Final Report
on NASA Grant No. NAG-1-233

ON THE ENGINEERING OF CRUCIAL SOFTWARE

Submitted to:
National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23665
Attention: A. O. Lupton
Mail Stop 477

Submitted by:
Terrence W. Pratt
Professor

John C. Knight
Associate Professor

Samuel T. Gregory

Department of Applied Mathematics and Computer Science
RESEARCH LABORATORIES FOR THE ENGINEERING SCIENCES
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA

Report No. UVA/528208/AMCS83/102
February 1983
On the Engineering of Crucial Software

John C. Knight and Samuel T. Gregory

University of Virginia
Thornton Hall
Charlottesville, VA 22901

National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23665

This report discusses the issues involved in building software for crucial applications. These are applications in which failure could endanger expensive equipment or threaten human life.

The conventional software development cycle is discussed and various enhancements suggested based on recent results in software engineering. It is argued that the conventional software development cycle is inadequate for crucial applications even if enhanced.

An alternative approach is proposed in which human creativity is removed from software development as far as possible and replaced by computer based program synthesis. This technology is relatively immature but offers great potential for improving reliability of software. For those parts of systems which cannot presently use this approach because the technology is inadequate, fault tolerance is proposed as a supplement to the conventional software cycle.

This work was undertaken to provide suggestions to the sponsor about promising research areas. Specific research suggestions are made as well as suggestions for experiments using the AIRLAB facility. The report contains extensive bibliographies on various related topics to provide sources of further reading for those areas not covered in sufficient depth in the report.

software reliability, reliable software development, fault tolerance, automatic programming

Unclassified

Unclassified

Unclassified

For sale by the National Technical Information Service, Springfield, Virginia 22161
ABSTRACT

The most significant shortcoming of all software development processes lies in the fact that humans are involved.
# TABLE OF CONTENTS

Table of Contents ................................................................. ii

List of Figures ........................................................................ vi

1 Introduction .......................................................................... 1

2 THE SOFTWARE DEVELOPMENT CYCLE ................................. 3
  2.1 Requirements Specifications ............................................ 5
    2.1.1 State of the Art ......................................................... 5
    2.1.2 Contribution to Reliability ......................................... 6
    2.1.3 Activity Centers ...................................................... 8
  2.2 Design Methodology .......................................................... 11
    2.2.1 Contribution to Reliability ......................................... 11
    2.2.2 State of the Art ...................................................... 12
  2.3 Programming Languages .................................................. 16
    2.3.1 Introduction ............................................................ 16
    2.3.2 Modula-2 ............................................................... 16
    2.3.3 HAL/S .................................................................. 19
    2.3.4 Ada ...................................................................... 23
    2.3.5 Summary ............................................................... 27
  2.4 Testing .............................................................................. 30
    2.4.1 State of the Art ......................................................... 30
    2.4.2 Contribution to Reliability ......................................... 32
    2.4.3 Testing Techniques ................................................... 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Programming Environments</td>
<td>41</td>
</tr>
<tr>
<td>3 Enhancements To The Conventional Software Development Cycle</td>
<td>50</td>
</tr>
<tr>
<td>3.1 Overview</td>
<td>50</td>
</tr>
<tr>
<td>3.2 Software Prototypes</td>
<td>52</td>
</tr>
<tr>
<td>3.3 Software Components</td>
<td>55</td>
</tr>
<tr>
<td>3.4 Integrated Environments</td>
<td>58</td>
</tr>
<tr>
<td>3.5 An Improved Conventional Software Development Cycle</td>
<td>60</td>
</tr>
<tr>
<td>4 The Inadequacy Of The Software Development Cycle</td>
<td>65</td>
</tr>
<tr>
<td>5 Fault Tolerance</td>
<td>68</td>
</tr>
<tr>
<td>5.1 Recovery Blocks</td>
<td>69</td>
</tr>
<tr>
<td>5.2 N-Version Programming</td>
<td>71</td>
</tr>
<tr>
<td>5.3 Reliability Improvement</td>
<td>73</td>
</tr>
<tr>
<td>6 Verification</td>
<td>74</td>
</tr>
<tr>
<td>7 Automatic Programming</td>
<td>76</td>
</tr>
<tr>
<td>7.1 Introduction</td>
<td>76</td>
</tr>
<tr>
<td>7.2 The Issues in Automatic Programming</td>
<td>81</td>
</tr>
<tr>
<td>7.3 Automatic Programming Systems</td>
<td>84</td>
</tr>
<tr>
<td>7.4 Conclusions About Automatic Programming</td>
<td>87</td>
</tr>
<tr>
<td>8 A Comprehensive Approach</td>
<td>88</td>
</tr>
<tr>
<td>8.1 Overview</td>
<td>88</td>
</tr>
<tr>
<td>8.2 Requirements</td>
<td>91</td>
</tr>
<tr>
<td>8.3 The Monitor</td>
<td>92</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 2.1 ................................................................. 4
Figure 3.1 ................................................................. 61
Figure 8.1 ................................................................. 90
Figure 8.2 ................................................................. 99
ACKNOWLEDGEMENTS

UNIX is a trademark of Bell Laboratories,
Ada is a trademark of the U.S. DoD.
and NonStop is a trademark of Tandem Computers.

It is a pleasure to acknowledge financial support for this work received under NASA grant number NAG-1-233.
SECTION 1

Introduction

Crucial software is any software whose failure could endanger human lives or threaten the safety of expensive equipment. For example, the software in computers providing active controls for aircraft is crucial.

Software is defined to be reliable if it complies with its requirements specification most of the time. Conversely, software is said to have failed when it no longer complies with its requirements specification. We choose not to define 'most' because that leads to an attempt to quantify software reliability and the goals of this grant do not include probabilistic and statistical analysis of software failures. Rather, we assume that any increase in reliability is desirable and any methodology which may bring about an increase is worthy of consideration. We assume that the determination of whether an increase has been achieved is ascertained by experiments using conventional statistical methods.

The purpose of this grant was to examine and extend a preliminary approach to the engineering of crucial software which was presented in the original grant proposal. The goals were to prepare a comprehensive approach together with recommendations of those areas of software technology which are most likely to produce a substantial improvement in software quality if vigorously pursued. Our primary conclusion from extensive reviews of the literature and discussions with numerous
experts is that it is inappropriate at this time to propose a single comprehensive approach to crucial software development. Rather, we find several complementary technology areas which seem to offer the potential of major increase in software reliability yet which are not sufficiently mature that a clear choice can be made as to which is most appropriate.

This report is divided into ten sections. In Section 2, we examine the various aspects of the conventional software development cycle. This cycle was the basis of the augmented approach contained in the original grant proposal. We have formed the opinion that this cycle is inadequate for crucial software development, and the justification for this opinion is presented in Section 3. In Section 4 several possible enhancements to the conventional software cycle are discussed. Software fault tolerance is a possible enhancement of major importance and is discussed separately, in depth, in Section 5. Formal verification using mathematical proof is considered briefly in Section 6. Automatic programming is a radical alternative to the conventional cycle and is discussed in Section 7. Our recommendations for a comprehensive approach are presented in Section 8, and various experiments which could be conducted in AIRLAB are described in Section 9. Our conclusions are presented in Section 10. Finally, we present extended bibliographies on the topics covered in this report. They are intended to provide the reader with starting points for exploring further any of the subjects addressed in this report.
SECTION 2

THE SOFTWARE DEVELOPMENT CYCLE

In the short term, the only feasible way to construct crucial software is to use all of the best available tools and technologies, and to apply them in the classical software development cycle. Even then, they may not yield the required quality, but this determination is specific to the system and the people involved in its creation.

The software development cycle which we are discussing in this section is shown in Figure 2.1. It consists of only those steps typically used at the present time in the development of software systems. As such, it is a starting point for discussion and is simpler than the approach contained in the original proposal for this grant.

In our review of the present state of the art, we have formed certain conclusions which relate to elements of the classical software development cycle, each of which is discussed briefly in the following sections.
Conventional Software Development Cycle

Figure 2.1
2.1. Requirements Specifications

A requirements specification is a formally written statement of what a software system is supposed to do rather than (as in a design document or the actual code) how the system is to do it. Here, what is required of the system is explicitly written down and can be reviewed with the customer at the earliest stages to verify that the system to be built actually reflects what is wanted. The creation of such a document affords an early opportunity to review the consistency and completeness of the idea so problems can be corrected before their consequences proliferate. If the requirements specification is written in a formal requirements language, it is possible to perform some consistency checks automatically.

2.1.1. State of the Art

Many projects, such as the original development of the A-7E software [1], surge ahead into design without ever finding out what the software system is supposed to do. Others do attempt to organize the requirements specification in English prose, producing large documents in which it is easy to get lost, which are often incomplete or wrong (e.g. not specifying functions which the customer wants), and which are often never read nor kept up to date (this was the case with the original 2500 page BMD [2] requirements specification).

There is a great deal of current activity in the development of requirements languages and analyzers. Some of the older attempts are merely text organizers which are incapable of much more than cross
referencing usages of words in the document [3]. An apparently successful method [1] provides suggestions of how to design and use forms to be filled out about the project rather than a language per se. There are those who contend that requirements languages should be especially designed for restricted application areas, thus we find people working on requirements language generators [4]. Work has also been done towards developing programs which will automatically perform consistency and completeness checks on machine-readable requirements specifications [5].

It seems to be somewhat easier to write down the requirements for business systems and for purely mathematical software than for real-time systems. As a consequence, languages for those areas are much more advanced. In the area of real-time software, hardware interfaces and timing limitations must be specified, and priorities of goals must be stated in anticipation of necessary optimizations. Just how best to express a requirements specification continues to be an area of investigation.

2.1.2. Contribution to Reliability

By definition, reliability of a software system involves the degree of its fidelity to its requirements specification. The requirements specification should be written before further work on the system is started. The requirements specification can be used to verify with the customer that the developers understand exactly what is required, and then can be used as a reference by the developers in making all deci-
sions regarding the project. The continual reference to an explicit statement of what the product is to do cannot but help to ensure the product's fidelity to those requirements.

If at all possible, the document should be written in a requirements language. When requirements analyzers become available, this would allow automated completeness and consistency checks, this is especially important for changes during the post-delivery phase of the software system's life. Requirements languages are designed to avoid some of the problems of natural languages. Part of the power of the English language lies in its ambiguity and the extensive use of context to convey meanings. Ambiguity is inherently unsafe. For example, although not a requirements specification, the Ada Reference Manual [6] has for several years been a source of controversy over the meaning of the language it is supposed to describe. Further, in a natural language it is too easy to omit parts of the requirements specification, and the very structure of the language prevents explicit connection of a network of interrelationships.

A statement of requirements serves as a reference of what the software system is really supposed to do, thus it serves as a "contract" with the customer and with the eventual user, and can guide decisions during design and coding. This helps to prevent "guesses" by design analysts and programmers. It is far easier to detect and remove basic concept faults before design than after much of a software system has been designed (& coded!) to depend on them. A requirements specification which is organized by the use of a requirements language can be analyzed for such faults before a design exists to be infected by them.
A requirements specification helps in the creation of tests which will actually "verify" the software product since it is explicitly stated what the product is to do in every situation. Thus, each test can actually contribute to knowledge about the system's reliability, and none need be superfluous. If the software's response to an input is unspecified, whatever response it gives is as valid as another. Problems of cross-accusations of what should have been assumed by whom could be avoided if completeness checks are performed on a requirements specification. In this context we completely reject the notion of "robustness" in software [7]. Robust software is supposed to act "sensibly" when it receives unexpected input in the event that nothing was in the requirements specification. During the post-delivery phase of a system's life, the document continues to serve as a reference to guide proper or permitted revisions. Moreover, it is an excellent place to document the whys and wherefores of the changes, and the altered set of requirements can be checked for consistency and completeness as before. Actually, since the system would have been built around the requirements specification, any changes during this period should be due to changes in what is required of the system, which makes it appropriate to amend the requirements specification.

2.1.3. Activity Centers

For a survey of work in this area, see the May 1982 issue of IEEE Computer Magazine, in particular the chart by R.J.Lauber comparing 11 requirements languages and analysis systems on page 40 [8].
Parnas and Heninger, for the Naval Research Laboratory, Washington, D.C. while at UNC Chapel Hill, developed a requirements specification methodology which is applicable to flight software, since their project was to build a system duplicating the functionality and time and space efficiency of the A-7E aircraft operational flight program using modern software engineering techniques [1,9]. There had been no previous requirements statement for the A-7E and the document resulting from this project is being used by the "maintenance staff" for the original software. It is unclear how much of their success was due to the fact that they were writing requirements specifications for an existing system [1].

PDL [3] is a text organizing method with limited cross referencing capabilities and, although intended for design documents, has been used for requirements specification. A problem is that garbage text is perfectly acceptable to its processor.

PSL/PSA (Problem Statement Language/Analyzer) [5] is an older system which seems to have had some success as we find many projects have used it and there have been favorable comments about it in the literature (see the Bibliographies).

SREM (RSL/REVS) (Software Requirements Engineering Methodology) [10,11,12,13] is available from the Ballistic Missile Defense center in Huntsville. This system has actually been used in specifying the requirements of a large real-time project.

