UNCLASSIFIED

AD NUMBER
ADB028228
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies only; Proprietary Information; 31 MAY 1978. Other requests shall be referred to Army Engineer Topographic Laboratory, Fort Belvoir, VA 22060.**
AUTHORITY
usaetl ltr, 12 mar 1979

THIS PAGE IS UNCLASSIFIED

AD B028228

LEVEL II

Report ETL-0144

A UNIFIED APPROACH TO MAPPING, CHARTING AND GEODESY (MC&G) DATA BASE STRUCTURE DESIGN

William K. Sharpley
John F. Leiserson
The Analytic Sciences Corporation
6 Jacob Way
Reading, Massachusetts 01867

Allan H. Schmidt
Laboratory for Computer Graphics and Spatial Analysis
Graduate School of Design
Harvard University
Cambridge, Massachusetts 02138

31 May 1978

Final Report

Distribution limited to
U.S. Government Agencies only;
Proprietary Information: 31 May 78
Other Requests for this
Document must be referred to
Commander and Director
U.S. Army Engineer Topographic Laboratories
Fort Belvoir, Virginia 22060

Prepared for
U.S. ARMY ENGINEER TOPOGRAPHIC LABORATORIES
Fort Belvoir, Virginia 22060

NOTICE

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such product.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ETL-0144
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): A UNIFIED APPROACH TO MAPPING, CHARTING, AND GEODESY (MC&G) DATA BASE STRUCTURE DESIGN.
5. TYPE OF REPORT & PERIOD COVERED: Final Report, 30 Oct 1977 - 31 May 1978
6. PERFORMING ORG. REPORT NUMBER: TR-1101-1
7. AUTHOR(s): William K. Sharpley, John F. Leiserson, Allan H. Schmidt, Harvard University
8. CONTRACT OR GRANT NUMBER(s): DAAK70-77-C-0265
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The Analytic Sciences Corporation, 6 Jacob Way, Reading, Massachusetts 01867
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: CLIN 0001 & 0002
11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA 22060, Code ETS-CS-A
12. REPORT DATE: 31 May 1978
13. NUMBER OF PAGES: 177
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): Unclassified; Appendix C, published separately, classified Top Secret.
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: IMPDET

16. DISTRIBUTION STATEMENT (of this Report)

Distribution Limited to U.S. Government Agencies only; Proprietary Information: 31 May 1978: Other Requests for this Document must be Referred to Commander and Director, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia 22060.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

MC&G Archives, MC&G Data Base Hierarchy, Image Data Bases, Image Archives, Topological Data Bases, Data Base Design, MC&G Data Bases.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is the final report for an investigation of a possible unified approach to Mapping, Charting and Geodesy (MC&G) data base structure design. The purposes of this investigation were to analyze the implications of various image archive structures in support of MC&G production and to demonstrate the potential value of recent results in Topological Data Base structures in MC&G applications. This report presents the results of the investigation. It includes analysis of the merits and deficien-

cont

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

78 07 06 024

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 565

cies of the various approaches and the results of an experimental application of Topological Data Base Structuring principles in an example MC&G data base processing problem.

SUMMARY

This document is the final report for an investigation of a possible unified approach to Mapping, Charting and Geodesy (MC&G) data base structure design, under Department of the Army Contract DAAK-70-77-C-0265. The purposes of this investigation were twofold: (1) to analyze the implications of various image archive structures in support of MC&G production, and (2) to demonstrate the potential value of recent results in Topological Data Base (TDB) structures in MC&G applications. The latter objective was accomplished in the context of a demonstration of the special capabilities supported by TDB structures. It was designed to serve as a proof-of-principle demonstration, which could lead to more intensive investigation of the application of TDB structures in concert with MC&G data bases (imagery, elevation matrices, feature files, etc.) to help resolve broad classes of problems.

This study report presents the results of the preliminary investigation. It includes analysis of the merits and deficiencies of the various approaches, and provides recommendations for further research and for demonstration of the applicability of those approaches to existing digital MC&G processes.

PREFACE

This report was produced under contract DAAK70-77-C-0265. The report was prepared for the U.S. Army Engineer Topographic Laboratories (ETL), Ft. Belvoir, Virginia 22060. The Contracting Officer's Representative was Mr. Carl S. Huzzen.

This report was prepared by William K. Sharpley and John F. Leiserson of The Analytic Sciences Corporation (TASC) and Allan H. Schmidt of Harvard University's Laboratory for Computer Graphics and Spatial Analysis (LCG), with significant contributions by the following:

J.L. Mannos (TASC)

N. Chrisman (LCG)

J. Dougenik (LCG)

S. Morehouse (LCG)

D. White (LCG)

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF TABLES

1. INTRODUCTION

1.1 BACKGROUND

The Analytic Sciences Corporation (TASC), for the past several years, has been investigating a number of applications of modern digital computer technology to Mapping, Charting and Geodesy (MC&G) processes, under the auspices of the Defense Mapping Agency (DMA). As an outcome of these investigations, the central nature and importance of digital data bases and the data base management operations, in both the near-term and the future evolution of DMA capabilities, have been identified. Foreseeable MC&G technology changes will impact both a demand for digital processing flexibility and a requirement for improved digital processing capacity. These technology changes will include:

- New, all-digital sources of MC&G data
- New and improved digital MC&G production approaches
- New user/product profiles, with emphasis on user processing of DMA-produced digital data
- Major dependence on computer-accessible data bases, to meet shorter data dissemination deadlines
- Increased demand for user-defined information products, possibly expressed via interactive user queries, as opposed to DMA-defined map and chart products.

In the same time frame as the TASC DMA activities, Harvard University's Laboratory for Computer Graphics and

PRECEDING PAGE BLANK-NOT FILMED

Spatial Analysis (LCG) has been engaged in National Science Foundation (NSF)-sponsored research on the subject of Topological Data Base (TDB)* structures. This work is based on applications of topological principles to the problem of geographic data base structuring. It extends and complements earlier efforts by the U.S. Census Bureau and the U.S. Geological Survey. As part of this activity, LCG has developed a series of computer programs (the ODYSSEY system (Chapter 3)) which constitute significant advances in such applications.

The result obtained in the NSF-sponsored research indicate that TDB structuring concepts offer a strong potential for improving certain MC&G processes of interest to the DMA. The following processes were deemed to be particularly likely to benefit from the improved data processing efficiency:

- Cartographic data base creation from manuscript sources
- Cartographic data base storage/retrieval processing
- Geography-based, automated and semi-automated analysis of MC&G source data
- Implementation of a flexible, product-independent package of MC&G data bases and supporting applications software.

These factors were carefully considered in a series of discussions among TASC, LCG and DMA personnel in the spring and summer of 1976. It was concluded that TDB structuring principles could provide a basis for unifying and enhancing the

*Topological data bases are cartographic data bases characterized by the explicit inclusion of data describing the spatial relationships of adjacency and connectivity of the map features. Further definition is provided in Chapter 3 of this report.

DMA-wide utility of a significant number of cartographic data bases. It was also conjectured that joint use of digital planimetric data with data bases of other primary forms (e.g., raster-type imagery) could be facilitated by topological structuring. The study reported on herein was therefore conceived as a means of exploring that potential and experimentally assessing the performance of applied TDB concepts.

1.2 SCOPE OF STUDY

The purposes of this investigation were twofold: (1) to analyze the implications of various image archive structures in support of MC&G production, and (2) to demonstrate the potential value of recent research results in TDB structures in MC&G applications. The first objective was conceived as a theoretical examination of such questions as whether or not an image archive is fundamentally required for MC&G operations, and if so, at what level in the MC&G data base hierarchy it should be established.

The second objective was conceived to be an experimental examination of the special capabilities supported by TDB structures. This part of the study was designed to be a proof-of-principle demonstration, at ETL, which could lead to more intensive investigation of the application of TDB structures in concert with more common MC&G data bases (e.g., imagery, elevation matrices, etc.) to help resolve broad classes of problems.

These two subinvestigations, although decoupled in the approach taken and specific issues addressed, are <u>compatible in the overall context of identifying a set of unifying principles for developing an MC&G data base hierarchy</u>. An important secondary objective of the study was to illustrate this compatibility, by showing that a natural hierarchy of MC&G data bases could be established which would efficiently support MC&G production with data from all levels of abstraction.

1.3 RELATION TO OTHER ETL ACTIVITIES

The study activities described herein complement major ongoing ETL investigations in the area of image processing, by (1) outlining guidelines for structuring the automated processing of imagery to provide a suitable MC&G image archive, and (2) developing a context and background for further work in the area of information measures for MC&G data base applications. The TDB structuring applications investigation is directly aimed at broadening the perspective of ETL's automated cartography research. It has particularly important implications for automating operations on fundamental MC&G data bases (i.e., point, line and polygonal (areal) feature data) to develop output cartographic files for map product generation.

Another major ETL activity to which this study is related is the development at ETL of the DMA Digital Data Handling System (DDHS) Test Bed, which will support experimental investigations of long-range digital MC&G production system requirements. In this context, the image archive investigation results and those of the TDB study have direct application in two ways:

- Specific application to guide the design of the Test Bed itself
- Providing guidance for experimental DDHS activities to be performed on the Test Bed.

This study is therefore seen to be fully consistent with both current and planned ETL activities in support of the DMA, and the conclusions and recommendations developed in the course of the study should be of continued interest and utility in future MC&G-related work at ETL.

1.4 REPORT OVERVIEW

The work performed and described herein has concentrated on two major activities which bear on the concepts discussed above. They are:

- Analysis of the utility of digital imagery data bases as a basis for MC&G production processing, including consideration of alternative archival storage approaches
- Demonstration of the potential of Topological Data Base structuring techniques for improving MC&G production processing, through development and experimental application of an ETL version of ODYSSEY software.

These two subjects are treated separately in Chapters 2 and 3, respectively, and the results obtained are evaluated in Chapter 4, in the context of the introductory discussion given above. Some supporting information on image compression techniques is provided in Appendix A, and the TDB demonstration test results obtained at Harvard University and at ETL are included as Appendix B. An analysis of some important source data attributes is provided in Appendix C (Classified), and Appendix D explains how to set up execution of the demonstration software on the ETL CDC 6400 computer.

2. IMAGE ARCHIVES AS A BASIS FOR MC&G PRODUCTION

2.1 MC&G SOURCE DATA CONSIDERATIONS

Present-day MC&G technology and all of its foreseeable projections depend overwhelmingly on imaging sources for fundamental data collection. These are currently represented by a wide range of film-based (aerial photography) sources, and on an emerging class of sources of direct digital images (e.g., LANDSAT). The rapid advancement of digital technology assures continuing and expanding MC&G coverage in the form of raw digital images. This both poses problems and offers challenges to the MC&G community.

The problems posed by the availability of raw imagery in digital form are primarily due to the scale of the digital processing resources required to handle a high-resolution digital image. For example, it may require 10^{10} bits, or more, to encode a full-resolution aerial photograph (Ref. 1). Storage, retrieval, communication and processing of these large data sets entails memory size, data channel bandwidth and digital computation speed requirements beyond the range of available general purpose computer capabilities. The extent of these problems has been well demonstrated in the literature (e.g., Refs. 1, 2, 3) and the application of special purpose hardware to solve the computational problems has been the subject of much research and development (R&D) effort in recent years (e.g. Refs. 4, 5). On the other hand, the assembly and long-term (archival) storage of digitally-encoded aerial photographic images has become an issue only recently, as indicated in Ref. 6.

The significance of the digital image archiving issue derives from the preeminence of image-based operations in current MC&G production. For example, highly advanced skills in aerial photointerpretation, coupled with analog (optical) image processing capabilities, support almost all DMA products at the current time. The fact that the input photographic imagery and the output cartographic products are easily and naturally related in the cartographer's perception has made recourse to hard-copy imagery an organic part of the bulk of MC&G production operations. The perceived necessity for such recourse to imagery has therefore resulted in accumulation of massive archives of raw and processed imagery.

Consequently, current consideration of all-digital MC&G processing for the DMA must include the possibility that human operators will require convenient, repeated access to archived versions of imagery over a long image lifetime. This consideration will hold at least until a wide range of experimentation (e.g., Ref. 7) has demonstrated that relaxation of the image archive size and archival lifetime can be accomplished without loss in product accuracy and without compromise in the responsiveness of the DMA to its user community.

Continued R&D activity in the DMA centers and supporting research laboratories has made steady progress in the areas of automating image-based MC&G production processes, and improving associated human performance. The most promising of these improvements are made possible by the use of digital computers, operating on digitized versions of photographic images, in either whole-image form or in the form of on-line digitizer output streams. The availability of intrinsically-digital input data promises to make these processes even more useful and efficient, since the digitizing (graphic to digital conversion) step is not required. If the source data can be rapidly and completely utilized, and if it can be flexibly obtained as required to meet DMA production demands, the necessity for long-term maintenance of imagery in archives may be reduced. The next section addresses some of these considerations in the context of DMA MC&G product line trends.

2.2 TRENDS IN MC&G PRODUCT-USER PROFILES

An important result of the proliferation of digital computers (and associated graphics peripherals) has been the increasing awareness, on the part of the MC&G user community, of the advantages of user processing of digitized MC&G data files, as opposed to printed maps. At a recent symposium, representatives of this international community of users and producers of geography-based information were able to demonstrate convincingly the broad range of analytical and informational processes that could be supported by a complete and comprehensive MC&G data base (Ref. 8). The DMA has observed similar interest and growth of capabilities in both its civilian and its DoD user community. Senior DMA management has therefore recognized an impending need for flexibility of output product types, formats and media, and for improved communicability and generality of these data products. For certain strategic and tactical applications, improved DMA responsiveness is also required, in terms of both the speed with which existing products can be generated or updated, and the speed with which special-purpose compilations of data can be provided.

These shifts primarily reflect the increasing user understanding of their need for information products, as opposed to only map products. This is not to say that map sheets are an obsolete means of packaging information. Rather, it is recognized that the familiar and powerful cartographic modes of information display must remain available to users. The evolution in DMA products will thus be aimed at:

- Improving the range and selection of data which can be displayed to the user
- Supporting inquiry-based, user-interactive display systems for cartographic data elements, in addition to standard static maps and charts
- Providing digital geographically-based information sets suitable for computerized applications (such as navigation of unmanned vehicles)
- Supporting an increased demand for limited-production, special-purpose MC&G products.

In addition, on-going internal evolution in DMA production techniques and satisfaction of the currently planned production requirements (Ref. 9) will lead to important changes in the DMA mode of operation. These changes will become dominant in the mid-to-late 1980's, and will include:

- A major shift toward product (and data base) revision, rather than original compilation
- A general upgrade and improvement in DMA digital data processing capabilities such that unified data bases can support agency-wide access to source data
- Integration and coordination of production systems to avoid proliferation of special-purpose, non-interchangeable data bases and production resources.

An important goal to be attained in accomplishing these changes is the planning and development of a generalized, coordinated set of MC&G data bases for the DMA. Not all such data necessarily will be digital, but efficient operation will demand digital computer data base management for the analog (film) repository. The following discussion of a conceptual MC&G data base hierarchy is provided to establish a framework for the investigations covered in this report.

2.3 CONCEPTUAL MC&G DATA BASE HIERARCHY

The concept of a "data base" is much broader in the context of MC&G operations than in most digital computer-based systems.* In this context, a data base may consist of any collection of MC&G information; a stereopair of film transparencies comprises an output data base, for example. Similarly, a collection of ship's navigation worksheets, with depth soundings annotated by hand, may constitute an input data base for a bathymetric chart. However, in the context of the study discussed herein, the idea of a data base is more narrowly defined, and is taken to mean digital-computer-compatible assemblages of MC&G data. (The single exception occurs in Section 2.5.2 where some of the "pros" and "cons" of storing digital images in analog (film) media are discussed.)

Furthermore, imagery-based operations are the point of concentration for the required study of image archive implementation alternatives, as is made clear in the remainder of this chapter. The investigation of topological data base structures applicability (Chapter 3) is not constrained by this concentration on imagery as source data, because it is keyed to examining efficient and innovative ways to analyze and display

*In this and the succeeding discussion it is assumed that the reader is familiar with terms and concepts related to modern MC&G operations, including the associated digital processing terminology.

derived MC&G data. That is, the postulated TDB applications package is located at some level of unified, largely product-independent data bases whose content is derived from all sources.

Figure 2.3-1 illustrates a hierarchical data base concept which provides a common basis for these two rather disparate investigations. In the figure, MC&G data for production applications are considered to exist in several successively more refined forms. Starting at the bottom of the figure, the levels of refinement are:

- Raw digital imagery (source data), assumed to undergo some minimal quality screening and then to remain in archival storage until needed
- Normalized digital imagery; i.e., photogrammetrically (and otherwise) corrected imagery suitable for MC&G feature extraction
- Unified data bases, which comprise extracted data and data structures appropriate to the data type, and which serve as a joint product-independent data source for DMA product generation
- Digitized product files, which may either directly constitute products, or which directly support automatic cartography, etc., for generating specific map and chart products.

Digital data processing capabilities required to transform data from one level of the hierarchy to the next are indicated in the figure. Also shown are functional interactions of digital and manual operations required to manage/maintain the hierarchy and data processing resources. Figure 2.3-2 offers an overview of the operation, in which it is indicated that the key attributes of the hierarchical concept are (1) control of data state transitions and (2) promoting processing efficiency by adding data base structure.

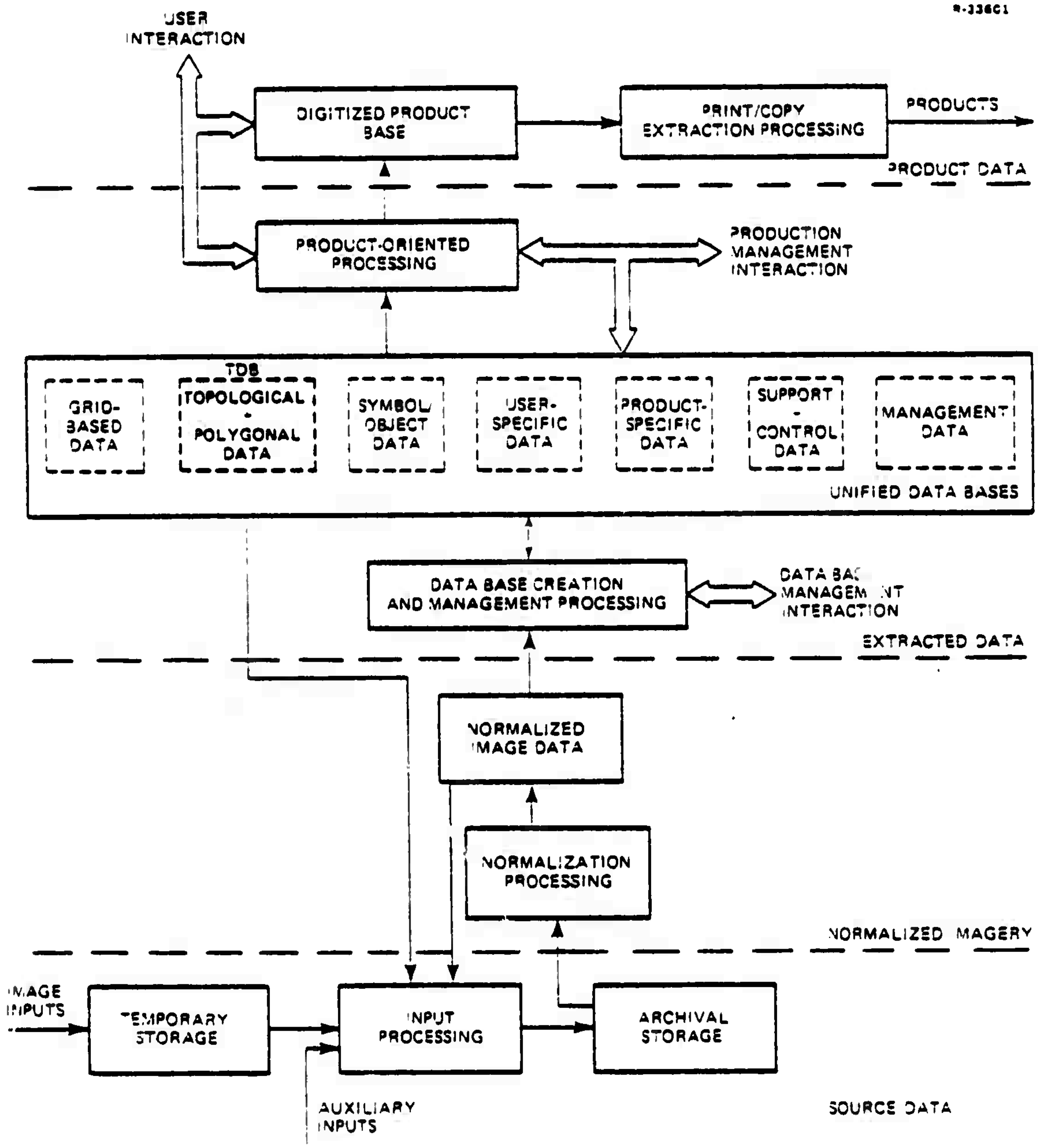


Figure 2.3-1 Conceptual MC&G Data Base Hierarchy

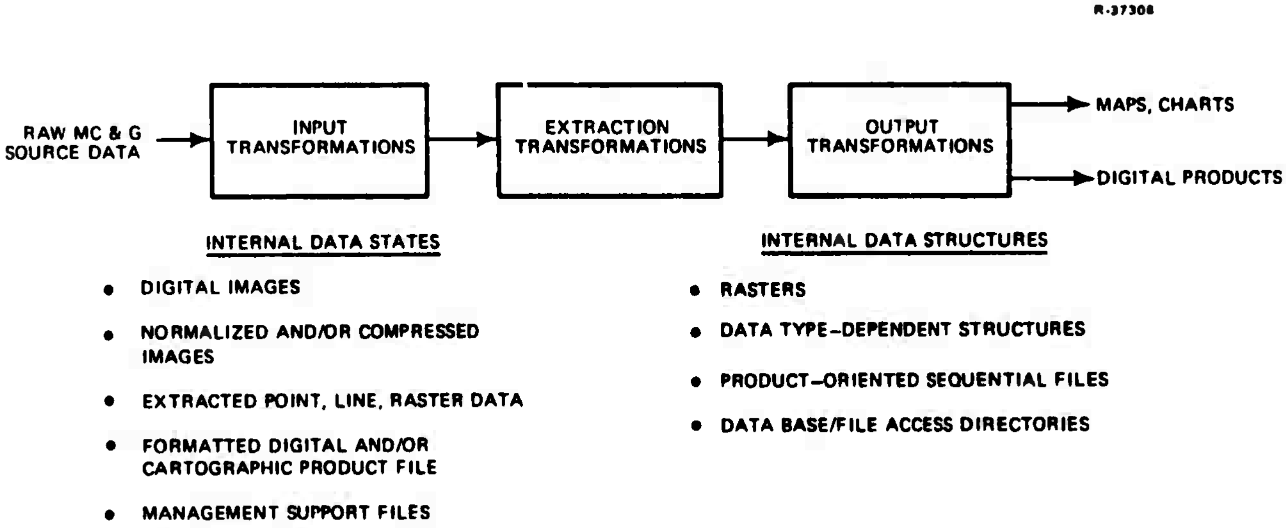


Figure 2.3-2 Data Base Hierarchy Functional Overview

2.3.1 Archive Considerations

Early in the course of this study, it became apparent that the fundamental definition of what should constitute an "archive" should not be based only on data type or data density considerations. A better approach is to consider, in addition, the current heavy dependence on raw imagery in MC&G production activities. This implies that future operations may exhibit similar dependence.

The following definition is therefore made, in preparation for later discussion: the MC&G archive is the lowest (i.e., the least transformed) level of data hierarchy which contains all of the data required for generating any given MC&G product at any given point in time. This definition allows consideration of the implications of not maintaining a raw image archive, for example, but instead developing rapid transformations for generating a normalized imagery data base.

