organizational structure apparently has been developed by the Navy for its new artificial intelligence laboratory at the Naval Research Laboratory (NRL) in Washington, D.C. A rough outline of the structure is shown in figure 6.3. The major point of interest regarding the NRL effort is that it is organizationally constructed to maximize scientific productivity. The head of the organization is the Chief Scientist, who is expected to contribute in a meaningful scientific way to the work of the organization. The CS&T study group was able to find out rather few details regarding the proposed operation of the NRL facility, but it appears to be sufficiently interesting that NASA may wish to explore this alternative during the system analysis phase.

6.6.6 Programs for Excellence

Perhaps the most fundamental requirement in maintaining one's technical excellence is to maintain active relationships with peer researchers. This will involve both formal and informal interfaces with standards organizations, other government agencies, universities, industrial R&D programs, and professional societies. A good set of computational facilities and communications capabilities as proposed in section 6.6.4 will facilitate this process.

Participation in joint government/industry/academic programs such as institutes and consortiums can formally provide not only a mechanism for applying more leverage to technical problems but also, potentially, a very appropriate forum for technical interchange. "Visiting Scientist" programs, where NASA sends selected individuals to major research environments such as MIT, SRI, and Xerox/PARC for periods of 6 months to a year can be very effective in transferring state-of-the-art concepts and technology into NASA programs. The Agency may also wish to consider Scientist Exchange Programs, in which NASA scientists perform research in university or industrial environments while their counterparts work at NASA for 6 months to a year. In some cases, close relationships between field centers and local universities may be mutually beneficial, and

TABLE 6.8.– POSSIBLE PERSONNEL POSITIONS FOR CS&T RESEARCH

Pay, status	Line management	Applied research and consulting	Independent research
	Division Chief	Senior Staff Scientist	Senior Research Fellow
	Branch Head	Staff Scientist	Associate Research Fellow
	Section Head	Research Scientist	Assistant Research Fellow
		Responsibility	Freedom

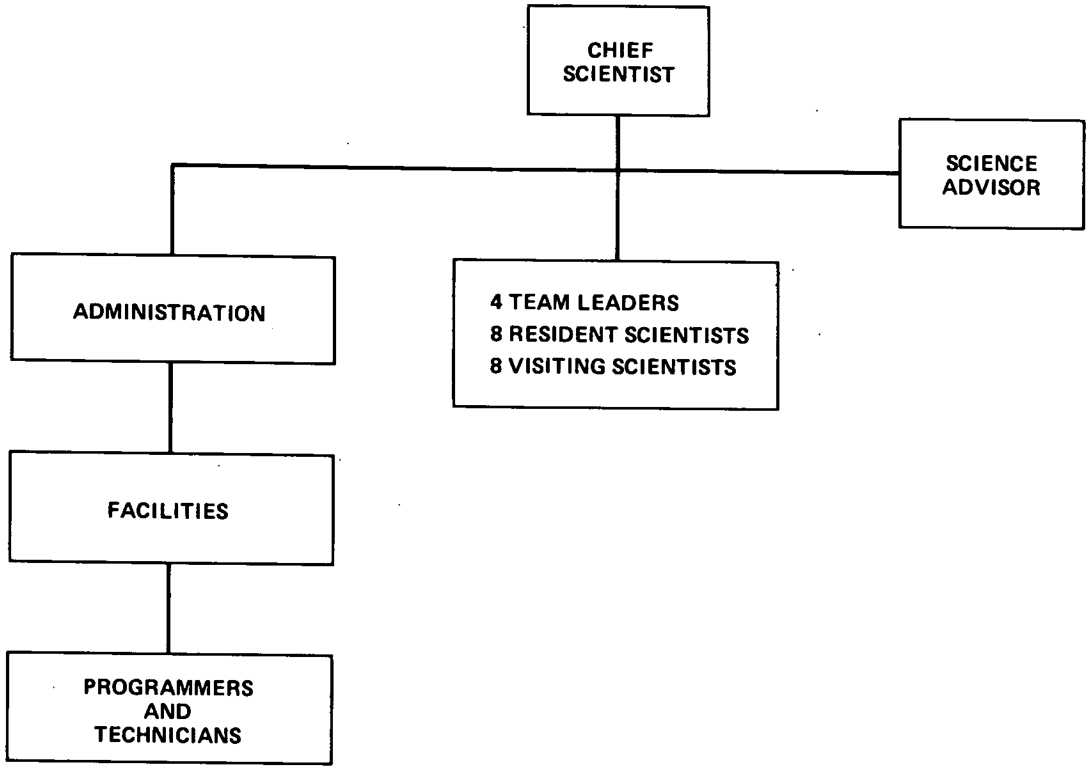


Figure 6.3.– NRL Artificial Intelligence Laboratory organization.

could include adjunct professorships as well as sponsoring graduate student thesis research at NASA facilities. This could prove to be an effective recruiting device. The agency may also wish to consider student loan programs for graduate students, wherein part of the loan is forgiven if the student completes an advanced degree and comes to work for NASA. A final suggestion is to sponsor Ph.D. thesis competitions, in which a "Space Technology Award," say, of perhaps $5000 is awarded annually to the best thesis relating to problems relevant to NASA. In general, the CS&T study group believes that there are many institutional programs which will cost NASA very little, yet can do much to maintain a NASA capability once it is obtained. Within NASA are examples of how programs such as these have succeeded particularly well in physics and the space sciences.

6.6.7 Summary and Conclusions

The primary technical components of a NASA program in Computer Science and Technology include research in machine intelligence, real-time systems, information retrieval, supervisory systems, and computer systems. NASA's computing and communications facilities require substantial upgrading in order to perform the proposed research and to attract competent personnel. An interactive, on-line programming environment appears essential, as well as a move by NASA to provide extensive communication capabilities among its computers and evolve toward an agency-wide computer network. It is further recommended that NASA consider the potentials of office automation technology for routine administrative work and document management, and explore the utility of machine intelligence in project management and policy analysis.

The organizational structure required to perform both state-of-the-art research and to apply modern CS&T was considered. The team concluded that this topic deserves a complete organizational analysis of alternatives, a task which can most effectively be done within NASA itself. Such a study should be given high priority in consideration of responses to requirements for implementing an advanced machine intelligence based program within NASA.

6.7 References

Amarel, S.: On Representations of Problems of Reasoning About Actions. In Machine Intelligence, D. Michie, ed., vol. 3, Edinburg Univ. Press, Edinburg, 1968.

ASC, "NASA Data/Information System-Related Technology Trends," prepared for NASA Goddard Space Flight Center by the Analytic Sciences Corporation (ASC) under contract NAS5-25739, May 1980.

Balzer, Robert et al., eds.: Proceedings of the First Annual National Conference on Artificial Intelligence, Stanford University, 18-21 August 1980. American Association of Artificial Intelligence and Stanford University, Menlo Park, Garcia-Robinson, 1980.

Baum, Michael: Giving a Robot the Eye, Dimensions, vol. 63, no. 4, April 1979, p. 3.

Bejczy, Antal K.: Advanced Teleoperators. Astronautics and Aeronautics, vol. 17, May 1979, pp. 20-31.

Bobrow, D. G.; and Winograd, T.: Experience with KRL-O: One Cycle of a Knowledge Representation Language. Proc. of the Fifth International Joint Conference on Artificial Intelligence, MIT, Cambridge, Massachusetts, 1977, pp. 213-222.

Bradley, William E.: Telefactor Control of Space Operations. Astronautics and Aeronautics, vol. 5, May 1967, pp. 32-38.

Bradley, William E.: Immediate Program to Provide Telefactor Systems for Demonstration and Test in NASA Centers. Personal communication, 5 August 1980.

Breckenridge, R. A.; and Husson, C.: Smart Sensors in Spacecraft: Impact and Trends. Progress in Astronautics and Aeronautics, vol. 67, AIAA, New York, 1979, pp. 3-12.

Brown, J. S.; and Burton, R. R.: Multiple Representations of Knowledge for Tutorial Reasoning. In Representations and Understanding, Daniel G. Bobrow and Allan Collins, eds., Studies in Cognitive Science. Academic Press, New York, 1975, pp. 311-349.

Buchanan, B. G.; and Lederberg, Joshua: The Heuristic DENDRAL Program for Explaining Empirical Data. IFIP Congress, 71, Ljubljana, Yugoslavia, Aug. 23-28, 1971. Information Processing 71, Proceedings. C. V. Freiman, Editor. Amsterdam, North-Holland Publ. Co., 1972, pp. 179-188.

Chin, C. H., ed.: Joint Workshop on Pattern Recognition and Artificial Intelligence. Institute of Electrical and Electronics Engineers, June 1-6, 1976, Hyannis, Mass. 1976.

Computing Reviews: Categories of the Computing Sciences, vol. 17, no. 5, May 1976, pp. 172-195.

Corliss, William R.; and Johnsen, Edwin G.: Teleoperator Control, NASA SP-5070, 1968.

Duda, Richard O.; and Hart, Peter E.: Pattern Classification and Scene Analysis. John Wiley & Sons, New York, 1973.

Duda, R. O. et al.: Development of the PROSPECTOR Consultation System for Mineral Exploration. Final Report, SRI International, Projects 5821 and 6415, Menlo Park, Calif., October 1978.