As can be seen, an explicit requirements specification is highly desirable in an effort to produce reliable software. However, the technologies of languages for its expression and analyzers of its
consistency and completeness are not yet well established. Further, there is nothing to assure that the document is actually used or kept up-to-date as the life of a project progresses.
2.2. Design Methodology

Software Design methods are largely disciplined ways of thinking through the problems the software is to solve. What the design stage is to accomplish is the translation of the "what" description of the requirements specification (most of the methodologies assume the existence of a requirements specification) into an overall plan for implementation—an overview of "how". This plan is to be written in what has come to be known as a design language; a specialized notation for accurately communicating what is to be accomplished to the individual programmers who will be implementing the system. Most of the work in discovering design methods occurred in the early to mid Seventies under the umbrella term "Software Engineering." Often, the process is seen as a continuum with only a vague distinction between "gross design" (which we are calling design) and "detailed design" (which we are calling implementation); in such cases, the design language is effectively the implementation language.

2.2.1. Contribution to Reliability

There are several motivations for preparing a design:

a) A thorough examination of the requirements specification for an implementation strategy affords the opportunity of ascertaining whether the project can be accomplished at all.

b) This same thorough coverage allows the design team to determine the most vital areas for allocation of implementing personnel. It also
allows the establishment of milestones for the development process.

c) A large system which is to perform a wide variety of functions needs a great deal of organization and planning. Creation of a design forces a disciplined approach to a problem and the resulting document serves as a guide at every stage of development. A document aimed at directing the implementors by limiting their scope of concerns can serve testers as well, indicating which areas of the software are intended to correspond to parts of the requirements specification.

d) A design document provides a mapping from the requirements specification to the coded software to limit the search for the modules affected by later revisions to the requirements.

e) This documentary evidence can be used early on as a check point for compliance with the requirements specification.

Unfortunately, the entire design process also provides more opportunities for faults to be introduced; hence, the attempts at devising analysis tools for design languages [5]. The factor which makes a design worthwhile is that faulty decisions may be detected as they are made rather than later when too much work depending upon them is at stake to do more than patch.

2.2.2. State of the Art

This section provides some warnings about those methods considered more likely than others to aid in creating effective designs. None of
them is a panacea. Indeed, it has been observed that most of these methods are those which have been unconsciously employed by the best programmers for years [14,15].

Most techniques are still at a stage in which they require lots of "magic" [16] and often are described in very vague terms by their devisers (see practically anything in the Design bibliography). Two people using the same method on the same problem (requirements specification) will rarely come up with the same design (this, the result of experiment [16]). Thus software design is still a game of skill, and quite prone to human error.

The Jackson methodology [17] views a program as a transformer of the structure of its input data to that of its output. Its area of application has traditionally been in business data processing; otherwise, it has not been applied in practice to large projects. Whether the complexity of resolving structural conflicts can remain manageable has not been determined. This is representative of the "data driven" design methods.

In Dijkstra's Programming Calculus [18], the Floyd/Hoare [19,20] axioms (augmented with later developments [21]) are used to formally derive a program from its requirements specification rather than to prove an existing program. This method is not necessarily a separate step from coding, and has been found difficult for the "average" programmer to understand. This method works, in the context of algorithms involving only integers and logicals, and is included within the basis for the recommendations below, but it can easily be mis-used through inattention to strict logical detail, a failing for which humans are
notorious. The Stepwise Refinement [14] strategy (also known informally as "structured programming", "structured design", "top-down programming", and "top-down design") is often incidentally employed to limit complexity.

The trend towards including data abstraction mechanisms in programming languages reveals a renewed respect for Parnas' Information Hiding [22]. This method has also been widely misinterpreted [23]. Other terms informally used concerning methods in this category are "functional decomposition", "modularization", and "object oriented design" [24].

The Data Flow design methods [25, 26, 27] attack a project by analyzing the necessary itinerary of various items of information through a network of transformations which gradually evolve the outputs from the inputs. The choice of division among nodes of the network, however, can often change application of this method into application of Functional Decomposition.

Iterative Enhancement [28, 29] is a simultaneous design and implementation method in which a small portion of a system's functionality is carried through to completion. This program is then given more functions piecemeal in the same manner as the original chunk. Occasionally, the original part may have been the prototype model.

SRI's Hierarchical Design Methodology [30] provides a set of tools and languages which together allow the consistent use of a combination of the above methods: top-down or hierarchical partitioning of the system (requirements specification, design, and implementation) into multi-level abstract machines, separation of the functions provided by
each machine at its level, and verification of the consistency of requirements specification, of design with requirements specification, and of the implementing code. The code-level formal proof is (or was, until recently) based upon the Boyer-Moore theorem prover [31].

The state of many design languages is evidenced by the fact that Basili's "system" is simply a means of rapidly changing the syntax of his "generic" design language [4].

A recommended overview of the more viable categories of methods, with examples, appears in [16].

A good design is vital to reliable software, but the technology for assuring production of or adherence to good designs is not there. We have not progressed far beyond explicit statement of what good programmers have always done unconsciously. The technology of design languages and analyzers is not very far advanced, nor is there any way of preventing their misuse. The apparently contained system of the HDM still only allows consistent use of design methods. There is little to require appropriate application of the system.
2.3. Programming Languages

2.3.1. Introduction

Crucial systems usually operate in real time. Modula-2, HAL/S, and Ada are high-level languages intended for real-time programming. In this section, we examine some of the facilities in each of these languages which have met with appreciation from real-time programmers and those which have been found unsatisfactory. This examination reviews the state of the art in programming languages.

2.3.2. Modula-2

Wirth claims that ordinary parallel languages contain all that is needed in a real-time language [32]. He proposes a discipline for their use in which a correct program is built first and then optimized to timing constraints. All time dependencies are confined to interrupt handlers and the program should not depend on any particular strategy for process scheduling. There has been some disagreement about this practiced ignorance of scheduling, and Wirth's second try at designing Modula, which produced Modula-2, forces the user to design his own scheduling algorithm. Confining all time dependencies to interrupt handlers cannot be done other than in programs which merely monitor devices. A program's computational processes can produce correct results, but if those results are not available for output when needed, the program is useless as a real-time program. Wirth also suggests avoiding many timing problems by adding more processors. This is fine if we have the money and space for the extra processors and necessary wiring.
However, the suggestion ignores the added problems and overhead of inter-processor communication. One suggestion that seems to get agreement from real-time programmers is that compilers should tell how long each statement and overhead operation will actually take.

Holden and Wand have programmed a loosely constrained real-time application (an operating system) in Modula [33]. They label as good the ability to give an absolute address to a variable at its declaration and complain about the difficulty of writing disk drivers without some generic parameter type. The latter problem is fixed in Modula-2 with the types WORD and ADDRESS which match almost anything. Wirth [34] claims a variable address declaration is extraneous with these "magic" types, but was included in Modula-2 at his colleagues' insistence. Holden and Wand point out that Modula's design calls for a uniform hardware I/O scheme of memory addressable "device registers" and may have problems on a different architecture such as ports with special I/O instructions. Modula-2 does assign static priorities to processes and procedures and these priorities are defined to be associated with those of interrupts in the hardware, but a dynamic priority effect can be achieved since procedures have the option of always being executed at their own declared priority rather than inheriting the priority of the process executing them. Thus, Modula-2 allows the user to determine whether his application will have priorities assigned statically or dynamically. Also, of these three languages, only Modula-2 defines what process priorities mean in relation to the environment they must deal with in real time.
Holden and Wand say that Modula's design limits its range of applications since all processes must cooperate in sharing the processor. In Modula (which Wirth [34] considers only a preliminary design for Modula-2 in the sense of Preliminary Ada and Ada), a process must consent to sending a signal and in Modula-2 a process must execute procedure TRANSFER before a process swap can take place. Complaints about lack of pre-emptability in Modula-2 seem suspect in light of the fact that pre-emption is generally achieved via interrupts as it is in Modula-2. These complaints seem to ignore the fact that Modula-2 is intended to be used in implementing facilities such as pre-emptive schedulers.

Modula [33] was a basis for the YELLOW candidate in DoD's search for a real-time language, i.e. it was a candidate design for Ada. It was found lacking in that it does not have a fixed point/floating point option, it does not provide for machine code inserts in the high-level language code, it has no exception handling capabilities, it has no facilities for specifying the machine representation of data objects, and cannot express, in one program, operation of a multiprocessor system.

Certain facts tend to cast doubts on the inherent efficiency of Modula-2 [34]. Wirth designed his Lilith machine especially for the language. Lilith is microcoded so that the instruction set is the Modula-2 specific M-code. Also, his most time-critical device, the high-resolution display, has its own bus to memory and that bus has four times the bandwidth of the CPU's bus.
Other obvious concerns with Modula-2 as a language for crucial systems are the relatively low level of typing in the language and the lack of a systematic approach to constraints. Much of this was corrected in Yellow, but that language was abandoned. Modula-2 is certainly better than assembly languages but was not designed for, and does not aid the development of crucial systems.

2.3.3. HAL/S

HAL/S [35] was adopted as a NASA standard flight language when an implementation was demonstrated to have a ten to fifteen percent inefficiency in size and speed over assembly language. We point out that this is a ridiculous metric. Efficiency is program dependent and compiler dependent. The most important issue is reliability and that is ignored.

The language itself puts the periodicity of process scheduling, control via wall clock time, events (hardware interrupts), and error conditions under explicit programmer control. These things are achieved via a large run-time library support system, and the HALMAT intermediate language operators for many of these facilities are mnemonic for IBM OS/360 supervisor calls. In contrast to Modula-2, HAL/S does not provide basic, low-level, facilities for tailoring an entire system to an application but tries to assume the class of real-time programs known as flight software and to provide a full underpinning for the user to build on. Where the user needs access to the hardware, the language provides the SUBBIT operator for bit manipulation and an implementation provides $MACRO's rather than allow assembly code insertion. This allows com-
piler checks on usage while providing high-level access to machine idiosyncrasies.

From the literature, HAL/S does not seem well known outside of INTERMETRICS and NASA. A brief description of some of its real-time related constructs follows. The words in upper case are keywords of the HAL/S language.

Outside of implementation-specific %MACRO's, there is no absolute addressing. Data storage may be AUTOMATIC (allocated only as long as a procedure is activated), STATIC (allocated as long as the program executes), or TEMPORARY (allocated only while a few statements execute). Data may be DENSE (packed), ALIGNED on unspecified "appropriate" hardware boundaries, or RIGID (laid out in memory exactly as described in the declaration). ACCESS rights may be associated with data objects and they may be grouped into LOCK groups for mutually exclusive access through UPDATE blocks by tasks. Events are boolean-like variables which may be LATCHED or not (able to hold a true value for more than an instant or not). All communication among tasks is through shared variables. Separately compiled entities access data via a FORTRAN COMMON-like facility known as COMPOOL's. Procedures and functions may be expanded INLINE or may be specified to be REENTRANT or not. A degree of optimization for common flight software applications is achieved by virtue of the special VECTOR and MATRIX operators and data types. A task may be stopped by another task by two methods: CANCEL allows the current instance of the task to continue to completion but prohibits any scheduled future instances of it, whereas TERMINATE destroys the current instance as well. A task may WAIT UNTIL a certain wall clock time, WAIT
for a certain length of time, WAIT FOR a combination of events to become true, or WAIT FOR DEPENDENT's to terminate. A hardware interrupt or a task may SET an event variable true or RESET it false or may make it true momentarily via SIGNAL. When an event variable changes values, every event expression which has been reached by any task must be fully re-evaluated to determine if the task is eligible to proceed. Error conditions within a class or entire classes of errors may be raised (SEND ERROR...), and the set of error handlers may be dynamically changed by declaring and removing them (ON ERROR... statement; and OFF ERROR...). As a special case, errors may be ignored or passed to the support system with an optional change to an event variable (ON ERROR... SYSTEM... or ON ERROR... IGNORE...).

The most attractive statement in HAL/S for the real-time programmer is the SCHEDULE statement. A task may be scheduled to begin execution AT a certain time, within (IN) a certain time interval of the current time, or ON the occurrence of true evaluation of an event expression. It is required to be started with a priority, and may be made DEPENDENT on the continued existence of the task executing the SCHEDULE statement. Execution of the task may be made to begin anew EVERY so often or a certain amount of time AFTER it completes. Such repetition may continue WHILE an event expression holds true or UNTIL an event expression becomes true or UNTIL a certain time. All this may be specified in a single SCHEDULE statement, and once started, a task's priority may be changed by the UPDATE PRIORITY statement.

On the surface, HAL/S seems to provide everything a real-time programmer could want; particularly if a compiler could guarantee the
scheduling requested in each SCHEDULE statement. Garman [36], however, describes several problems with HAL/S in the Space Shuttle project.

On occasion the project was forced to take risks by changing shared variables outside of UPDATE blocks. This casts some doubt on the utility of any language which prohibits shared variables or their unprotected update. Either the implementation (one of three [37]) of HAL/S used by the project did not support or the project did not use the following features: DEPENDENT, REPEAT AFTER, TERMINATE, WAIT UNTIL, and ON ERROR. UPDATE PRIORITY was rarely used, which implies that the need to change a process' priority is rare in real-time programs but probably vital when it does arise. Also, the implementation imposed severe limits on the complexity of event expressions that could be used. This last rule was probably imposed to cut down on overhead since all event expressions must be re-evaluated on any event change.

The original coding of the Shuttle software [36] turned out to be plagued with throughput problems. For example, the I/O via READ/WRITE statements or MACRO's was too expensive. The project called for the various machines to synchronize at most support routine calls. And there were too many processes, resulting in scheduler queue overflows. It was also apparent that, even with a SCHEDULE statement, timing constraint calculations had to be made by hand or with the aid of FSIM, a functional simulation tool. The solution chosen was to break up certain tasks into procedures and change the support executive to call these procedures in an order determined by table-lookup, a technique employed in many assembly language real-time programs [38].
The generalized scheduling constructs of HAL/S, a language designed for flight software, were found to be too inefficient in practice and some parts were not implementable. Tripathi, Young, Good, and Brown, in describing a verifiable subset of HAL/S and before completion of a verifyability study of Ada, concluded that a project should choose Ada over HAL/S, noting that Ada has all the capability of HAL/S and more [39].