2.3.2 Data Base Creation/Management Processing

An important attribute of the data base hierarchy shown in Fig. 2.3-1 is its specific provision for a class of unified, naturally structured data bases to support both flexible interrogation by users and product-specific processing. "Unified" in this context means that the data bases:

- Are referenced to common datum sets
- Are standardized, for their data type, in all production activities
- Include data base structuring information, as appropriate to their data type, to support query-based data retrieval and to minimize applications software redundancy
- Support intercenter data/software exchange
- Support the raw image processing function.

All but one of the data bases represented at the unified data base level in Fig. 2.3-1 are exemplified by existing DMA data bases, with the additional unifying constraints discussed above. The exception is the Topological Data Base, labelled "TDB", which is structured according to the concepts discussed in Chapter 3. These concepts are expected to aid significantly in the unification and interrelation of the various MC&G data bases, with concomitant data processing efficiency improvement.

The main tasks of the data base creation and management function are thus to support the data extraction process, provide the necessary data base structure and structural interrelations, and support the maintenance of data base currency and accuracy. This function also supports production management in terms of assuring that necessary data is available when a project order comes in, and providing the requisite production control information for DMA operators.

2.3.3 Production Processing

The generation of specific digital and chart-form DMA products, based on the unified MC&G data base level, will be accomplished by product-oriented software under control of a production management system. The projected unification of the data bases is expected to foster a much more comprehensive sharing and commonality of these applications programs.

Furthermore, the intrinsic data base structuring information should provide a basis for highly interactive, relational, query/response interactions, to benefit both DMA production management and DMA users. As a result, the development of special-purpose, limited-scope ad hoc information products will be expedited. Requirements for new forms of high-volume DMA product lines will also be inexpensively piloted and tested, offering an important improvement in DMA's responsiveness to new demands by the users.

2.4 DESIRABILITY OF AN IMAGE ARCHIVE

The preeminence of imagery in MC&G production processes, as discussed earlier, will continue as more imaging sensors are developed for these applications. For example, Synthetic Aperture Radar (SAR) is beginning to provide useful images of terrain in tactical low-altitude applications with sufficient resolution for many MC&G applications, and with near immunity from weather. Digital data processing is the key to successful SAR operation, and to most of the other new sensor designs (e.g., LANDSAT Thematic Mapper). Consequently the outputs of these sensors are generally available in digital form.

The availability of imagery in digital format provides an excellent opportunity application of digital image processing research results. As stated in Ref. 14, the application of digital techniques provides more photogrammetric accuracy and greater process flexibility than do analog and optical operations. In fact, a survey of the literature reveals that much of the newer analog and optical photogrammetric equipment actually consists of analog/optical-to-digital transducers, followed by digital processing to obtain the desired outputs from aerial photography (Ref. 10). Direct digital image input, when widely available, will make obsolete the analog/optical part of many of these devices.

The map update process shown in Fig. 2.4-1 is typical of current MC&G image utilization. The central position of imagery archives in the process is apparent, as is the fact that multiple access to the archive and multiple examination of each image are implicit in the processing. The capability to display raw or processed imagery to the operators, and their resulting capability to screen, evaluate, select and operate on the imagery, are key factors in maintaining efficiency in current operations. Any application of digital imagery must offer at least equivalent capabilities and operator convenience, and lower operating cost, if it is to be an acceptable operating mode for the DMA. Provision of that kind of digital capability requires a great deal of storage, digital processing power and digital communications power. These are potentially expensive commodities, and it is extremely important to realistically balance the benefits of a largely digital approach against the costs entailed.

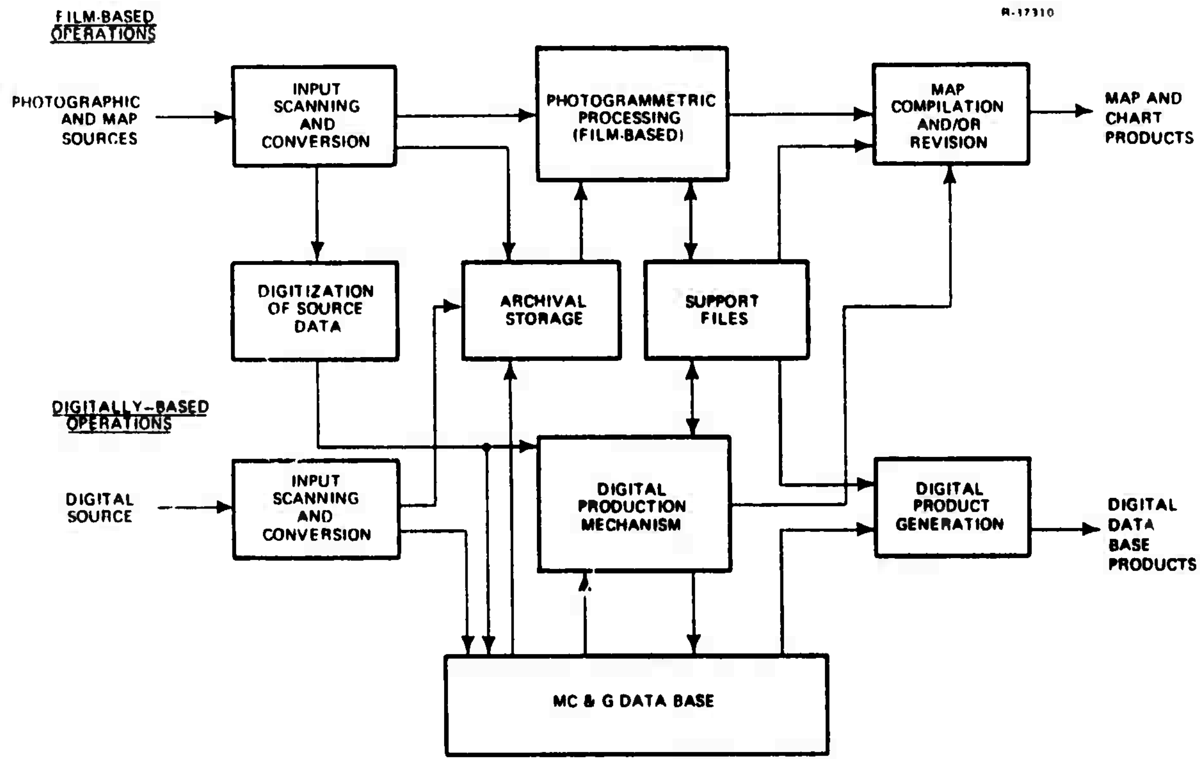


Figure 2.4-1 Overview of Current MC&G Compilation and Revision Processes

Modern digital technology has made its greatest recent strides in logic circuit design and packaging. Great advances in digital communication have also been made, to the point where extremely high-bandwidth links (up to 1,000 megabits/sec) can be purchased off the shelf. However, advances in Very Large Mass Storage Systems (VLMSS), of capacity greater than 10^{12} bits, have only recently become available (Ref. 6) and the ultimate price levels have not yet been approached. Thus the provision of a digital archive, analogous to the DMA's present film-based archive, is of major interest because of its cost/risk implications. This problem is compounded by the fact that a certain degree of redundancy of coverage is required for MC&G purposes, including:

- Temporal redundancy, accomodating seasonal variations in shorelines, foliage cover, etc.; this results in approximately three-season coverage of most areas of interest
- Obscuration redundancy, accomodating areas where climatic conditions do not permit cloud-free or haze-free photographs to be easily obtained
- Culture change redundancy, required for maintenance of accurate small-scale maps of urban (or developing) areas and locations of isolated installations
- Scale redundancy, which may be required for convenience in processing data for small-scale *vs* large-scale maps.

This redundancy affects archive size, update frequency and image obsolescence rates. It must be factored into archive cost tradeoffs for any hierarchy scheme based on imagery archives.

Concepts of multi-spectral redundancy also must be allowed for in future MC&G applications (Ref. 11). Current activities related to LANDSAT data exploitation indicate a high probability that that data may support such processes as automatic surface classification, which could apply to many current or planned DMA product lines (e.g., Digital Land Mass Simulation (DLMS) Culture Files). Because of the difficulty in predicting the total coverage subject to each of these factors, the effect of redundancy is lumped in with other uncertainties in archive size in the parametric analysis to follow.

The concept of a fundamental archive consisting of images (i.e., rasters of radiometric data) is an ingrained part of DMA operations, and continued availability of such an archival resource must be given extra weight in devising a data base hierarchy for future applications. In other words, there must be a very strong advantage achievable by some other recommended approach in order for it to be acceptable to MC&G production personnel. This comment applies, irrespective of storage medium questions (e.g., film vs magnetic tape), because of the following considerations:

- The current DMA MC&G product list cannot cover all uncertainties relative to the feature, surface classification and terrain information which will be important in the future to DoD users; some recourse to imagery will almost certainly be required
- New image analysis techniques, such as are now being actively pursued in research laboratories, may automatically reveal important residual information in previously processed images which has not been perceived by the photointerpreter
- As an extension of the previous remark, it is also possible that interactive or semi-automated image interpretation may make such re-examination of previously processed image advisable, at any particular geographic location
- Analyses of the relative costs of collection vs storage for imagery and other large data sets, although sketchy, have indicated that the total cost to the Government is less if data is stored and re-used rather than collected each time it is needed (e.g., Ref. 12).

It is to be noted that the above comments are not strongly impacted by questions of image storage medium selection, e.g., film vs magnetic tape. The specific equipment provided for displaying the imagery to the photointerpreter/cartographer is also not of key importance. What is important is that the availability of final recourse to imagery, in a more or less limited manner, should probably be provided by any MC&G data archiving scheme designed for the DMA.

2.5 IMAGERY DATA BASE PRE-PROCESSING CONSIDERATIONS

An important issue to be considered in developing an image archive is the degree of pre-processing required to be performed on images, prior to their acceptance into the archive. This pre-processing must satisfy the requirement for bringing the raw imagery up to the level of abstraction appropriate to the hierarchical level of the archive; in an important sense, the pre-processing implements the archive level definition. Table 2.5-1 summarizes the rationale for this point of view, for the three image archive levels implied in Fig. 2.3-1; i.e., a raw image archive level and two subcases of a normalized image archive, extending the concept slightly beyond the discussion in Section 2.3. Before addressing these concepts, the important pre-processing concepts of an Information Measure must be discussed.

2.5.1 Information Measure and Data Utility

Purpose of an Information Measure - Incorporation of any method of storing data into an MC&G production scheme will carry implications for the procedures used in generating the cartographic products. The primary data storage - the archive - is designed and configured to support a specific production goal and a specific method of organizing the production task. Any of the other possible levels and purposes for storage of MC&G data have similar implications, regardless of the actual storage mechanism.

In order to use the stored data efficiently and to make the best use of the storage space and the production facility, it is necessary to evaluate the content of a given image. The form of this evaluation varies with the form of the stored data and with the intended purpose. In a completely

TABLE 2.5-1

PRE-PROCESSING CONSIDERATIONS FOR POSSIBLE IMAGE ARCHIVE LEVELS

T-2274

IMAGE ARCHIVE LEVEL	INPUT PROCESSING STEPS	CANDIDATE ARCHIVE ACCEPTANCE CRITERIA (INFORMATION MEASURE)	LEVEL OF UTILITY
RAW IMAGES	• Image Quality Screening • Cataloging • Organizing • Compression (Limited)	Based on Descriptors • Image Obtained by Request, OR • Fills Coverage Gap, OR • Provides Temporal Update	Uncorrected: Non-Metric Uses Only, Any MC&G Feature Extraction Requires Preliminary Processing.
NORMALIZED IMAGES (Case 1)	Above, Plus: • Compression (Adaptive) • Geometric Correction • Radiometric Correction • Routine Enhancement • Rectification, Registration	Based on Image Above, Plus: • Changes From Previous Images, Especially New Features • Improved Image Quality (Subjective) • Improved Residuals in Correlation, Rectification Processing, etc	All Applications, But Generally Requires Image-Specific Processing to be Used With Other Data for Feature Extraction
NORMALIZED IMAGES (Case 2)	Above, Plus: • Standard Enhancement Processing • Image-Specific Enhancement • Resampling Onto Common Grid • Mosaicking/ Merging	Above, Plus: • Contains All Old Info in Covered Area • Improved S/N Ratio in Covered Area	Supports Direct Application of Auto/Semi-Auto Feature Extraction

manual, non-automated installation, all evaluation of images and decisions about processing are operator activities, either fully specified and conscious or implicit in other tasks. In a system which handles only some form of extracted MC&G data, an implied decision has been made about the utility of the elements of the image, in that only certain quantities are stored and used. The evaluation of the utility of the extracted data becomes a simpler task, however, due to the additional structure imposed by the extraction process.

The important issues for a measure for information and utility of imagery in any contemplated MC&G production scheme are the simplification of the processing stream and the elimination of labor-intensive operator interaction. Since there is no single utility measure which will perform all such tasks (short of a full mathematical theory of picture understanding), specific tools for different tasks are needed. Such tools would answer questions like these:

- When has all the useful data been extracted from the image?
- When is the image redundant?
- When is the image different from other images which have the same nominal subject?
- Are differences in a pair of images due to flaws in the image or to content?
- What is the quality of the image?
- Has a processing task on the image worked correctly?
- Where are the gaps or inconsistencies in the data?
- What is the density of useful or new information in the image (given a set of data types and an existing image file)?

The tools for answering such questions need to be considered as part of the data structuring and storage design tasks, since the problems interact so strongly.

<u>Problems and Issues in Developing an Information Measure</u> - As a first step in developing a tool for appraising the information content of an image, the aspects of the image which are of interest must be specified. This is the beginning of a formal description of the important attributes of the

image. In all likelihood, the same tool will not suffice for answering all the questions listed above. Part of the problem lies in defining the criteria that are appropriate for each type of appraisal; such specific tools do not profess to be general descriptors of the meaning in an image, but only answers to localized questions about certain types of information. The questions which a human operator would ask - specific searches or evaluations - need to be formulated narrowly, in a more mechanical manner, for automatic or semi-automatic implementation. Structural aids - like the topological encoding which is one of the subjects of this study - are the key to this task. Some of the applications of such tools are shown in Fig. 2.5-1.

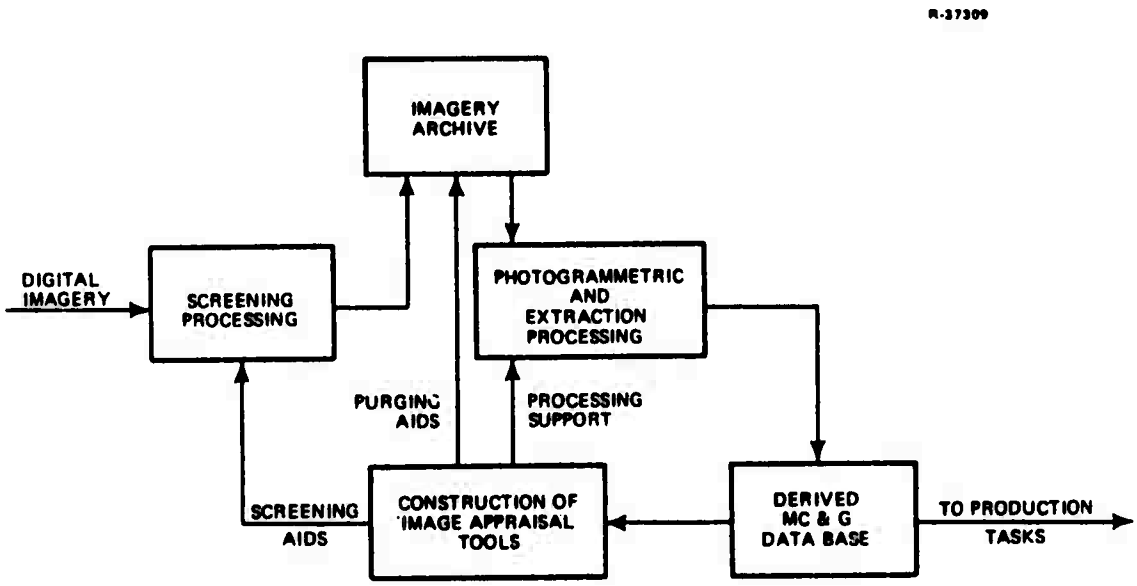


Figure 2.5-1 Potential Use of Image Utility Measures in MC&G Production

Application of Structural Analysis to MC&G Data - There are three separate ways to which structural approaches to cartography can be applied, not all of which are equally relevant to the DMA MC&G task:

1) As a description of the inherent properties of the object - the concrete level

2) As a description of the inherent properties of the map - the abstract level

3) As a cognitive model of the application of the map and the understanding of the user.

The demonstration of topological data structures application (Chapter 3) is an application of the second type, since selected features of interest are encoded and abstracted in a completely map-oriented format. The processing tasks (which demonstrate the utility of the topological constructs in responding to data-retrieval questions about the mapped area) use a model of a map; i.e., an encoding of it.

An initial effort to broaden the use of the set of topological constructs would combine the topological descriptors with another data format or another processing stage: e.g., "using an image of the same area and the graphical floodplain coverage description, determine if floodwaters lie within the 100-year floodplain." In this manner derived data is used to support an image processing task, aiding the evaluation of digital imagery before data extraction takes place.

The important future requirement for an aid to MC&G data handling is the need for a similar tool which can be applied directly to the digital image. The structural description of the image content would contain the topological data (as in this study) and the attributes required for the various DMA products. Automation of this task will require the implementation of pattern recognition and change detection techniques, as discussed in the DDHS General Test Plan (Ref. 7).

Symbolic, derived representations of digital images provide greater power for the analysis and manipulation necessary to judge their information content. Reference 7 describes a change detection task which uses this approach on various sorts of images, although without including any topological data in the derived data record. Processing using such symbolic data to support change detection has additional power to cope with new or missing objects which might cause difficulties for other methods.

The following subsections return to consideration of the implications of Table 2.5-1 and Fig. 2.3-1 in order to discuss the concepts of digital archival data bases consisting of raw images and two subcases of normalized images.

2.5.2 Raw Image Archive

Maintaining a raw image archive has two purposes:

- No complex processing is applied to any image unless and until the image is determined to be potentially useful for a specific MC&G project; this decision is assumed to be based on the catalog (descriptor) information obtained during input processing
- Very little degradation (or none at all) is produced in the archive contents, so that no future use of the images, which the original image data would have supported, is precluded by the archiving process.

The fundamental information measure required to assess new data for acceptance into an archive at this level is simple to express, as in Table 2.5-1, but somewhat subjective to apply. This implies that an intensive man-machine interaction is required, which cost must be factored into the operating cost of the system.

It is to be noted that the raw image archive concept is the basis for the conceptual DDHS architecture (Ref. 6). In the reference, costs for archival storage are shown to be controllable if a purging regimen is established and an erasable medium of storage is provided. When either of these conditions is not true, the archive system costs will be dominated by medium costs at some point, determined by the technology employed. Figure 2.5-2 illustrates the sensitivity of archive system costs to the technology employed, based on the observations that:

- Magnetic high-density tape costs approximately the same as high-resolution film, on an areal basis (0.67¢/in^2 TBM tape; 0.72¢/in^2 for 9" film at current prices)
- The key to low archive medium cost is increasing data packing density; optical and particle-beam technologies offer the best promise.

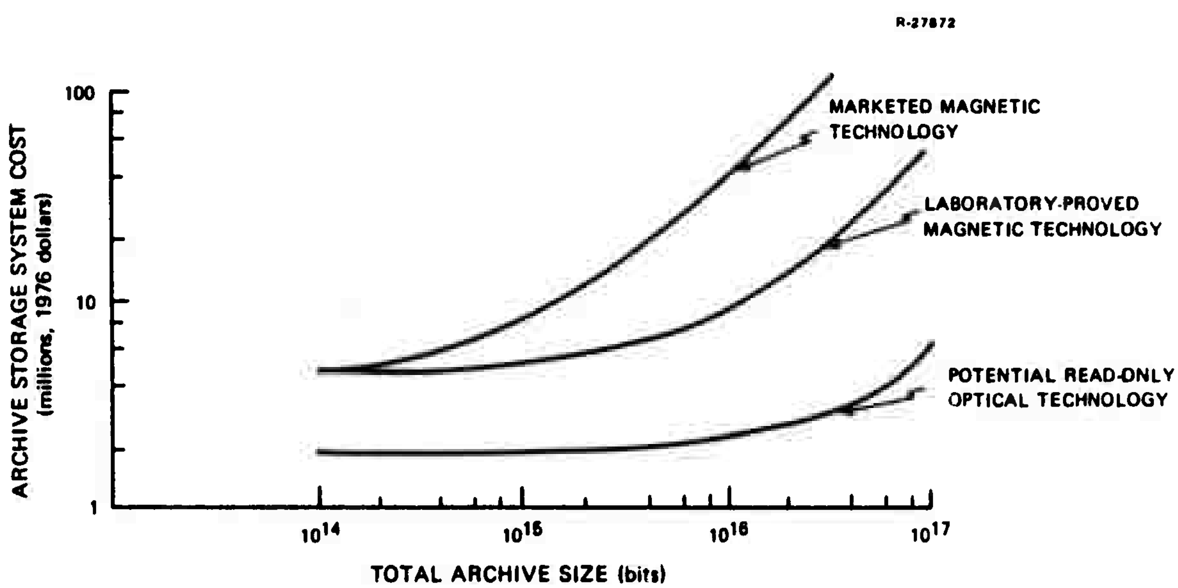


Figure 2.5-2 Approximate Raw Image Archive Cost Sensitivity

However, some important factors which do <u>not</u> show in Fig. 2.5-2 will become important in Section 2.5:

- Laboratory models of high-density magnetic tape units are currently working at densities comparable to effective (analog) film image data densities (3-5 x 10^7 bits/in^2)*
- A reasonable level at which to <u>purge</u> a raw data archive is approximately 5 x 10^{15} bits (Ref. 6)
- Anticipated rates of receipt of new data (Appendix C, CLASSIFIED) will bring non-reusable medium costs up to the level of advanced reusable medium costs in about 10 years.

Thus it can be concluded that: (1) a raw image archive scheme will be fairly expensive (approximately $5 million 1976 dollars); (2) a reusable archive medium should be employed; (3) a size limit should be imposed on such an archive, by purging.

2.5.3 Normalized Image Archive (Case 1)

The concept of a normalized image archive for MC&G purposes is based on the fact that almost all such uses of aerial photography require the following:

- The image must be restored to display the radiometry <u>as sensed</u>; i.e., all distortions and processing artifacts must be removed
- The image must be registered to a common earth-referenced MC&G datum

*Based on LANDSAT projections for Thematic Mapper scenes, using standard 9" x 11" film format.

- The image must be processed to display uniform contrast, haze reduction, etc. (standard image enhancement).

The Case 1 concept consists of performing the above processing on all images which pass the quality screening and needs-assessment screening steps, and then archiving the results. An archive is then formed which can directly support <u>almost all</u> MC&G production operations currently applicable. Further processing of such images, such as automatic or semi-automatic feature extraction, would be facilitated by the more consistent quality of these archive images.

In some important respects, however, this concept offers little improvement over the raw image archive. The economics of processing resulting when an image is used more than once, but processed only once, are undercut by the possibility that a given image may never be used at all. That is, the image may be superseded or made temporally obsolete before it is required for production, so that <u>both</u> the archive storage cost and the processing cost are wasted.

Such an archive also requires consideration of purging as a means of controlling growth and medium cost. However, at this level, due to the normalization of the input as well as the existing archive, some automatic or semiautomatic information measures can be conceived (Table 2.5-1). These may be applicable to help reject redundant input, or select purgeable archived images, to control the archive size. Without such aids the archive maintenance process would be identical to that required for a raw image archive.

It is important to note, however, that the proposed information measures included in Table 2.5-1 are currently being pursued only in a laboratory environment. Progress on these has been disappointingly slow, and much of the work is essentially stalled at the level of two or three years ago (e.g., Ref. 13). The key problem areas are:

- Discrimination of significant changes, from one image to another, against a background "noise" due to insignificant radiometry differences, shadows, etc.
- Difficulties encountered in translating laboratory model, small-scale algorithms to the large-scale processing demands of real images
- For archives, further difficulty in applying the measures when the input image overlaps several archived images.