EER: Information Science Program Plan for NASA/ Goddard Space Flight Center. Prepared by Engineering and Economics Research, Inc. (EER), under contract NAS5-73942-B, June 1980.

Erman, L. D.; and Lesser, V. R.: A Multi-Level Organization for Problem Solving Using Many, Diverse, Cooperating Sources of Knowledge. Proceedings of the Fourth International Joint Conference on Artificial Intelligence, Tbilisi, Georgia, USSR, Sept. 1975, pp. 483-490.

Ernst, George W.; and Newell, Allen: GPS: A Case Study in Generality and Problem Solving. Academic Press, New York, 1969.

Fann, K. T.: Peirce's Theory of Abduction. Martinus Nijhoff, The Hague, 1970.

Feigenbaum, E. A.: The Art of Artificial Intelligence – Themes and Case Studies of Knowledge Engineering. Proc. Fifth International Joint Conference on Artificial Intelligence, MIT, Cambridge, Massachusetts, 1977.

Feigenbaum, E. A.; Buchanan, B. G.; and Lederberg, J.: On Generality and Problem Solving: A Case Study Using the DENDRAL Program in Machine Intelligence, B. Meltzer, D. Michie, eds., vol. 6, Halsted Press, New York, 1971, pp. 165-190.

Fikes, R. E.; and Nilsson, N. J.: STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. Artificial Intelligence, vol. 2, 1971, pp. 189-208.

Gilmore, P. A.; Batcher, K. E.; Davis, M. H.; Lott, R. W.; and Buckley, J. T.: Massively Parallel Processor. Phase I Final Report, prepared by Goodyear Aerospace Corporation for Goddard Space Flight Center, July 1979.

Goldstein, I. P.; and Roberts, R. B.: Nudge: A Knowledge-Based Scheduling Program. Proc. of the Fifth International Joint Conference on Artificial Intelligence, MIT, Cambridge, Massachusetts, 1977, pp. 257-263.

Goulomb, S.; and Baumert, L.: Backtrack Programming. J. ACM, vol. 12, 1965, pp. 516-524.

Hajeck, P.; and Havranek, T.: Mechanizing Hypothesis Formation: Mathematical Foundations for a General Theory. Springer-Verlag, Berlin, 1978.

Hanson, Norwood Russell: Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science. Cambridge Univ. Press, Cambridge, 1958.

Haralick, R. M.: Texture-Tone Study with Application to Digitized Imagery. Final Tech. Summary, ETL-CR-74-17, Contract DADKO2-70-C-03888, August 1974.

Haye, R. M.: Smart Remote Sensor Needs for U.S. Coast Guard Ocean Environmental Missions. Progress in Astronautics and Aeronautics, vol. 67, AIAA, New York, 1979, pp. 44-59.

Hayes-Roth, Frederick: Theory-Driven Learning: Proofs and Refutations as a Basis for Concept Discovery. Paper delivered at the Workshop on Machine Learning, Carnegie-Mellon Univ., July 1980.

Heer, Ewald: Toward Automated Operations in Space: An Introduction. Astronautics and Aeronautics, vol. 17, May 1979, pp. 16-18.

Hewitt, C.: Description and Theoretical Analysis (Using Schemata) of PLANNER: A Language for Proving Theorems and Manipulating Models in a Robot. Artificial Intelligence Laboratory Report AI-TR-258, MIT, Cambridge, Massachusetts, 1972.

Ho, Y. C.: Team Decision Theory and Information Structures. Proc. of the IEEE, vol. 68, June 1980, pp. 644-654.

IEEE Proceedings on Pattern Recognition and Image Processing, May 1979.

Kalush, R. J., Jr.: The Problem of Resolution in the LANDSAT Imagery. Remote Sensing Quarterly, vol. 2, 1980, pp. 3-48.

Kowalski, R. A.: Predicate Logic as Programming Language. Proc. IFIP 1974, North-Holland, Amsterdam, 1974, pp. 569-574.

Kraiss, F.: Decision-Making and Problem Solving with Computer Assistance. NASA TM-76008, January 1980.

Krone, R. M.: Systems Analysis and Policy Sciences, Theory and Practice. John Wiley & Sons, Inc., 1980.

Kuipers, B. J.: Representing Knowledge of Large-Scale Space. Artificial Intelligence Laboratory Report AI-TR-418, MIT, Cambridge, Massachusetts, 1977.

Lakatos, Imre: Falsification and the Methodology of Scientific Research Programs. In Criticism and the Growth of Knowledge, Imre Lakatos and Alan Musgrave, eds., Univ. Press, Cambridge, 1970a, pp. 91-196.

Lakatos, Imre: The Changing Logic of Scientific Discovery. University Press, Cambridge, 1970b.

Lakatos, Imre: Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge Univ. Press, Cambridge, 1976.

Lighthill, Sir James: Artificial Intelligence: A Paper Symposium, Science Research Council, Great Britain, 1972.

Lippay, Andrew L.: Multi-Axis Hand Controller for the Shuttle Remote Manipulator System. Paper delivered at the 13th Annual Conference on Manual Control, June 1977.

Martin, W. A.; and Fateman, R. J.: The MACSYMA System. In Proc. of the ACM Second Symposium on Symbolic and Algebraic Manipulation, S. R. Petrick, ed., Los Angeles, Calif., March 1971, pp. 23-25.

Matsushima, H.; Uno, T.; and Ejur, M.: Image Processing by Experimental Arrayed Processor. 6th International Joint Conference on Artificial Intelligence, Tokyo, 1979, pp. S13-S15.

Meade, E.; and Nedwich, B.: Manipulator Development Facility Description. NASA JSC-11029, Rev. A, January 1978.

Miller, R. H.; and Smith, David B. S.: Extraterrestrial Processing and Manufacturing of Large Space Systems, Volumes 1-3, NASA Contract NAS8-32925, NASA CR-161293, September 1979.

Minsky, M.: A Framework for Representing Knowledge. In The Psychology of Computer Vision, P. H. Winston, ed., McGraw-Hill, New York, 1975, pp. 211-277.

Minsky, M.: Toward a Remotely-Manned Energy and Production Economy. MIT Artificial Intelligence Laboratory, A.I. Memo No. 544, September 1979. 19 pp. See also "Telepresence," Omni, vol. 2, June 1980, pp. 44-52.

Mitchell, O. R.; Meyers, C. R.; and Boyne, W. A.: A Max-Min Measure for Image Texture Analysis. IEEE Transactions on Computers, vol. C-25, 1977, pp. 408-414.

Morgan, C. S.: Hypothesis Generation by Machine. Artificial Intelligence, vol. 2, 1971, pp. 179-187.

Morgan, C. S.: On the Algorithmic Generation of Hypothesis, Scientia, vol. 108, 1973, pp. 585-598.

Murphy, L. P.; and Jarman, J. W.: The Role of Smart Sensors in Earth Resources Remote Sensing Programs. Progress in Astronautics and Aeronautics, vol. 67, AIAA, New York, 1979, pp. 101-109.

NASA Advisory Council: Future Directions for the Life Sciences in NASA. Report of the Life Sciences Advisory Committee of the NASA Advisory Council, November 1978.

NASA SP-387: A Forecast of Space Technology, 1980-2000, 1976.

Newell, A.; and Simon, H. A.: GPS: A Program that Simulates Human Thought. In Computers and Thought, E. A. Feigenbaum, J. Feldman, eds., McGraw-Hill, New York, 1963, pp. 279-293.

Nilsson, Nils J.: Problem-Solving Methods in Artificial Intelligence. McGraw-Hill, New York, 1971.

Nilsson, Nils J.: Artificial Intelligence. Information Processing 74, vol. 4, Proceedings of IFIP Congress 74, 1974, pp. 778-801.

Nilsson, Nils J.: Principles of Artificial Intelligence. Tioga, Palo Alto, California, 1980.

Office of Aeronautics and Space Technology (OAST): NASA Space Systems Technology Model, volumes 1-3, NASW-2937, NASA Headquarters, Washington, D.C., May 1980.

Peirce, Charles Sanders: Collected Papers, vols. 1-6, Charles Hartshorne and Paul Weiss, eds. The Belknap Press of Harvard Univ. Press, Cambridge, 1960.

Peirce, Charles Sanders: Collected Papers, vols. 7-8. Arthur W. Burks, ed. The Belknap Press of Harvard Univ. Press, Cambridge, 1966.

Pohl, I.: Practical and Theoretical Considerations in Heuristic Search Algorithms. In Machine Intelligence, vol. 8, E. W. Elcock, D. Michie, eds., Ellis Horwood Ltd. (Halsted Press), 1977, pp. 55-72.

Pople, H. E., Jr.: The Formation of Composite Hypotheses in Diagnostic Problem-Solving: An Exercise in Synthetic

Reasoning. Proc. of the Fifth International Joint Conference on Artificial Intelligence, MIT, Cambridge, Massachusetts, 1977, pp. 1030-1037.

Radner, R.: Team Decision Problems. Ann. Math. Stat., vol. 33, 1962, pp. 857-881.

Raibert, M. H.: Autonomous Mechanical Assembly on the Space Shuttle: An Overview. NASA CR-158818, July 1979.

Reiter, R.: The Use of Models in Automatic Theorem-Proving. Technical Report 72-09, Dept. of Computer Science, Univ. of British Columbia, Canada, 1972.

Reiter, R.: A Logic for Default Reasoning. Artificial Intelligence, vol. 13, 1980, pp. 81-132.

