

W. T. HARDGRAVE

E. H. SIBLEY

DEPARTMENT OF INFORMATION SYSTEMS MANAGEMENT

UNIVERSITY OF MARYLAND

1.0 Motivation

The major agencies which fund large scale systems research all tend to be worried by the difference between "basic research" (which is good, and not funded by operating units of government and industry) and "applied research" (which is tainted with commercialization). This short article tries to show that the "applied research" of ordinary software systems is really more "basic research" when applied to large scale data systems.

The problem is one of size and the difficulty of finding an adequate laboratory for large systems: it therefore requires some new views on the question "What IS basic research?".

1.1 The Traditional Approach to Software Research.

The traditional philosophy of research activity involving software (e.g. programming language/compiler) design and development has partitioned the process into the following categories (See Figure 1):

- Basic Research
- Applied Research
- Commercial Development

Basic Research in this philosophy is considered synonymous with "thinking research". That is, the investigator(s) think their way through the entire design process and polish it through discussions, working papers and feedback from other thinkers. Only when subproblems of moderate scale are isolated can they be committed to an analysis by machine. The terminal point of the basic research phase is a set of documents that comprise the specifications for a prototype system.

Applied Research begins with the development of a prototype system implementation. This begins by hiring programming personnel to code and make preliminary tests of the specifications produced during the basic research phase. In many cases, the basic research and the applied research are performed by different people. Once a version of the prototype is constructed to execute correctly (according to the preliminary tests), the next step is to perform detailed testing using "real-world" data. If the results are satisfactory, then (usually) a commercial venture arises to carry the approach from research into a product. This, of course, is the desired outcome. If the results are not satisfactory, then the specifications are refined and another prototype produced. This process, shown in Figure 1, is continued until either a satisfactory prototype is produced or the idea is discarded.

1.2 The Problem of Large Scale Systems

The methods just discussed as the traditional approach are not adequate for large scale systems, especially in the applied research state, because:

- Implementation of a prototype generalized database system is costly.
- Construction of a sample database is costly.
- Access patterns, data flow, and economic factors are generally unknown or only partly assumed in the applied research stage. In fact, the design of any large system is a series of compromises based on assumed utility of the system.

The implementation of a prototype generalized database system (GDBMS) is, in general, quite costly. Effective GDBMS systems consist not only of modules handling data structures and associated operations,

but I/O modules, integrity/recovery/restart modules, security modules, special communications modules, and a package of software for the database administrator. The implementation of a fully specified system covering all of the subareas of database analysis is necessarily costly. If a system is fully specified and documented in the Basic Research phase, it would not be attractive for the applied research phase, because a large bulk of implementation is necessary before any tangible results can be examined.

The construction of a representative test database is an even more costly item. In order to test a generalized database system, it is imperative that several representative databases of some magnitude be constructed. Furthermore, representative operation and transaction patterns must be defined and submitted to the prototype to provide some ability to benchmark. This is a complex problem and it certainly is costly. Thus the construction of a suitable test environment is another factor that makes applied research a costly and generally unattractive process.

At this time, very few organizations, public or private, have converted to a truly integrated system of data operating under a generalized database management system. A typical organization may be seen from different viewpoints: the number of viewpoints is potentially larger with a computer-based information system than with a manual system. As a result, the data flow, and the various economic factors that determine user access patterns are seldom known (or derivable) from manual systems or non-database-oriented computer-based systems.