Because of these problems it appears likely that application of the information measures listed in Table 2.5-1, or others to be devised, will be done in semi-automated processing with strong human interaction. It is to be noted that the trained cartographer/photointerpreter has a large repertoire of information measures available within his capabilities which are not well understood, but on which much current MC&G operations depend. Application of computer-based algorithms will therefore most likely be used to prompt the operator, to record decisions (with respect to features identified), and to provide advisory assessments of image quality.

Finally, it is important to state that images which have undergone normalization processing as described above are unavoidably degraded to a greater or lesser extent. The actual impact of the processing depends on the severity of the original distortions and look-angle obliquity which must be corrected. In general, the images cannot be restored to their original state, so that the normalized archive concept permanently loses some information content.

2.5.4 Normalized Image Archive (Case 2)

Carrying the idea of a normalized image archive one step further, it is possible to conceive of a fully mosaicked and integrated image archive. This would be a single raster-type data base of pixels, derived from many input images, which together constitute a composite image of the earth's surface area of interest to DMA. Such an archive would require significantly more input processing than would the Case 1 archive, but it would have a natural upper limit in size (and therefore storage cost) of the order of 2-3 x 10^{15} bits (Ref. 6). (This figure allows for necessary MC&G redundancy.)

The availability of such an archive would provide DMA cartographers with advantages quite similar to those expected to accure from the world-wide elevation data bases now being accumulated (Ref. 14). Principal among these are:

- Simplicity of coverage analysis, due to non-overlapping of images
- Ease of access to a desired area's coverage; indexing by geographical coordinates would suffice
- Ease of examination/evaluation of archive contents, since the digital archive is already mosaicked.

The eventual (i.e., complete) archive might even be maintained in orthophoto format, given that the Ref. 14 world-wide elevation data bank exists. However, for the first 10 years or so of its buildup, the development of this archive would be uneven. Source data from many input systems would have to be integrated, and the probable approach would be to maintain the data as a Case 1 archive until some carefully-determined threshold had been reached in a particular area, permitting the mosaicking process. The mosaicking process is well understood (Ref. 15), but it does require careful control of such parameters as sun angle and the seasonal relationship of adjacent images. These parameters would have to be considered in early input screening.

The update process implied for such an archive could be quite different from current practice. Simple pixel replacement for updating portions of the data base with portions of a new image would be most nearly analogous to current mosaicking practice. However, in areas devoid of culture or isolated man-made features a merging process akin to digital filtering might be applied. In this case a weighting process would be applied to the results of information measures similar to those indicated for Case 2 in Table 2.5-1. Where a new image's weighted "information" content relative to the existing mosaicked data base exceeded some threshold, the new image could (1) directly replace the affected pixels, or (2) be merged by an averaging process of some kind.

These normalized image data bases should be the subject of further research. The impact of a mosaicked data base on automatic feature extraction operations, for example, would be of vital interest. However, the possibility of automatically maintaining a complete and up-to-date composite image of the earth's surface is an interesting one. This is especially true since current archive buildup is tending toward a greater amount of redundant coverage (e.g., inadvertent image overlays) without providing the composite view required for mapping. If the very difficult problems alluded to can be solved, the mosaicked image archive could provide a truly product-independent, flexible basis for all known image-based MC&G operations.

2.5.5 Possible Hierarchical Situations for Image Data Bases

In summarizing the above discussion, it can be seen that MC&G data bases consisting of digitized aerial photographs may usefully occupy several of the hierarchical positions indicated in Fig. 2.3-1. Furthermore, any one of those levels could serve as the archival level. That is, a digitized image archive could be maintained at any level of abstraction from raw imagery through mosaicked orthophotos. Certain important tradeoffs must be made to choose the best level, if an image archive is desired.

- Average production utilization rates must be balanced against average input rates, to determine the probability of processing any given image
- The availability of ancillary information (e.g., elevation data) in conjunction with imagery must be assessed to determine the highest practical level of processing for archival purposes
- The effectiveness and cost of standardized normalizing processes must be assessed and compared to similar parameters for image-specific processing, to determine whether or not a normalized image archive is practical.

Most of these issues are currently expected to be addressed experimentally under DMA sponsorship (Ref. 7), because there is no large base of experience with digital MC&G production to draw upon. When the results of those experiments are available, the necessary cost-benefit studies can be performed. These study results will support determination of the best mode of use and hierarchical position for imagery archives, assuming that imagery archives are the basis for MC&G production. (This assumption is further addressed in Section 2.6.4.)

Having tentatively selected an image archive approach, additional questions must be addressed relative to maintaining the archive, i.e., purging unwanted images and making space for new ones. These processes are directly analogous to ordinary data base maintenance, but they are complicated by the volume of the digital image data sets and by the difficulty of determining the amount of useful (and/or unused) information contained in a given image. The application of information measures, if shown to be of use for archive input, should simultaneously support selection of imagery for purging. Such a process would be automatic for the Case 2 (Normalized) archive scheme discussed above, but somewhat more difficult to apply to the other forms of image archive. This is because "new" and "old" images could not be expected to overlap exactly, such that a complete replacement is possible. Some labelled partitioning scheme (or human intervention) would be required to detect when a given image has been effectively "replaced" by new images.

Finally, it is noted that archiving and purging processes associated with archiving at the level of the Unified Data Bases (Fig. 2.3-1) are discussed in Section 2.6.4. This concept leads to data base management processes very similar to the more advanced commercially available processes. However, the certain loss of information implied by the data extraction processing poses some problems for MC&G production. These are also discussed in the following sections.

2.6 ARCHIVE UTILIZATION MODES

The various possible positions of the designated archive in the MC&G data hierarchy each imply a mode of data utilization. That is, there is implied pre-processing and post-processing for the data in the archive which supports the development of MC&G data bases, from which products can be compiled. Furthermore, there are implied cost and processing complexity tradeoffs to be made in the final design stages of an image-based production system. These tradeoffs cannot be adequately evaluated until more information is obtained from ongoing research into automated MC&G data extraction processes.

On the basis of the cost analyses in Ref. 6 (and subsequent related materials, such as briefing notes), it may be assumed that digital archive medium costs will be a major driving parameter for system design. The implications of several image archive utilization schemes are therefore examined relative to this key parameter.

2.6.1 Holding Pen Concept

Definition - The idea behind the holding pen concept is that an archive *per se* is not required. That is, it is assumed digital and/or analog processes can be developed to the point where imagery is held in temporary storage (holding pen) only long enough for project planning to be completed; the process streams directly through to production of product-oriented files (Fig. 2.6-1).

Implementation - The mechanization options, relative to storage, for the holding pen concept are shown to include both analog film and digital storage of either a reusable or non-reusable type. It is quickly apparent, however, that any use of a non-reusable medium in this scheme defeats one of its primary purposes, which is to avoid the expense of archival image storage. Thus, the mechanization of choice is to maintain only enough temporary, reusable storage capacity to allow for

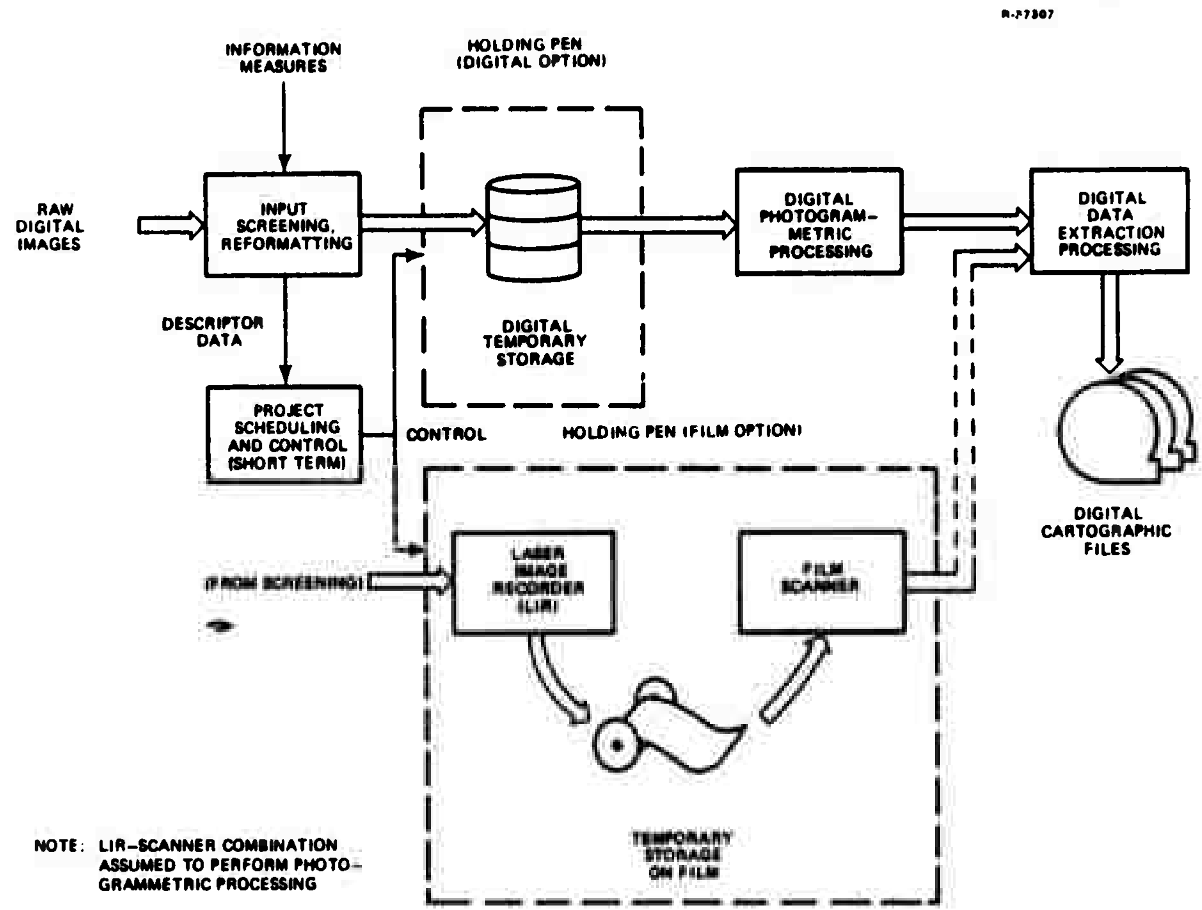


Figure 2.6-1 Holding Pen Concept

short-term scheduling of the extraction process. The total process should pass data of the order of 10^{12} to 10^{13} bits per day (Ref. 6), so the temporary storage volume should be of the same order.

An important implication of the holding pen concept is that the information measures gauging the utility of specific images must be applied at the source and screening levels. That is, the desired image parameters must be available a priori since there is no archive to be used for comparison with candidate images. Another implication is that digital processing, through the point of final data extraction, must be almost completely automated, since the data is to be streamed through at near-real time rates.

Advantages - The holding pen concept provides the following advantages not shared by other concepts addressed herein:

- Freedom from archive medium cost impact
- Utilization of simple, coverage-based information measures.

Disadvantages - In comparison to archive-based MC&G processing schemes, the holding pen concept has some severe constraints:

- Any change in user information requirements requires recollection of source data, since there is no archive to consult for possible extraction of new data types
- Little or no product-independence is available; the concept requires foreknowledge of the total product line form and content.

Other Implications - Very tight coupling and control of source data collection operations is required for this concept, in order that the specific needs of a given product compilation or revision project can be met. The concept also depends very heavily on positive results from current research on full automation of digital MC&G data extraction from images, or at least near-real time interactive processing. This arises from the lack of any archival storage which decouples product generation from input data acceptance. It is to be noted that with provision of buffering capability the concept becomes almost identical to the MC&G data base archiving concept analyzed in Section 2.6.4.

2.6.2 Fundamental Image Archive

Definition - The digital image archive concept which most closely matches current DMA image utilization involves storing images in essentially the form in which they are received. Figure 2.6-2 illustrates this concept for an all-digital implementation. The application of information measures is shown to exist principally in the archiving operation; i.e., as an archive entry mechanism. These measures are therefore likely to consider primarily the coverage available in the current archive, including the degree of planned redundancy, the quality of that coverage for MC&G purposes, and the need for new coverage based on current and future production needs.

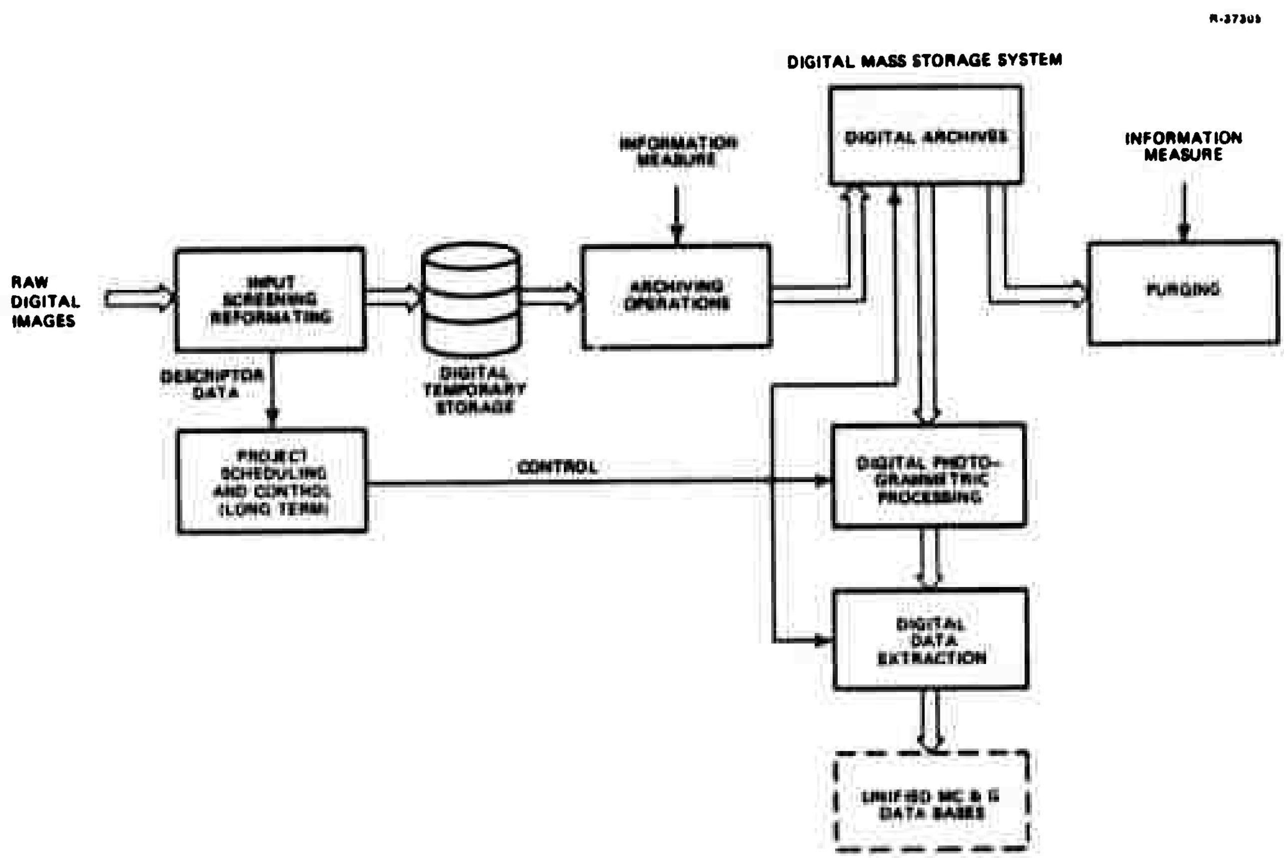


Figure 2.6-2 Fundamental Image Archive Concept

The most important aspect of the fundamental image archive concept is that the imagery is always available in its least degraded form. Not all image preparation operations are necessarily redundant, however; e.g., the necessary rectification and the registration processing parameters can be saved after their first derivation and may be reapplied in future use of the same image. The availability of all images is essentially "raw" form provides assurance that no hitherto unneeded or unnoticed feature data is lost due to the archiving process, and that no loss in resolution is induced by the process. The penalties associated with this assurance are threefold, affecting archive medium cost and processing costs:

- The archive size grows continually to accommodate new images, except as it may be maintained at a given maximum size by purging
- Many image processing tasks will have to be repeated for each use of the image
- The difficulty of automatic and interactive feature extraction processing is increased by the image-to-image variation in contrast, shading effects, defilading, etc.

Implementation - The mechanization of the archive will depend primarily on the storage density supportable by the technology at the time the mass storage system is selected. As discussed in Section 2.5.2, archive medium cost is the key parameter to be minimized. This dictates selection of either a reusable medium technology providing at least 3×10^7 bits/in^2, or a digital optical bit-by-bit system yielding 10^8 - 10^9 bits/in^2, to make non-reusable medium cost effective. The desire for non-degradation of the raw images reduces data compression capabilities to the area of 2:1 or 4:1 (Appendix A).

The option of direct image recording on film, with scanner re-conversion to digital format, is not viable for this concept because of the potential size of the archive and the relatively low density of information which can be effectively used (Sec. 2.5.2). Again, the non-reusable medium costs dominate the system costs with such an approach.

Two important implications of this concept are derivable from Fig. 2.6-2:

- The principal information measure applications are included within the processing system, and do not directly bear on the source activities
- The concept allows for the development of unified, largely product-independent MC&G data bases.

Advantages - The fundamental image archive concept incorporates most of the desirable features outlined in Sections 2.4 and 2.5. In particular, it provides:

- Recourse to essentially non-degraded raw images, if required
- Flexibility of application of information measures, allowing for adjustments of archiving rules to cover special cases
- Decoupling of production demands from source activites.

Disadvantages - The concept of maintaining a reconstitutable raw image archive has some significant weaknesses:

- Archive storage costs, as well as implied archive management processing costs, are relatively high

- Archive size and costs are strongl· affected by external factors such as input data rates and coverage redundancy requirement

- Cost control by purging requires complex information measures and probable labor-intensive activity.

Other Implications - A key consideration which favors the fundamental image archive concept is its functional similarity to present production practice. Training, procedures development and system and product acceptance by users would all be easy to accomplish if this scheme were adopted.

2.6.3 Merged and Updated Image Archive Concept Definition

Based on the idea of normalizing an image archive discussed in Sections 2.5.3 and 2.5.4, an MC&G imagery archiving concept can be formulated as indicated in Fig. 2.6-3. In this concept, it is assumed that the archive is not only size-limited, but also that further savings are obtained by incorporating only those portions of new images which represent new information.

As mentioned in the earlier discussion (Sections 2.5.3, 2.5.4), this approach could be based on only a slight conceptual extension (Case 1) to the fundamental image archive concept. That is, by normalizing each new candidate image, improved ability to measure the new information it provides against that already in the archive could be expected. This would allow improved ability to reduce the archive input rate and to support the purging process, with resulting improvement in archive costs.

The second alternative (Case 2) offers more power, in that mosaicking is invoked to permit merging into the archive

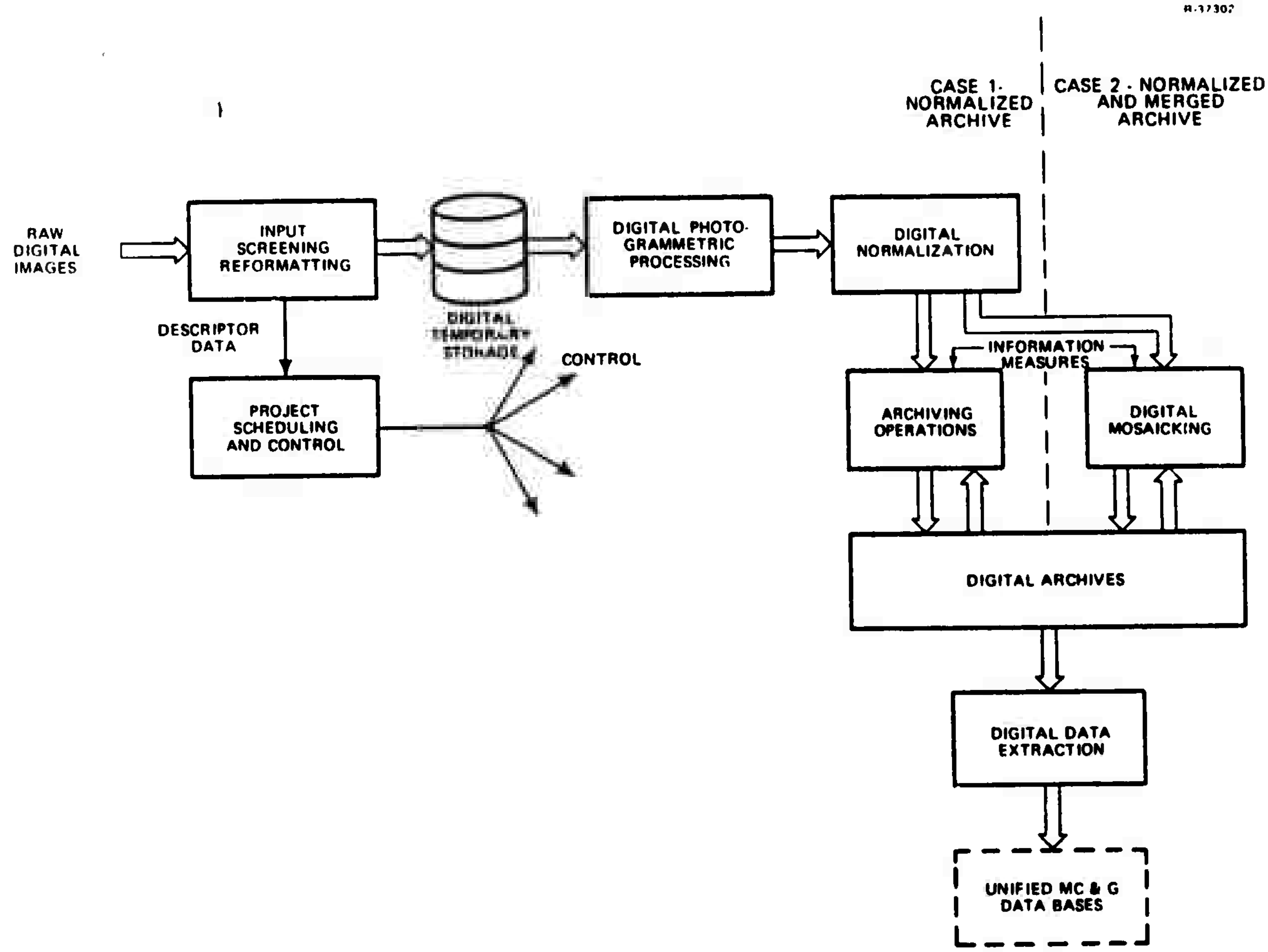


Figure 2.6-3 Merged and Updated Image Archive Concept

only those portions of new image in which the objective and subjective information measures indicate new information resides. This process would result in an archive of constant size, and currency in detail. No reconstruction of an exact raw image could be obtained from the archive, but an up-to-date mosaic, current almost to the pixel, could be obtained at any time for any location.

Implementation - It is apparent from previous discussions about medium costs that the merged and updated archive concept demands a reusable medium, that is, one in which image records or portions of image records are capable of being modified.

This requirement is less stringent for Case 1 than for Case 2, and in fact a <u>very</u> inexpensive read-only medium might be tolerable for Case 1.

The major application of information measures is again at the archiving level (Fig. 2.6-3), with two key points to be noted:

- Significant processing is devoted to candidate images prior to the acceptance decision
- The image information measures are really based on <u>change detection</u> and <u>change evaluation</u> relative to the archive contents.

The information measures adopted may therefore be simpler to develop, since they may be based primarily on comparison of new data with old, with less of a subjective component required. In Case 1, acceptance evaluation will generally require assessing a new image against several archived versions. In Case 2, a view of the precise area of the new image's coverage can be assembled from the mosaic, allowing one-to-one comparison by what should be relatively uncomplicated procedures.