Rich, E.: Building and Exploiting User Models. Proc. of 6th International Joint Conference on Artificial Intelligence. Tokyo, 1979, pp. 720-722.

Sacerdoti, E. D.: Planning in a Hierarchy of Abstraction Spaces. Artificial Intelligence, vol. 5, 1974, pp. 115-135.

Sacerdoti, E. D.: Problem Solving Tactics. Proc. Sixth International Joint Conference on Artificial Intelligence. Tokyo, Japan, 1979, pp. 1077-1085.

Sagan, Carl, chmn.: Machine Intelligence and Robotics: Report of the NASA Study Group. NASA-JPL Report No. 715-32, March 1980.

Sandell, N. R.; Varaiya, P.; Athans, M.; and Safonov, M. G.: Survey of Decentralized Control Methods for Large Scale Systems. IEEE Trans. Automat. Contr., vol. AC-23, April 1978, pp. 108-128.

Sandford, D. M.: Using Sophisticated Models in Resolution Theorem Proving. Lecture Notes in Computer Science, vol. 90, Springer-Verlag, New York, 1980.

Schappell, R. T.: A Study of Early Intrusion into Algorithm Development Utilizing Information Processing Advanced Technology. Martin Marietta Corp., P-80-48052-1, 1980a.

Schappell, R. T.: FILE Technology Developments for Remote Sensing. SPIE Annual Int. and Tech. Symposium and Exhibit, San Diego, California, 31 July 1980b.

Schappell, R. T.; Polhemus, J. T.; Lowrie, J. W.; Hughes, C. A.; Stephens, J. R.; and Chang, Chieng-Y.: Applications of Advanced Technology to Space Automation. NASA CR-158350, January 1979a.

Schappell, R. T.; and Tietz, J. C.: Landmark Identification and Tracking Experiments. Progress in Astronautics and Aeronautics, vol. 67, AIAA, New York, 1979, pp. 134-151.

Schappell, R. T.; Vandenberg, F. A.; and Hughes, C. A.: Study of Automated Rendezvous and Docking Technology: Final Report. Martin Marietta Corporation, 1979b.

Schlienn, S.: Geometric Correction Registration and Resampling of LANDSAT Imagery. Canad. J. Remote Sensing, vol. 5, 1979, pp. 74-89.

Shin, C. N.; and Yerazunis, S.: Data Acquisition and Path Selection Decision-Making for an Autonomous Roving Vehicle. Progress Report, Rensselaer Polytechnic Institute, TR MP-62, 1978.

Shortliffe, E. H.: Computer-Based Medical Consultations: MYCIN. Elsevier, New York, 1976.

Space Shuttle, JSC, 1976: Space Transportation User Handbook, JSC, July 1977; Space Shuttle System Payload Accommodations, JSC-07700, vol. XIV, Revision F, 22 September 1978.

Spann, G. W.: Satellite Remote Sensing Markets in the 1980's. Photogrammetric Engineering and Remote Sensing, vol. 46, 1980, pp. 65-69.

Srinivasan, C. V.: The Architecture of Coherent Information Systems: A General Problem Solving System. IEEE Trans. on Computers, vol. C-25, 1976, pp. 390-402.

Srinivasan, C. V.: Knowledge-Based Learning, An Example. DCS-TR-90, paper presented at Workshop on Machine Learning, Carnegie-Mellon Univ., July 1980.

Srinivasan, C. V.; and Sandford, D. M.: Knowledge-Based Learning Systems, Part I: A Proposal for Research; Part II: An Introduction to the Meta-Theory; Part III: Logical Foundations. DCS-TR-89, Dept. of Computer Science, Hill Center, Busch Campus, Rutgers Univ., New Brunswick, New Jersey, March 1980.

Sussman, G. J.: A Computer Model of Skill Acquisition. American Elsevier, New York, 1975.

Teitelman, W.: A Display Oriented Programmer's Assistant. Int. J. Man-Machine Studies, November 1979, pp. 157-187.

Tenenbaum, Jay M.: Reconstructing Smooth Surfaces. In Proc. Image Understanding Workshop, Univ. of Southern California, November 1979.

Thorley, G. A.; and Robinove, C. J.: Current and Potential Uses of Aerospace Technology by the U.S. Department of the Interior. Progress in Astronautics and Aeronautics, vol. 67, AIAA, New York, 1979, pp. 15-26.

Vahey, D. W.: Smart Remote Holographic Processor Based on the Materials Characteristics of $LiNbO_3$. Progress in Astronautics and Aeronautics, vol. 67, AIAA, New York, 1979, pp. 352-366.

Van de Brug, G.; and Minker J.: State-Space, Problem-Reduction, and Theorem Proving – Some Relationships. J. ACM, vol. 18, 1975, pp. 107-115.

Van Emden, M.: Programming with Resolution Logic. In Machine Intelligence, E. W. Elcock and D. Michie, eds., Ellis Horwood Ltd. (Halsted Press), vol. 8, 1977, pp. 266-299.

Warren, D.; and Pereira, L. M.: PROLOG – The Language and its Implementation Compared with LISP. Proc. Symposium on Artificial Intelligence and Programming Languages, ACM, 1977.

Whitney, M. W.: Processing and Storing Information. In A Forecast of Space Technology, 1980-2000. NASA SP-387, section IV, 1976.

Williams, O. W.: Outlook on Future Mapping, Charting and Geodesy Systems. Photogrammetric Engineering and Remote Sensing, vol. 46, 1980, pp. 487–490.

Wilson, G. A.; and Minker, J.: Resolution, Refinements, and Search Strategies: A Comparative Study. IEEE Transactions on Computers, vol. C-25, 1976, pp. 782–801.

Winston, Patrick H.: Progress in Artificial Intelligence. Vols. I and II, 1978.

Woods, W. A.: What's in a Link: Foundations for Semantic Networks. In Representation and Understanding, D. G. Bobrow and A. Collins, eds., Academic Press, New York, 1975, pp. 35–82.

Woods, W. A.; Kaplan, R. M.; and Nash-Webber, B.: The Lunar Sciences Natural Language Information System: Final Report, BBN Report No. 2378, Bolt, Beranek and Newman, Cambridge, Massachusetts, 1972.

CHAPTER 7

CONCLUSIONS AND IMPLICATIONS OF AUTOMATION IN SPACE

During the 1960s NASA proved that access to space is feasible for both manned and unmanned systems. During the 1970s NASA demonstrated that important scientific exploration and applications missions could be conducted in orbit. Simultaneously, imaginative and worthwhile future space missions were conceived and studied. However, two major constraints limited implementation – cost and technology. A great many proposed missions could be accomplished through application of current technology but at unacceptable costs. New technology is needed which is not only mission-enabling but also cost-reducing. The Space Shuttle is NASA's first major technology project to address these twin objectives. In years to come, advanced automation will play a major role in achieving similar objectives.

A space mission life cycle may be divided into three phases: (1) conception, design, development, test, and evaluation; (2) procurement of mission flight and ground articles; and (3) mission operations. At present, procurement is only about 10% of life-cycle costs. Most facility support equipment already is in place, also on the order of 10% of mission life-cycle costs. The first phase through conception, development, and test, and the last phase of flight operations, each is on the order of 40% of life-cycle costs. Consequently, reducing space hardware disbursements has small effect on total life cycle costs.

The present dominance of the first and last phases in mission accounting is due in large measure to their people-intensive character. Cost reductions are possible by focusing on two specific goals. First, increased personnel productivity can help make space affordable, in part by using computer technology to organize and integrate knowledge, for information extraction and retrieval, decisionmaking, scheduling, and for automatic problem-solving. The efficiency of human action may also be improved through advanced teleoperations and robotics. Second, costs may be cut by decreasing the requirements for human interaction and the need for terrestrial materials. This ultimately can be accomplished through more complete *in situ* machine intelligence and robotics.

Advanced automation can substantially contribute to both approaches. Applicable techniques range from intelligent computer assistants for enhanced human productivity to, ultimately, autonomous self-replicating systems utilizing extraterrestrial materials and energy. These latter automatons could be materially self-sufficient and produce immense economic returns if employed in production or service capacities.

The Mission Goals Symposium which took place at Pajaro Dunes in June 1980 (sec. 1.2.3) addressed a specific question: "What bold new NASA space missions could high levels of automation make possible 25-50 years hence?" In their deliberations the participants postulated levels of automation capability that might be achievable given adequate funding and a clear focus, and also a range of mission types that such capabilities could, at least in principle, make possible.

The Symposium concluded that if certain (very difficult) new levels of automation capability could be achieved, a whole new set of space missions having high economic and scientific value would become possible. In each case a decision to pursue one of these long-term goals would demand focused research beginning decades earlier, each having a series of rather sharply defined short-term goals of its own. Such subgoals provide valuable focus and stimulation for automation research generally, and suggest a natural stepping-stone developmental sequence of graded complexity in the areas of command and control, robot dexterity and repair capability, sensing and reasoning, and multirobot system organization.

7.1 Space Facilities and Programs Overview

The missions considered in this study are based on a broad array of activities which have been proposed to achieve various scientific and technical objectives. If cost as a factor were excluded, there would be little question of the impetus for doing most if not all of these missions. The costs involved, however, are such as to require an orderly progression of activities so that needed technologies can be developed in an affordable manner over the next several decades. The scenario that has emerged from this study is logical, with an orderly progression from early Shuttle operational phases to the establishment of self-replicating lunar factories and (possibly) space colonies. An underlying premise is the commitment to an ongoing program of space exploration and utilization.