2.0 A New Definition of "Basic Research"

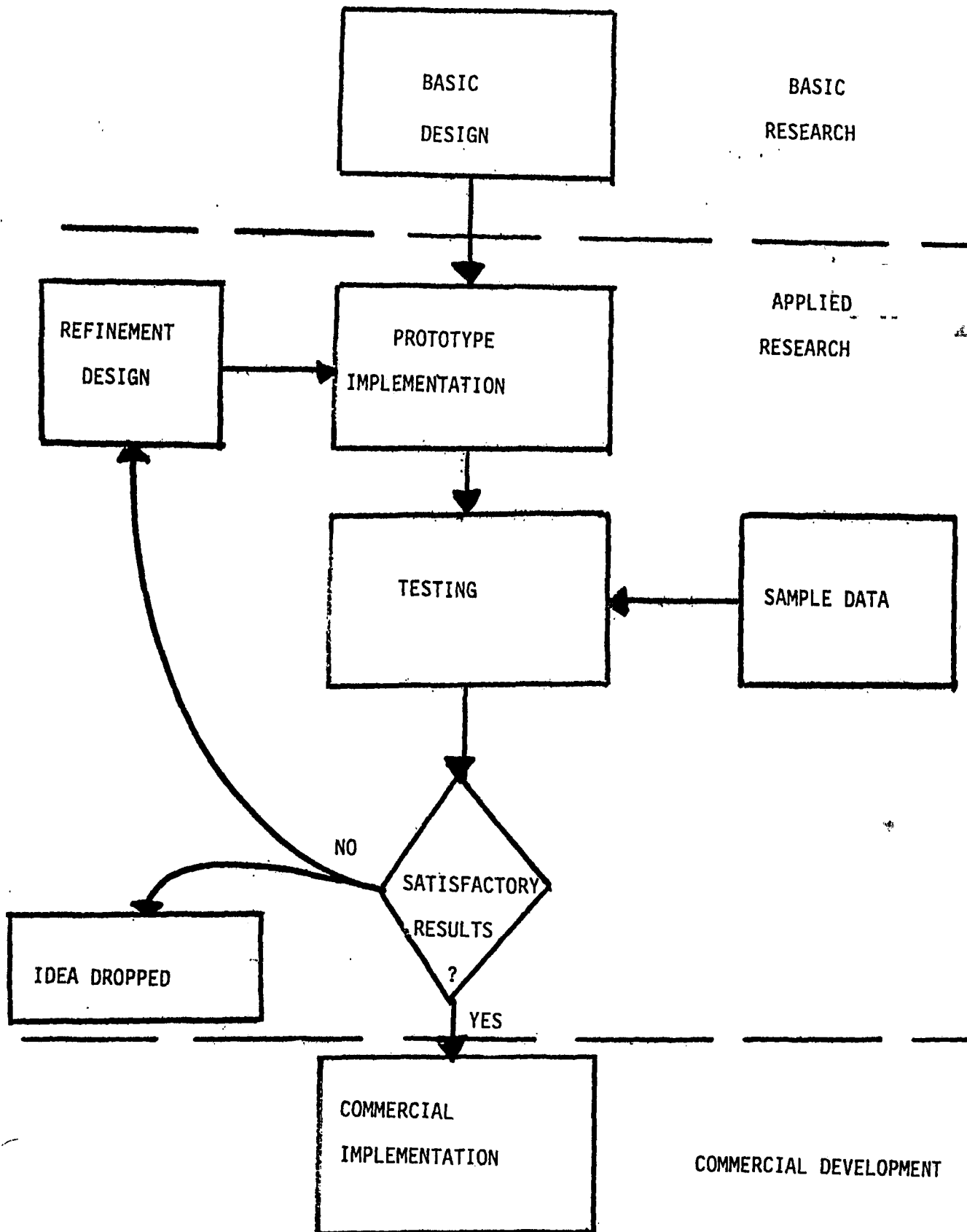
In order to address the problems in very-large-database-system research, it is necessary to revise our definition of "basic research"

as shown in Figure 1. Largely for economic reasons, the new definition must include the production of an initial prototype and the introduction of a comprehensive modeling effort. The modeling effort would parallel the development of a prototype using data produced by the prototype for input and/or validation. The modeling effort must include, besides the database software itself, comprehensive modeling of the hardware environment and comprehensive modeling of the application area. The latter would include a detailed model of the sample database and the associated access patterns as they are discovered. This process is shown in Figure 2.

Upon completion of the initial development and testing of system capabilities using the "toy" database, the detailed model can be used to extrapolate the results to predict performance when operating on a large database with higher dataflow and access rates. If the results are encouraging, then the research can progress to the applied research stage where a prototype can be implemented using a fully developed database. This is depicted in Figure 3.