<u>Advantages</u> - The key advantages of a merged and updated image archive would be:

- Constant archive size, or at least constant within narrow bounds, eliminating archive storage cost growth
- Simplification and possible automation of information measures, relative to a raw image archive

- Ability to reconstitute an up-to-date, full resolution mosaic of any area of interest
- Simplification of data extraction processing due to normalization and the ability to access only the portion of interest.

Disadvantages - The key disadvantages of this concept are that implementation will require:

- Extensive development of image normalization, mosaicking and data management routines
- Application of expensive processing to images which are finally rejected
- Possible increased investment in processing power, tending to reduce the cost advantage of constant archive size.

Other Implications - The operating concept and capability discussed herein are different from DMA production processing currently in force, and would require significant evaluation and experimentation before implementation. Significant development effort will be required for many of the processes alluded to herein, with concomitant technical and cost risks. However, the concept successfully addresses all of the problem areas discovered in the previous analyses.

2.6.4 MC&G Data Base Archive Concept Definition

It has been conjectured that, given sufficient digital processing support, images might be processed rapidly enough to permit on-line extraction of all MC&G-related features (Fig. 2.6-4). The MC&G archive in this concept would consist of extracted feature data, in the form of unified MC&G data

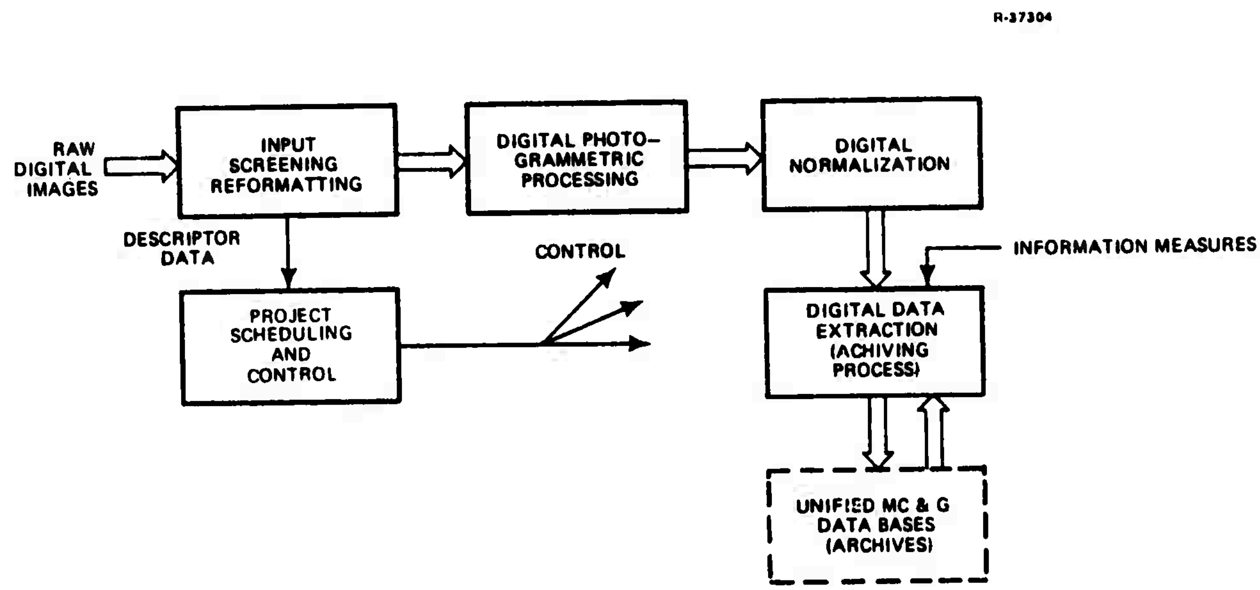


Figure 2.6-4 MC&G Data Archive Concept

bases. The objective is to take advantage of the large (factors of 100 to 300)* effective data compressions resulting from the data extraction process, to avoid the cost of an image archive.

The information measures employed for controlling the contents of an MC&G data archive would be relatively straightforward and primarily based on change detection, with respect to the existing data bases. The key data processing implication of Fig. 2.6-4 is thus that <u>all</u> images passing screening would be processed to completion, prior to the archiving decision. This in turn requires that very fast, accurate and comprehensive feature extraction processes exist, whether automatic or interactive with cartographers in the loop.

---

*Based on an estimated 10^8 bits required to digitize a 1:50,000 scale topographic map (Ref. 16) versus 3 x 10^{10} bits per LANDSAT Thematic Mapper scene (Ref. 11).

The principal difference between this concept and the holding pen concept defined in Section 2.6.1 is the presence of the unified data bases/archive level of storage. This level, as discussed in Sections 2.1 - 2.3, allows a high degree of product independence in the data bases and a concomitant high degree of flexibility for generating new product data compilations.

In both the holding pen and the MC&G data base archive concepts there is implicit the idea that recourse to old imagery need not be provided, if sufficiently rapid processing of new images can support extraction of new feature classes. This idea is based on two assumptions:

- Data extraction processing is sufficiently modular to allow timely design, implementation and installation of new feature extraction processes when new requirements are imposed
- Means are available to obtain coverage including the new feature classes within a reasonable response time.

Implementation - Given that the above-mentioned data processing capabilities can be achieved, no special concerns exist relative to the archive in this concept. Such an archive has a natural upper bound in data volume, based on the desired geographic coverage, because the individual MC&G data bases are naturally updated as required and do not have to hold old data on permanent file.

At least an order of magnitude reduction in data volume can be achieved over, say, a mosaicked image archive, leading to a maximum data volume of the order of 10^{14}-10^{15} bits. This is based on the idea that any new feature classes demanded in the future will be relatively sparse compared to the very large terrain data files currently being accumulated. (Significant ongoing research exists which aims at greatly reducing terrain data files by developing digital terrain models; e.g., Refs. 17, 18.)

Data bases of this range can reasonably be handled by current and projected Very Large Mass Storage System (VLMSS) technology, as indicated in Ref. 6, especially if the bulk of the data files are located in off-line shelf storage most of the time. There is a definite need for a read/write (reusable) medium, however. The fact that these unified MC&G data bases are not composed of huge, inseparable blocks of data (as are individual images) further assures the applicability of VLMSS techniques, and advanced data base management approaches.

However, as can be seen in Fig. 2.5-2, with reasonable extrapolation of VLMSS technology to either advanced magnetic or optical technology, an archive in this range of storage volume is not especially cost-sensitive to volume. That is, the bulk of the VLMSS cost for volumes of 10^{15} bits or less is due to the read-write equipment and data module handling equipment costs. Medium costs should not drive the system to any great extent, so that cost savings over an image archive of say 10^{15} bits could only be realized in ancillary costs. These would include those realized by not having an image archive to manage, in addition to the MC&G data base. Since the MC&G data bases are at least an order of magnitude less voluminous than the postulated image archives, it is reasonable to suppose that they would be implemented and maintained in the same VLMSS equipments as the imagery in the image-based concepts of Sections 2.6.2 and 2.6.2. The resulting ancillary cost savings realizable by the MC&G data base concept therefore do not include significant savings in VLMSS costs.

Advantages - The direct digital extraction of MC&G feature data from digital imagery offers significant advantages in terms of archiving considerations:

- The resulting archive is bounded in data volume, and maintainable by well-understood data base management techniques
- The required archive storage volume is much less than for an equivalent image archive; more than an order of magnitude saving is readily realizable
- The information measures applicable to admit new feature data to the postulated unified MC&G archives should be relatively simple, and based on reasonable extrapolations of current MC&G data editing practices.

Disadvantages - Several difficult problems must be solved before the MC&G data base archive concept can be made practical, and some of them may not have ready solutions. These difficulties include the following:

- There is no recourse to imagery for recovery of previously unextracted data
- Current progress in automated and semi-automated feature extraction has not been promising, except in very specific cases; the required general capability for this concept is well in the future
- The implied requirement for rapid, DMA-directed collection of large volumes of new feature data poses serious cost and procedural problems.

2.7 COMPARATIVE EVALUATION OF APPROACHES

2.7.1 Recourse to Imagery

In Section 2.4 it is pointed out that large advantages would have to be realizeable in order to offset the advantages of recourse to imagery for MC&G production. Of the two concepts discussed above in which recourse to imagery is denied (holding pen and MC&G data archive), only the holding pen concept offers the possibility of large savings, since no VLMSS equipment would be needed for archiving. However, this concept cannot provide product independence, and it is very inflexible with respect to changes in user requirements.

The holding pen and MC&G data concepts both require significant automation progress in the difficult field of feature extraction, because they cannot provide an "old" imagery data base for human operators to examine in concert with new images. This capability is critical to image understanding, which is in turn critical to the human operator's ability to perform interactively with the feature extraction system. The possibility of using previously extracted MC&G feature data to support examination of a new image should be explored; it might go a long way toward making the MC&G data archive more viable. However, these two concepts have not yet been shown to provide sufficient compensating advantages to merit further consideration.

Of the three image-based archiving concepts discussed, considering the non-mosaicked and mosaicked normalized data bases separately for the moment, it is apparent that the first two offer recourse to imagery in very much the same sense as is currently available to DMA operators. The mosaicked normalized data base would not provide access to "real" images, in that any image produced for a given area would probably be a composite of several real images. The effect of the mosaicking should be minimal, since cartographers are used to working with images mosaicked on a somewhat larger scale. However, it appears that the fundamental image archive and Case 1 normalized image archive concepts are superior to the other three concepts in this regard.

2.7.2 Data Processing Implications

All of the archiving concepts discussed above except the fundamental image archive concept require significant digital image processing advances to be made for their implementation. The most radical advances, to either full automation or very complex interactive processing, are demanded by the holding pen and MC&G data archive concepts. The mosaicked normalized image archive concept demands some significant advances in digital mosaicking, to make it applicable at the level of small groups of pixels.

The normalization processes, on the other hand, include only reasonable improvements and standardization of processes which are currently well understood:

- Gray-scale adjustment
- Contrast adjustment
- Noise and haze suppression
- Image enhancement for specific processing advantages, e.g., edge sharpening.

2.7.3 Data Storage Implications

The only concept addressed which offers major storage savings is the holding pen concept, which provides no archive, but only product-oriented files. The mosaicked normalized image concept requires a well-bounded archive, and the MC&G data base concept a much smaller, bounded archive. As is pointed out in Section 2.6.4, however, the archive cost differential between these components is not large, and they are essentially equivalent in terms of archive cost, except that the MC&G data archive offers a simpler total data base management problem.

The fundamental image archive approach demands the most storage capability, and could incur heavy medium costs if purging were not imposed, or if an <u>extremely cheap</u> non-reusable medium were not found. The projected availability of reusable media at reasonable recording density, with the concurrent ability to support economical purging, indicates that this approach can be made economically viable.

2.7.4 Concept Ranking

On the basis of the above evaluation and the detailed discussions earlier, it is concluded that:

- The fundamental image archive concept is the best approach for DMA to take in the near- and mid-future (10 years)
- The potential of the mosaicked, normalized image archive (Case 2) concept should be explored for possible implementation in the 10-year time frame
- The normalized image archive (Case 1) offers insufficient advantages over the other two mentioned thus far, and can be ignored
- The holding pen and MC&G data base archive concepts offer high risk, relatively little advantage and a great disadvantage in that they provide no recourse to imagery. These should be rejected.

3. TOPOLOGICAL DATA BASE STRUCTURES APPLICATION EXPERIMENT

3.1 BACKGROUND

A major portion of this work was defined by ETL to be an experimental study to evaluate the potential utility of Topological Data Base (TDB) structuring techniques. Applications of TDB structures by the U.S. Census Bureau since the late 1960's and by the U.S. Geological Survey since the mid-1970's have suggested that the use of topological principles to structure a geographic data base can provide a number of valuable capabilities. The key relations applied are those of spatial adjacency, connectivity and left-right ordering of geographical entities (e.g. map features). TDB structuring results from specific encoding of these relations in the data base of interest, as discussed in Section 3.2.

The Laboratory for Computer Graphics and Spatial Analysis (LCG) at Harvard University has been engaged in research on the subject of topological data structures for the past four years. As part of this work, LCG has developed a series of computer programs (the ODYSSEY system) (Ref. 19) designed to support creation, manipulation and retrieval of geographic data using TDB structuring techniques. LCG has represented the ODYSSEY system to be the state-of-the-art in software development involving the use of topological data structuring principles. In particular, it has been asserted that these principles are applicable to the design and implementation of large cartographic data bases, with the following advantages:

- Structuring of the data base at the polygon level permits design and use of "local" processing algorithms, which do not require rapid access to the entire data base. This reduces computer speed and memory requirements

- TDB structures are close analogs to human perception of geographically-based data, and thus easily support relational queries and data aggregation in a geographic context

- TDB structures are at the same time readily processed by advanced computer graphics and data base management software.

The problem presented for solution in this study reflects DMA and ETL interest in developing a means for integrating information descriptive of many different subjects for a given region. It involves the merger of cartographic data from several different maps into a single data base and the extraction of specific information from the combined data base. The individual maps provided by ETL, on separate Mylar overlays, describe land use, elevation contours and 100-year floodplain boundaries for the northwest quarter of the Healdsburg Quadrangle, at a scale of 1:24,000. The operations to be performed upon the combined data base include selection and retrieval of map features according to single or multiple criteria. (E.g., "plot all land use polygons having a code of AVV," or "plot all land use polygons having a code of AVV <u>and</u> lying within the floodplain <u>and</u> at an elevation below 100 ft." Figure 3.1-1 shows typical results of such query).

This study requires consideration of data base maintenance and updating according to separate descriptions of land use changes, which are to be added to the combined data base after it is constructed. Additional operations involving map updates, edge matching and scaling are included. The topological data base research upon which this project is

Figure 3.1-1 Plot of Land Use Polygons Within Floodplain and Below a Given Contour Line

based has focused upon applications involving the representation of map features as polygons. Therefore, in this experiment, data describing land use, terrain height contours and floodplain boundaries are all treated as polygonal features, to allow the application and evaluation of topological data structuring procedures as embodied within existing software at LCG. It should be noted, however, that features defined as networks rather than polygons can also be described within a topological data base. Examples of such networks, used to describe topographic featues, are given in Ref. 18.

3.2 TOPOLOGICAL DATA BASE STRUCTURES AND APPLICATIONS

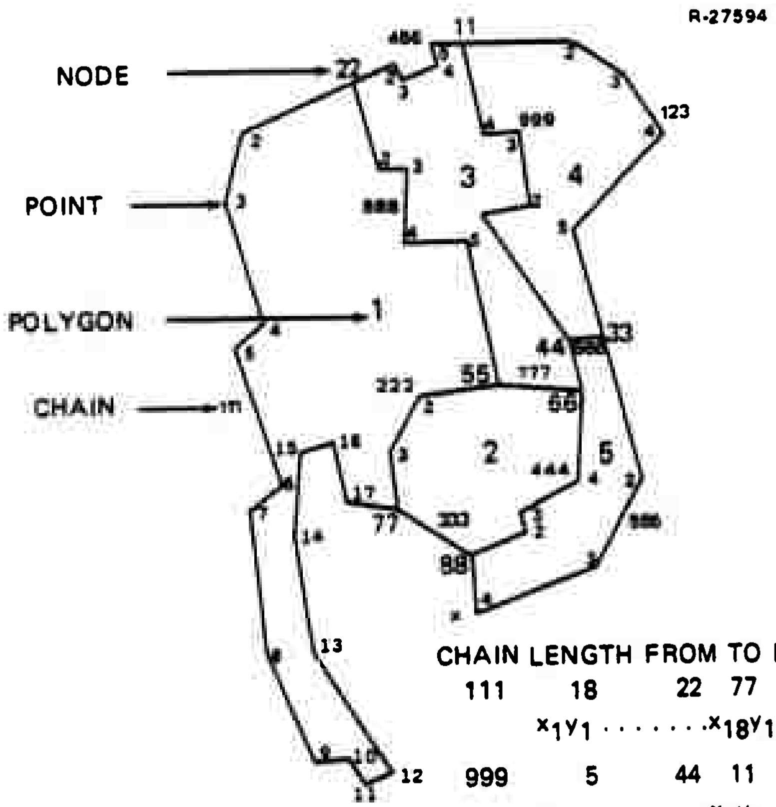
3.2.1 Concepts

Three important concepts support TDB structuring applications:

- The use of topological information to define spatial relationships between geographic entities
- The use of polygon overlay techniques to define Least Common Geographic Units (LCGUs)
- The use of highly efficient local, rather than global, data processing algorithms.

Use of Topological Information - A topological data base incorporates elements of graph theory and elementary topology for the purpose of defining geographic entities. Of primary importance is the *explicit* definition of *connectedness* and *adjacency* for the entities being described. Having included such information, it is possible to develop linkages via software which allow for automated error detection, integration of different map overlays into a combined data base, and flexible retrieval of geographic entities. For example, such a combined data base may be queried to retrieve individual data elements, as originally stored, or composite entities resulting from Boolean operations upon the combined data base.

In constructing a topological data base, polygons representing areal or lineal feature sets are described in terms of their common edges. (E.g., political boundaries might be represented as a polygon net similar to Fig. 3.2-1.) Each edge is described in terms of its geometrical and topological characteristics. This data is then manipulated by



CHAIN	LENGTH	FROM	TO	LEFT	RIGHT
111	18	22	77	1	0
	$x_1y_1 \ldots\ldots x_{18}y_{18}$				
999	5	44	11	3	4
	$x_1y_1 \ldots\ldots x_5y_5$				
222	4	55	77	2	1
	$x_1y_1 \ldots\ldots x_4y_4$				

Figure 3.2-1   Topological Chain File Components

software which uses data structures and algorithms specifically designed to provide the required editing, retrieval and display capabilities.

Geometrical characteristics of each edge are described in terms of sets of X-Y coordinates which define its relative location on the map. The string of coordinates for an edge are referred to as a "chain". The end points of each chain are called "nodes." X-Y coordinate locations lying between the nodes are simply called "points". Figure 3.2-1 illustrates this nomenclature. In the figure, chain #111 represents the edge of polygon #1 which runs between nodes #22 and #77; the other points of that chain which specify the location of the boundary are identified as #2 through #17.

Topological characteristics of each edge are recorded to provide information which allows for the construction of linkages with other edges as well as reference to the polygons bounding either side of a given edge. Thus, a geographic region can be described, for a given set of attributes, by partitioning it into a single set of zones ("polygons"). Each polygon can be described, i.e., encoded, in terms of its edges ("chains"). Finally, the key concepts of spatial adjacency and relatedness are inserted as part of the chain encoding, in that each chain is described in terms of:

- Two end points ("from" node and "to" node), which also define a direction of traversal for the edge
- A reference to the two polygons which have a given chain as a common edge ("left" and "right" polygons).

Typical results of this TDB encoding process are illustrated in Fig. 3.2-1, as can be seen by tracking the implications of the given chain descriptions through the associated polygons. The logical relations among these topological chain file components are illustrated in Fig. 3.2-2. The power of the TDB structure derives from that hierarchy of structures, as is discussed in the following sections.

Polygon Overlay Operations - (Ref. 20) consider that within a given geographical area there are two different polygon networks to be overlaid, the "solid" network and the "dashed"

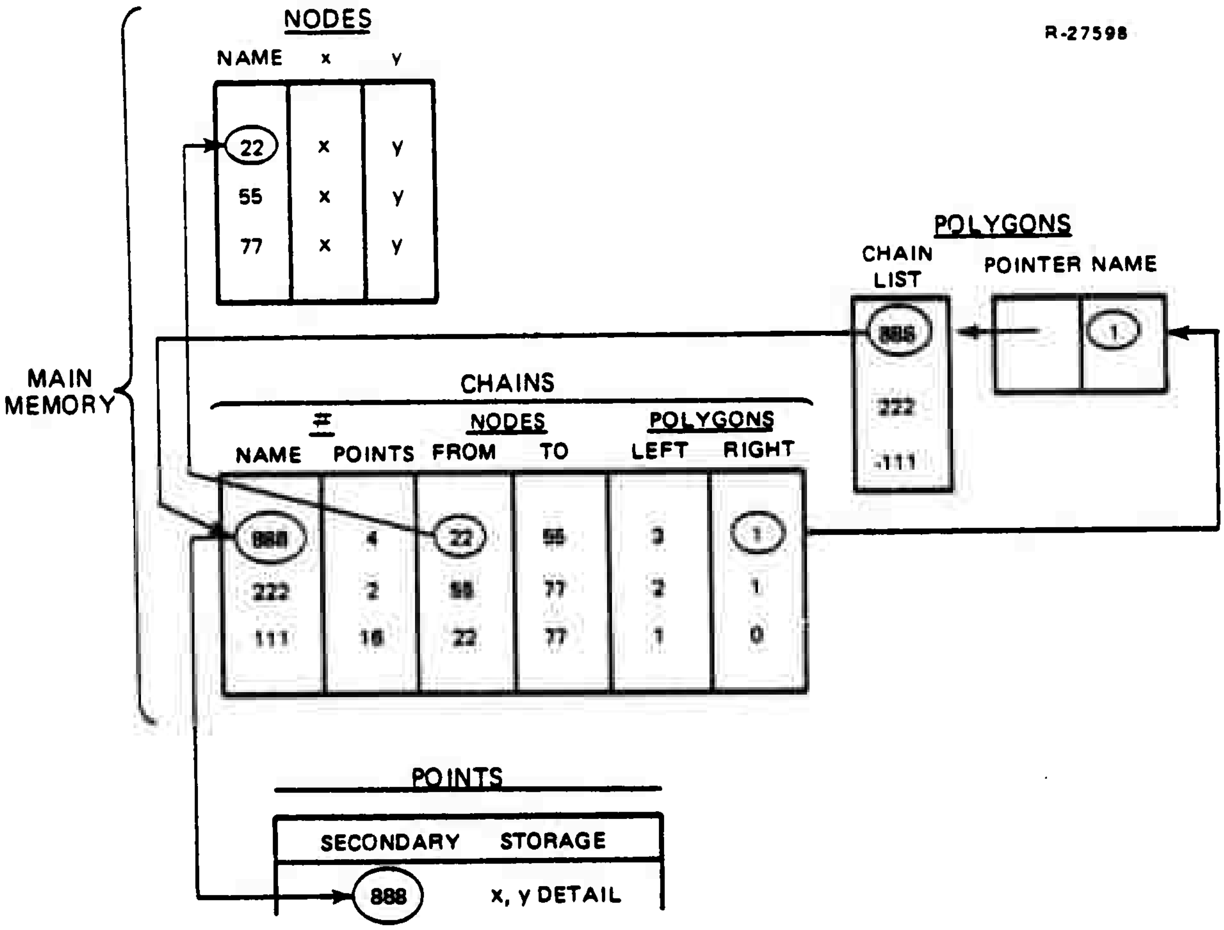


Figure 3.2-2 TDB Structure Logical Relations

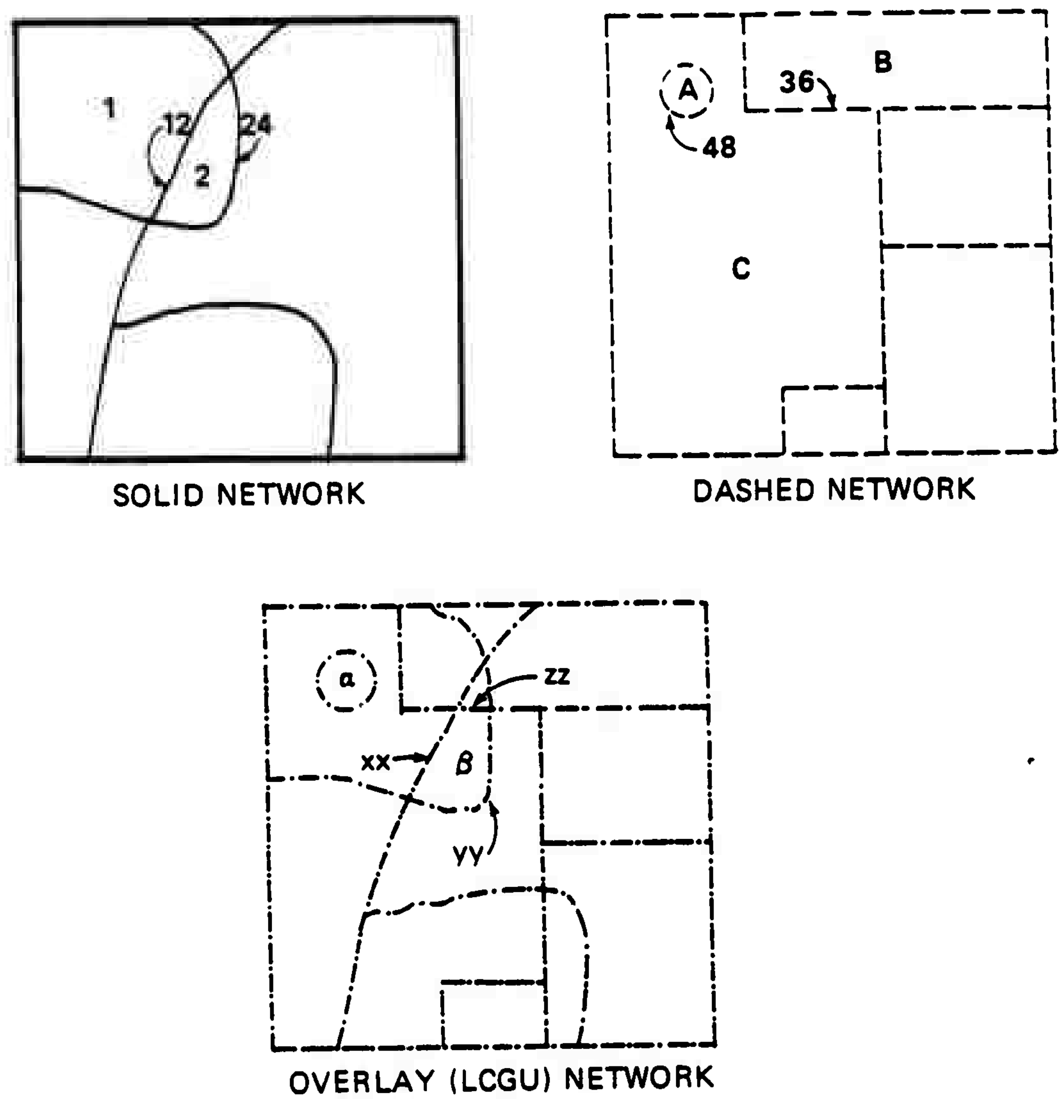
one, the composite of which will be an overlay network (Fig. 3.2-3); these might represent land use and political boundaries, for example, There are five operations to be performed:

1) Given a polygon of one network, find the polygons of the other network which it may possibly overlap

2) Given possible polygon overlaps, find points of actual intersection

3) Recognize when a new polygon has been created (i.e., a new closed polygon formed by intersection or inclusion, referred to as a "Least Common Geographical Unit (LCGU)")

R–36255



EXAMPLE OVERLAY ATTRIBUTES LISTS

LCGU POLYGON	PARENT POLYGON	CHAINS	ATTRIBUTES (TOTAL OF PARENT ATTRIBUTES)
α	1, A	48	Att_1, Att_A
β	2, C	xx, yy, zz,	Att_2, Att_C

NOTE: NEW CHAINS xx, yy, zz FORMED FROM 12, 24, 36 BY INTERSECTION TO DEFINE LCGU β

Figure 3.2-3 Example Polygon Overlay

4) Identify the solid and dashed "parent" polygons of each LCGU (i.e., the polygons from which it was derived)

5) Test all new LCGU's for topological consistency; e.g., closure, proper left-right adjacency, proper edge tracing direction information.

It may seem that the first two tasks are very similar; however they are treated separately here to illustrate that, in fact, different solutions are used for each. The first task consists of determining which lines may intersect, and the second task consists of finding points at which lines do intersect.

The way the first task is handled sets the framework within which some or all of the other tasks are accomplished. The overlay process is controlled by a geometric orientation; i.e., a conceptual passage through two parent networks from the same starting point and in the same direction. The two networks are concurrently searched for possible intersections, but only in the region being passed through, via local processing techniques as discussed later in this section.

After the set of candidates for intersection has been chosen, the actual points of intersection must be found. In concept this process is similar to selecting candidates, namely, a search of the edge representations of each polygon to find points of intersection.

The creation of new composite polygons, LCGUs, is at the crux of the overlay process. The problem to be solved is partly an accounting task consisting of aggregating pieces of boundaries when they are broken by intersections of parent polygons. As illustrated in Fig. 3.2-3, the typical LCGU is comprised of pieces of polygons from both networks. These must be assembled in some manner so that the LCGU can be accessed and mapped.

In addition to assembling the coordinate description of each LCGU, its parentage must be recorded. This can be established at the time that intersections are made, because the data structure for the coordinate representation of the boundary (chain) has labels for polygon names, and composite labels can be created when lines are intersected.

The last task required to create a LCGU is to verify its completeness by making sure that its boundaries are continuous, i.e., that it can be traversed back to the given starting node. Such "cycling" of the polygon is an important check on the validity of the composite network, and it is specifically enabled by the topological encoding of <u>direction</u> in a chain and <u>left</u>, <u>right</u> orientation of adjacent areas.

<u>Local/Global Data Processing Algorithms</u> - Local, versus global, data processing algorithms demonstrated in this project are designed to allow for the efficient processing of very large data bases while using only very limited amounts of main memory at any one time. This is accomplished by pre-sorting the input data on its y-axis in order to allow for the subsequent processing of the data in a <u>band-sweep</u> fashion. Such a technique requires that only a very limited subset (i.e., that portion lying within the band) need be resident in core memory at any one time (Fig. 3.2-4).

After the information contained within the band has been processed, it may be put back in mass memory. The location of the band may then be advanced and data lying within the

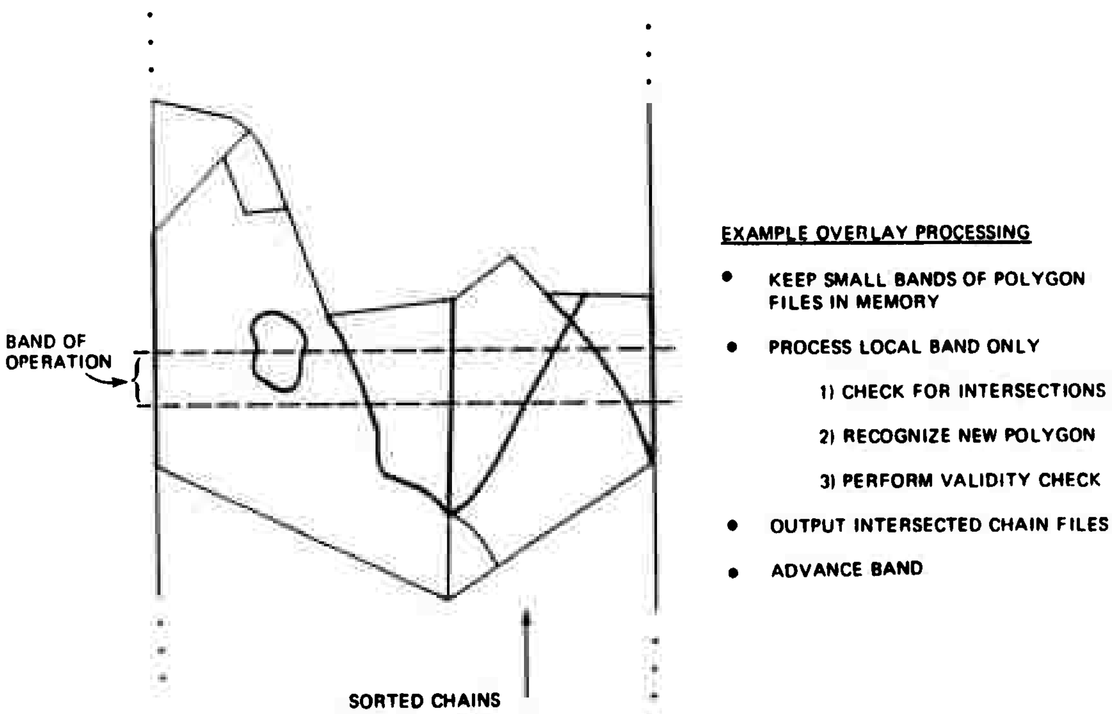


Figure 3.2-4 Local Processing of Polygonal Entities

region defined by the new location of the band may be called in for processing, and so forth across the entire area.

An additional benefit of local processing is that it is able to perform its operations using sequential, rather than random, file processing techniques. Data defining the region lying within the band will be able to reference temporary random access files, but bulk data file input and output involve only the use of pre-sorted, sequential files. The data base which is <u>created</u> by using local processing algorithms is composed of records which describe the edges of all polygonal areas resulting from the overlay of two or more input data bases. Each record describes an edge of an LCGU in terms of its geometric (X-Y coordinate) properties as well as its topological properties of <u>adjacency</u> and <u>connectivity</u>. Adjacency for any given edge is defined by recording the names of LCGUs which lie on either side of that edge. Connectivity is defined by recording the names of the nodes at each end of the edge.

Further Illustration of Polygon Overlay Applications - Assume that a geographic region can be described in terms of two or more different sets of zones, or polygons, each of which defines the geographic distribution of specific subjects (e.g., land use and elevation); these may be combined by use of polygon overlay techniques to create an integrated composite set of zones. Consider the following example:

Given region x of known location, covered by Map 1 and Map 2 (Fig. 3.2-5):

- Map 1 shows the location of subject A within a region x. E.g., land use which may consist of several classes: A_1 = residential, A_2 = commercial, etc.
- Map 2 shows the location of subject B within region x (e.g., 100-year floodplain boundary).

The problem is to create a map which identifies the location, within region x, which has residential land use (A_1) and is also in the floodplain (B_1).

Polygon overlay operations serve to integrate the geographic description of subject A with that for subject B. As a result, a new description is obtained which defines their combined geographic distribution; for example, the location of residential and within the floodplain for region x is polygon A_1B_1.

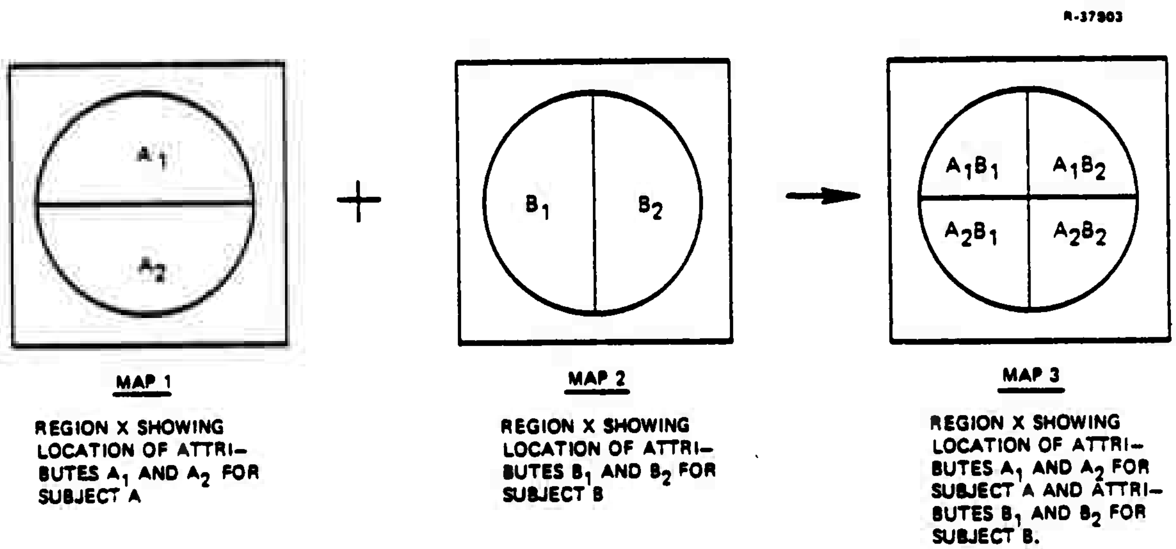


Figure 3.2-5 Hypothetical Mapping Problem

To accomplish this, one topological data base would be prepared describing the location of subject A and another describing the location of subject B. Each of these files would be described as a set of polygons in terms of their edges. When combined via polygon overlay techniques, a new topological data base composed of LCGUs is automatically created. Note that in this example four polygons have been created and are described in terms of their edges, where once again these edges are defined as topological chains. Each of these chains includes in its left/right identifier references to the parent polygons (A_1, A_2, B_1, or B_2) from which it was obtained. As a result it is possible to retrieve either of the original subjects, or to retrieve only those LCGUs which have the desired combination of attributes, by simply collecting LCGU edges on the basis of their left/right identifiers.

3.2.2 TDB Structures Application Experimental Approach

The software used in this project is a subset of the ODYSSEY Geographic Information System developed by LCG (Ref. 14). The complete ODYSSEY system constitutes a family of geographic information processing modules. A common file structure (Ref. 21); language processor and other programming utilities provide ease of data transfer from one process or another. Program modules combine system facilities and implement internal data structures and algorithms.

Specific ODYSSEY program modules which were applied to the ETL-specified data bases include the following:

DIGIT:	Interactive digitizing for file creation
WEED:	Line generalization, detail filtering
CYCLONE:	Verification of topology, correction of nodes and chains
FRED:	Chain editor
PSYCHE:	Heapsort, polyphâse merge sort
WHIRLPOOL:	Polygon overlay to create Least Common Geographic Units
DMA:	Display, mapping and analysis of overlaid files.

Topological data structure concepts were used in this demonstration to construct a combined data base incorporating land use, elevation, and floodplain boundaries for the project study area. Additional topologically structured files were prepared for purposes of recording land use update and map sheet revision data.

A user-oriented retrieval language is included as part of the software. This langauge supports extraction of components of the combined data base as required to satisfy a given user request.

The specific operations performed involved construction of a combined floodplain-land use-topography data base. This required the overlay of separate polygonal data files for land use, elevation and floodplain boundaries, encoding the basic map features in the combined data base as a set of LCGUs, as discussed in Section 3.2.1. As a result, a user at ETL can retrieve any set of LCGUs defined in terms of several different categories of land use and/or elevations and/or floodplain boundaries. The retrieved LCGUs are used to automatically construct an output file suitable for plotting the selected map features.

It should be noted that the process demonstrated in this study is one of "exact" polygon overlay. One result of this process is the creation of many small LCGUs which are actually artifacts, and do not represent real map features. These "sliver" polygons are caused by minute differences in the point-chain definitions of the same feature, such as a river boundary, derived from separately digitizing from two or more different maps, e.g., elevation and floodplain. The sources of these differences include the relative errors produced by separately digitizing different maps, as well as fundamental inconsistencies in the maps themselves.

Clearly, it would be desirable to eliminate such slivers. Several different approaches could be used to deal with this problem of line congruence:

- Each source map could be given a priority rank in relation to other source maps with which it is to be combined. When a very small sliver results from minor differences in the definition of the same map feature, then the line as described on the higher ranked map would be used

- A new line could be generated midway between the two similar lines and substituted for both
- A special line symbol to represent the fuzzy or uncertain definiton of the line based upon different descriptions of its precise location.

None of these solutions was implemented in the course of the present study, due to the importance of being able to first perform an exact overlay of several different data bases. The need for addressing problems resulting from the exact overlay process (such as line congruence) is an area requiring additional research and development.

3.3 BASELINE PROBLEM DESCRIPTION

3.3.2 Source Data

The source data provided for this study consists of four hardcopy map products covering a common area. These maps describe (1) land use, (2) 100-year floodplain and land use revision, (3) topographic contours, and (4) land use segments to be merged. The first three maps (Figs. 3.3-1a, b and c) were provided on Mylar and the fourth as a photocopy. Land use segments for merging are included in Fig. 3.3-1b with the full set of land use areas.

3.3.3 Outputs

Outputs include topologically-structured data bases and software capable of performing the following operations:

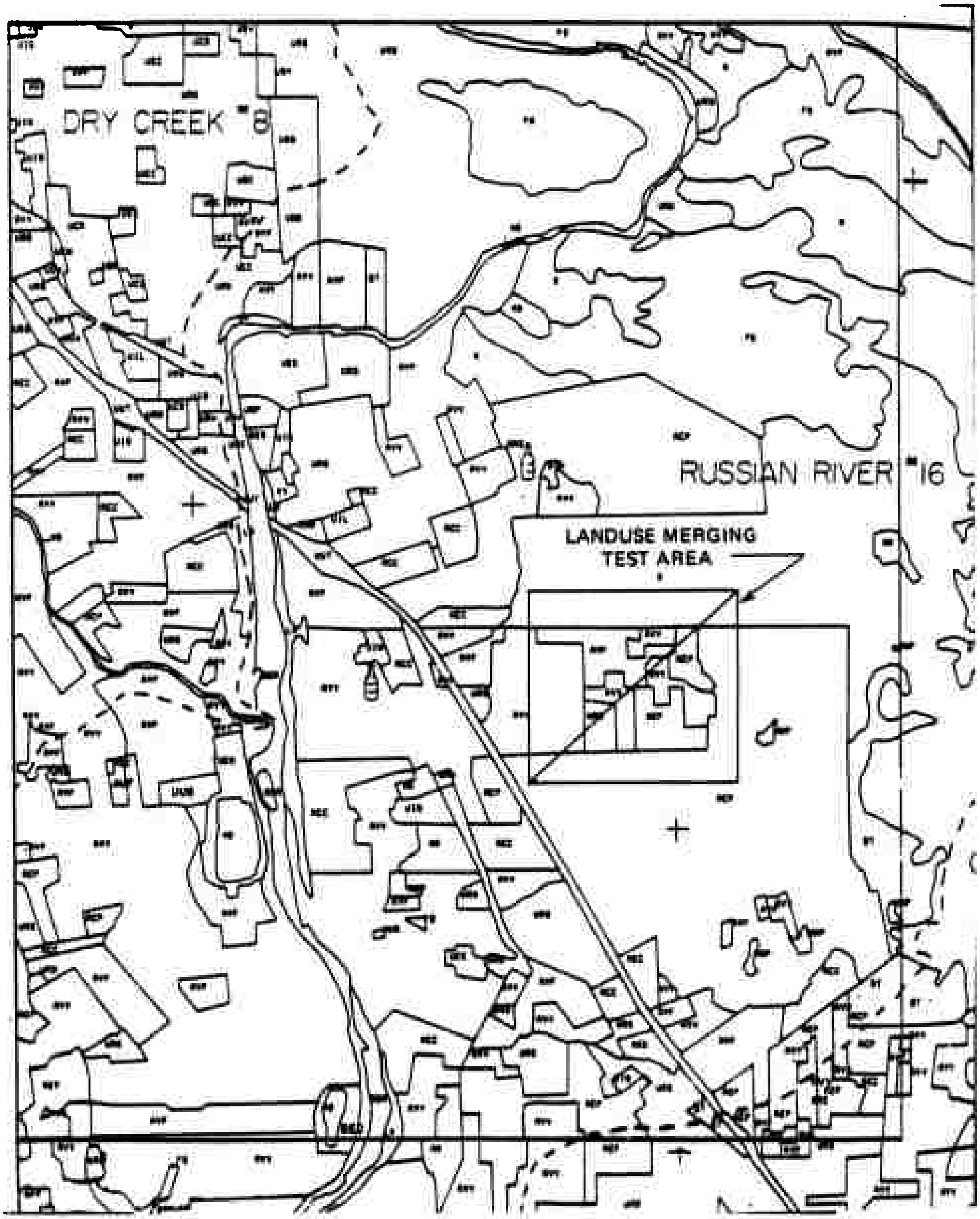


Figure 3.3-1a Original Land Use Map With Map Edges Inserted

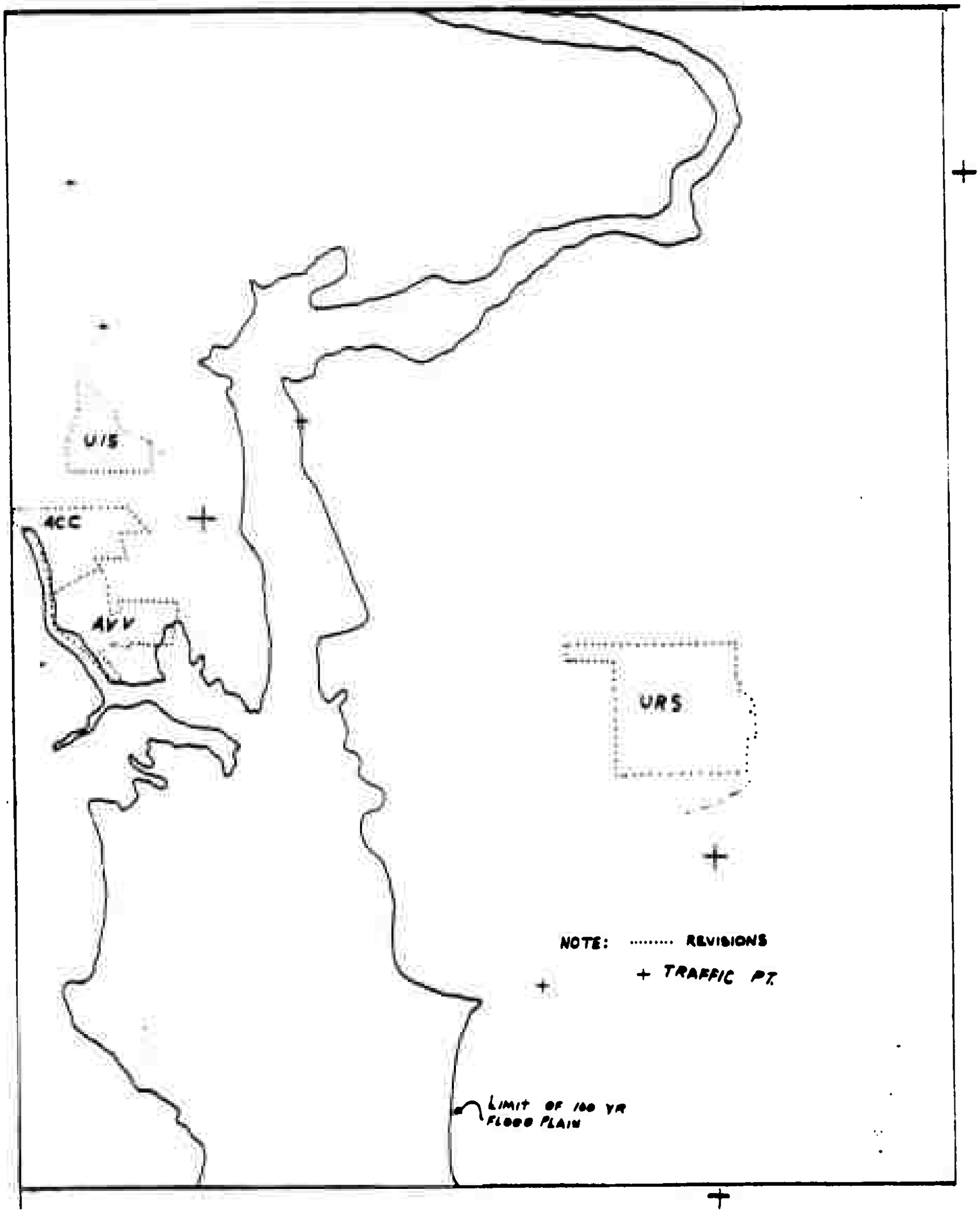


Figure 3.3-1b Original Floodplain Map With Land Use Revisions

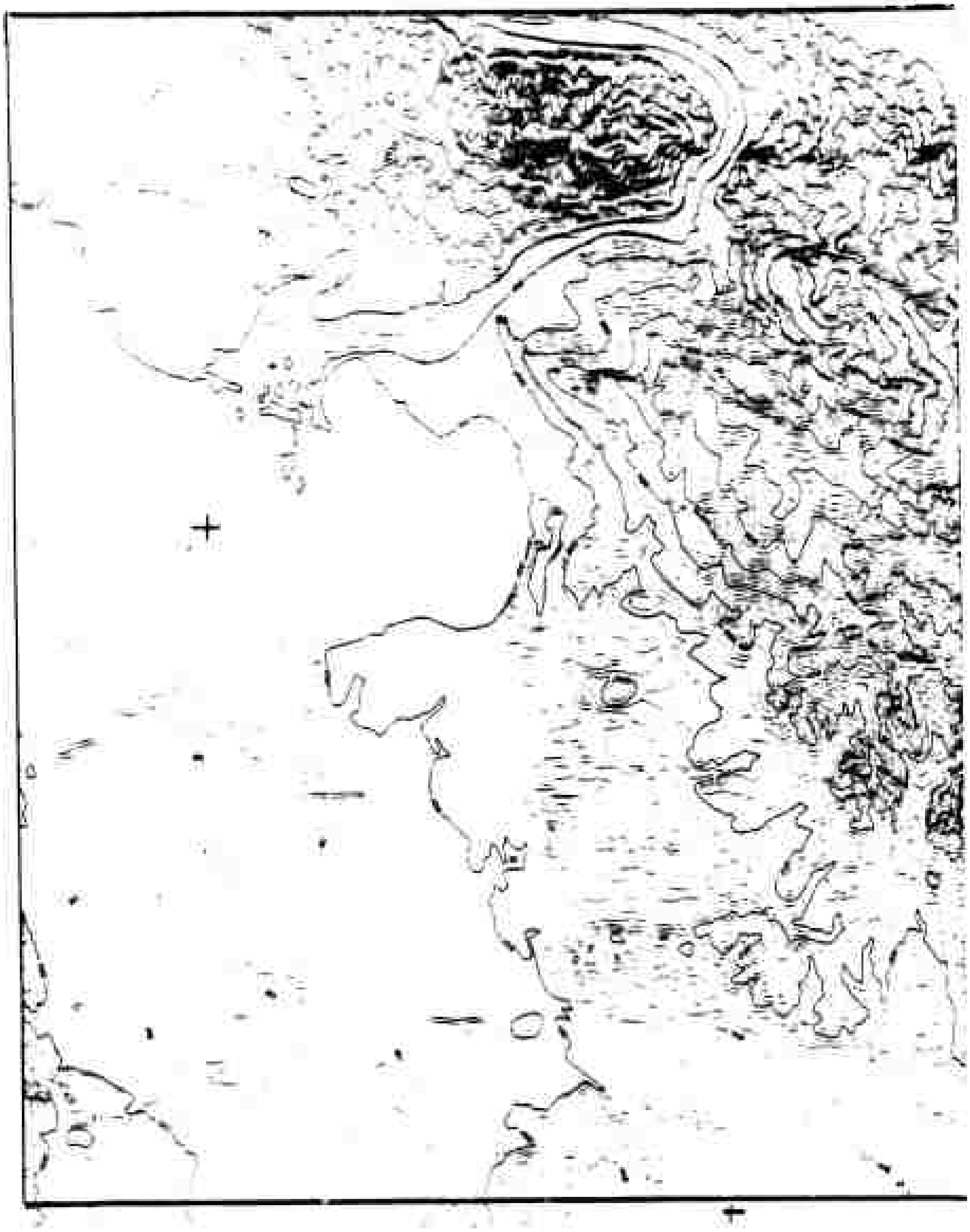


Figure 3.3-1c Original Contour Map

- Identify the individual land use areas within the floodplain and below the 100-foot contour line