While space exploration can be accomplished largely using unmanned, highly automated craft, space utilization

involves a wide range of activities where human intelligence and versatility are invaluable. It can be argued that any activity involving human participation ultimately can be preprogrammed and accomplished by machine, but it is equally true that at present total automation would be prohibitively expensive. There is a tradeoff made between full- and nonautomation, as suggested by figure 7.1. There is no single optimum level of automation, but rather a range of performance hybrids from which the mission planner must choose. As a technology base develops, incorporating advances in computer science, artificial intelligence, and robotics, the cost of autonomous operations should decrease, thereby reducing mission costs and giving planners more options.

The Space Transportation System as described in the NASA Technology Model represents a well developed technology base amenable to future progress in advanced automation. Indeed, the scenario developed in the present study leads to a logical development of this capability. This is grounded on a rationale of blending man and automation for maximum productivity, together with the parallel evolutionary emergence of fully automated systems. The scenario provides a roadmap for actualizing NASA's commitment to advanced programs.

The study group considers a LEO base critical to later programs. Thus the facilities and programs plan can be considered in three phases: (1) early operations at LEO, (2) establishment of a permanent LEO base and extended operations at LEO, and (3) operations beyond LEO. Program development plans consistent with this strategy are continuously refined and updated within NASA. The transition from one phase into the next is not chronologically precise or even distinct, depending as it does upon the availability of skills and techniques permitting development of the next phase. Further, there is some flexibility in selection of activities within a given phase (see fig. 7.2). Priorities or technological breakthroughs may reorder or modify some programs. However, the development of technologies required to support this or some similar plan should progress roughly according to the timetable shown in figure 7.3 to maintain an orderly space program development.

7.2 A Consistent Space Program Strategy

This report has addressed several missions and numerous technologies and problems related to space activities that may be undertaken during the next several decades: the integration of satellite technology into an intelligent network capable of answering broad or narrow questions concerning Earth resources; the exploration of Titan by an automated, intelligent probe; the construction of an automated factory on the Moon which self-replicates and delivers useful products, such as energy via solar power satellites and other means; and the development and growth of a material economy independent of Earth supply using

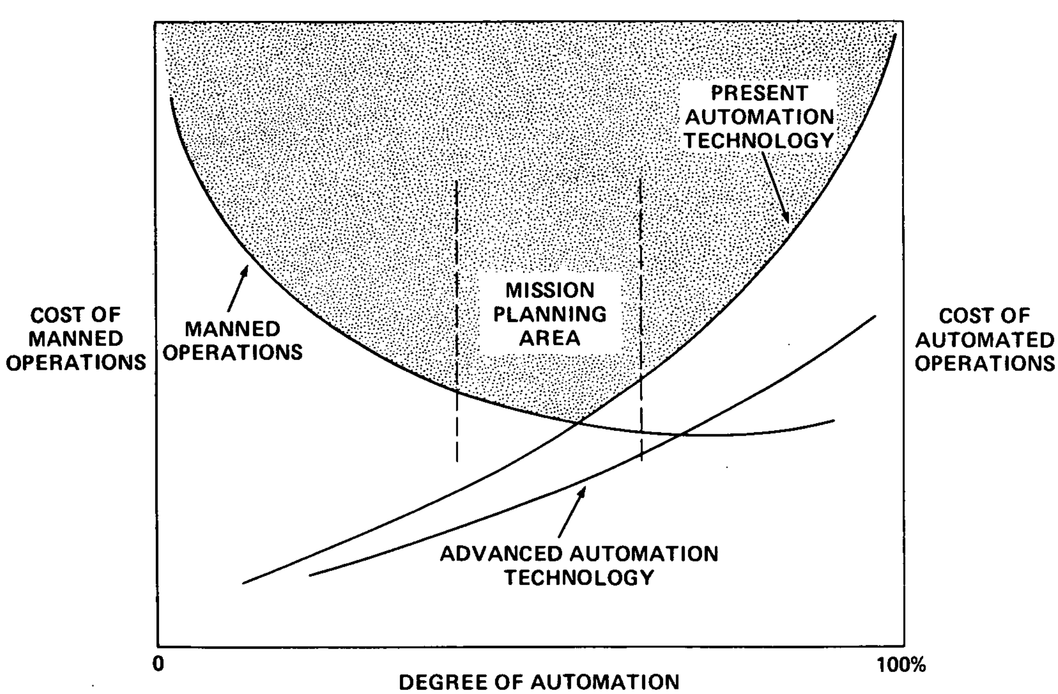


Figure 7.1.– Cost-effectiveness of automation in future NASA space missions.

extraterrestrial resources from the Moon and the asteroids. The Epilogue provides an opportunity to assess the broader perspective of space development – how to start, and how to grow in skill, knowledge, resources, and energy to accomplish long-term goals discussed in this report.

Studies of ecosystems as diverse as those of bacteria and whales suggest the maximization of information content and energy flow in all living systems (Miller, 1978; Odum, 1971). Examination of many species shows that diversity and adaptation to numerous habitats leads to survival. Humanity appears poised to accept the challenge of the frontier of space, and for the same fundamental reasons: knowledge, energy, resources, and room to grow; in short, for survival.

The space program has in the past been pursued for reasons of exploration, scientific knowledge, national security, and pride. In the future the *utilization* of space will take precedence over pure *exploration*. New resources from space can help put NASA programs on a sounder footing in providing benefits and services of great economic and national importance. If NASA clearly realizes the tremendous opportunity and makes these goals known to the public, it will accomplish feats and gain popular support far greater than ever thought possible.

The aim of the present discussion is to show how space activities undertaken by NASA in the immediate future and over the next several decades can help solve three major problems facing the United States today:

- Energy independence
- Material independence
- Increased productivity

The tools for solving these problems include:

- Space technology
- Teleoperators
- Automation and robotics

The American people have great pride in their technology and ability to confront challenges. Presently the United

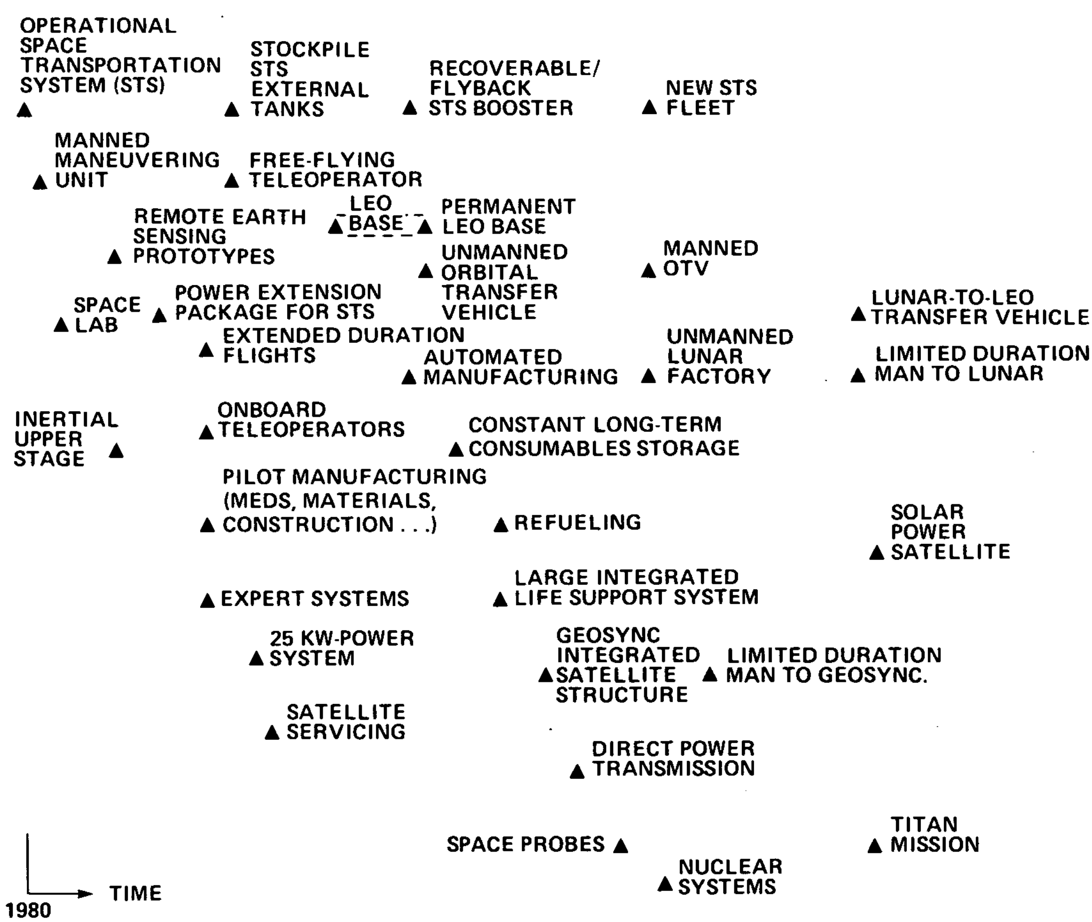


Figure 7.2.– Space facilities and programs.