Apart from the functional criticisms of HAL/S, there are major deficiencies relating to reliability. The language offers relatively poor typing (no programmer defined types, for example). The process communication mechanism, which relies on shared variables, is archaic and very error prone. It is not amenable to automatic checking for deadlock and similar difficulties. The control structures and expression structures of the language are also very poor. They are oriented more towards ease of programming than reliable programming.

2.3.4. Ada

Ada was chosen as meeting the DoD's specified requirements for a real-time language. It, like Modula, has gone through at least one redesign after public comment. These comments came in a wide variety. Some were objections to necessary features on purely aesthetic grounds, e.g. the ELSE within a SELECT statement was found "nasty", although it is needed for proceeding in the face of communications breakdowns or time-critical processing [40]. Some were specific suggestions about preliminary Ada which were included in "final" Ada, while others were disagreements other about whether it was easy to program a favorite
solution to some pet problem. We use the word "favorite" since it is often not the case that "you can't do X in language Y" but instead "you can't do X in language Y by method Z" [41].

Boute [42], in a study of preliminary Ada on representative communications control problems found it "very satisfactory", noting that the complexity and structure of the solutions matched that of the problem statement. On the other hand, Roberts, Evans, Morgan and Clarke [43], also looking at communications control and claiming experience in that area, say that the rendezvous mechanism is overly general and a potential time waster for message passing within or among processors. Specifically, a message that does not even need acknowledgement cannot be sent without at least four scheduling operations and that the sender is tied down until the receiver is finished reading the message. They state that Ada's philosophy is wrong for this application in that data rather than processes should be queued.

Mahjoub [44], also in the area of distributed processing, is more concerned with the asymmetry of the rendezvous. A task cannot know the sender of a message and messages cannot be broadcast. The concern with the asymmetric rendezvous seems to be a common one in resource allocation and scheduling [43, 45], although there is a solution to this problem, involving creation of a resource task. An early problem [46] with scheduling was fixed in "final" Ada with task types so that manipulable structures of processes could be created. But problems with scheduling persist. Haridi, Bauner, and Svensson [47] and Mahjoub [44] favor static assignment of priorities by the user but, as we have noted, there are applications in which dynamic priorities are necessary. People
examining preliminary Ada [43] (before introduction of families of entries) found the rigid FIFO queue organization prevented urgent requests and tended to flatten different priorities to one level. Mahjoub [44] says that real-time programmers need to be able to write their own schedulers since different algorithms will be optimal for different applications. Roberts et al [43] agree and declare that, to build a scheduler in Ada, one is building one scheduler on top of another, thus multiplying the overhead in what, in practice, is already a tight situation. Different applications have different ranges of speed requirements, some of the more highly constrained of which need radically different organizations. They conclude that Ada offers the wrong level of granularity of parallelism.

The method of inclusion of interrupt handling in Ada met with mixed response. Bennett, Kornman, and Wilson [48] and Haridi, Bauner, and Svensson [47] were in favor of it, but Mahjoub [44] was concerned with response time in that the handler task might not be scheduled right away or worse, might take a very long time to reach an accept for that entry.

The semantics of several Ada statements could result in bad states in a distributed system [44]. Between initiation and termination of an ABORT statement, a task might be able to communicate with another which, by virtue of being on another machine, has not been destroyed yet. Alternatively, a centralized knowledge base of what is alive and what isn't which had to be interrogated at every call would present a bottleneck which could easily bring a system down. The semantics have been revised in ANSI standard Ada to alleviate such situations [49]. Other potential overhead problems for real-time systems involve the
"Page missing from available version"
implementation level, the machine code insert capability was found useful [48] but dangerous if used unnecessarily.

There have been a few experiments and analyses of the potential efficiency of Ada implementations. Haridi, Baumer and Svensson [47] created a model intermediate language for Ada and ran (it may have been interpreted) programs hand-translated into it against a real-time version of C. The results of this experiment were deemed favorable for Ada's efficient implementation. Eventoff, Harvey, and Price [52] did an analysis of a generalized monitor based language vs. Ada's rendezvous on multiprocessor shared memory systems. They concluded that each approach was better suited for its own set of classes of applications. The monitor approach imposed less overhead for problems involving asynchronous communications and buffered synchronous communications while the rendezvous was better for problems requiring direct synchronization and problems which exhibited any degree of contention.

2.3.5. **Summary**

Programming languages have received a great deal of attention over the last thirty years and yet new ones continue to be designed. The reason is that no programming language yet devised is perfect. The design of languages is not a suitable problem for the short term, but the proper choice of an existing language to use is. There are many languages that are suitable for describing crucial software. Ada, HAL/S, and Modula-2 are examples.
The difficulties lie in finding a language:

a) which is of modern design,
b) which received sufficient care and analysis during its design,
c) which has a precise, formal definition,
d) for which compilers exist for the machines of interest,
e) for which validation of compilers and run-time support systems (within the current state of the art) is available,
f) and for which rigid configuration control of the language exists.

In the short term, these apparently minor issues are the really important issues. Differing opinions on what a language construct means, or subtle faults in compilers are major causes of faults in programs, but which have nothing to do with the programming language itself.

In practice, the only programming language which has faced all these issues and attempted to solve all of them is Ada. In addition, Ada is the only widely known and soon to be widely available language to include facilities for data abstraction. These facilities make the more modern design methodologies (such as Information Hiding, the Jackson method, and the Yourdon and Constantine system) far easier to use, and far easier for their use to be enforced. We conclude that Ada is the only choice of programming language for constructing crucial systems in the short term, and that language design is such a massive project that it is inappropriate for NASA to consider it. However, there are inadequacies in Ada and in the description of Ada. Short term investigations of the use of Ada and into its formal definition are appropriate in support of crucial software development.
Although we prefer Ada to the other extant candidates for programming languages for crucial real-time software, we still bemoan the fact that Ada was not designed with that purpose unwaveringly in mind. Ada, despite the original goals, was designed to do "everything for everybody". Hence, there are many aspects of the language which are not verifiable. Ada provides facilities which the community has deemed necessary to the creation of reliable software, but practices which lead to unreliable software cannot be prevented in any language with current technology without removal of features which are truly necessary.
2.4. Testing

Programmers have been running their programs against sample inputs to see if they "work" since the first fault was ever found in a program, yet no one has managed to move testing out of the world of ad hoc methods. The situation seems best described by the following quote:

We know less about the theory of testing, which we do often, than about the theory of program proving, which we do seldom [53].

As long as humans are involved in the transformation of specifications of ideas into programs, we cannot be sure that no faults have been introduced without testing the resulting programs. The problem lies in choosing the set of tests which will uncover any faults in a given program. There are kinds of faults which we know about and can categorize, but there are also faults of a very much more subtle nature which are heavily involved with the semantics of the individual program and which we do not know any general way of detecting.

2.4.1. State of the Art

Despite some attempts [53,54], no one has yet completed a formal theory upon which to base the activities we call testing. Many of the proposed methodologies appear to be attempts to systemize the ad hoc methods of experienced program testers and to find systematic means of detecting types of faults which it is known that programmers commonly introduce. This may be in the hope that some formalism will fall out of such efforts and that an organized approach will help avoid wasted test-
ing effort in the mean time. Some directions concentrate on categories of faults while others tend to concentrate on the input spaces of the programs under test. One thing which must be remembered about testing real-time software is that one of the dimensions of the input space is time in that the behavior of the program usually changes over time for the same inputs. This complicates any testing strategy since the potential exists for, say, a program which reads two input values to require an infinite number of tests with different spacing of the inputs in time. Just when is enough enough? Statistically based reliability estimation and, of course, the exhaustive testing method, however, seem to be the major offers of a strategy for telling when to stop testing a given program [55,56,57]. Yet, there is a great deal of controversy within the reliability estimation camp about which basic theory of statistics applies, and exhaustive testing for real-time programs can be impractical.

There are known types of faults which seem to evade these efforts:

With most testing methods, missing path errors are only detected by mere chance. In fact, missing path errors cannot be found systematically unless a requirements specification is available. A correct requirements specification would describe all the cases that should be handled by the program [58].

The allusion in the above to the unavailability of requirements specifications brings up a point of difficulty in testing. Due to the fact that in practice a program often reaches the testing stage without anyone having bothered to create a requirements specification, testers often have nothing but their own intuition to use in determining whether a program run against a test case has passed or not.
Another difficulty in testing is that programmers often try to "cover" themselves by including redundant conditionals in their programs. This fact often makes it difficult for a tester to determine whether a section of code is mistakenly unreachable or whether the conditions being evaluated are simply impossible. Further, it seems to be as difficult to create tests which create exceptional situations which the software is supposed to recognize as it is to create test cases which are intended to "stump" the software.

2.4.2. Contribution to Reliability

Without a formal theory, testing will only do two things for us:

a) It will assure us that, for the statistically meaningless set of inputs which we have tried, a program or system of programs "works."

b) It will give an unjustified increase to our subjective feelings of "confidence" in our software systems.

With a formal theory of testing, a set of tests performed in line with the theory would give a level of assurance of the program's correctness comparable to that given by a formal proof of correctness (without human mistakes in the proof). Short of a formal theory of testing, exhaustive testing (when possible) is by definition a proof of a program. Without a formal theory and with no possibility of exhaustive testing, the activities now pursued give a wholly unjustified confidence in programs.
2.4.3. Testing Techniques

In this section we give a list and brief explanations of the testing techniques which have been proposed in the literature.

a) Execute Every Line

Since it is impossible to have tested everything that a program does without trying each statement in it, it at first seems reasonable to create a set of test cases which together cause the execution of each statement in the program. This does generate a good number of test cases but it does not follow that executing each statement in a program exercises all of the program's functions.

b) Branch Testing

One of the ways functionality can be missed by simply executing every line in a program is for the program to contain a simple conditional branch around a statement, call it 'S'. The strategy of executing each statement would generate a test case which caused evaluation of the conditional to allow the statement 'S' to be executed, but would not generate the test case which took the other side of the branch. Branch testing is designed to make certain that each statement is executed and both possibilities are tried for each conditional branch in the program.

An example of a fault which Branch testing can miss is as follows. Suppose a statement being guarded by a conditional branch is supposed to be performed only under condition 'A' yet the program as
coded mistakenly allows the statement to be performed under either of conditions 'A' or 'B'. A test of both sides of the branch might be created containing two test cases, one in which 'A' was false, and one in which 'A' was true. If 'B' happened to be false in both test cases, we have a situation in which, although both sides of the branch would be exercised, the fault in the conditional expression would not be detected.

c) Path Testing

The idea here is to execute each possible path in the code as a method of checking the program's functionality. Executing each path is different from taking both sides of each branch. For example, if the code contains a loop for which it is possible to execute the loop 0, 1, 2, or 3 times based on particular input values, that loop contains 4 paths and thus requires 4 test cases. Should that loop be nested within a similar loop, the number of test cases required to test all paths in the loops is multiplied. Path testing cannot consistently detect paths which the requirements specification (if it exists) calls for but are missing in the coded program. For programs of a practical size, the number of possible paths approaches the size of the input space, so, to keep from testing forever, limits need to be made on loop executions or a closed form for loops needs to be proven to make this method practical.
d) **Structural Testing** (Also called White Box Testing)

The internal structure of the program as coded is used as a basis for choosing test cases. In using such test cases, the entire functionality of the program as coded is supposed to be revealed and that is to be compared with the specified requirements.