- Update the individual land use areas within the data base where affected by the land use revisions on the floodplain map. This is to be done without redigitizing the basic data

- Merge data and perform overlay operations within a specified revision area. For this purpose the land use overlay, with the specified area outlined in red, was provided in the photocopy

- Perform search and retrieval operations by allowing a variety of queries to be processed against the data base, including: 1) a plot of all land use polygons with a code of ACC and 2) a plot of all 300-foot contour lines.

The topological data structuring routines used in this study were intially implemented on a DEC PDP-10 computer at Harvard University. They were then converted to run on a CDC 6400 computer at the Smithsonian Astrophysical Laboratory in Cambridge, MA. The converted software was then installed on the CDC 6400 computer at ETL. Documentation was provided which will allow for continued use of the software after completion of this contract.

A briefing was given at ETL describing manual operations related to the encoding of graphic products and computer operations capable of performing the operations described above. This demonstration consisted of the following steps:

1) Describe and illustrate procedures used to digitize maps provided, to prepare files of the following data:

   - Land use polygons (with polygon identifiers and land use codes) from the land use map

- Floodplain boundary, from the floodplain and land use revision map
- Land use revisions, from the floodplain and land use revision map
- 100-foot contour lines, from the topocontour map
- Land use polygons within each red triangle, from the land use merge area map. (Each triangle's land use polygons to be separately digitized)

2) Convert digitized files to ODYSSEY file format

3) Plot outputs:

   - Plot all digitized maps at same scale as original (see Fig. 3.2-2a, b,c; Appendix B includes full scale plots)
   - Plot all land use polygons having a land use code of ACC. Allow for possible plot of other land use codes (see Fig. 3.3-3)
   - Plot all contour lines representing an elevation of 300 feet. Allow for possible plot of other 100 increment contour intervals (see Fig. 3.3-4

4) Plot and tabulate areas for individual land use polygons within the flood plain and below the 100-foot contour lines (see Fig. 3.3-5)

5) Plot and tabulate areas for individual land use polygons on land use map as affected by land use updates from the land use revision map (see Fig. 3.3.6a,b)

6) Merge land use polygon data digitized separately from each red triangle; on the photocopy; i.e., as though two adjacent map sheets were being merged