States is faced with a situation in which national economic independence is held hostage to the critical resources of minerals (manganese, cobalt, chromium, titanium, and tantalum) and energy (OPEC petroleum) controlled by unstable or unfriendly nations. Space can provide energy and mineral resources to sustain steady growth of both our economy and that of the world for the foreseeable future. The notion that space can produce benefits that directly affect the way ordinary people live has the potential to mobilize strong public support and provide the funding and planning stability NASA now so desperately needs.

The requirements of energy, knowledge, and resources create markets for specific services. Already the communications and Earth resources satellites are experiencing rapid market growth. NASA gradually must develop new markets and new constituencies in the business and public user communities. Lift costs remain expensive and necessitate reducing Earth resupply. In the past this pressure has led, for example, to satellites able to survive unattended for years. In the near future, it will mean on-orbit assembly and checkout, and on-orbit repair and refueling. In the long-term, satellites can be manufactured directly from extraterrestrial materials in space.

According to United Nations' estimates, some 70% of humanity is poor, underfed, and undereducated. Most recent analyses of the future have concluded that the outlook for humanity is dark – an increasingly grim world of limited living space and resources on a finite planet. To our knowledge, *none* of these gloomy studies has considered the liberating potential of space, either the advanced technology necessary to master it or the possibilities for long-term solutions using extraterrestrial resources (Vajk, 1978).

7.2.1 Specific Goals for Growth Scenario

Some studies of space stations start by assuming full-scale activity or by ignoring the process of growth entirely. A central conclusion of the present study is that growth must be incorporated into policy planning, and must proceed from current or easily foreseeable capabilities toward desired goals in such a way that two principles are observed:

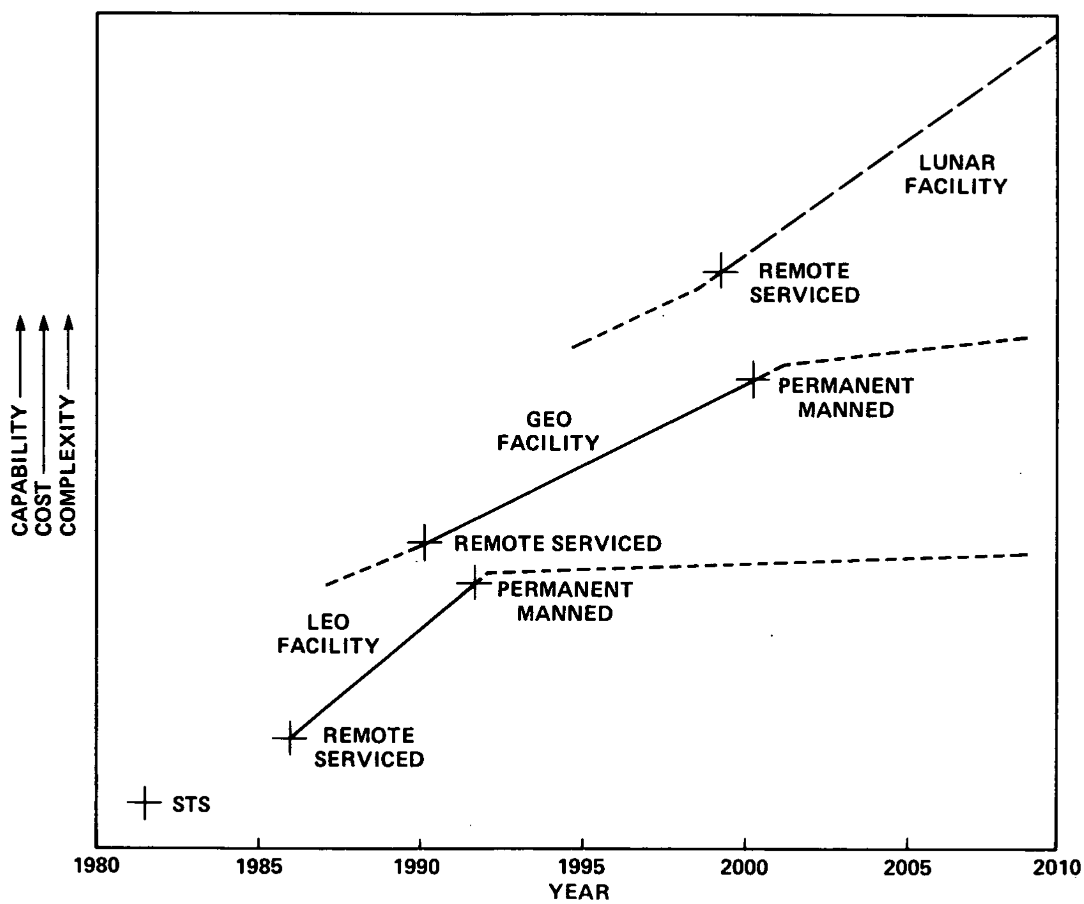


Figure 7.3.– Timetable for development of automated activities in space.

(1) Each growth step must be justifiable in its own right.

(2) Each growth step must lead clearly toward the defined goals.

The starting points for the growth scenario have been outlined in preceding discussions. The final goal, human operation in space with greater independence from Earth, is pursued through a series of interim goals – in particular, the development of the abilities to:

- Modularize equipment
- Tend co-orbiting free satellites
- Build, test, and transfer to orbit large complex systems
- Reduce station Earth dependence for control and resupply
- Automate space operations in support of in-space human activity
- Advance deep space exploration

In pursuit of these goals a program scenario is required. Figure 7.4 shows an overview of space activities starting

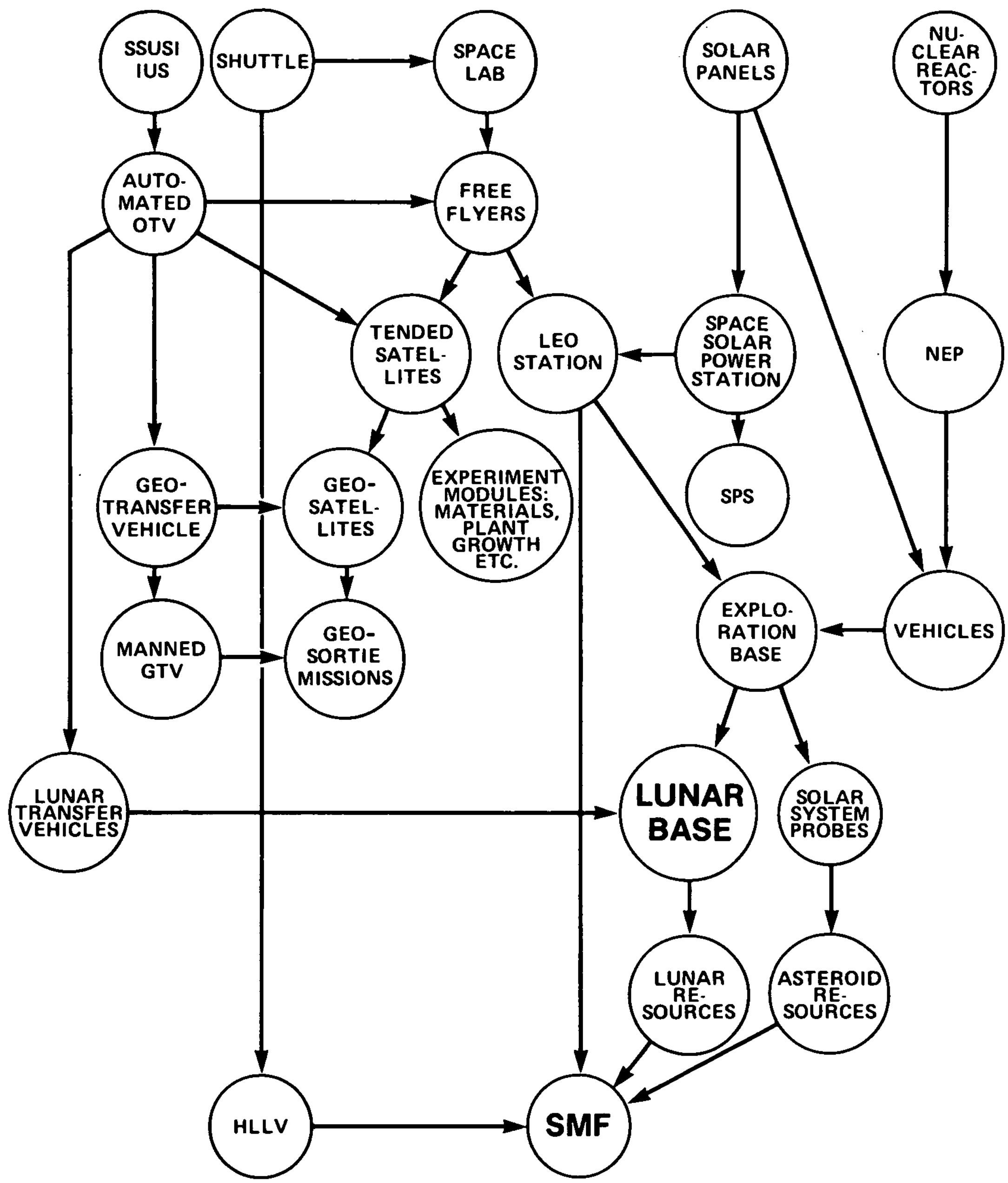


Figure 7.4.– Overview of space activities development.

from the present and continuing through the development of an increasingly independent space manufacturing capability. Figure 7.5 provides additional details on the possible development of a low Earth orbit station. Both suggest the kinds of facilities expected to be required to attain the desired goals. The use of a permanently manned LEO station as a Solar System exploration base and as a support station for tended free-flying unmanned satellites means that toward the end of this century many characteristics of manned and unmanned space endeavors will blend together.