Several other methods fall under this category. At one level the structure of a program is given by the conditional branches and call structure. However, one can also see a program's structure in other components. Geller [51] attempts to formalize structural testing.

e) **Functional Testing** (Also called Black Box Testing)

Functional testing attempts to test against the requirements specification for functionality. If the requirements specification states that the program should function in a certain manner when confronted with a category of inputs, it is tested with instances from that category. Test cases are chosen as if nothing other than the required behavior were known about the coded program being tested. It has been noted that this method cannot catch all faults since the method does not know anything about the coded program's internal structure i.e. the program may check out perfectly well but may behave properly only for the inputs used in the test and branch off into code which does something else entirely for other inputs.
f) **Exhaustive Testing**

All possible inputs are tested. One might think this impractical if possible, and improbable if (as in most cases) there is a large input space, but on current computers even the representable number of 'real' numbers is finite. With VLSI technology, it may become reasonable to create a large array of chips to generate test cases and run tests to exhaustion of the input space for a program. For truly crucial software, the cost of creating and running the VLSI chip array for years if it takes that long may be justified if a formal theory of testing is not found which can definitively give a more limited set of test cases for each program. Note that an exhaustive test of a program is by definition a proof of the program. All of the other test methods are capable of missing serious faults while the only problem with exhaustive testing is the large number of cases which must be run.

g) **Error Seeding** (Also called Mark-Recapture Testing)

In the error seeding strategy, a predetermined number of known faults are deliberately introduced into a program and arbitrary test cases are applied to the program (preferably by someone who does not know how many and what faults were seeded). At any point during testing, the percentage of seeded faults found is supposed to approximate the proportion of the naturally occurring faults which have been found so far by those tests. There is no reason to believe that the number of seeded faults is anywhere near the number of natural faults in any given program, nor that they occur
with a similar distribution. Also, the seeded faults would be manufactured by humans and as such would reflect the kinds of faults humans expect themselves to introduce. This biases the distribution of the seeded faults toward the first few kinds of tests the testing staff would try anyway. All of the seeded faults would be found quickly whereas the truly subtle and difficult faults would remain hidden.

h) **Statistical Testing**

Test cases are chosen via statistical sampling of the input space. Because real-time programs usually deal with the physical world, statistical testing is not likely to generate a realistic set of tests. Changes in the real world are smooth and gradual whereas a random sample from the input space is likely to vary widely.

i) **Error-Based Testing**

Experience with programming computers tells us that there are certain kinds of faults which we, as humans, commonly introduce. Error-Based Testing is an approach to testing in which test cases are designed especially to detect these kinds of faults. Unfortunately, we do not have a complete list of faults which humans can introduce, so no such set of tests is likely to detect all faults in a given program. Subtle faults are difficult to classify and more difficult to ferret out with classification-oriented testing strategies. This represents the brute-force approach to learning from experience.
j) **Mutation Testing**

Mutation testing derives from error-based testing, but as a methodology, seems to contribute more indirectly through evaluating the effectiveness of the test set than directly through testing the program. The required program and the coded program are thought of as being instances within a "cloud" of similar programs each of which differs from the others only slightly. The idea is to repeatedly transform the coded program $P$ into similar programs $P'$ by changing small parts of $P$. The set of test cases is run through $P'$ to see if the test set is complete in its ability to distinguish between $P$ and $P'$. If not, the tester must find a test which will distinguish outputs from the two. Each mutation transform is said to correspond to a class of faults. Among advocates of mutation testing, there seems to be a consensus that no more than a one "change" difference between $P$ and $P'$ is necessary to test the test set's effectiveness i.e. each $P'$ is created via one small alteration to $P$. This method seems to call for combinatorically many more "runs" of tests than the size of the program being tested. It is difficult to tell how this process is supposed to determine whether the coded program $P$ is the required program. For example, mutation testing cannot detect errors of omission where some part of the requirements specification is not satisfied.

k) **Partition Testing**

Goodenough and Gerhart [53] explain this and offer basic definitions and theorems which seem to be acknowledged as a good basis
for a formal theory. Briefly, the requirements specification is analyzed to determine a set of equivalence classes (partition) on tuples of the input space. Running as a test case one tuple from any equivalence class of the partition is completely equivalent to running as test cases all tuples in that equivalence class. Thus, exhaustive testing can be achieved by running one test case based upon one tuple from each equivalence class of the input space.

This seems like an ideal test method since it limits the total number of tests needed and is equivalent to exhaustive testing. The problem with this method lies in determining the equivalence classes. For realistic programs, this is not a solved problem.

1) Domain Testing

This is a refinement of Path Testing in conjunction with partition testing. Tests are devised to make sure that the set (domain) of inputs driving each path is correct, i.e. that the partition of the input space defined by the requirements specification and the partition of the input space effected by the coded program are one and the same. Some of the limiting factors are that this method with current technology cannot handle other than simple conditionals and that it cannot detect mutually canceling faults. There seems to be some merit in this approach as a lead-in to a testing formalism [58].
m) **Boundary Value Testing**

Test cases are created to exercise each conditional in the program as close as possible to the point where it changes between True and False. This is a limited form of component testing where the components are conditional expressions controlling branches.

n) **Range Testing (Also called Stress Testing)**

Same as Boundary Value Testing except the extrema of the ranges of values of each variable and input are exercised as well.

o) **Component (Unit) and Integration Testing**

Each component of the system is tested as a separate unit using whatever method is preferred, and test the combination of components (the entire system) for functionality as an assemblage of known-to-be-correct parts. Some people seem to have the idea that this can be done recursively.

Psychologically, we need to test our software before entrusting the safety of ourselves or our equipment to it. Practically, we find that the methods we use in testing are inadequate to the task. The hopes for formal theories upon which to base testing strategies worthy of our trust have not yet come to fruition, and may well never do so.
2.5. Programming Environments

A Programming Environment is the group of tools employed by humans to develop and, later, revise software. Much of the paperwork involved in such things as version control, data dictionaries, and management reports on programming projects can be considered drudgery and as distractions from the task at hand (building software). It seems reasonable to try to migrate that work onto computers as we have migrated much of the bookkeeping of programs to compilers of high-level languages. Although reasonable, this has seldom been done.

What work in the area has been done in the past seems to be disorganized and skewed toward the initial coding section of the software life-cycle. The reasons for this seem to be summarized by the following observations:

The financial structure of many software producers is that production costs are a liability but maintenance costs are an asset or income ... In academic environments, using a portion of another person's code is often considered cheating. No credit is given for producing reusable software [59].

A few attempts, notably the Programmer's Workbench and the National Software Works, have been initiated to collect and implement on computers some of the tools which designers and programmers typically use [60,61,62]. More recently, with skyrocketing software costs (both in development and later revisions) and increasing complexity of systems, the DoD has become concerned about both automating the tools and integrating them. The DoD has commissioned the construction of integrated Ada Programming Support Environments (APSE's) [63,64] and the NBS [65], in the Spring of 1981, studied what can be done with today's
technology for medium and large software projects and what directions funding of research should take in the next five to fifteen years in the area of integrated environments for the entire software life-cycle.

We have had limited language specific computerized environments for years. Interactive BASIC, APL, and LISP systems have usually had their own file systems and editors, have been able to detect and notify the programmer about syntax and context-sensitive syntax errors as programs are entered. They have run-time systems capable of indicating errors in terms of source lines or statements. Some are capable of backing up, allowing source changes at execution time, and otherwise suspending execution while the programmer does other things, and have uniform and omnipresent sets of commands so that a programmer, for instance, does not have to "leave" the editor in order to "get" another file. A particular system, INTERLISP, has included many of the other capabilities and features to be described below [66,67]. Two of the primary qualities these systems have in common, and which are seen to be the enabling qualities of other planned environments, are that programmers deal with the systems interactively and that the tools in the systems "know" about each other and about the programming language.

Working from this starting point, much of the effort which has been put into programming environment research has gone into "smart" editors and source level debuggers [68].

Noting that we have been using text editors or other context-ignoring systems (e.g. CDC's UPDATE) to enter and alter programs in computers, and noting the success interpretive interactive systems have had in detecting errors as they are entered and the fact that the trend in
language translators has been toward syntax-directed translation, it becomes reasonable to consider syntax-directed, language-specific editors for entering program text. The use of such editors could eliminate compilations which are used only to detect and remove syntax errors. Since we are dealing with high-level languages, it is silly not to debug in terms of high-level language statements. Thus the idea of syntax-directed interactive editing is extended to source-level debugging in which one is able to interpretively "run" partial programs using such techniques as "stepping" through statements, substituting values, backing up, forcing branches, and making source-code patches while debugging.

As currently implemented [69,70], many such editor-debugger systems do not actually deal with a source code "file" but immediately internalize the input characters so that they use a data structure directly analogous to the syntax. In combination with CRT's they automatically "prettyprint" the source display as it is entered and might flag erroneous text in "reverse video" characters. Some even use color [71]. What does this buy in terms of reliability of life-critical software? We save syntax error debugging runs, and individual programmers on large projects can try out decisions in early stages without waiting for later testing stages when such decisions and any possible alternatives may have been forgotten.

The Cornell Program Synthesizer [70] allows a programmer to "hide" sections of code to abbreviate the source so all of the currently interesting parts can be displayed on a CRT at once. It also moves the CRT's cursor around from statement to statement on the screen as its
debugger "executes" them and presents a running display of variable-value pairs on part of the screen; the speed and direction of such "execution" can be controlled by the programmer and suspended or altered at any time. Statements are entered by selecting and filling in templates and errors are tolerated but flagged until corrected.

Work on improving the hardware of programming environments is being performed [72]. Noting that humans usually refer to several documents and several areas of a source program at once, this research is concentrating on how to partition screens and provide for multiple screens and still provide portable software and coherent, easily learned controlling commands as part of the command language of the environment.

Of course, the computerized tools already in use would not necessarily be abandoned. Since the editor would parse and internalize programs, a complete compiler is not needed. Rather we would need code improvers, code generators, and simulators for host and non-host target machines. The internalized form from the editor could also be fed into a static analyzer.

One ingredient considered essential to an integrated environment is a uniform command language. The language should be well human engineered with extensive help facilities which could, in advanced systems, even be anticipatory. The UNIX approach of keeping manuals online is seen as a large step in the right direction but has the failing that one must know (the name of) what one is looking for in order to find it. Although experimental systems are geared (consciously or otherwise) to be used by experts (their creators), actual environments must be able to serve novices with equal ease. The command language should
also be omnipresent: within reason, any command should be valid at any time. If the programmer is in the middle of using a tool when he remembers he needs to start up another, he should be able to call upon that other tool without abandoning the current tool, and the command language interpreter should be able to figure out the object and change of viewpoint without being explicitly told.

There has been considerable discussion on the degree of granularity or tool size and the amount of integration desired in an environment. Experiments with programming environments have ranged from the monolithic (a single gigantic program) such as INTERLISP [66], to the tool box approach provided by UNIX [73,74]. The monolith is seen as being less flexible and as hindering creation and inclusion of new tools provided by programmer-users. The tool box can be a jumble of bits and pieces so that a programmer must expend great effort just in picking out and properly composing the tools needed to perform even a simple operation. The trend seems to be toward small tools which can be composed, but for the environment to figure out which ones are applicable and how to compose them (i.e. for the tools to compose themselves), and for the environment to be easily told about and include new tools. There is also a strong trend toward having many tools running continuously as independent processes unseen by the programmer.

File systems and systems for keeping up with what is in each file play a major role in large projects. Does a given file contain a requirements specification, design specifications, source text, compiled binaries, executable code, implementor documentation, or test data for a given module? Where are all of the source modules for a given project
as they existed six weeks ago? Where is all material relevant to a certain paragraph of the system design document? Once a project gets to the stage of needing versions of modules, especially for revisions long after the original teams are gone, the picture gets even more complicated.

Most, if not all, environments involve a well-coordinated database. The database should manage objects (files), remembering properties about each one and managing relationships between objects with like properties and between objects whose properties exhibit dependencies. For security and information hiding in large complex systems, it should also provide and use access controls on its objects. All tools can be seen as creating new objects with properties relating to pre-existing objects. It is suggested that, in concert with the command language help facilities, the database could also serve as a kind of “Ann Landers” to field programmers’ questions about policies, relationships among objects and groups of objects, and even “how to” questions to prevent people from constantly having to “re-invent the wheel”, all the while avoiding violations of information hiding and security rules.

In heavily integrated systems, tools might monitor other tools’ transactions with the database and initiate still other tools automatically when changes occur in objects which are related to other objects by dependency relationships. For example, a change and recompilation of an Ada package could trigger automatic recompilation of units which use it. Such tools might also insist that the original change be related to some report or test failure or try to aid the system documentation by obtaining some other sort of verification stamp.
Other tools (such as the Programmer's Assistant [75]) might be able to undo a programmer's mistakes. This has implications for a database since a mere trace of a programmer's transactions is insufficient: the database must remember everything, all versions of all objects that ever existed for a project and their properties and relationships. A line of research arises here into how to compact this tremendous amount of information. One proposal is that, rather than keeping redundant information, the environment should keep a history of all objects and re-generate individual objects when needed.

Integrated programming environments are envisioned to have everything from the original requirements document in machine-readable form. Some distant prospects exist for specialized editors which "know" about the various kinds of documents in the system as the above mentioned smart editors "know" about programming language source. There is also the suggestion that such a requirements editor or design editor might feed into a quick prototyping tool which eventually might evolve into a program generator needing only a small amount of human "help" in the form of answering questions about ambiguities in the requirements specification.

Configuration management tools might monitor various releases of a system: who got it, did each recipient get all "fixes", etc. Such tools would track complaints, making sure someone handled them, and following them through changes and re-testing of modules and being sure the new configurations were actually released to the correct sites. All tests should be kept automatically by the system from the first test of a partial code segment on the source level debugger through to system
verification tests and the system should automatically re-run all of them (which are still applicable) as part of any change before release. In line with testing, there might also be automatic theorem provers whose results could be used to keep down the numbers of necessary tests.

One proposal would have management select a methodology or set of methodologies, based on the project's application domain, which thenceforth would drive the environment with respect to the project [76]. This suggestion is consistent with the goal of not having a particular methodology inherent in an environment yet guarantees that all programmers abide by management's rules. An environment could provide further management tools by automatically keeping track of who is working on what project/module, the amount of time and money being spent, and when the person moves on to something else. For instance, he might mark a module "complete" or signal that he has dealt with changes necessitated by some complaint or design, etc. change. The environment could also generate reports about these activities for purposes such as scheduling personnel and monitoring the progress of the project. Other reporting tools might include redundancy reporters and schedulers of review sessions based on some combination of elapsed time, percentage of the system that has been changed, and faults reported, etc.

An important consideration for environments for large projects is that often they are scattered over great distances and among many organizations. It has been proposed [77] that environments be designed flexibly enough to themselves be distributed with parts communicating with each other, or to adapt to dealing with other, perhaps manual, environments in a secure manner. UNIX has mail and news systems which can
serve for communication among individuals and groups at different sites or on the same site, but is considered by this proposal to be too general. A mail system is desired in which not just any text can be sent and in which receipt must be acknowledged and, for action requests, in which the acknowledgement must include an agreement or suggested alternative routing.