3.0 Concluding Remarks.

A revised definition of "basic database research" is necessary because applied research is generally too costly using the traditional definition of basic research that precludes implementation of prototypes. The revised definition incorporates implementation of elementary research-oriented prototypes and comprehensive modeling of the database system, the hardware environment and the application areas. Using this approach, some predictions can be made about database performance in a very large database environment without committing such large implementation resources.



THE TRADITIONAL APPROACH

FIGURE 1

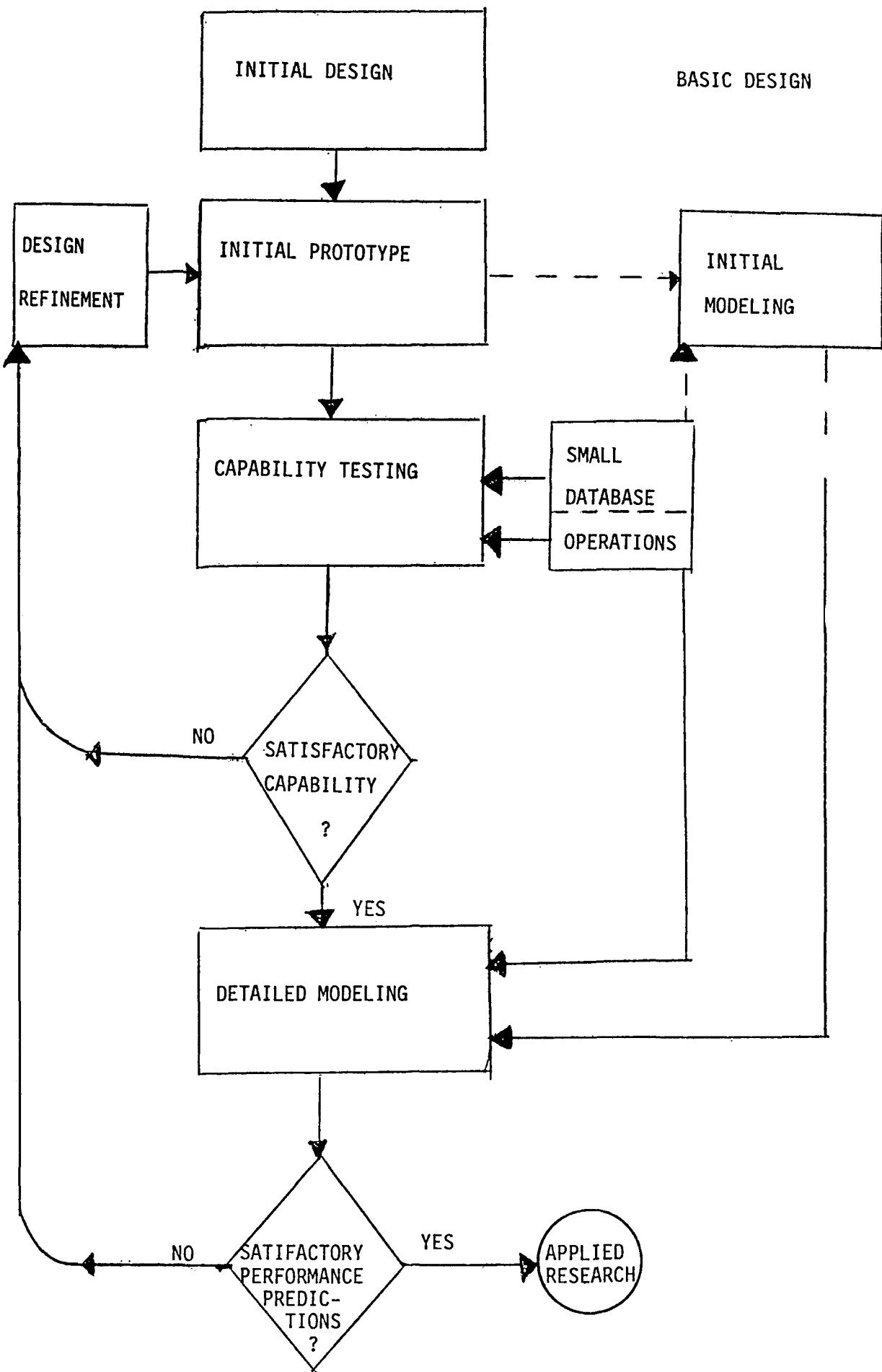
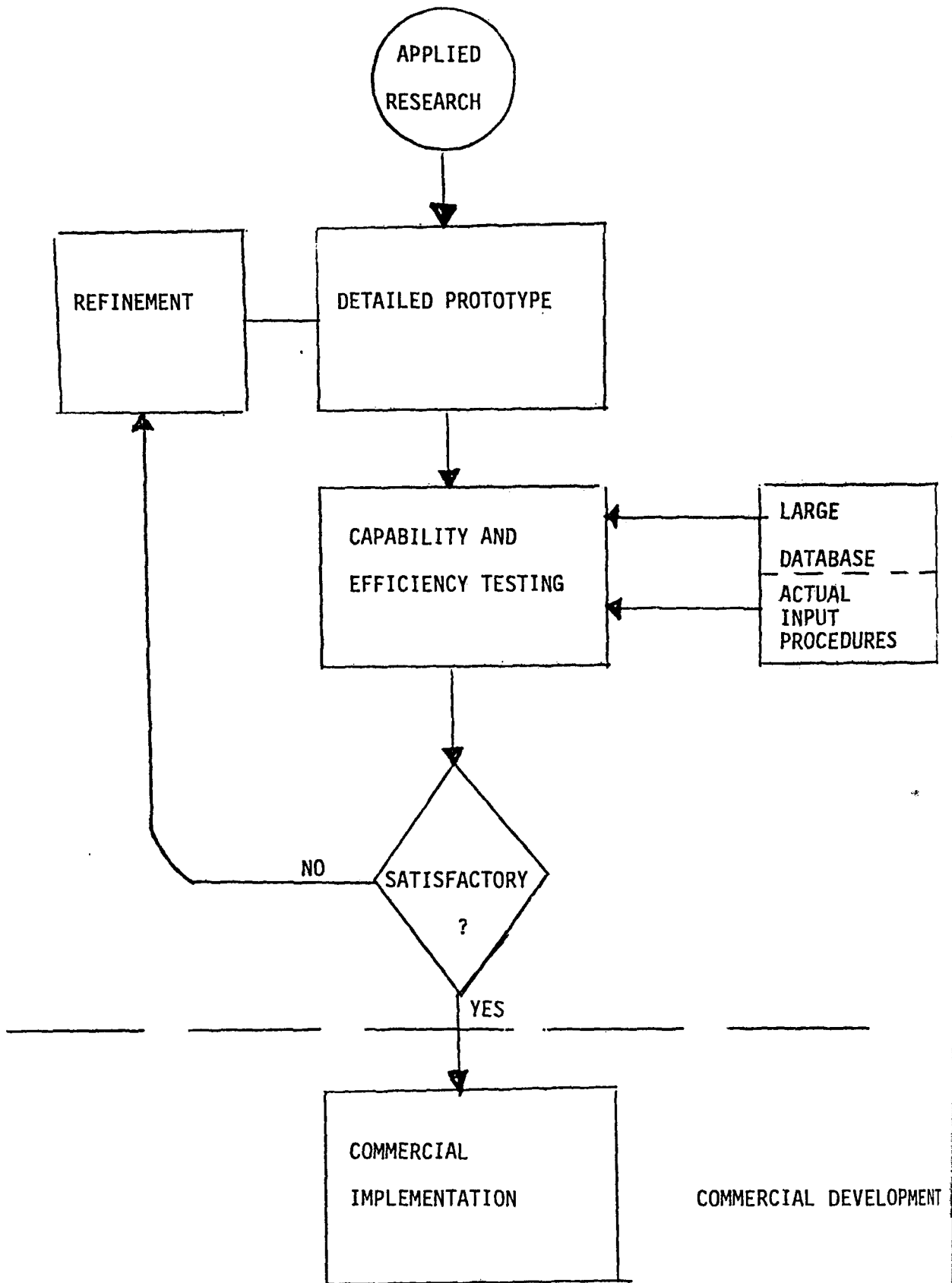


FIGURE 2

NEW DATABASE RESEARCH

Library of Washington Libraries



APPLIED RESEARCH

FIGURE 3