Modularize equipment. Modularization of equipment reduces the mass of material that must be transported to and from orbit for repair and replenishment of orbital facilities. This is because only defective or depleted modules need be handled rather than the entire system. Overall system mass is higher for modular designs, but system maintenance mass is significantly reduced. The modification of an existing facility to accept new capabilities is greatly simplified through the use of plug-in modules. Parts installation and replacement already are prime candidates for early automation and these procedures are streamlined by modularization. Both station growth and the multiplication of capabilities become more feasible since the entire system need not be replaced as parts become inadequate or obsolete. The sizes of Spacelab and the Shuttle payload bay suggest a practical size limit for modules.

Modularization is not intended to replace space fabrication of stations and equipment but is to be used in parallel with on-orbit construction.

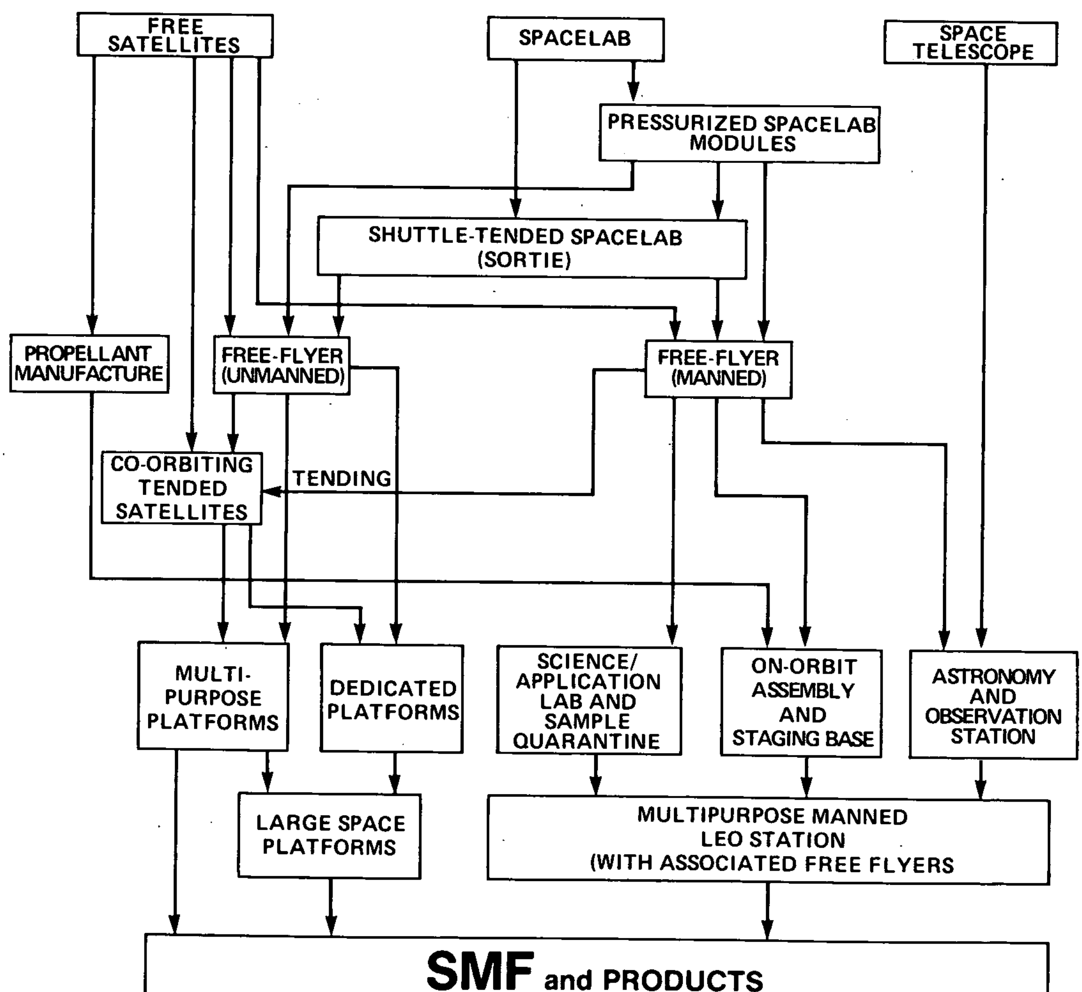


Figure 7.5.– Details of LEO station development.

Tend orbiting satellites. These satellites are emplaced by and are under the control of the main station. They fulfill a single major purpose or a related series of purposes such as Earth sensing, plant growth experiments, or optical astronomy. The use of tended free-fliers solves the problem of conflicting priorities (e.g., pointing a telescope at a star and a sensor at the ground simultaneously). It enables scientific experiments or other tasks to be performed without continuous human supervision but which do require very low movement or very low containment levels. The satellite module can be resupplied, repaired, or given new tasks and then left undisturbed. The satellites should themselves be designed on the modular principle, and tending them will drive development of technologies required to repair satellites in LEO and GEO, to assemble large satellites and space probes, and to provide flight support for missions conducted farther from the Earth.

Build, test, and transfer to orbit large complex space stations. This capability derives from Shuttle beam building experiments and the assembly of modular structures and produces many of the products of this stage of development. These products include:

- SPS proof-of-concept and prototype devices
- Communications platforms
- Large antennas and antenna forms
- Clustered satellites and multipurpose platforms

The ability to construct very large space structures such as the solar power station depends heavily on automated construction and assembly techniques developed at this stage.

Reduce station dependence on Earth for control and supply. Solar and nuclear power modules, modules for atmospheric recycling and renewal, food production and waste recycling capabilities, and the ability to assemble platforms and station sections from supplied parts and then to make those parts from supplied raw materials, all contribute to the growing independence of the space industrial complex. Increasing independence demands progressively increasing capacity to generate electricity (energy self-sufficiency), to recycle air and water in an increasingly closed ecosystem, to monitor crew and station health (life-sustaining independence), to process materials and fabricate products in space (economic and manufacturing independence), and finally to acquire and utilize nonterrestrial resources (material independence).

The stages of independence may be pursued in parallel, and the progress toward self-sufficiency is quite gradual. There is no requirement that a facility be fully or even mostly self-sufficient at the outset.

Automate space operations to support in-space human activity. Automation will not eliminate human activity in space. Human-built machines are not people and space in the long run will have little meaning for humanity unless people are living in space. The purpose of automation is to make human tasks easier and less hazardous, to remove tedious and repetitive tasks from human responsibility, and to enable each individual to accomplish more. The term "automation" is loosely used to encompass the full range of autonomous or semiautonomous systems including robots and free satellites that perform a set of tasks and exercise judgment when faced with unforeseen developments, automatic processing machines with little or no judgment, and teleoperators controlled entirely by humans who provide the required judgment. Prime candidates for early automation are the assembly of stations and OTV modules, satellite assembly, emplacement, replacement and repair, and space processing techniques.

Advance deep space exploration. Advanced future deep space probes can build on the experience of earlier probes such as Voyager and Viking, but the new generation of machines may be larger and considerably more autonomous. They can be greater in size because of modular construction in LEO. On-board artificial intelligence will enable the probes to perform largely autonomous scheduling, sequencing, and contingency planning. Automation may also include on-board analysis, data correlation, and perhaps, even mechanized, hypothesis-formation development of models describing a remote planet. All of these capabilities, along with robotic systems, are required for the ambitious Titan mission described in chapter 3.

In the shorter term, space probes will serve as the automated prospectors of the utilization-oriented space age. Asteroid rendezvous missions currently being considered will only be the first of many missions to seek out possible extraterrestrial resources for in-space materials processing. Multiple asteroid prospecting in the asteroid belt, comet rendezvous and sample return, Mars rover and sample return, planetary and satellite lander/rover/sample return missions are believed credible but are not yet in mainstream space mission planning.

7.2.2 Recommended NASA Space Systems Technology Model Updates

The NASA/OAST Space Systems Technology Model (OAST, 1980) is intended to serve as a reference for planning technology programs and options, identifying technologies required for planning and potential future missions, assessing ongoing technical programs, and providing a technology reference source for mission planning. Ascertaining requirements for planned and future missions ensures that the focus of technology programs is coupled to the overall goals and missions of NASA.

The three volumes of the Model treat systems, programs, and technology from the present to the reasonable limits of projection. Volume I describes those systems and programs which the NASA program offices endorse as being within

their 10-year planning horizon. Volume II contains trends and forecasts of space technology. Volume III provides information about innovative systems, programs, and technology. Part A presents novel systems and programs validated by program offices but which are either beyond the 10-year planning period or are still deemed speculative. Systems and programs in Part B derive from sources other than the program offices and, as such, are considered more speculative than Part A missions. Part C presents emerging technologies with little or no historical trend, and provides best possible forecasts of their potential.