This has been an extremely brief survey of some of the things researchers are trying to do and are thinking about doing with programming environments. For more depth, it is suggested that one read [65], and for an analysis of the prospects of introducing efficient environments into the everyday world of programmers in the field, that one read [59].

The technology of programming environments as currently implementable [65] does not go far beyond collections of "good" toolsets appropriate to general software construction. The prospects for the future are brighter for well-organized, cooperating systems which may have a chance at enforcing adherence to those methodologies deemed more likely to produce reliable systems. Unfortunately, that day is not here. The current toolset approach has the same failing noted in the other areas examined in this section: The approach allows rather than enforces practices which may lead to the development of reliable software.
SECTION 3

Enhancements To The Conventional Software Development Cycle

3.1. Overview

The conventional process of developing software might be made more reliable through the inclusion of several advanced techniques and a controlled reorganization. Appropriating the prototyping concept from other fields permits rapid feedback from customers on the accuracy of the specified requirements and opens the producers' eyes to the problems which will present themselves during full development. Re-use of the work of others in the form of components limits the effort required in implementing and verifying a new system. An integrated environment can organize and enforce the flow of activities in the process, carry out some transformations itself, and provide the "memory" necessary for life-cycle-long configuration and enhancement control. Closing the gap between requirements specification and implementation languages through development of very-high-level languages (VHLL) would enhance the ability to simulate proposed designs before committing to them and would lessen the chances of introducing faults into design and implementation due to improper semantic mappings.

The cycle itself needs reorganization to place the decision-making and checking in the proper order and relegate to their proper roles less beneficial activities. Often, in the conventional process, implementation decisions are made during the design phase, no checking for design
validity is done until test cases are run on the implementing code, and the current state of testing methodologies is such that developers place unjustified importance on that part of the process. One proposal for automatically enforcing the ordering on activities in the cycle is embodied in the SAGA system. Here, a program enforces previously-defined rules governing which commands (such as "EDIT design document" or "COMPILE module") are valid at any given point in the development process.
3.2. **Software Prototypes**

In most engineering fields, a full-scale product is never attempted before a pilot plant or prototype version has been built and operated to the satisfaction of both producers and customers. This has rarely been the practice in software development projects. Software prototyping has grown out of experiences in which software systems have been completed only to be immediately scrapped because the customer realized too late that the product specified and built was not what was wanted [15].

Software prototyping technology is becoming a useful tool that should be pursued with a view to applying it to crucial software. The New York University implementation of Ada using the SETL system is a superb example of prototyping. The prototype implementation proved that Ada could be translated, to counter the arguments of those who could not design compilers for it. It provided early feedback to the language designers about things which were indeed unimplementable. And it allowed the Ada Compiler Validation Capability (ACVC) development to proceed in parallel with other translator development projects, since proposed validation suites could be tried out on the prototype translator before anything else existed. There has been substantial criticism of the NYU Ada translator because it executes very slowly. The critics are missing the key point that in this prototype, speed has been routinely sacrificed for functionality.

SETL is not particularly application specific although it is clearly more appropriate for prototyping compilers than control systems. Systems oriented to control systems' prototypes could probably be constructed on the SETL model.
An overview of the issues and work being done in software prototyping can be found in [78]. There has also been an NBS workshop on rapid prototyping a report on which is to be included in an issue of Software Engineering Notes in the Spring of 1983 [79].

To be reliable, a software system must conform to its requirements specification, but it cannot be built to meet requirements which are not known. A prototype model enables the customer to notice the absence of, and make explicit, requirements which had been assumed but not previously specified or the presence of things specified unintentionally. Large systems' requirements tend to change while they are being built. The early use of a prototype can serve to stabilize system goals sooner in the cases where changes to requirements were due to capabilities previously "left out" of the requirements specification. Often the originators of the requirements specification will not have experience with making explicit such things as a system's desired behavior. So it is difficult for analysis of the requirements specification to produce an accurate depiction of the behaviors wanted. Such problems can be ameliorated by allowing the eventual users to exercise a rapidly built prototype of the system. As in the above example of the SETL Ada implementation, a prototype may be used to experiment with possibilities for dealing with novel problems. Thus, production of a prototype serves as a means of verifying the transformation of the original idea to machine readable requirements specification. In that they will most likely build the prototype quickly while examining the original requirements specification, it gives the highest people on the production staff a chance to foresee some of the problems to be encountered later on.
Often, the most risky or uncertain aspects of a problem are placed into the prototype while more pedestrian aspects are ignored for the sake of cost savings on this version which will be thrown away. The analysis of what to leave out in the simple prototype helps to establish a basis for later application of functional decomposition. The Irvine report [78] offers several examples or real prototypes and the kinds of functions emphasized or left out to enable their rapid construction. One example involved the user-friendly interface and estimates of computational load for an automated FAA flight service station.
3.3. **Software Components**

A software component is a routine or set of routines with their own private data which have been written to provide a service useful to a variety of larger projects. A component which is in a portable form, and has been proven to actually provide the service it claims to provide, can be of great use in building crucial software. The persons responsible for the project can limit design efforts at higher levels to matching the components' interfaces. There is also less project-specific code to view with suspicion should faults be detected.

A limited form of software components has been with us for many years in the form of mathematical subroutine libraries. However, frequently other kinds of components have not been included in such libraries because of the difficulty of specifying what functions are performed and of writing understandable interface specifications. A more important reason is that, previously widespread languages which could interface to routines in libraries had to be able to access everything within a library; there could be no information or auxiliary routines hidden from the user. Further, the desire for highly optimized code has led to users' reluctance to use anything they did not tailor to individual applications.

These otherwise valid reasons do not apply to Ada. The Ada package mechanism can provide portable abstractions of higher-level concepts and structures in which external interfaces are fully specified yet with internal workings inaccessible to users. As for the optimization problems, to make the language usable in real-time Ada compilers must perform extensive optimizations including those which apply across
procedure boundaries anyway. "Optimization" via algorithm selection can be done by providing a set of components for each function with additional specifications describing the types of situation most likely to benefit from each component in the set. These qualities are not specific to Ada. The technologies were not available in earlier languages or had not all been brought together in one system before.

One system for using components in the development of efficient and correct software is described in [80]. The view taken in that system is that a software component can be seen as a part which itself is composed of parts depending on level of abstraction. The traditional or craftsman approach, through an expensive and time-consuming process, produces efficient software requiring custom "maintenance" in the same way as any "hand-made" item does. The parts-and-assemblies or components approach produces cheaper software with a common "language of discussion" and allows the parts to be studied for the ways in which they can fail and be repaired in all applications. The component approach does not eliminate the craftsman since he is needed to build good, reusable parts, and a system can rarely be built entirely from such reusable parts. The relative costs of the approaches depend on the numbers of like programs to be eventually produced. Since components represent implementation choices, a fully coded and compiled part cannot be seen as an assembly which can be optimized in a manner which would make software components usable directly. Thus that system represents components as designs or input/output specifications and enabling conditions which can influence the choices of an automatic coding system in optimizing for particular applications. In that system, libraries of components were built for
specific application domains but were able to make use of components in libraries for other domains which had already been built. The components contained several alternatives with individual enabling conditions so that an alternative could be chosen based upon "goals" for development as specified by a human interacting with an experimental transformation system. The components were relatively small but the system could build up larger programs by combining them and using some components for selective replacement within the text of other components. The author describes this as "A domain's software components map statements from the domain into other domains which are used to model the objects and operations of the domain ... Each object and operation in the resulting program may be explained by the system in terms of the program specification." The examples actually presented in the text are necessarily small and textually oriented, but include the construction of a natural language parser-generator and a natural language relational database.
3.4. Integrated Environments

Environments were discussed in Section 2. However, as was noted, the current state is not as advanced as it could or should be. Much research needs to be done to create programming environments which actively take part in the production and correct revision of software systems rather than passively offering individual unrelated and inadequate tools. This active participation characterizes the concept of an integrated environment. Where the environment knows about the forms and processes of software projects in general and of an individual project in particular, it can institute and impose appropriate checks and documentation policies. Thus, the minimum amount of human work/ineptness need be applied in a system's development.

The idea of environments needs to change from the box of tools approach to active participant. An integrated environment needs to recognize and save potential components for future use, and recognize places where a previously developed component can be used and insert it. The environment also needs to be able to generate a prototype model from any level of "document" such as requirements specification, design language "program", or partially implemented software to allow exercise by users or simulation at any stage of the project. The order of events needs to be controlled and enforced. For example, discovery of a fault should trigger re-examination of the requirements specification before design, and that before implementing code.

The HOST [81,82] system is to be such an integrated environment. By using H-Graphs as a standard form for internal manipulation, all of the system's "tools" can deal with the semantic basis of the project.
Thus the project's requirements specification, design language, and implementation language can all be reduced to H-Graphs, or an H-Graph form can be entered directly. This allows comparisons for compatibility and consistency among all forms of the software, and the use of components developed for other projects, perhaps in a different language. Finally, prototyping can be achieved via interpretation of the H-Graph representation of the requirements specification.
3.5. **An Improved Conventional Software Development Cycle**

We propose an enhanced software development cycle to include all the techniques mentioned above. It is shown in Figure 3.1.

The entire process is controlled by an integrated environment which, among other things acts as a determiner of the "next valid activity". All of the tools interact with a database which is pervasive, supplying the appropriate information where needed. The database is made explicit in the figure at the interface between the idea and the requirements specification for two reasons.

The first is that during post-delivery, as needs change, additions to the original idea can re-enter the system in the process normally termed "maintenance" (more accurately called "revision"). The original idea and requirements specification are retrieved from the database and fed through a consistency checker along with the additions. The consistency checker should insist that conflicts be explicitly overridden. (As a part of configuration control in the environment, the original requirements specification is not overwritten but a new one for the revised project is created.)

The second reason that the database is made specific is that during original entry of wants, previous projects' ideas can be compared and suggestions made for clarifications. Also the consistency checker can play a part in amendments during a project's development. The database as described resembles what has often been called an "expert system", and it is intended to be at least a primitive version of one.

All but the final path in the figure lead back to the requirements specification. Any fault detected during testing, analysis, or exercise
Improved Conventional Software Development Cycle

Figure 3.1
of a prototype and any item found unimplementable during coding or
design must be traced back to its origin in the requirements specifica-
tion as prime suspect with other areas becoming suspect if the require-
ments specification is found innocent.

Prototyping follows the requirements specification since we must
have some specification from which to build the prototype no matter how
vague. The prototyping step occurs between the requirements specifica-
tion and design on every iteration. Any revision of the requirements
potentially invalidates the previous requirements specification and its
approval which was derived from exercise of a prototype. This can be
seen as an instance of needing a rapid prototyping capability, and if
the prototype can be automatically generated or revised in real time
during the human-expert interaction that would be even better. In the
case of a fault being found and the requirements specification being
found innocent, the prototyping and analysis box may only entail the
check that the requirements specification really is correct.

Note here that the figure represents a general plan and the details
of each box may be complex with internal path control directed by the
environment and with the amount of complexity dependent upon the partic-
ular methodology chosen. For example, from Section 2 we saw several
requirements language-analyzer pairs and several design methodologies, a
choice of any of which would radically alter the appropriate box over
the others.

A prototype can be viewed as a model or simulation but we distin-
guish between the vehicle for "verifying" the mapping of ideas to
requirements specifications and that for verifying the mapping of
requirements specifications to designs, just as testing and static analysis are distinguished as steps verifying the mapping of designs to implementations. A device for running a test case may be termed a simulator, but more accurately a target environment emulation device.

Fault tolerance is brought in as early as the design phase. Such capabilities should be designed into a system rather than appended after full development. We have more to say on fault tolerance in Section 5.

After a design has been verified through simulation or other analyses, its implementation should be made formal wherever possible. This includes formal derivation where the state of that technology is applicable, the use of previously verified components, and proving of coded portions of the program to the maximum extent possible. A formal semantic definition of the implementation language and a formal semantic representation of the design can be used to direct and guide the implementation process to the extent that matching semantic definition languages may allow the design "document" to select statements or routines.

Note that the components box involves some give and take with the implementation box. Design information and specified requirements information influence the choice of components from the database and, in a system such as [80], influence automatic optimizations on the components chosen. Recognition, automatic or otherwise, of newly created items in a project which could themselves be used elsewhere as components can be made to trigger inclusion of these items into the components database.
The problems with testing were described in Section 2. However, any exercise of a program has the potential for detecting some fault, and humans have a psychological need for some sort of testing of any newly developed product. If and when a formal theory of testing is developed, the figure has reserved a place for it.

The stages of the software life-cycle are often said to be conceptual, actually taking place in parallel or in an overlapped manner. The controlling environment should insist on an order which separates the concerns of requirements specification, design, and coding. Just as separation of concerns in a design limits complexity and enables an accurate mapping of specified requirements, separation of the stages in the development cycle limits the amount of information needing to be dealt with at one time and prevents premature decisions in one part of a system from having undue influence on the rest of the system. The fact that we have paths back to the requirements specification does not change this position. Any alterations to requirements specification, design, or implementation necessitated by traversal of such a path should cause the replacement of the affected parts, no matter how widespread.
Given the stringent reliability requirements of crucial software, can the conventional software development cycle or the cycle described in Section 3.5 be used to build software of the desired quality? The answer of course is yes, but rarely and unpredictably. There may be circumstances in which reliable software is developed using conventional methods. The problem is knowing that the software is sufficiently reliable. Crucial applications will certainly not use software which contains known faults. However, what is required is assurance that either there are no residual faults or that the unknown number of residual faults will not lead to failure. We claim that the conventional software development cycle cannot meet these requirements. We will attempt to justify this claim with some experimental evidence and expert opinion.