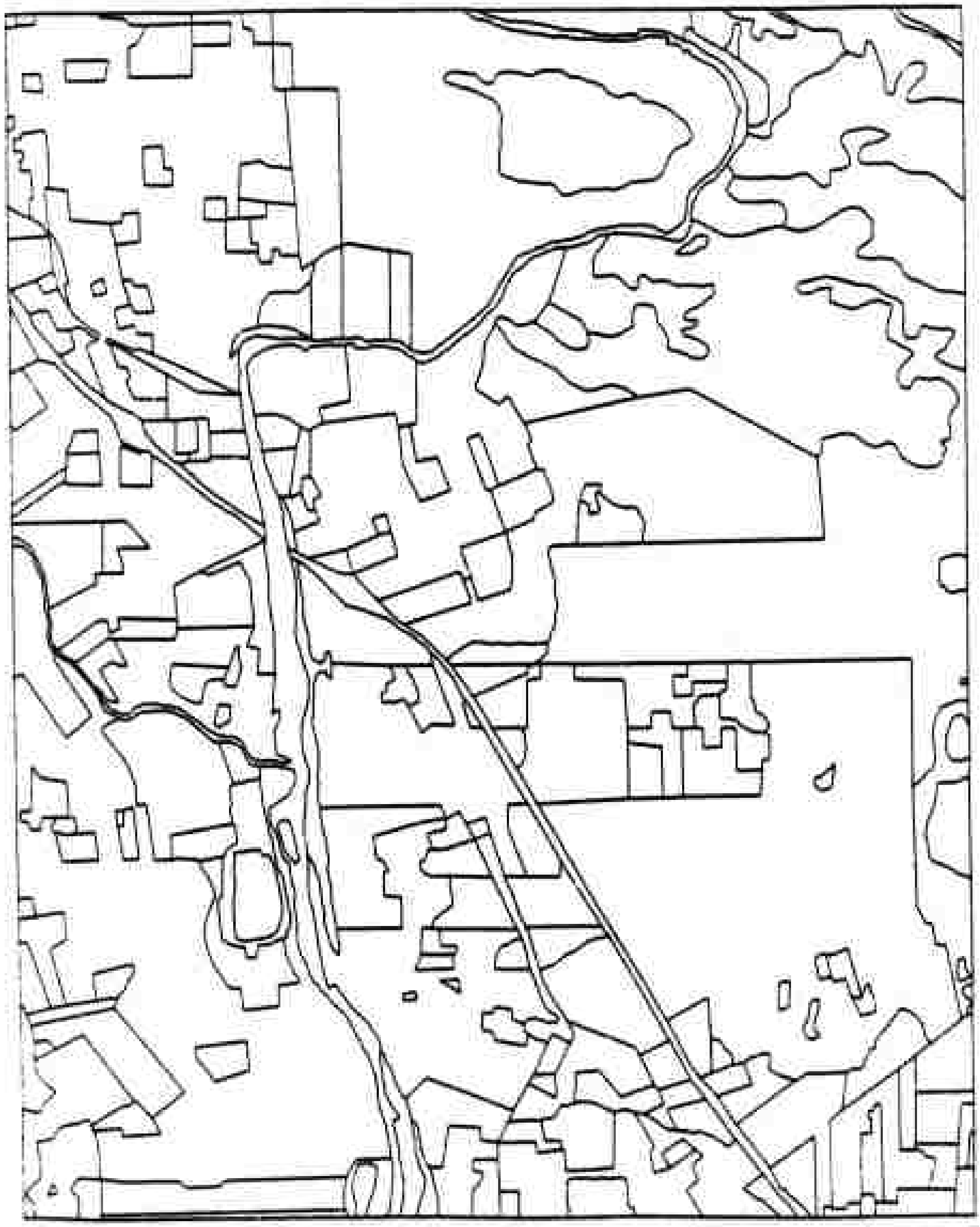


Figure 3.3-2a Plot of Digitized Land Use Map

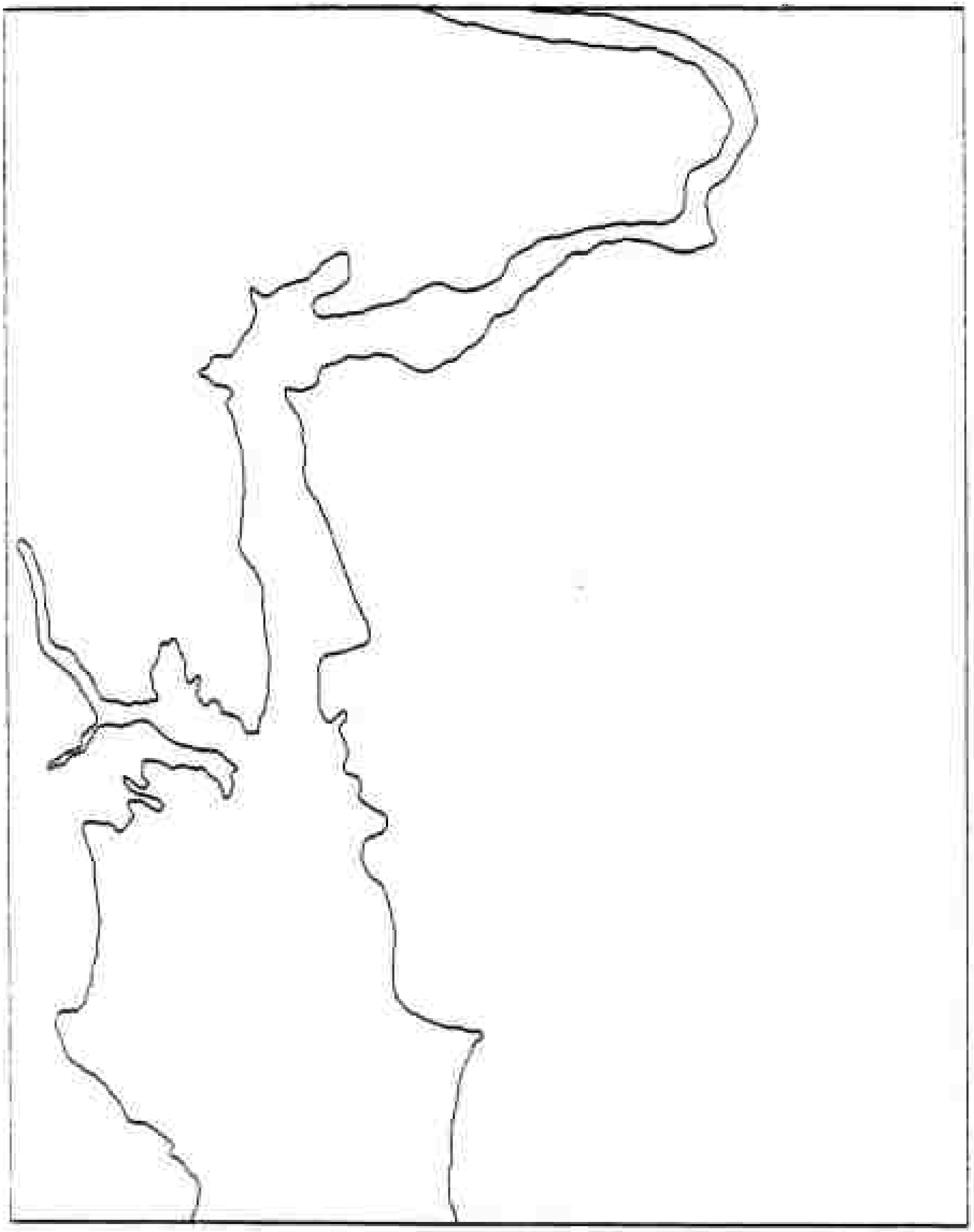


Figure 3.3-2b Plot of Digitized Floodplain Map

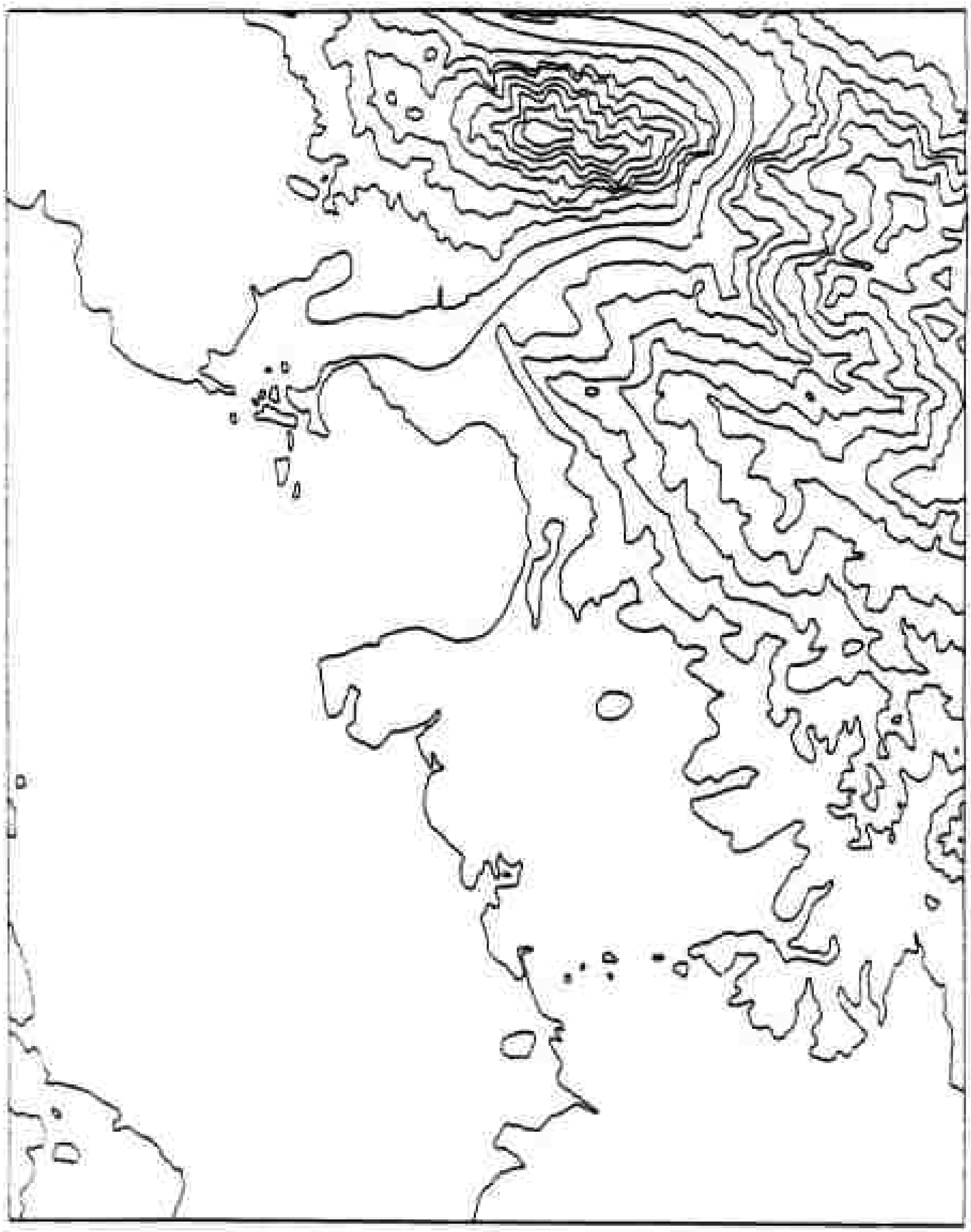


Figure 3.3-2c Plot of Digitized Contour Map

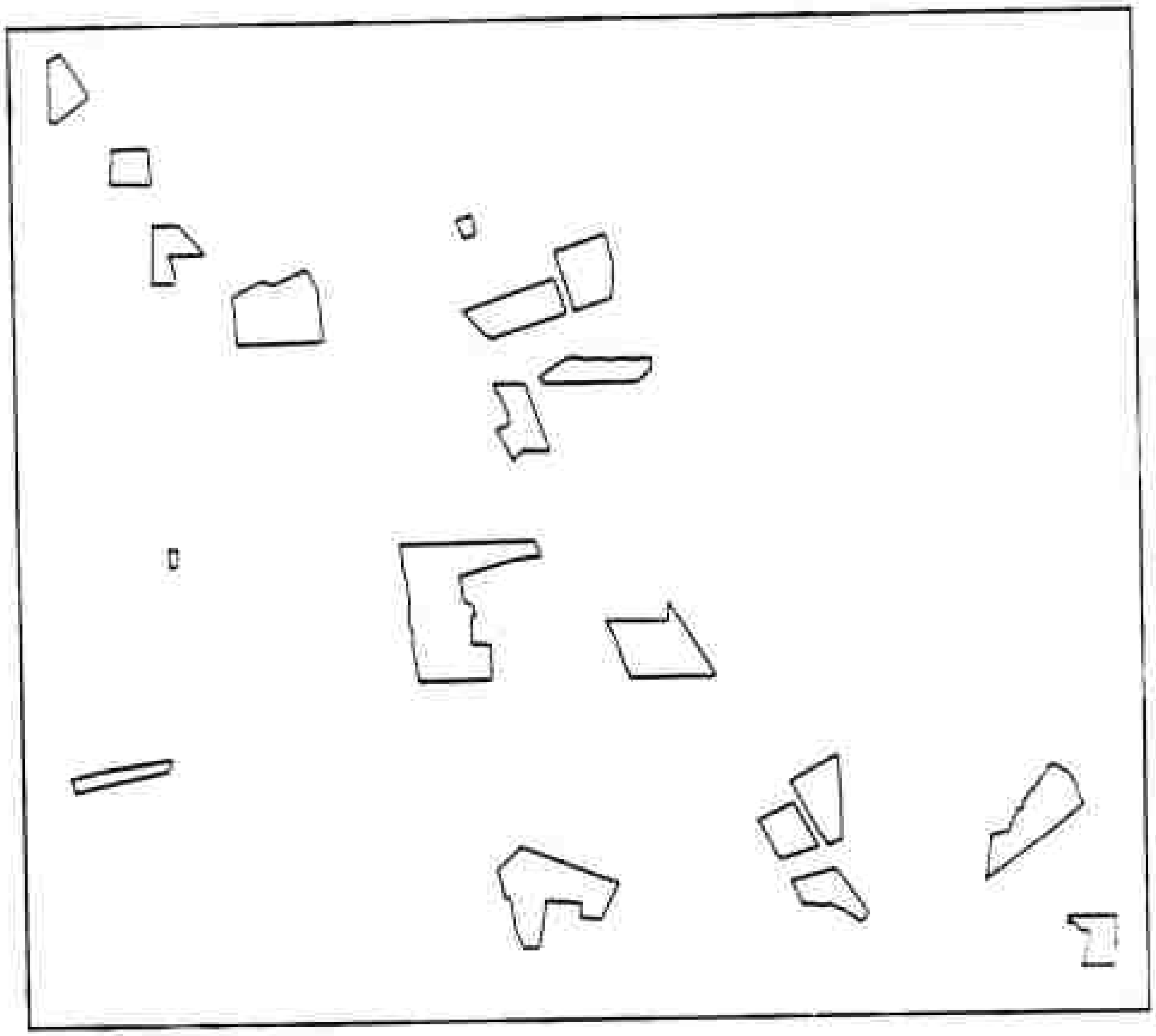


Figure 3.3-3 Plot of Land Use Polygons Having a Land Use Code of ACC

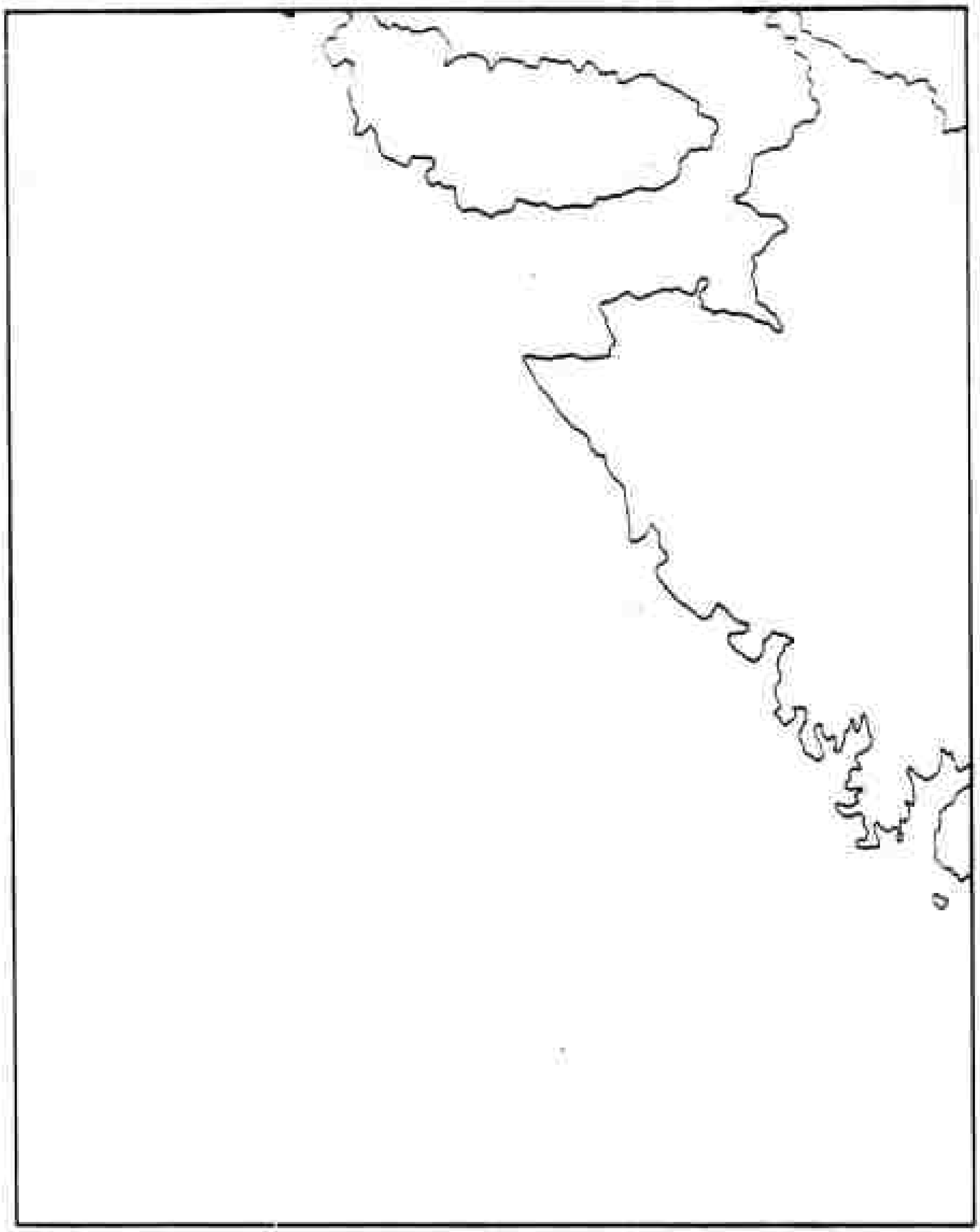


Figure 3.3-4 Plot of Contour Lines Representing an Elevation of 300 ft

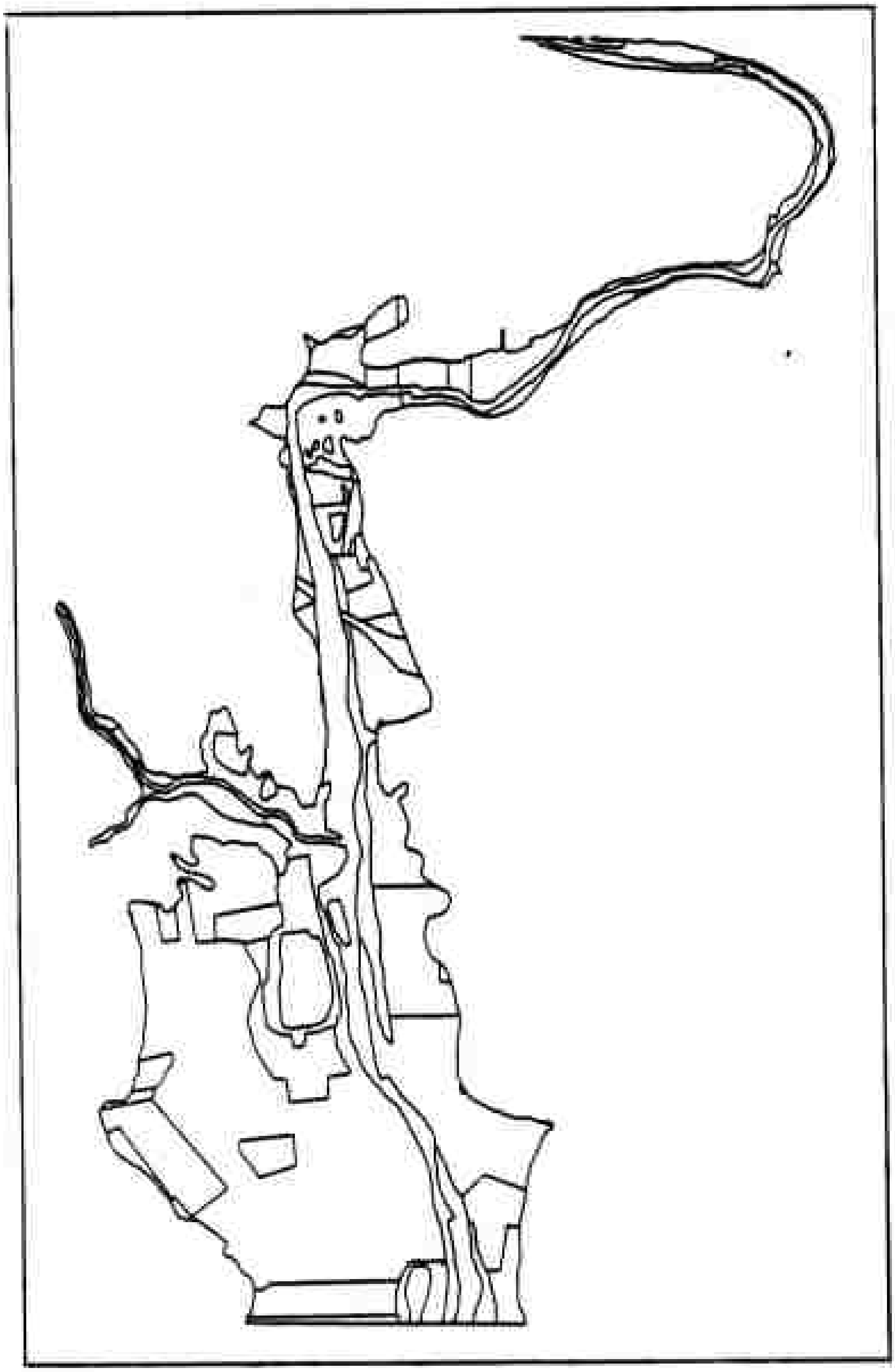


Figure 3.3-5 Plot of Land Use Polygons Within the Floodplain and Below the 100-ft Contour Line

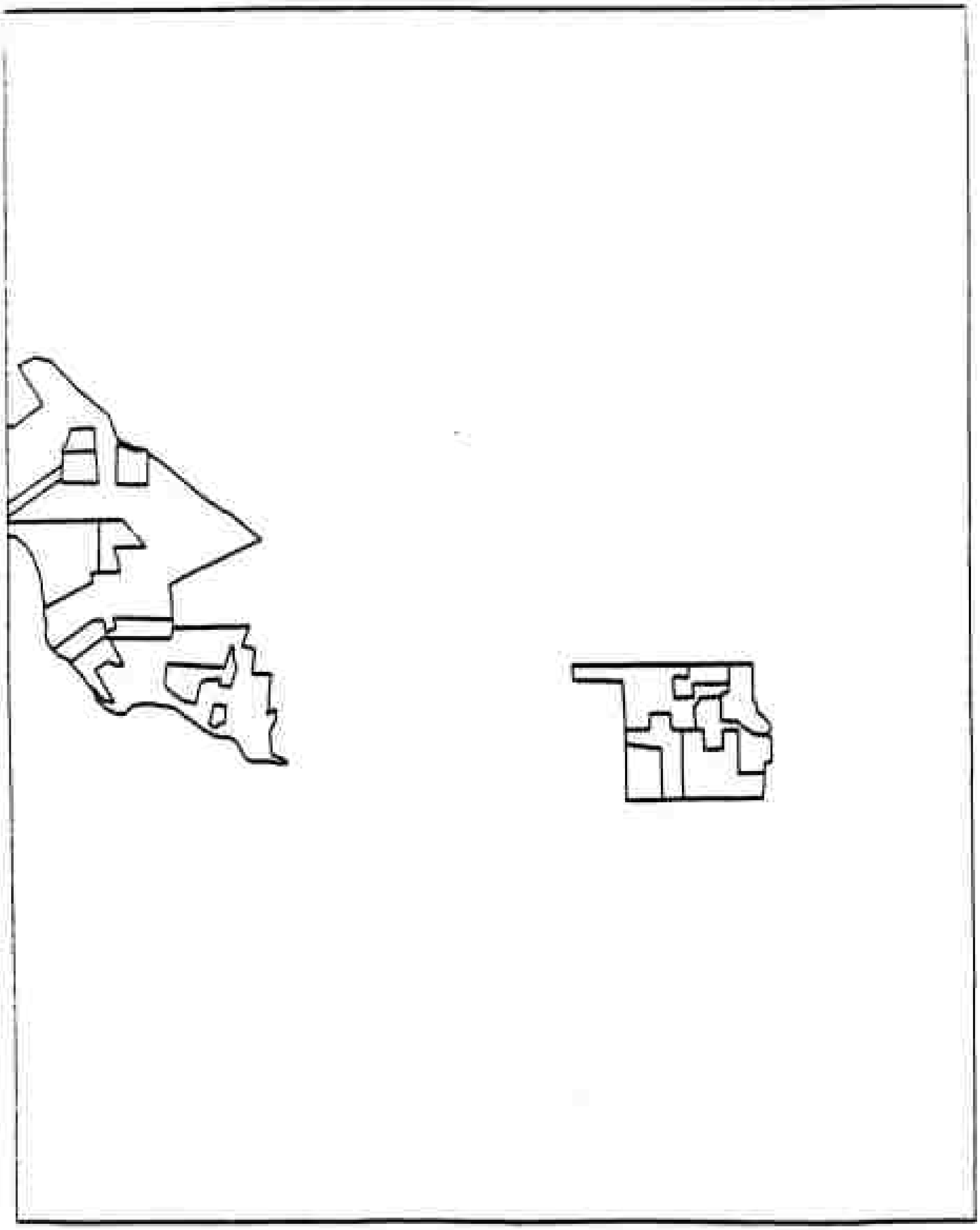


Figure 3.3-6a Plot of Land Use Polygons Affected by Updates

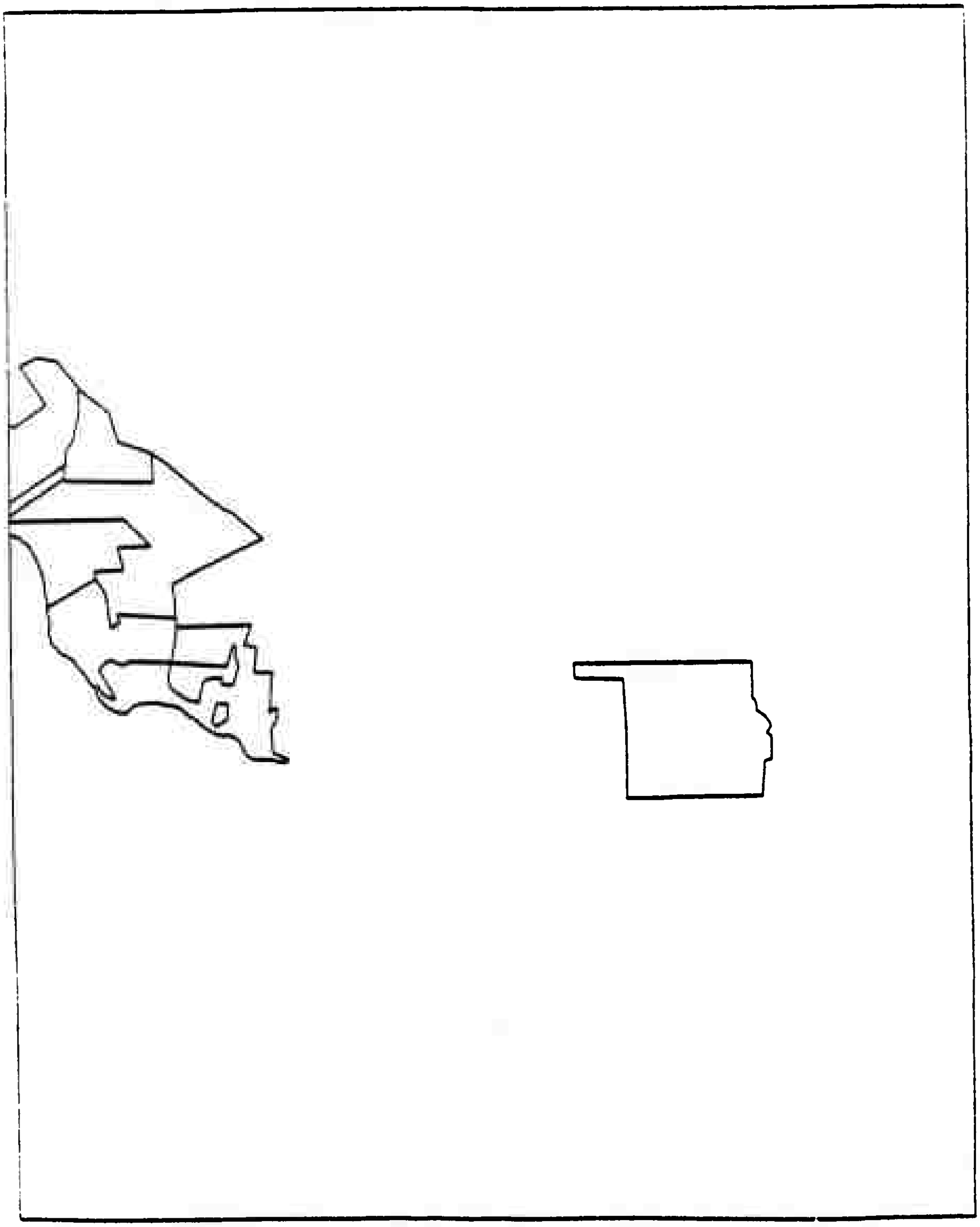


Figure 3.3-6b Plot of Affected Land Use Polygons After Updating

7) Plot and tabulate areas for individual land use polygons resulting form #6 above

8) Plot and tabulate land use areas resulting from merger of adjacent polygons having similar land use codes from #7 above

3.4 DOCUMENTATION OF PROCEDURES AND RESULTS

3.4.1 Map Encoding and Digitizing

The basic problem of data capture was to produce an edited, standard ODYSSEY digital file from the materials provided by ETL, as indicated in Fig. 3.4-1. The methods employed and described below do not represent an optimal production digitization and edit system. More sophisticated tools (e.g., PASTA (Ref. 22)) could significantly reduce the amount of manual labor involved in encoding source materials in the ODYSSEY format (see Section 3.5). The following discussion indicates the actual steps undertaken for this demonstration.

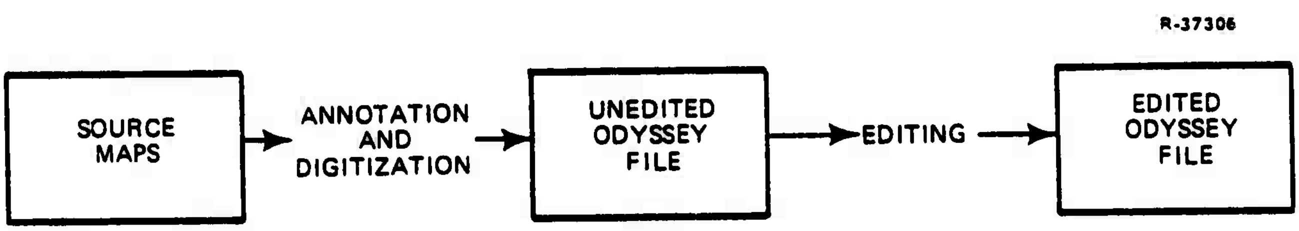


Figure 3.4-1 ODYSSEY File Preparation Steps

Map Annotation - This step consisted of manual preparation of the source materials for digitization (e.g., Fig. 3.4-2). This involved delineation of a study area boundary, identification of control points, and manual encoding of the map topology. The procedure used in annotation is as follows:

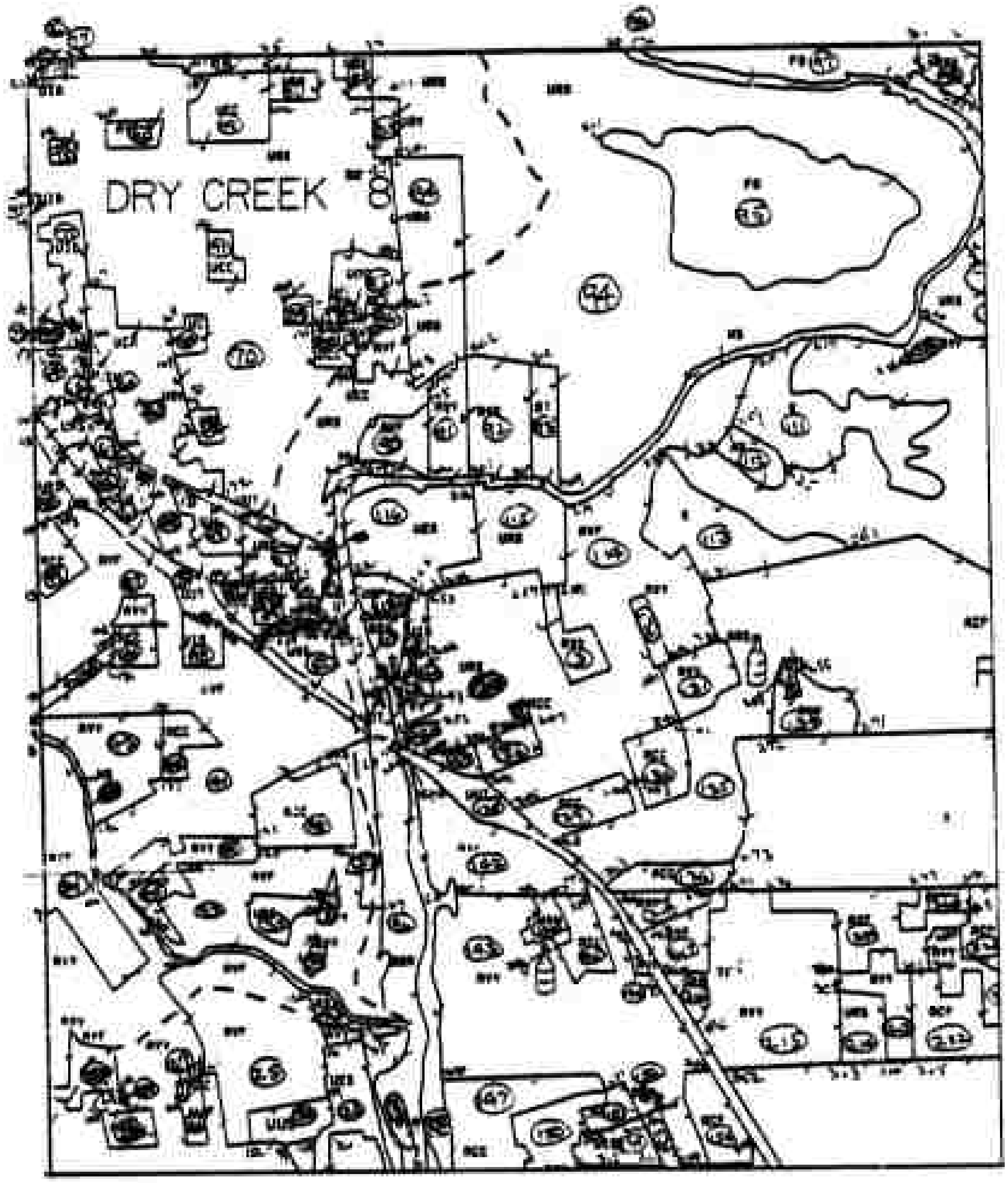


Figure 3.4-2 Portion of Annotated Land Use Map Prior to Digitizing

1) Number all nodes along the map border

2) Find and number a polygon

3) Find and number all nodes around the polygon

4) Repeat steps 2 and 3 until the entire sheet is annotated.

<u>Digitizing</u> - Following map annotation, the topology and coordinate map description were digitized and a standard ODYSSEY chain file was produced. This was accomplished using a Tektronix 4954 digitizing tablet controlled by a time-sharing DEC PDP-10 computer. For each digitized line segment (chain), the topology information (<u>from</u> node, <u>to</u> node, left polygon, right polygon) was keyed in at the terminal, then the coordinates describing the line were digitized using the tablet. Transformation from digitizer coordinates to UTM coordinates is accomplished by a least squares fit to UTM control points on the original map.

<u>Editing</u> - The input to the editing step was the series of point chain files produced as output from the digitization step. The first task in the editing step involved merger of all files of a given coverage (e.g., land use) into a single file, which was then smoothed to filter out the excessive detail produced by the digitizing tablet when operated in "stream" rather than "point" mode. Editing then proceeded through two phases: a coordinate edit and a topological edit.

Coordinate editing for this experiment was an iterative process. Each file was plotted to scale on a CalComp 11-inch drum plotter. This plot was visually compared with the source map. Spike errors (from bit errors in data transmission), duplicate chains, missing chains, and inaccurate line tracing were identified at this stage. A chain file editing program was then run to correct these errors. This program has the ability to delete entire chains, delete points within a chain, or replace a point. Chains seriously in error were deleted and noted for redigitizing. These new chains were digitized and merged with the existing file and the process repeated.

When the obvious coordinate errors had been removed from the file, the topology of the file was edited using the CYCLONE program. For each node in the input file, CYCLONE assembles all chains which reference that node. A node is considered error-free if the chains can be assembled such that there is a coherent sequence of left and right polygons around the node; such a node is called "cycled". Once all topological and coordinate errors were removed from the digitized files, they were passed to the data base construction stage of the process.

3.4.2 Data Base Construction

Overlaying the Land Use, Elevation and Floodplain Files - Overlaying of the edited files was performed in three steps:

1) Preprocessing each of the three files

2) Overlay of the Land Use file and the Floodplain file

3) Overlay of the Topography file and the combined Land Use-Floodplain file (output from step 2).

Preprocessing the Files - This consisted of running a program called SLICE on each file. SLICE takes each chain which contains more than a specified number of segments and cuts the chain into smaller chains with fewer segments. In preprocessing each of the three files, every chain with more than one hundred segments was "chopped" into shorter chains using SLICE. This preprocessing step is not required to perform polygon overlay. It was done in this instance in order to decrease the amount of core memory required in the subsequent overlay process.*

Overlaying the Land Use and Floodplain Files - The following process is performed for each pair of files to be overlaid. First, the input files are separately sorted on the Y-axis from south to north. PSYCHE, a special program which sorts ODYSSEY format files, is used for this phase. Once the files have been sorted, they are input to the overlay program, WHIRLPOOL (described in Ref. 20), for the second step. This produces an ODYSSEY file of the combined coverages. All polygon intersections between the two input files are detected, their intersecting chains are broken, and resulting new chains are assembled into polygons and topologically verified.

Statistics generated by the overlay program indicate that the interaction between the files was as expected, following results derived from past experience overlaying coverage files. The average number of resident chains during the overlay process was slightly less than twice the number of output chains, and the number of output chains was slightly less than twice the number of input chains. (The implications of this normal ODYSSEY behavior are addressed in Section 3.5).

Overlaying Topography and Combined Land Use Floodplain - To perform further overlays the two-phase process is repeated again, as many times as required to produce the desired output. In general, to overlay n coverages requires $n-1$ operations of the overlay process. The results of the overlay process for the three files used in this experiment are shown in Fig. 3.4-3.

*Long chains (300 or more segments), when they are approximately parallel to the direction of sort, occupy a great deal of memory due to their large numbers of coordinates. The coordinates are distant from the band-sweep intersection activity (Section 3.2.1).

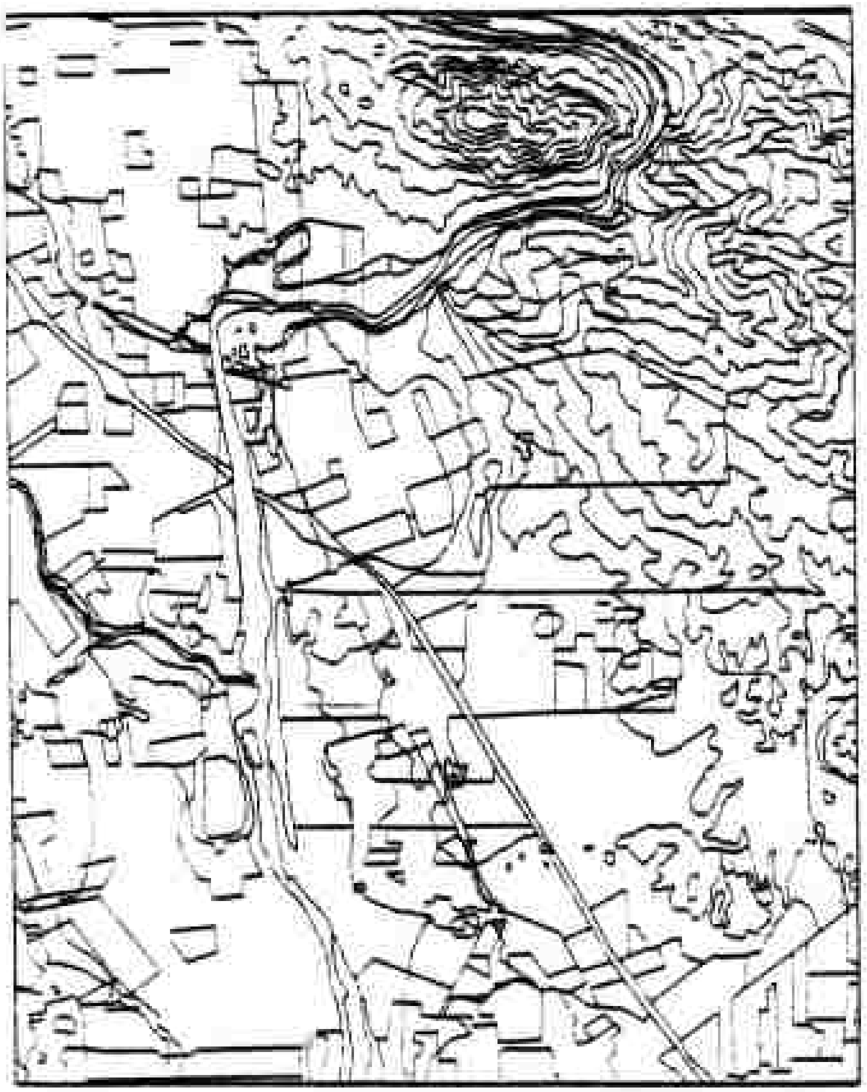


Figure 3.4-3 Overlaid Land Use-Elevation-Floodplain Data Bases

3.4.3 Revising the Land Use File

The revision task consisted of modifications of the original land use file to incorporate changes in the study area as shown on the land use revision map. These modifications included:

- Boundary changes of existing polygons
- Deletion of polygons
- Creation of new polygons.