An addition to the Model as a result of the present study should be made to Space Technology, volume II, under "Information Systems." Present categories include:

4.1 Sensors
4.2 Data processing
4.3 Communications

The Study Group recommends the following addition to section 4 of volume II:

4.4 Computer science and technology (including computer systems, software, management services, and systems engineering)

Further additions to volume II of the 1980 Model should be made under "Automated Operations." Present categories include only:

6.1 On-board automation
6.2 Automated problem solving
6.3 Machine vision

The study group recommends the following additions to section 6 of volume II:

6.4 Automated "World Model" based information systems (including land and ocean modeling, Earth atmosphere modeling, planetary modeling, automatic mapping, intelligent image processing and information extraction, "smart" sensors, plan formation and scheduling, and global system management);
6.5 Automated learning and hypothesis formation (including analytic, inductive, and abductive inferences);
6.6 Natural language and other man-machine communication (including machine understanding of keyed natural language, machine participation in natural language dialogues, machine recognition/understanding of spoken language, machine generation of speech, and visual and other means of communication such as iconic formats);
6.7 Automated space manufacturing (including automatic extraction and purification of raw materials, forming of product components, product component assembly and inspection, autonomous system control, and self-replicating machine systems generally);
6.8 Teleoperators and robot systems (including the remote manipulator system, the teleoperator maneuvering system and other free-flying teleoperators, on-board teleoperated "walkers" and mobile workbenches, robot devices, "telepresence" operator sensory environments, and replicating telefactor systems); and
6.9 Self-replicating systems (including automata self-reference and self-reproduction methodologies, materials and parts and assembly closures, man-machine divisions of labor, and large complex hierarchical system control techniques).

Finally, four additions should be made to Opportunity Systems/Programs, Volume III of the 1980 Model. Under the categories "Resource Observation" and "Global Environment" there are no entries. The Intelligent Earth Sensing Information System (IESIS), developed by the Terrestrial Applications Team, may be inserted in either category as it fulfills the mission descriptions of both. A System/Program Summary may be assembled from information provided in chapter 2 of this report.

Under the category of "Planetary Missions" should be added the Autonomous Titan Survey Demonstration mission conceived by the Space Exploration Team as a precursor to interstellar-capable exploratory systems. A System/Program Summary of the proposed Titan demonstration mission may be generated from information provided in chapter 3.

Under the category of "Utilization of the Space Environment" two additions should be made. First is the SMF mission devised by the Nonterrestrial Utilization of Materials Team as a self-contained, evolving automated orbital manufacturing capability eventually using extraterrestrial material resources. Full details are provided in chapter 4, which may be used to assemble a System/Program Summary. Second is the Self-Replicating Growing Lunar Manufacturing Facility proposed by the Replicating Systems Concepts Team as a prototype for an autonomous general-purpose factory able to reproduce its own substance from arbitrary raw material substrates. A System/Program Summary may be prepared from information provided in chapter 5 of this report.

All four missions should be entered in Part B of volume III since they are opportunity programs unsupported at present by NASA program offices. As such support materializes they may be upgraded to Part A.

7.3 Conclusions and Recommended Technology Priorities

Many detailed conclusions and recommendations regarding technology needs and development requirements have been identified and discussed elsewhere in this report. An effort is made here briefly to highlight the major themes and milestone recommendations of the entire study activity having highest priority.

An evolutionary NASA space program scenario was developed by the study group, based on various relevant planning documents and other information. The major scenario premise was that coordinated developmental initiatives would be undertaken by NASA in the next 20 years to establish the basis for an aggressive, multidisciplinary program of space exploration and utilization early in the next century. Although the specifics of such a program can vary significantly, several generic characteristics were thought probable for any intensive space exploration and utilization effort. These could be used as meaningful guides for the mission problems selected by the study group to identify future automation technology requirements, and include:

- A major Earth resources observation program
- Intensive exploration of the Solar System and beyond
- Major low Earth orbit activities requiring the continuous presence of man as troubleshooter, supervisor, and operations coordinator
- A significant capability for acquiring and utilizing nonterrestrial materials for products to be used in space, such as large structures, power systems, antennas, expendables, etc.
- An advanced mobile communications system. (The importance of this program element was recognized by the study group but was not specifically addressed by any of the selected mission teams since the automation requirements were not considered unique.)

Advanced machine intelligence and automation technology as described in this report is believed to be essential in evolving toward a major space program capability for exploration and utilization within realistic resource limits. To this end, the following general conclusions and technology recommendations are worthy of special consideration:

(1) Machine intelligence systems with automatic hypothesis formation capability are necessary for autonomous examination of unknown environments. This capacity is highly desirable for efficient exploration of the Solar System and is essential for the ultimate investigation of other star systems.

(2) The development of efficient models of Earth phenomena and their incorporation into a world model based information system are required for a practical, user-oriented, Earth resource observation network.

(3) A permanent manned facility in low Earth orbit is an important element of a future space program. Planning for such a facility should provide for a significant automated space manufacturing capability.

(4) New, automated space materials processing techniques must be developed to provide long-term space manufacturing capability without major dependence on Earth resupply.

(5) Replication of complex space manufacturing facilities is a long-range need for ultimate large-scale space utilization. A program to develop and demonstrate major elements of this capability should be undertaken.

(6) General and special purpose teleoperator/robotic systems are required for a number of space manufacturing, assembly, inspection, and repair tasks.

(7) An aggressive NASA development commitment in computer science is fundamental to the acquisition of machine intelligence/automation expertise and technology required for the mission capabilities described earlier in this report. This should include a program for increasing the number of people trained in the relevant fields of computer science and artificial intelligence.

7.4 References

Miller, James G.: Living Systems. McGraw-Hill Book Co., New York, 1978. 1102 pp.

Odum, Howard T.: Environment, Power, and Society. Wiley-Interscience, New York, 1971.

Office of Aeronautics and Space Technology (OAST): NASA Space Systems Technology Model, Volumes 1-3, NASW-2937, NASA Headquarters, Washington, D.C., May 1980.

Vajk, J. Peter: Doomsday Has Been Cancelled. Peace Press, Culver City, California, 1978.

GLOSSARY

FOR MACHINE INTELLIGENCE AND AUTOMATION IN SPACE

Algorithm – A procedure for accomplishing a given result by proceeding on a logical step-by-step basis. Computer programs and N/C routines for machine tools are developed in this way.

Analog – Computers of this type are designed to respond and control continuous process operations such as flows, temperatures, or other infinitely variable-type operations. Digital computers process only discrete digital data.

Automatic – Functioning in a predefined manner with a minimum of reprogrammability; possesses only limited process information closure.

Autonomous – Functioning independently of other components or systems; self-governing or self-controlling; possessing virtually complete information closure in normal operation.

Axis – A general direction of relative motion between N/C machine cutting tool and the workpiece.

Bit – A binary digit of either 0 or 1; the smallest unit of information.

Bootstrap – A technique for loading the first few instructions of a computer program into active memory and then using them to bring in the rest of the routine.

Buffer Storage – A place for storing information in either a computer or a control unit so that it is immediately available for action once the previous instructions have been completed. Buffers eliminate the need to wait for information to be transferred from a slower bulk storage medium into active memory.

Byte – A series of computer binary digits organized to represent an alphanumeric symbol; sometimes called a "word" of memory; 4-, 8-, and 16-bit bytes are common in computing.

CAD – Computer-aided design; the use of computers to aid in product design and development

CAM – Computer-aided manufacturing; the use of computers to assist in any or all phases of manufacturing. N/C is one form of CAM.

Cartesian Coordinates – A system of two or three mutually perpendicular axes along which any point may be located in terms of distance and direction from any other point.

CAT – Computer-aided testing; the use of computers to aid in the testing of manufactured output.

Chip – Small piece of semiconductor material upon which electronic components and subassemblies are formed. Integrated circuits, LSI and VLSI are made on chips.

Closed-Loop System – A system whereby signals from a control unit are acted upon by the machine effector or teleoperator, and a monitoring unit then returns the acted upon signals for comparison; operates using feedback from errors, thus achieving some level of self-correction; opposite of open loop.

Closure – Exists when system function or output exceeds system structure and input requirements. Closure may involve quality, quantity, or throughput rate, and may apply to mass (parts, materials), energy (power, collectors), or information (assembly operations, repairs).

Cognition – Programmed models which approximate the behavior of natural cognition, in the context of robotic and artificial intelligence systems.

Compatibility – The degree to which tapes, languages, and programming can be interchanged among various computer and computer-controlled systems.

CPU – Central processing unit; the basic memory or logic center of a computer that includes the circuits controlling the processing and execution of instructions.

CRT – Cathode ray tube; an electronic vacuum tube containing a screen on which graphic or alphanumeric information may be displayed.

CS&T – Computer Science and Technology.

Dedicated Computer – A computer devoted exclusively to a single application.

Degrees of Freedom – The state of a mechanism can be described by specifying the current value of each variable parameter, particularly rotating or sliding elements, of robot systems.

Digital – Information and values are expressed in discrete terms. In a digital computer such terms are generated by a combination of binary on/off or positive/negative signals, the opposite of analog wherein a fluctuating signal strength determines the fluctuations of values.

Digitize – The process of converting a scaled, but non-mathematical, image into digital data.

Disc – A random-access storage component of a computer system.

DOD – Department of Defense.

DOC –Department of Commerce.

Dump – The removal of all or part of the contents of a computer storage medium such as memory or disc and its reproduction in some other medium such as hard copy printout, tapes, or cards.

Feedback – Information returned from the output of a machine or process intended for use as input in subsequent operations or for purposes of automatic control.

GEO – Geosynchronous Earth Orbit; also Geostationary Earth Orbit.