Two crucial applications relying on digital computers are the control of manned spacecraft and the control of nuclear weapons. Software failure in either case could be catastrophic. Both applications presently rely on conventional software development methods, and both have experienced failures in production software systems. For example, the first launch of the Space Shuttle was delayed for two days [36] by a software fault. Fortunately the consequences were not serious. In another example, the launch control system for the Trident missiles on
board a Trident submarine went into an infinite loop when an operator attempted to "launch" all 24 missiles in sequence during an exercise [83]. All the missiles were disabled, and had this been a battle situation, none could have been launched. The diagnosis of this problem was operator error since the missiles are supposed to be launched in three sequences of eight each.

The software which operates the SIFT computer [84] must be regarded as crucial since the correct operation of the computer relies on the correct operation of this software. The designers of SIFT did not use the conventional software development cycle but chose instead to use a formal verification method. They feel that faults were found this way which would not have been found by conventional methods [85].

The software which supports communications of classified data is crucial in the sense that failure might allow compromise of classified data. Note however that failure which causes loss of service is acceptable provided security is maintained [86]. This is far less stringent a requirement than is imposed on crucial software.

The workshop on The Production of Reliable, Flight-Crucial Software [87] was asked to discuss the issues involved in crucial software development and make recommendations on research areas which should be pursued. The first conclusion reached (which was agreed upon unanimously) was:

There is serious doubt that it is presently possible to produce flight software systems having the stated level of reliability and to assure that they have that level of reliability.
Finally, Winograd [88] has argued the case for major changes in software development methodologies for various reasons, and Wasserman et al [89] have pointed out that, in crucial applications, the consequences of software failure may extend beyond the normal concerns for human life or expensive equipment to legal actions against the programmers involved.

Taken together, these points convince us that the conventional software development cycle is inadequate. There may be examples of programs running which have been created by conventional means and which appear to be reliable. The key word here is "appear". It is necessary to show scientifically that the software is sufficiently reliable.
SECTION 5

Fault Tolerance

Although fault tolerance has been applied extensively in hardware, it has received relatively little use in software. There is an important distinction between hardware and software faults which must be born in mind in discussing fault tolerance. The majority of hardware faults are the result of physical degradation of components whereas software faults have the characteristics of design faults. This precludes the use of parallel executions of identical software to guard against faults but, in contrast to hardware, software does have the potential for being permanently fault free.

In this section we review the state of the art in software fault tolerance. We assume the reader is familiar with the basic principles of the various methods. In general we feel that software fault tolerance has the potential to increase reliability dramatically. It can be considered part of the conventional software development cycle. It is considered here in a separate section because of its importance.
5.1. **Recovery Blocks**

Recovery blocks were proposed as a technique for providing tolerance to faults in sequential programs. A very strong theoretical background has been developed for recovery blocks. Provided erroneous states are detected, damage assessment and state restoration are totally reliable, and continued service can proceed from a secure starting point. Two disadvantages are the need for hardware support (the recovery cache) for state restoration and the fact that this is backward error recovery. Despite the fact that the recovery cache was patented ten years ago, there are no commercially available machines with recovery caches and so there is no opportunity to use recovery blocks in practice. Backward error recovery could be a problem in real-time systems and has to be taken into account.

Attempts to extend recovery blocks to concurrent programs led to the problem of the domino effect and to the conversation technique as a solution. Conversations are theoretically quite simple but rather surprisingly no syntax has been chosen for their inclusion in programming languages (in contrast to recovery blocks). Some proposals have been made [90, 91] but none has gained even modest acceptance and none has been implemented. The reason for this situation is twofold. Firstly, although conversations seem simple, integration of their semantics into a language supporting concurrency is a major effort. Concurrent languages are still in their infancy and there are many very difficult issues in their design. Incorporating conversations just makes a very difficult problem even harder. The second difficulty with conversations is that once again hardware support is required. In
contrast to recovery blocks however, in order to implement a conversation, a recovery cache is required for every process involved. Thus in principle, many logically separate cache's have to be provided.

It has been observed [92] that many real-time systems have properties which allow fault tolerance using backward error recovery to be included fairly easily. A framework has been proposed which allows fault tolerance to be included in cyclic real-time systems with no special hardware provisions. It has been pointed out that this work does not cater for real-time systems which are interrupt driven and this is a serious weakness. The work is being extended to include interrupt driven systems to provide a comprehensive approach to fault tolerance in real-time systems.
5.2. **N-Version Programming**

With obvious analogy to hardware techniques, N-version programming [93] has been proposed as a method of providing software fault tolerance. It relies for all aspects of fault tolerance on the execution of multiple versions of a program and comparison of their results. This is somewhat weaker theoretically than recovery blocks. Damage assessment is handled by the **assumption** that damage will be limited to the versions in the minority when the vote is taken. To ensure that this is true, the versions must be physically separated. Clearly this is not easily achieved for parts of programs such as subroutines. In practice, this limits the application of N-version programming to the system level and precludes its inclusion in technologies like software components.

A further difficulty is the treatment of state restoration. Again, this is handled by the **assumption** that the different versions do not interfere and that the states of the versions in the majority after the vote are consistent and ready for continued service.

It is important to note that any versions in the minority after voting must be **assumed** to have failed. Thus they cannot participate in any further system activities. If the system is required to continue operation, there must be sufficient versions remaining for voting to be possible.

Voting presents another problem for N-version systems. If the versions are implementing some form of arithmetic, the results may not be in bit-for-bit agreement. In such cases, have there been failures? Probably not, but to avoid detecting failures in these cases it is
necessary to use ranges rather than exact inequality tests. How wide should the ranges be? If they are too wide, failed versions will not be detected, and if they are too narrow, successful versions will be rejected.

An advantage of N-version programming is that it can be readily applied to concurrent and real-time programs since it does not rely on backward error recovery. Indeed, it has already been applied to a crucial application [94]. Hardware support is required for N-version programming in the from of provision for physical separation (usually multiple processors) and for voting.
5.3. Reliability Improvement

A major area of concern in all aspects of software fault tolerance is a lack of data showing that reliability is improved by using it. No major demonstrations have been performed which show that fault tolerant crucial systems can be built (although one such experiment is underway at the University of Newcastle upon Tyne in England [95] ), let alone that they will be adequately reliable.

It is intuitively reasonable to expect software reliability to be improved by using software fault tolerance. Intuition is often wrong, and it is necessary to resolve the remaining issues in the technology of both forms of fault tolerance and to obtain reliable data on reliability improvements that can be expected before the technology can be recommended for inclusion in crucial software.
By verification we mean the technology of establishing a mathematical proof that an executable computer program complies with its requirements specification. We have not spent a great deal of time on this topic because of the substantial experience already in Langley's Fault-Tolerant Systems Branch. The SIFT project and the contact with the SRI verification group is extensive and provides a far better assessment of that technology than we could obtain from the literature. For the sake of completeness, we have included an extensive bibliography on verification.

We make several observations of a cautionary nature because we feel that it is important that verification not be viewed as a panacea. First, if a program is to be proved, its requirements specification has to be in machine readable form which is amenable to analysis and this is not always easy. For crucial applications it could be required but that means that the engineer and the computer scientist will have to communicate in an informal language (English) or the engineer will have to learn (and be comfortable with) the formal notation. Another difficulty with verification is the complexity of the proof process. Theorem provers are a help but there is still a need for human guidance and inspiration. This makes the proof process long and tedious, and contributes to the fact that program proofs are not a routine matter and
proofs of programs more than a few hundred lines long are very rare.

Perhaps the biggest danger with verification is the prospect of the proof being wrong, i.e. a proof being produced for a program containing faults. There are numerous examples of this in the literature. One example is by Geller [54] in which two proofs are presented for a program which is wrong. It must also be noted that there are major areas where verification has had no success whatsoever. These areas include floating point calculation, concurrent programs, and until recently real-time programs.

Despite these reservations, there have been some remarkable successes in verification technology. The proof of a simple real-time program [96] is very encouraging. The recent proof of a program that is more than 4000 lines long is also a major accomplishment. This program and its associated proof were constructed at a measured productivity rate of four lines of code per programmer per day [97]. This compares very favorably with the productivity obtained using conventional methods.

Provided the problems are kept in mind, verification appears to be a technology that is almost ready for application in some parts of crucial systems. The comprehensive approach to crucial software engineering that we propose in Section 8 incorporates verification.
SECTION 7

Automatic Programming

7.1. Introduction

We have come to the conclusion that in the long term, major improvements in the reliability of software will only be achieved if the ad hoc methods of construction in which humans are involved can be eliminated. The problem in the classical software development cycle, or any enhanced version of it, is that humans make decisions at every stage and thereby introduce errors at every stage. As noted in Section 6, verification can provide substantial reliability improvements. It relies, however, on human-generated programs and human-generated proofs for the most part, although there may be extensive computer checking. Despite impressive success with verification, it is only an intermediate step. The long term goal has to be the removal of unchecked (or uncheckable) human decision making from the software generation process.

The creation of a software requirements specification is the only step in software development where human decision making is required. It is the link between the "idea" or "concept" for a system which exists in a human's brain and computer processing of that idea. Once a complete, formal requirements specification exists in machine readable form, it is amenable to many formal methods of analysis. In principle, these methods can be used to build an executable computer program directly from the requirements specification with either no human...
intervention or just human guidance. Thus it is potentially possible to derive a program from its requirements specification and thereby "prove" that the resulting program complies with its requirements specification (this is the definition of reliability). Note that no proof in the classic sense of program proving is needed. Where formal methods do not yet exist, or are not yet sufficiently powerful (such as program design), additional research can be expected to yield satisfactory new or improved techniques.

Unfortunately, since the requirements specification is the first machine readable version of the "idea" or "concept", the translation from the "idea" to the requirements specification cannot be automated and subjected to completely formal methods. Thus it will never be possible to prove that the requirements specification corresponds precisely to the original "idea". Many faults are introduced because of the necessarily informal (and thus inadequate) translation of the "idea" into a requirements specification.

The ideal situation would be one in which the requirements specification is entered into a computer by a human at the highest practical semantic level and the process of producing an executable program would be left to the computer. The only testing that would be needed would be that which convinced the human that the requirements specification as initially entered corresponded to the "idea" in his/her head. Emphasis must be placed on notations which allow requirements specifications to be expressed in a form where the semantics can be determined by processors which will be responsible at least for analysis and possibly for constructing the executable program.
As noted above, no proof of correctness would be needed for programs which are automatically derived from their requirements specifications. Similarly, no fault-tolerance methods would be required or desired. Another major advantage of this approach is the simplified procedure needed for enhancement or modification. Changing a program should always begin with changes to its requirements specification though it rarely does. Changing a program may involve a substantial redesign of algorithms and data structures although since this is so time consuming, quicker methods involving "patches" are often used. For programs which are automatically derived from their requirements specifications, these problems go away. The requirements specifications have to be changed but the rest of the process is automated. Although it may require very large amounts of computer time, the derivation can proceed automatically.

The practical implementation of these notions is termed automatic programming. We have been reviewing this technology at some length in order to determine its feasibility in the long term as a method for building crucial software. A modest version of the technology has been in use for some time in the form of high-level languages. Programs written in high-level languages are really requirements specifications for machine-language programs. These machine-level programs are not written by humans but are derived automatically from the requirements specifications by a computer program; namely a compiler. It is not unreasonable to state that most programmers never write programs; they write the requirements specification in a non-executable language (a HLL) for a machine language program which is synthesized automatically.
This application of automatic programming is readily accepted and needs to be extended to higher-level constructs to allow more of the translation process to be automated.

The state of the art in automatic programming at the level needed to eliminate human programmers from all but the requirements specification phase is very far from practical use, as would be expected. However, we have performed an extensive literature search, and the example systems that have been built and reported are quite impressive. For example, with a little human guidance, a program to solve the eight queens problem has been derived from its requirements specification.

There are several approaches to automatic programming and there are related research projects which contribute to the goal of eliminating human creativity from programming. There are many excellent surveys of this field and they will not be duplicated here. The interested reader is referred to the bibliography section on Automatic Programming, to the survey by Biermann [98], and to the survey in the Artificial Intelligence Handbook [99]. These latter two papers are excellent surveys of the state of the art in automatic programming. Both are quite long (63 and 110 pages respectively) and summarize the theory behind the methods as well as describing the major operational systems. They are both very readable and the second one is very recent (1982).

In this section we will discuss briefly the various approaches to automatic programming and the associated technologies. Although under the general heading of automatic programming, two related research projects are mentioned. They are SAFE and the Programmer's Apprentice. These two systems are applications of artificial intelligence which help
reduce human error but are not complete program synthesis systems.
7.2. The Issues in Automatic Programming

Any automatic programming system will require as its input a program's requirements specification. We have already noted substantial difficulties in this area in the conventional software development cycle. Precision, freedom from ambiguity, and so on are all very hard to achieve. For automatic programming, the situation is more difficult since the requirements specification has to be amenable to machine analysis. This seems to eliminate natural languages which are very convenient for humans but very difficult for computers to process. The predicate calculus is usually suggested as a suitable notation, but it is quite difficult for most humans to deal with.

This conflict has lead to two important lines of research. One is automatic programming systems based on the predicate calculus [100], and the other is an effort to build a processing system for the English language [101]. In limited ways, both have been successful and we recommend the products of this research in our comprehensive approach (see Section 8).

Another technique for specifying requirements is the use of examples. The intent is that the user gives the system examples of the computation required and the system builds a program which satisfies all the examples. Two different approaches to defining examples are used. In one approach, the output expected for each input is given. In the second, the user works through the desired algorithm with sample inputs and the system is required to infer the algorithm.

Programming by example has been studied in depth and implemented in two commercial systems by IBM [102,103]. Unfortunately, a program which
works for all the examples may fail on the first application. Specification of requirements by example and hence programming by example does not seem like a viable technology for building crucial systems and will not be discussed further in this report.

The output notation used by an automatic programming system is called the target language. Different systems use different languages but in most cases the target is some kind of high-level language. The output of an automatic programming system can be translated into an executable program by some form of compiler, thus completing the synthesizing process.

In principle, it is not necessary to be aware of the existence of the target language or the fact that one form of the desired program is written in this target language. In practice it is important and quite useful. We have noted that automatic programming may not be able to build programs of the size or complexity that we need. In section 8 we propose an approach in which part of the program is synthesized automatically and part is written by conventional means. The parts will have to be merged and this can best take place at the level of the target language.

To reduce the complexity of program synthesis, most existing systems restrict the application area that they deal with. Since automatic programming is basically a research area, this approach is appropriate. The goal of most researchers is to develop algorithms which will synthesize something rather than something specific. Thus existing systems are impressive (in some cases) but not particularly relevant to crucial software development. This means that it will probably not be possible
to use any existing system, or minor variant thereof, in a practical crucial software development system. It also means that it will probably not be possible to use a single automatic programming system, even if it were specially developed, in crucial software development. In practice, it will probably be necessary to use several complementary systems; each working on part of the problem.

There are several fundamentally different methods of operation used in the various automatic programming systems. The different methods have various advantages and disadvantages but externally the major difference is in the size of programs which can be synthesized. At the time of writing, the program sizes vary from a few lines in the approach of Manna and Waldinger [104] to many tens of lines in the system of Balzer105. It must be kept in mind that there are other issues to consider in comparing systems. For example, the Manna and Waldinger approach is potentially more general although though there is no clear limit to the transformation approach of Balzer.
7.3. Automatic Programming Systems

In this section we discuss some of the more significant automatic programming systems and consider their relevance to crucial systems. This is not a general survey. The interested reader is referred to [99].

SAFE

The SAFE (Specification Acquisition From Experts) system [101] is an attempt to build a program which can interact with a user in natural language in order to determine the requirements for a program. Its input is in a subset of English and its output is a requirements specification written in a formal notation. Thus SAFE is trying to solve the difficulties known to exist in the phase of software development where requirements are specified.

SAFE expects its input to be ambiguous and incomplete. The goal of the system is to determine these problems and resolve them by interaction with the user. With this goal, and the apparent successes of the system, it appears to be an ideal candidate for use in crucial system development.

The literature on SAFE indicates that its primary focus is on requirements specification. In fact in recent work [106] the SAFE system has been coupled to a system supporting transformational implementation and a complete automatic programming system is being built. This was the original goal of the SAFE system designers. SAFE is just an intermediate step but a very important one.

Programmer's Apprentice
The Programmers Apprentice is described by its designers [107,108,109,110] as an intermediate point along the road to the desired goal of automatic programming. It is not an automatic programming system. It is an application of artificial intelligence methods in a system designed to help a human programmer by checking his work. It has a substantial "knowledge" of programming maintained in a library which it uses to help validate that human generated code is consistent with the specification for that code.

The system is intended to operate interactively, conversing with a programmer as an apprentice would. The examples of the system in use which appear in the literature are very impressive but apparently describe what it would do if fully implemented rather than what it is able to do as currently implemented. It also appears that the system has not been the subject of active work for some time.

The ideas behind the Programmer's Apprentice and the capabilities that it apparently provides seem very suitable for use in crucial system development. This technology is probably applicable in the short term.

PSI

PSI is a system built at Stanford by Cordell Green and colleagues [111]. It is very large and apparently powerful. Some confusion can easily occur in examining the literature since two major parts of PSI (PECOS [112] and LIBRA [113]) have been described in separate papers and appear to be separate systems. PECOS and LIBRA are both capable of some independent operation but are basically parts of PSI. In fact, PECOS is the "coding expert" and LIBRA is the "efficiency
expert" of PSI. We will not discuss PECOS and LIBRA separately.

As far as we can tell, PSI is the most comprehensive automatic pro-
gramming system that has been built and it is the only system to have
addressed certain issues. For example, its method of operation is basi-
cally transformational but it can reach a state in which several dif-
ferent transformations are valid. Note that any system based on
transformations can reach such a state. A choice at such points could be
made by a human, or by the system at random, but PSI invokes its effi-
ciency expert (LIBRA) which searches the valid transformations for the
one which will yield the most efficient program. Efficiency could be
terms of time or space.

PECOS is the "coding expert" for PSI and uses a database of rules
to make decisions about program synthesis. An impressive aspect of
PECOS is the way in which certain simple, "well-known" rules of program-
ing are contained it is database. For example, the following rule
which is frequently used by programmers can be included in the database:

If a collection is input, its representation may be converted
into any other representation before further processing.

In papers describing PSI and its subsystems, various examples are
given of programs that have been synthesized. Although we are impressed
that anything can be synthesized, we find the semantic level of the
input to be very low. The input definition of the problem in many cases
seems to contain too much detail and in fact is virtually a complete
program.
7.4. Conclusions About Automatic Programming

Several experimental systems have been built (many have not been mentioned here but see the bibliography), and a great deal of research has been performed on automatic programming. Though the subject is far from ready for general use, it does hold great promise and there do not seem to be any fundamental, theoretical reasons for thinking that progress will not be made in automatic programming.

The advantages of synthesizing programs from their requirements specifications are many but the most important here is the potential increase in reliability. This approach seems to be the only viable way of ensuring that programming is carried out in a scientific way, and does not rely on human fallibility and the associated introduction of faults.
8.1. Overview

Recall from Section 1 that no effort is made to quantify reliability or reliability improvements in this grant. Thus in this section we describe an approach which we feel will produce the most reliable software product but we make no claims as to the degree of reliability which might be achieved.

Given the inadequacy of the conventional software development cycle, the difficulties with verification, and the infancy of automatic programming, how should crucial software development proceed? Moving from conventional methods to formal verification will yield an improvement in reliability of software. Similarly, moving from formal verification to automatic programming will yield another improvement, and automatic programming probably represents the best that can ever be done. An ideal solution would be the use of automatic programming for the entire development of software for crucial applications. This ideal is far from possible at this point so a less desirable, more practical approach must be sought.

Basically, we propose that a combination of these three techniques be used. For a given application, those parts which can be synthesized by an automatic programming system should be. Of the remainder, those parts which are written by humans but are amenable to verification
should be verified. The remainder of the application (which may well be large parts of it) will have to be built with conventional methods. To provide some hope that this latter part is adequate, we propose the extensive use of fault tolerance throughout this part of the software.

An overall environment needs to be produced with three clearly defined but cooperating paths for the three development techniques. A monitor is needed which would interact with the user to allow different parts of the requirements specification to be guided down different paths. The requirements specification for some clearly defined part of the software could be presented to an automatic programming system. If the system failed to synthesize the necessary software, the user would then have to write the software and attempt to verify it. If that failed, the user would be required to restructure the software to include the necessary fault tolerance. The monitor would be required to keep track of these various activities and assemble the final program from the various synthesized, verified, and fault tolerant parts.

Figure 8.1 shows this proposed approach in rather limited detail. Essentially, each of the three major aspects of the method is a software development approach in its own right and will be described below in more detail than shown in Figure 8.1.
The Comprehensive Approach

Figure 8.1
8.2. Requirements

It is difficult to choose any one of the various requirements languages that have been developed; each has its advantages and disadvantages. For reliability, extreme formalism is best. The predicate calculus is a good choice because there is such an extensive body of theory supporting it, and it is precise. It is also quite difficult for the average computer scientist to use. For ease of use, natural language is a good choice but there is no supporting formal theory, and natural languages are imprecise and ambiguous. The impressive work of Balzer on the SAFE system leads us to suggest that some form of restricted natural language be used for crucial system requirement specification and processed by the SAFE system to produce formal requirements specifications.

We have not had any direct experience with using the SAFE system. The papers which have been published about the system are rather limited, but the system seems to be very impressive. A major concern is whether it can handle a natural language with sufficient expressive power for crucial systems.
8.3. The Monitor

The monitor is responsible for coordinating all the system activities once the requirements specification has been produced by the SAFE system. It requires a database for maintaining data, source files, reports, etc, as development proceeds. It is not dissimilar from the control systems of existing advanced environments. However, since there are three parallel development paths, the interactions between the paths will have to be handled very carefully. Some of these interactions are touched upon in further subsections. It is probably the case that no existing or currently proposed environment could handle all of these interactions.
8.4. The Automatic Programming System

Which automatic programming system should be used? We cannot select a single method because the technology is so immature. Given the current state of the art however, we suggest the methods of Balzer, and Manna and Waldinger.

There is a conflict between the use of automatic programming systems and the other two major parts of this system. In principle, an automatic programming system is supposed to do everything, including design. The other two parts require human input for everything including the design phase. If the automatic programming system cannot handle the entire development, it may be able to synthesize part of the software. The part which remains may not be in a convenient form for any of the traditional design methods.

Another issue is the difficulty of building automatic programming systems which can handle arbitrary problem domains. The search space that this implies is very large and this is a major limiting factor in the ability of automatic programming systems to synthesize programs.

Both of these problems can be solved in the following way. The automatic programming aspect of this system can be implemented as a series of automatic programming systems operating in parallel and each tackling a small, well-defined part of the crucial software applications domain. For example, most crucial systems operate in real time and an automatic programming system capable of synthesizing real-time schedulers, and nothing else, could be a component of the system. That part of the specification defining the real-time requirements could then be supplied to that module and a suitable scheduler output.
Major aspects of the software design phase would then take place at the monitor stage. Those parts of the software which could be synthesized by the automatic programming systems could be selected and code synthesized. What remained would be well defined and amenable to human detailed design and conventional development. Thus, we propose a set of automatic programming systems, rather than one, until technology reaches the point that a single system can cope with a complete crucial system. As technology proceeds and more powerful automatic programming systems become available, they can be added to such a design, and more of the development of a crucial system can be moved from verification and fault tolerance to automatic programming.
8.5. Verification

Verification is just one part of software development, unlike automatic programming which (in principle) covers the entire transition from requirements specification to executable program. Thus a program which is to be verified requires all the elements of the software development cycle to be present. That is assumed in this comprehensive approach.

The verification part of this system would operate as existing verification systems do. Other parts of the system would be required to assist the process. The monitor and associated database would be used to store proofs, control access to source code, and so on.

A theorem prover would be needed for verification. Many approaches to theorem proving exist and we do not comment on which might be used. However, we note that some automatic programming systems rely on theorem proving. Indeed, the Manna and Waldinger [100] approach derives a program from a proof. Thus a theorem prover is central to both verification and program synthesis, and proof techniques which can be shared by both technologies should be included in this comprehensive approach.
8.6. **Conventional Software Development Cycle**

The "improved conventional software development cycle" which was delineated in Section 3 is modified from a stand-alone system to fit into the context of our comprehensive approach. The monitor serves in the role of the enhanced programming environment. The SAFE system stands in the role of the requirements specification and analysis stages of the conventional approach (see Figure 3.1). The use of the SAFE system may eliminate the possibility that the requirements specification is incomplete or inconsistent. Gross design has been done by the SAFE system and the monitor. This entails the separation of concerns for the automatic programming systems and the human development effort. The portions of the crucial system developed by the automatic programming systems can (and are intended to) act as components for the human effort. Such portions may also enhance the overall system's prototyping capabilities in that an early prototype may consist almost entirely of the automatically programmable parts of the crucial system. It may occur that the automatic programming systems and verification paths "fail" due to human error in the human-guided gross design. For this reason, the conventional cycle's loop back to original requirements specification has not been eliminated.
8.7. **Fault Tolerance**

The monitor can require that fault-tolerance be designed into all human programmed software. The code built for attempted verification, if it passes, can serve as is. However, if it fails verification, it can serve as primary alternative in a fault-tolerant design with the monitor then requiring the creation of other alternates. The conditions used in the verification attempt can be saved for use as acceptance tests to fill out the fault-tolerant implementation.

A monitor which "knows" about Safe Programming [114] can aid both in verification and in providing fault-tolerance by enforcing limits on all loops. The monitor should also force the use of fault-tolerant techniques at all interfaces of the human-created code with the outside world and with other parts of the human-created code. All of these interfaces would be known to the monitor since it presided over the high-level design.

The application of fault-tolerance is not strictly limited to the non-verifiable human effort. The automatically programmed parts need fault-tolerant interfaces with the human's code, both when using a human-programmed component and when being passed parametric information from human-programmed components. One way to create a design fault is for the human to mis-use the automatically programmed code. Although in the strictest sense a requirements specification problem, the automatically programmed portions must be able to handle all possible situations in the critical system/world interface as well.
8.8. Testing

No matter how the software is built, if it is for a crucial application, it will be tested. There are many technologies which are involved in testing. In Section 2 we discussed only commonly-used conventional methods in which the program is executed on sample inputs and the resulting outputs are compared with those expected (recall however that it is frequently difficult to know what output is expected).

Naturally, this approach should be taken with crucial software also. As we have noted however, there is no theoretical basis for any of the practical, conventional testing methods and very little can be concluded about the software from the results of the tests.