Hardwired – Computer or computer-controlled system which functions by means of fixed and committed circuitry; reprogramming is possible only by altering the nature of or interconnections among physical components.

Heuristic – A heuristic computer program is one which begins with only an approximate method of solving a problem within the context of some goal, and then uses feedback from the effects of the solution to improve its own performance.

ICAM – Integrated computer-aided manufacturing.

IESIS – Intelligent Earth-Sensing Information System.

Integrated Circuits (ICs) – A very small single structure assembly of electronic components containing many circuits and functions on a chip.

Interface – The medium by which two separate elements of a computer system are joined to permit mutual interaction.

I/O Device – Input or output equipment or programming, used to communicate with a computer or control system.

LBM – Laser beam machining.

LEO – Low-Earth Orbit.

LSI – Large-scale integration; the organization of many integrated circuits on a single, very small substrate; the basis of microcomputers and minicomputer logic systems.

Manipulator Systems – A generic term for any mechanical device which a robot uses to directly manipulate its environment.

Motive Systems – A generic term for the mechanisms used to convey a robot around its environment.

NASA – National Aeronautics and Space Administration.

N/C Machine Tool – Numerical controlled machine tool; a mill, lathe, or other production machine driven by computer-generated instructions for manufacturing that are read from a tape or other input medium, which the machine follows to complete a given task.

NEEDS – NASA End-to-End Data System.

NSF –National Science Foundation.

NTM – Nonterrestrial manufacturing.

Offline Operation – Peripheral equipment operating independently of a central computer, to conserve expensive online central computer time.

Open-Loop System – A system which generates output signals but which relies upon the integrity of the system to execute them, without feedback for monitoring or comparison purposes. Open-loop systems ignore error signals and operate on the assumption that no errors occur.

OTV – Orbital Transfer Vehicle.

PROM – Programmable read-only memory; can be programmed only by special routines. Once programmed with permanent data, it becomes a ROM or read-only memory.

RAM – Random access memory; access time is effectively independent of the data location.

R&D – Research and development.

Real Time – The ability of a computer to function and control processes as they actually occur.

ROM – Read-only memory; see PROM.

SETI – Search for Extraterrestrial Intelligence.

SMF – Space manufacturing facility.

SRS – Self-replicating system(s).

STS – Space transportation system; the Space Shuttle.

Teleoperator – A mechanical device for following, displaying, or amplifying the motions of the human body (hands, legs, head) to perform some useful task.

Tukey Ratio – In a nonterrestrial manufacturing facility, a measure of closure defined as the fraction of all terrestrial materials supplied per unit mass of output product.

World Model – An information structure built up in the memory of a computer or robot, based on both initialization and heuristic interaction with the environment.

LIST OF PARTICIPANTS

CO-DIRECTORS

Healy, Timothy J.
Department of Electrical Engineering and Computer Science
University of Santa Clara

Long, James E.
NASA-Jet Propulsion Laboratory

TERRESTRIAL APPLICATIONS

Team Leader

Sanford, David
Department of Computer Science
Rutgers University

Members

Fay, Temple H.
Mathematics Department
University of Southern Mississippi

Fischer, Norman
Battelle Memorial Institute

Nazemetz, John
School of Industrial Engineering
Oklahoma State University

Strong, James
NASA-Goddard Space Flight Center

Taneja, Vidya
Mathematics Department
Western Illinois University

Walsh, Peter
Department of Physics
Fairleigh Dickinson University

Woodruff, John
NASA-Goddard Space Flight Center

SPACE EXPLORATION

Team Leader

Krone, Robert
ISSM Room 201
University of Southern California

Members

Bernabe, Barbara
Institute for Cognitive Studies
Rutgers University

Brackman, Elizabeth
Department of Astronomy
University of Massachusetts

Gravander, Jerry
Humanities Department
Clarkson College of Technology.

Isaacson, Joel D.
Department of Mathematics, Statistics and Computer Science
Southern Illinois University

Seaman, Timothy
Department of Electrical Engineering and Computer Science
University of Santa Clara

Strong, James
NASA-Goddard Space Flight Center

Wells, William
Science Applications, Inc.
Huntsville, Alabama

NONTERRESTRIAL UTILIZATION OF MATERIALS

Team Leader

Criswell, David
California Space Institute
University of California, San Diego

Members

Dunning, Jeremy
Geology Department
Indiana University

Fronberg, Eric
Department of Electrical Engineering and Computer Science
University of Santa Clara

Green, Jack
Department of Geological Sciences
California State University, Long Beach

Hyson, Michael
Bioinformation Systems
California Institute of Technology

Luegenbiehl, Heinz
Division of Humanities, Social and Life Sciences
Rose-Hulman Institute of Technology

Ross, David
NASA-Jet Propulsion Laboratory

Zachary, William
School of Business Organization and Management
San Jose State University

REPLICATING SYSTEMS CONCEPTS

Team Leader

Laing, Richard
Computer and Communication Sciences
University of Michigan

Members

Cliff, Rodger
NASA-Goddard Space Flight Center

Freitas, Robert A., Jr.
Space Initiative/XRI
Santa Clara, California

Von Tiesenhausen, Georg
NASA-Marshall Space Flight Center

SPACE FACILITIES AND OPERATIONS

Team Leaders

Lovelace, Woody
NASA-Langley Research Center

Pennington, Jack
NASA-Langley Research Center

Members

Jones, James
NASA-Johnson Space Flight Center

Larsen, Ronald L.
NASA-Goddard Space Flight Center

Livingston, Lou
NASA-Johnson Space Flight Center

McCoy, Gene
NASA-Kennedy Space Center

Schappell, Roger
Space Division
Martin Marietta Corporation

Turner, James
NASA-Marshall Space Flight Center

TECHNOLOGY ASSESSMENT COORDINATORS

James E. Long, Chairman

Jerry Gravander
Ronald L. Larsen
John Nazemetz
Jack Pennington
David Sandford
Peter Walsh

NASA HEADQUARTERS SUPPORT

William Bradley
John R. Carruthers
Robert A. Frosch
William Gevarter
Jan Heuser
Lee Holcumb
Stanley R. Sadin
Tony Villasenor
Dell Williams

SRI INTERNATIONAL CONSULTANTS

Don Walker, Coordinator

Harry G. Barrow
John Gaschnig
Peter E. Hart
Gary Hendrix
Nils Nilsson
David Nitzan
Earl Sacerdoti

ADDITIONAL NASA/INDUSTRY SUPPORT

Gilbreath, William
NASA-Ames Research Center

Kincaid, William
Lockheed Missiles and Space Co.

Naumann, Robert
NASA-Marshall Space Flight Center

Painter, Earl
NASA-Goddard Space Flight Center

Pittman, Bruce
NASA-Ames Research Center

Renner, Fred
Battelle Memorial Institute

Robinson, Alfred C.
Battelle Columbus Laboratories

Ruoff, Carl
NASA-Jet Propulsion Laboratory

Steuer, Wolfgang
NASA-Jet Propulsion Laboratory

Wang, Taylor G.
NASA-Jet Propulsion Laboratory

SECRETARIAL STAFF

Yvette Duboscq
Noreen Golden
Carol LeClair

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CP-2255		

4. Title and Subtitle

ADVANCED AUTOMATION FOR SPACE MISSIONS

5. Report Date

November 1982

6. Performing Organization Code

7. Author(s)

Robert A. Freitas, Jr. and William P. Gilbreath, editors

8. Performing Organization Report No.

A-8618

9. Performing Organization Name and Address

NASA Ames Research Center
Moffett Field, CA 94035
and
University of Santa Clara, Santa Clara, CA

10. Work Unit No.

K-1577C

11. Contract or Grant No.

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
and
American Society for Engineering Education

13. Type of Report and Period Covered

Conference Publication

14. Sponsoring Agency Code

15. Supplementary Notes

Report by participants of the 1980 Summer Faculty Fellowship Program in Engineering.
Point of Contact: William P. Gilbreath, NASA Headquarters, Washington, DC 20546.

16. Abstract

This document is the final report of a study on the feasibility of using machine intelligence, including automation and robotics, in future space missions. The 10-week study was conducted during the summer of 1980 by 18 educators from universities throughout the United States who worked with 15 NASA program engineers. The specific study objectives were to identify and analyze several representative missions that would require extensive applications of machine intelligence, and then to identify technologies that must be developed to accomplish these types of missions.

This study was sponsor

This study was sponsored jointly by NASA, through the Office of Aeronautics and Space Technology and the Office of University Affairs, and by the American Society for Engineering Education as part of their continuing program of summer study faculty fellowships. Co-hosts for the study were the NASA Ames Research Center and the University of Santa Clara, where the study was carried out. Project co-directors were James E. Long of the Jet Propulsion Laboratory and Timothy J. Healy of the University of Santa Clara.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Robotics Artificial intelligence Space missions	Unclassified - Unlimited Subject Category 19

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	393	A17

For sale by the National Technical Information Service, Springfield, Virginia 22161

National Aeronautics and
Space Administration

Washington, D.C.
20546

Official Business
Penalty for Private Use, $300

SPECIAL FOURTH CLASS MAIL
BOOK

Postage and Fees Paid
National Aeronautics and
Space Administration
NASA-451

U.S.MAIL

[redacted]

NASA

POSTMASTER: If Undeliverable (Section 158
Postal Manual) Do Not Return