Nevertheless, given that testing will occur, how should the test cases be selected and how should the tests be conducted? With no theory of testing, any and all the methods have merit. There is no reason why they cannot all be used. We show in the detailed version of our approach (Figure 8.2) a test case generator which is merely a piece of software designed to aid the programmer in correctly generating the desired combination of inputs. The monitor is shown connected to the testing tools because the tests will be driven by the formal version of the software requirements specification.
The Comprehensive Approach In Detail

Figure 8.2
We use the term AIRLAB here to mean the facility presently being constructed and the research objectives of developing technology to meet the reliability requirements of crucial systems such as digital systems for commercial air transports. We assume that NASA's major interest is in making large improvements in software reliability over the long term via essentially basic research. We also assume that resources are limited and that the most promising technologies need to be selected. Thus these recommendations are limited to those which we consider to be high risk and high payoff. We divide these recommendations into the categories of the enhanced conventional software development cycle, fault tolerance, automatic programming, and the comprehensive approach.
9.1. The Software Development Cycle

There are many areas of research in the software development cycle which could be pursued. It is likely, however, that good progress will be made in some of these areas independently of any NASA sponsored research. For example, there is a great deal of research on environments being funded by the Department of Defense. This work is not oriented particularly to highly-reliable systems, but should yield valuable results and demonstrations which will be of direct benefit to those interested in crucial systems.

In examining the conventional software development cycle and enhancements, suitable topics for research can probably be limited to those areas which are receiving substantially too little attention or where the goals of other researchers do not adequately address the difficulties of crucial software development. The latter is characterized by research motivated by cost reduction, improved programmer productivity, or faster and smaller software. None of these is a very great concern for crucial software where reliability is the dominating metric.

Given these criteria, we suggest the following areas from the enhanced software development cycle be considered for research support:

1. Rapid Prototyping. The technology is very immature and holds great promise for clarifying issues at the start of a software project.

2. Requirements Specification. Although there is active research in this area, it is not directed to crucial applications and the state of the art is really very poor.

3. Software Testing. Despite the amount of testing which is performed and the length of time we have been testing programs, there is
still no scientific basis for testing.

(4) Static Analysis. This is basically an immature technology which seems very promising but still has major problems. It is potentially very valuable in crucial software development because it is automatic.

As well as the above, we suggest that a monitoring project be started to examine and evaluate conventional software development methods. New techniques are continually being developed and reported. How good are they? What is their impact on crucial software development? These questions need to be answered by experts in the development of crucial systems. Many military and commercial crucial systems are presently being built with ad hoc collections of tools, very limited knowledge of the state of the art, and limited resources to follow technology as it evolves. A source of information and assessment of technology as it applies to crucial systems would be very valuable to the developers of these systems. It would also permit a clear assessment of research needs.
9.2. Fault Tolerance

As noted in Section 5 there are many open questions in the technology of fault tolerance. A high priority area of research has to be the resolution of these various issues in order to provide a complete framework for the construction of fault-tolerant crucial systems. Topics include:

(1) Design and construction of hardware to provide processors which include recovery caches and support for the voting necessary in N-version programming.

(2) Determination of a suitable syntax for the conversation technique and its incorporation into a general language structure for fault tolerance based on backward error recovery.

(3) The creation of a better theoretical background for N-version programming and the formulation of a framework which guarantees the atomicity of the versions.

(4) A comprehensive study of the voting issue in N-version programming.

(5) A study of the most appropriate way of combining recovery blocks and N-version programming in the construction of crucial software.

Intuitively, software fault tolerance seems like a good idea. There is precious little evidence, however, showing that it really is. In fact, there is very little evidence showing that software fault tolerance is even feasible. Ideas which seem reasonable in theory sometimes turn out to be impractical, especially in computer science. Some experiments have been done which have implemented fault tolerant systems but they were very limited in scope and not in the avionics or even
real-time field.

A major research area to which AIRLAB is ideally suited is the construction of realistic demonstration fault-tolerant systems. We propose that advanced applications such as active controls be taken as typical of the crucial systems which will be required in the near future and that fault-tolerant versions of these applications be constructed. We have no doubt that many significant issues will arise in such activities which have not so far been suggested or resolved.
9.3. **Automatic Programming**

Automatic programming is very far from practical use but seems to hold great promise. Clearly, it is the highest risk, highest payoff, longest term technology being considered in this report. It has to be understood that the payoff period is likely to be many years [98]. In view of its technical infancy, there are few clear-cut experiments which can be conducted in the AIRLAB framework. Experiments involving program synthesis would probably have to be extremely simple, reflecting the state of the art.

In general, we recommend that automatic programming be reviewed in more depth than has been possible in this study. This review should include detailed evaluation of specific systems by installing them in AIRLAB if possible, and evaluating them carefully in the context of crucial software. The results of these analyses would permit a coordinated research program to be planned. Specifically, we recommend:

1. A working group of leading researchers in the field be assembled to review the state of the art, compare and contrast systems, and discuss the applicability of the technology to crucial systems.

2. Install, test and evaluate the SAFE system. Based on published reports, this seems like a very powerful system which could be applied to crucial system requirements specification in the very near future.

3. Install, test and evaluate the PSI system. Based on published reports (which are extensive), this seems to be the most complete and general automatic programming system that has been built.
(4) Install, test and evaluate any other automatic programming systems which appear to hold any promise of being suitable for programming crucial applications.

Once some experience has been gained with the available automatic programming systems, research goals will become clearer. It may be appropriate to begin constructing an automatic programming system tailored to real-time control although this does not seem desirable at this point. It is important to review existing systems, get the opinions of experts in the field, and gather specimen problems before defining research goals in this area.
9.4. Comprehensive Approach

The comprehensive approach which we proposed in Section 8 attempts to mix several technologies which are not normally used together. A central experiment which we recommend is an attempt to build a version of the comprehensive approach to determine the feasibility of this integration. If this experimental system is built carefully, it should allow new tools, such as more powerful automatic programming systems, to be added and evaluated as they become available. A testbed for new tools or modified environments is essential to allow for the assessment of these technologies in the crucial software context.

As we have noted elsewhere in this report, some software engineering technologies are developed with a specific application area in mind. If this area does not include crucial software, the technology may be useful and it may not. To allow for uniform evaluation, we recommend the establishment of a collection of representative problems from crucial applications. These could be made available to researchers to assist them in evaluating their own work, and they could be used by NASA to evaluate new technologies as they become available.
SECTION 10

Conclusions

We have reviewed many areas of software engineering in an effort to determine which areas of technology could contribute to a major improvement in software reliability if research is pursued vigorously. We have formed the opinion that methods which are presently used for software development are inadequate for building crucial systems. Further, we feel that existing methods are so far from producing the desired level of reliability, and that the required level will not be reached by incremental improvements to commonly used techniques.

As a first step, we propose that the conventional software development cycle be enhanced substantially by integrating the new technologies of software prototypes, software components, fault tolerance, the Ada program language, testing based on the emerging theories of adequate test coverage, and machine-based methodology enforcement. Even using the best modern technology, there seems little hope of achieving the required level of reliability and certainly no hope of being sure that this level has been achieved. The flaws in the conventional development cycle (even if it is substantially enhanced) are the extent to which it relies on human decision making and the non-scientific basis of most of the methodology.

Fault tolerance is often proposed as a "safety net" for software. Supposedly, even if the software contains faults, fault-tolerant methods...
will prevent these faults from leading to failures. It may, but this has still to be demonstrated and for concurrent systems (including many real-time systems), fault tolerance still has to be shown to be practical let alone useful.

Formal verification has made remarkable progress and is able to deal with quite sophisticated programs. Unfortunately, there are still major areas where verification is not possible. As a second step therefore we suggest that verification be integrated into the software development cycle and that its use be required wherever possible. Informally, we expect to see systems developed in which those parts of the system amenable to verification are verified and the remainder build by conventional methods. These latter parts would be required to include fault tolerance so that there is some "insurance" against failure in the non-verified parts.

In the long term, really large improvements in reliability will be achieved only if human creativity and decision making are removed from software development. This leads us to suggest that the techniques of automatic programming might provide the source of major reliability improvements. Automatic programming is very limited in its capabilities now but the possibility of direct machine translation of requirements specification to executable program has obvious and major advantages. We propose therefore that automatic programming be pursued as a topic of basic research. It cannot be used in building crucial systems at present but as research advances the state of the art, it could be used to build gradually larger parts of crucial systems.
Our comprehensive approach is a combination of automatic programming, verification, and fault tolerance coupled to improved conventional methods. The approach involves a system in which all three paths are available. A crucial system would be constructed by synthesizing as much as possible (which may be very little) using an automatic programming system, building the remainder using conventional methods and verifying as much as possible, and finally employing fault tolerance for those parts which cannot be synthesized or verified.

This approach will not necessarily improve reliability, but, even if it does, it may be very difficult to ensure that desired levels of reliability have been achieved. However, as basic research on automatic programming and verification allow more of a crucial system to be built with these technologies, reliability will surely increase. When systems can finally be totally synthesized automatically, it may be possible to make definitive statements about reliability.
REFERENCES

The works below are referenced within this report. Most also occur within the subsequent bibliographies.


41. J. C. Knight, slides from presentation for Hampton chapter of ACM.


78. D. A. Smith, "Rapid Software Prototyping," 187, Department of Information and Computer Science, University of California, Irvine (May 1982).


83. A. Dean, in personal communication.


95. T. Anderson, in personal communication.


97. D. Good, in personal communication.


INTRODUCTION TO THE BIBLIOGRAPHIES

The following extended bibliographies are intended to provide the reader with starting points for exploring further any of the subjects addressed in this report. These bibliographies list books, articles, reports, and papers to be found in the literature on the subjects discussed herein. There is some minor duplication among the sections since authors often deal with more than one subject within a work. Further, some of the more obscure works were not actually located, but the references obtained from the references sections of works which were located. Hence, there may be some inherited inaccuracies. The topics covered are:

(1) Requirements Engineering.
(2) Distributed Specifications.
(3) Design Methodologies.
(4) Parallel Programming Languages.
(5) Testing Methodologies.
(6) Static Analysis.
(7) Symbolic Execution.
(8) Programming Environments.
(9) Software Prototyping.
(10) Functional Languages.
(11) Verification.
(12) Fault Tolerance.

(13) Automatic Programming.


COMPSAC (November 1978).


52. R. Loshbough, M. Alford, J. T. Lawson, D. M. Sims, and T. R. Johnson, "Applicability of SREM to the Verification of MIS System Software Requirements (Vol I and II)," NTIS reference AD-A100 720/2 Volume I; AD-A120 721/0 Volume II.


134


82. "Cascade/2 in Brief," in Runit, University of Trondheim, Norway.


BIBLIOGRAPHY ON DISTRIBUTED SPECIFICATIONS


BIBLIOGRAPHY ON DESIGN METHODOLOGIES


42. T. Gilb, in Technoscopecs ((MANUSCRIPT)).


56. B. H. Liskov, "A Design Methodology for Reliable Software Sys-


58. A. deMarco, in Structured Analysis and System Specification,


60. A. Mili, "Designing Software Without Backtracking, The Part of


62. G. J. Myers, in Reliable Software Through Composite Design,

63. P. Naur and B. Randall, in Software Engineering, NATO Scientific
Affairs Division, Brussels (1969).

64. P. G. Neumann, "Experience with a Formal Methodology for Software
Development," in Proceedings International Seminar on Software
Engineering Applications (1980).

65. D. L. Parnas and Darringer, "SODAS and Methodology for System
Design," pp. 449, in Proceedings AFIPS Fall Joint Computer Confer-
ence (1967).


Languages (1981).


BIBLIOGRAPHY ON PARALLEL PROGRAMMING LANGUAGES


7. P. Brinch Hansen, in *Edison Programs*, Computer Science Department, University of Southern California (Sept. 1980).

9. P. Brinch Hansen, in The Design of Edison, Computer Science Department, University of Southern California (Sept. 1980).


Corrigendum: CACM vol:21, #12, Dec. 1978


BIBLIOGRAPHY ON TESTING METHODOLOGIES


BIBLIOGRAPHY ON STATIC ANALYSIS


BIBLIOGRAPHY ON SYMBOLIC EXECUTION


34. in *STONEMAN* (Feb. 1980).
BIBLIOGRAPHY ON SOFTWARE PROTOTYPING


BIBLIOGRAPHY ON FUNCTIONAL LANGUAGES


BIBLIOGRAPHY ON VERIFICATION


77. N. Francez and A. Pnueli, "A Proof Method for Cyclic Programs," 


79. M. Geller, "Test Data as an Aid in Proving Program Correctness," 

80. S. L. Gerhart, "Correctness Preserving Program Transformations," 

81. S. L. Gerhart, "Limitations of Proving and Testing," Validation 
Group Meeting I, NASA Conference Publication, NASA Langley 
Research Center, Hampton, Virginia (1979).


120. L. Lamport, "'Sometime' is Sometimes 'not Never'," in Seventh Annual ACM Symposium on Principles of Programming Languages (1980).


Verlag (November 1968).


169. S. Owicki, "A Consistent and Complete Deductive System for the Verification of Parallel Programs," Proceedings of the 8th ACM


178. V. R. Pratt, "Semantical Considerations on Floyd-Hoare Logic," 


202. J. W. Thatcher, E. G. Wagner, and J. B. Wright, "More on Advice on Structuring Compilers and Proving them Correct," Semantica-


BIBLIOGRAPHY ON FAULT TOLERANCE


76. W. G. Wood, Recovery Control of Communicating Processes in a Distributed System, Computing Laboratory, University of Newcastle upon Tyne (Nov. 1980).

BIBLIOGRAPHY ON AUTOMATIC PROGRAMMING