The revisions affected both the coordinates and topological content of the original land use file.

The land use revision task was accomplished using the same basic tools and processes as for data base creation. For this reason, the revision task as performed involved more manual labor than would be necessary or desirable in a large scale production system. Other approaches to the revision task, based on topological data base concepts and using proposed ODYSSEY software or a modification to the WHIRLPOOL process, could automate many of the procedures used in this project.

Using the editing tools available for data base creation, the revision process was performed in three operations: (1) identify polygons affected by revision, (2) update the land use chain file to reflect revisions, (3) correct any coordinate or topological errors arising from the modifications.

The polygons affected by the revision were identified by a manual inspection of the revision overlay source map (Fig. 3.3-1b) together with the annotated land use source map (Fig. 3.4-2). This resulted in a list of affected polygons. The FRED file updating software was then used to produce a listing of the chains which describe the affected polygons, plotted in Fig. 3.3-6a. The listing was manually analyzed and annotated with the necessary topological and coordinate changes - the same operations as were used for the original map encoding and digitizing. At this point, a list of new chains required was also made. The FRED program was used to make the necessary changes to the original land use file. New chains were digitized using the digitizing procedure described in Section 3.4.1 and merged with the edited files; the resulting revised land map is shown in Fig. 3.3-6b. Finally, the CYCLONE software was used to verify the topology of the revised file and to coalesce redundant nodes created by the revision process.

The entire revision task required approximately 3 man-hours and 15 seconds of DEC PDP-10 CPU time.

3.4.4 Merging Adjacent Map Sheets

The WHIRLPOOL program can be used to merge map sheets as long as two simple requirements are met:

- The adjacent map sheets must be coded in a common coordinate system
- The same polygon numbering system must be used on both map sheets.

Given that these two requirements are met, the merger of the map sheets is performed in a manner identical to the overlay of two data coverages for a single map sheet. Each map sheet file is sorted so that its chains are in order of ascending minimum y-coordinate. WHIRLPOOL is then used to merge the sheets. After the files have been opened by WHIRLPOOL, but before the overlay process command is executed, the command "MAKE SINGLEFILE" is given. This alerts WHIRLPOOL to the fact that identical polygon identifiers on separate files refer to the <u>same</u> polygon, and causes it to alter its algorithm for naming output polygons.

A problem which can occur in sheet merging is that the sheets will not exactly line up at the edge. Portions may overlap slightly or small gaps may appear. These may be due to digitizer errors or small errors in coordinate system registration. WHIRLPOOL automatically handles small overlaps if the overlapping polygons have the same identifier. Small gaps can be handled by specifying a proximity tolerance, i.e, a distance defining a neighborhood within which two distinct points are to be considered by the same. By choosing a proximity tolerance equal to the width of the widest gap, WHIRLPOOL will close the gaps. However, if the gaps are large compared to the local detail, that is, if there exist chains or coordinates which should remain distinct but which are closer together than the specified tolerance, then use of the proximity option will produce errors. The distinct points <u>will</u> be merged if they are on separate sheets, and <u>may</u> be merged if they are on the same sheet and close to the sheet border.

The WHIRLPOOL procedure was followed in merging the land use polygons contained within the two triangles on the original land use map (see Fig. 3.3-1a). An initial run contained gaps due to digitizer imprecision (see Fig. 3.4-4). These gaps were eliminated by using a proximity tolerance of five meters. This was done by specifying "MAKE PROXIMITY: 5" before executing the process command in WHIRLPOOL. The resulting merged red triangles are shown in Fig. 3.4-5.

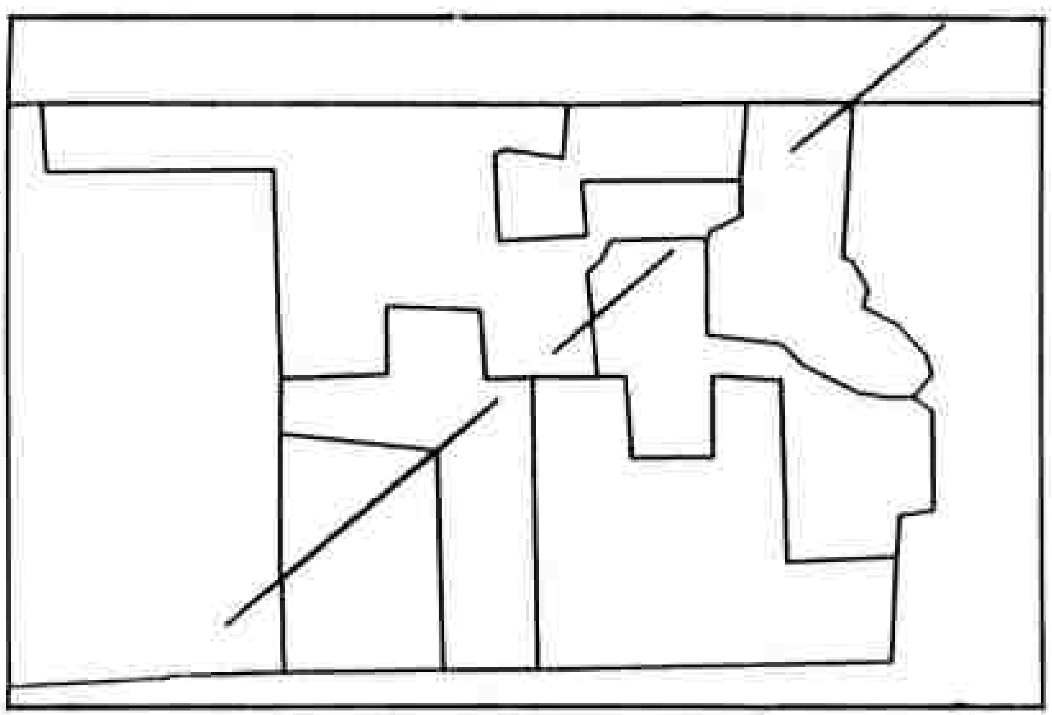


Figure 3.4-4 Failed Land Use Merger Due to Digitizer Imprecision

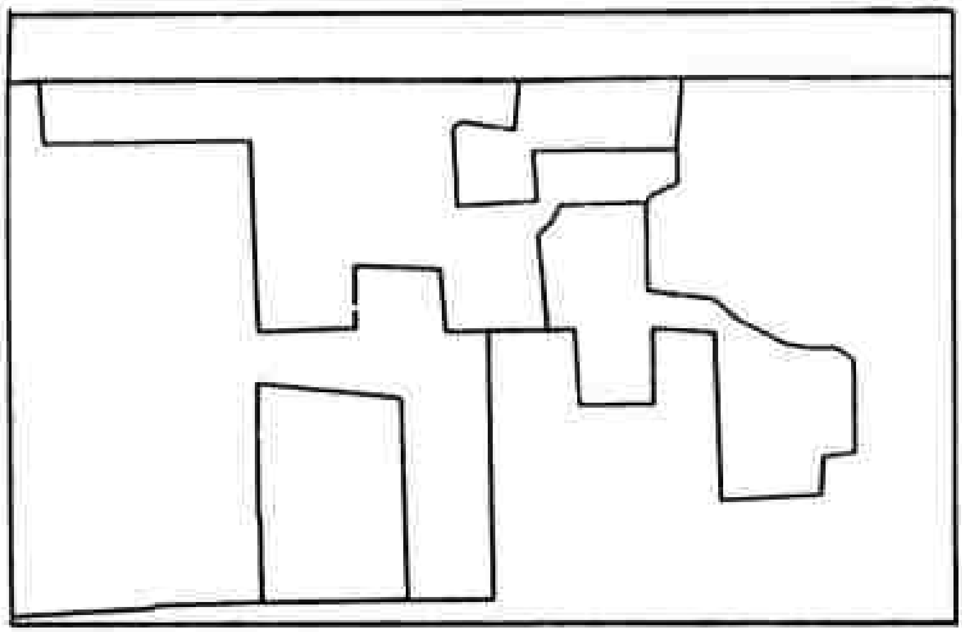


Figure 3.4-5   Successful Land Use Merger

3.4.5 Data Base Retrieval and Display

The bulk of the demonstration for ETL was concerned with selective retrieval from the overlaid data base established in the previously described process. Retrieval from such an overlaid data base is centered around the selection capability. The user may specify a subset of each coverage to be included using Boolean set notation. The inclusion set can be manipulated with union, intersection, and complement operators, and the resulting set may be saved with user-defined labels for use in subsequent selection expressions. Features from each overlaid coverage can be requested in a selection statement which specifies the area of interest, using the operators AND and OR (evaluated left to right) to define the resulting combination of coverages. For example, all three ETL-provided coverages may be mentioned in making a selection.

In addition to defining an area of interest, commands can be used to retrieve detail within that area. Each coverage may be chosen to be aggregated or presented in detail; this additional specification is independent of the selection set. When all coverages are set to aggregate, the zone of interest is retrieved as a single region. Plots will show the entire region, which may fall into any number of disjoint polygons. Similarly, only one value for the area will be tabulated.

When detail for a coverage is specified, all codes for that coverage which are present in the area of interest will be displayed as separate regions and as subtotalled areas. Some portions of the area of interest may include elements of a coverage which are not specified in an "OR" statement. These codes will be aggregated into a "nonselect" classification.

Control of detail is also extended below the level of the coverage classifications to the individual polygons created through overlay. The identifiers in the data base reflect the parentage of a polygon on the original coverages. This capability allows a variety of requests to be serviced through a simple modular, consistant language. The following discussion introduces that language.

Selection and Aggregation (Verbs: SELECT, INCLUDE, EXCLUDE, FORM) - The central ability of the display and analysis package is the ability to handle subsets of the data base. Subsets are specified through combinations of attributes of the geographic features represented in the composite data base, which for the ETL experiment consists of three different coverages of polygonal data. The coverages are referred to through sets named:

- LANDUSE
- ELEVATION
- FLOODPLAIN

The elements of these sets are the polygons shown on the original individual coverages. Note that elevation polygons consist of land between two successive contour levels. These are specified by the height of the lower of the contour pair, using the prefix "E". For example, the E100 zone defines the elevation band encompassing terrain heights in the range 100-199 ft. The land over 1000 ft is specified by E1K.

Coverage Sets - Each of the three coverages in the total data base is exhaustively partitioned by its attribute polygons, and no area belongs to more than one element of each coverage. Set operations (including complementation) on the coverages are therefore permitted. The mutually exclusive nature of elements within each set imposes natural limits upon the manipulations possible, e.g., there can be no polygon which is described by LANDUSE elements AVV and AVF. Hence, "coverage sets" can be selected from the coverages by combinations of inclusive OR operations. A given coverage set takes on values specifying which elements are "on" and "off." These values can be manipulated through set operations described above. Examples:

1) SELECT ELEVATION: [E0] ;

   makes the coverage set ELEVATION specify those areas below the 100-foot contour, i.e., between 0 and 100 ft. No other elevation zones are included in the coverage set.

2) SELECT ELEVATION: [E200, E400] ;

   makes the coverage set ELEVATION include both the zone from 200-299 and the zone 400-499 ft.

3) DEFINE URBAN: [URS, URH, UCW, UCR, UOC, UOG,...];

   DEFINE OPEN: [UOG, UOC, WS, BBR, R];

   These two sentences establish user-defined variables which contain a particular value for the LANDUSE coverage. These variables may be used directly to provide a value for the coverage in the place of temporary set assignment. This option provides easier interaction. For example,

   SELECT LANDUSE: URBAN;

   makes the current state of the land use coverage assume the value stored in the user-specified variable, URBAN, defined above.

4) User-defined coverage variables, original coverage variables and explicit sets may all be incorporated in set operations:

   SELECT LANDUSE: URBAN & # OPEN;

   makes the land use coverage contain those elements which are in the URBAN subset, and <u>not</u> in the OPEN subset. This manipulative capability avoids the necessity for element-by-element description of the coverage of interest.

   (& stands for inclusive OR and # for the complement option; read "NOT").

The language of this package provides the user with two procedures for altering the values of coverage sets. The procedure illustrated above uses set notation and the SELECT verb. For convenience, direct access to the individual elements of each coverage also is provided through the use of verbs INCLUDE and EXCLUDE. These two verbs affect only the elements mentioned in the sentence.

5) Example:

INCLUDE E0

puts the element E0 in ELEVATION without regard to other elements. This is the same operation as

SELECT ELEVATION: ELEVATION % E0

(% stands for inclusive OR)

The use of INCLUDE/EXCLUDE may be combined with the more flexible set notation system as needed by a user.

Selection - If a data base consists of a single coverage, then that single coverage defines the extent of the area of interest. With the overlay of many coverages, the complexity increases. An area of interest may be specified as before, based on a single coverage set, or it may be specified terms of combinations of different coverages. The SELECT verb has as its primary purpose the specification of the combination of coverages forming the area of interest. The SELECT verb, whenever used, replaces the previously defined combination with a new one.

The simplest selection sentence is "SELECT ALL;". This selects the whole data base to allow tabulations and plots unrestricted by selection. When using the complete inclusion of all the study area, the states of the various coverage sets remains unaffected. Particular attention to the FORM verb should be exercised while selecting ALL.

A coverage combination may mention any or all coverages. The coverage sets are combined by the operators AND and OR as required by the user. The expression formed is evaluated from left to right, and no parenthetical expressions are allowed.

The combination of coverages given in the SELECT statement specifies a total zone of interest. All areas of the data base which fit the current selection combination, given the current coverage sets, become part of the selected portion of the data base. (This is the area drawn (plotted) by DRAW and tabulated by TABULATE, as discussed later.)

Within the window defined by SELECT, different features of the data base may be requested using the verb FORM. FORM controls the aggregation of each overage independent from its selection state. A coverage may be defined as either AGGREGATE or DETAILED. (The set called GROUPING, available through the verb SHOW, contains the current state of this operation.)

Aggregation of a coverage is useful to remove its detail from a report or clutter from a map. For instance, if a coverage set of urban use has been created and the set intersection of LANDUSE and FLOODPLAIN = WET is selected, a map plotted with aggregate landuse will remove the detailed landuse divisions inside the selected region, allowing the general location to appear more clearly. By contrast, it is possible to use the same specification of selection with detailed landuse to tabulate areas of specific urban land use in the flood risk zone. The choice of aggregation or detailing is distinct from the matter of selection and closely linked to the particular product desired by the user. When a coverage is defined as AGGREGATE, no internal features appear within the selected zone. When all coverages are AGGREGATE, the selected zone becomes the only region to draw or tabulate.

The most detailed output is invoked by "FORM DETAILED INSTANCE" and turned off by "FORM AGGREGATE INSTANCE." This allows the user to obtain area tabulations for each of the least common geographic units (LCGUs) created by the overlay. LCGUs are named by a label indicating the LCGU parentage from the basic coverage files.

Tabulation of Polygon Areas (Verb TABULATE) - One of the primary functions of the display and analysis package is reporting areas for user-defined regions in the overlaid data base. The TABULATE verb calls on the program to list areas according to the current selection and aggregation control. The tabulation is generated in the form of a hierarchical table, in which each coverage is mentioned with progressive indentation, and the innermost coverage is varied the most rapidly. If only one coverage is detailed, it will be the innermost, and its element codes will appear on the tabulation. If a coverage is aggregated, it will only appear as AGGREGATE LANDUSE (or whatever) on the listing. For all coverages, regardless of aggregation, a NONSELECT region may appear, i.e., if all coverages in the selection were joined by AND, then the NONSELECT region is the total area outside the selection zone. When OR is used, NONSELECT portions can be mixed in with selected portions of other coverages. Examples of tabulation with differing selection and aggregation appear in Appendix B.

The TABULATE sentence can have two forms:

1) TABULATE; uses current units (default is ACRES)

2) TABULATE IN units;

   where units may be one of the following:

   ACRES
   HECTARES
   SQUARE MILES
   SQUARE KILOMETERS
   SQUARE METERS

The units which appear on the tabulation may be specified in the sentence or left at the last specified value.

<u>Additional Tabulation Controls</u> - Tabulation output has further controls which are adjusted using the MAKE or ROUTE verbs.

MINIMUM: This parameter specifies the minimum size area, in the current units, which will appear on the listing. This value is only effective when greater than zero. The default value is 0.01, the limit of the output format used. This parameter provides a useful filter to avoid cluttering the analyst's listing with insignificant areas. If set to zero, the filtering is disabled. Note that MINIMUM is always interpreted in current units, i.e., acres or square miles. For example:

```
MAKE MINIMUM 0.05
```

PRINTER UNIT, or FILE: Available through the ROUTE verb. The tabulation output is directed to the FORTRAN logical unit given as PRINTER UNIT. Where applicable, PRINTER FILE gives a file name for the output. For example:

```
ROUTE PRINTER UNIT: 6;
```

<u>Routing of Plotter Output (Verb ROUTE)</u> - Using the ROUTE verb the DRAW output may be prepared for a number of different devices. The default device is set at each installation depending on hardware availability. The possibilities for DEVICE are:

CALCOMP (an 11-inch drum plotter)

TEK4010 (smaller screen, 1024-address Tektronix CRT display)

TEK4014 (larger screen, 4096-address CRT display).

<u>Control of Window (Verb MAKE)</u> - By default the DRAW output presents plots which fit the whole data base onto the current device. Control is available to specify this process more carefully if required. All these parameters are accessible using the verb MAKE.

> WINDOW: Contains four numbers specifying the X, Y center of the window in meters on the ground and the X, Y lengths of the window. After the file opening has occurred, this is set to the value which will contain the whole data base.
>
> MAKE WINDOW: (X center, Y center,
> X length, Y length)
>
> ANGLE: A single number given in degrees specifies a rotation of the data base coordinates around the center specified in WINDOW.
>
> VIEWPOINT: Contains four numbers specifying the desired size of the drawn output in display units. If this remains zero, the maximum size of the device is available. As with WINDOW, VIEWPORT is given as (X center, Y center, X length, Y length).
>
> SCALE: A single number used to convert meters to display units. If zero, the WINDOW is fit to the VIEWPORT. When nonzero, it converts the meter values in the data base to display units without regard for VIEWPORT, up to the limitations of the plotting devices.

Error checks are provided for these parameters; if an error exists a message is printed and no plot is produced.

<u>Controlling a Run</u> - The language of the display and analysis package provides for control of execution. The OPEN verb provides access to different files during a program run. Since the reassignment of files is troublesome under CDC FORTRAN, the OPEN verb is not useful in the delivered CDC version. File access is instead performed on the SCOPE control cards and

provided to the package as positional parameters on the execution of the program.

The other control of execution is termination. The QUIT verb stops execution, closes plotter files if open, and returns control to the monitor. A QUIT statement should be at the end of every run.

Additional Language Features - The language of the display and analysis package affords a number of features which enable a user to develop more accurate control and reduced clerical work. The features described below are generic to the language system and available also with the other programs delivered. They are primarily useful in an interactive mode.

SHOW - The SHOW verb prints the current value of a parameter on the output listing. This is particularly valuable in the interactive mode on the DEC PDP-10, but is also can be used to document the activity of a batch run. SHOW takes as an object any entity in the language, such as coverage sets, individual coverage types, parameters and user-defined values and sets.

CLEAR - This verb takes no argument. It returns the program to its initial state. This is useful if current values of parameters are suspect. A particularly good use of the CLEAR verb is to separate runs submitted by separate workers or trainees. Each run will be executed as if it were run independently, but saving the overhead cost of restarting the program.

RESTORE - The RESTORE verb allows more selective re-initialization than CLEAR. RESTORE applies to all parameters accessible through SHOW. This returns that particular object to its initial state.

DEFINE - The DEFINE verb allows the user to create objects of his own. These objects have a name (an alphabetic string up to 31 characters long) and a value. The value may be of any type available such as a number, a string, or a list of types from a coverage set. A list of numbers may be enclosed in parentheses with commas between. The list of coverage types must be enclosed with brackets [,] and separated by commas.

The objects created by DEFINE may be included in any subsequent value expression where appropriate.

DELETE - This verb is used to get rid of a user-defined object.

3.4.6 ODYSSEY Data Files

Both the Chain File and the Polygon File are implemented using the ODYSSEY file system (Ref. 21). Every ODYSSEY file has two basic parts: (1) the data part and (2) the global part. The global part contains data within a single record summarizing the form and content of the file.

The data part consists of a set of records which may be accessed sequentially. Each data record describes a single chain or polygon. Each data record contains two sections. The first is called the fixed section, since it has a fixed length through the entire file. The fixed section is subdivided into groups of elements. The second part of the record is the variable length section. This section contains a single element, called the variable element, but this element may be repeated an arbitrary number of times in each record. Each element contains either an integer, real, or character value. Fig. 3.4-4 illustrates this general ODYSSEY file structure. The following discussion delineates the specifics of the Chain and Polygon files.

```
File
    Global Part (1 record)
         global summary information
    Data Part (1 record/chain or polygon)
        1st Record
          Fixed Section
            Group
               Element
                 .
                 .
                 .
            Group
              .
              .
              .
          Variable Section
            Variable Element
        Next Record
          .
          .
          .
          End of File
```

Figure 3.4-6 ODYSSEY File Structure

Chain Files - A Chain File contains the basic topological and geometric description of a cartographic data base. It consists of a sequence of chain records. The fixed section of the chain record contains the topologic definition of the chain as well as optional statistics summarizing the chain, while the variable section contains the geographic coordinates of the chain. The layout of the fixed section is shown in Table 3.4-1.

TABLE 3.4-1
FIXED SECTION OF ODYSSEY CHAIN FILE

T-2478

INDEX	NAME	TYPE	DEFINITION		
1	ID	1	the chain identifier	Digitizer Software (DIGIT) Output	Overlay Software (WHIRLPOOL) Output
2	NPNTS	1	the number of points		
3	FROM	1	the from node identifier		
4	TO	1	the to node identifier		
5	LEFT	1	the left node identifier		
6	RIGHT	1	the right node identifier		
7	LENGTH	2	LENGTH OF THE CHAIN		
8	AREA	2	area under the chain		
9	XMIN	2	COORDINATE BOX CONTAINING THE CHAIN		
10	YMIN	2			
11	XMAX	2			
12	YMAX	2			

By convention, the length of the chain file's fixed section is variable, defined in the global part as NFIXD, and may be six elements, 12 elements, or 14 elements long. Most programs require six element headers and some, such as WHIRL-POOL (the overlay program) require 12-word headers. The variable section of the chain file contains the chains coordinates.

The use of ODYSSEY files in this demonstration task, the data flow, and the order of the processing tasks are shown in Fig. 3.4-5.

3.4.7 ODYSSEY Software Modules Delivered to ETL

The following ODYSSEY software modules were delivered as object code for installation on the CDC 6400 computer at ETL.

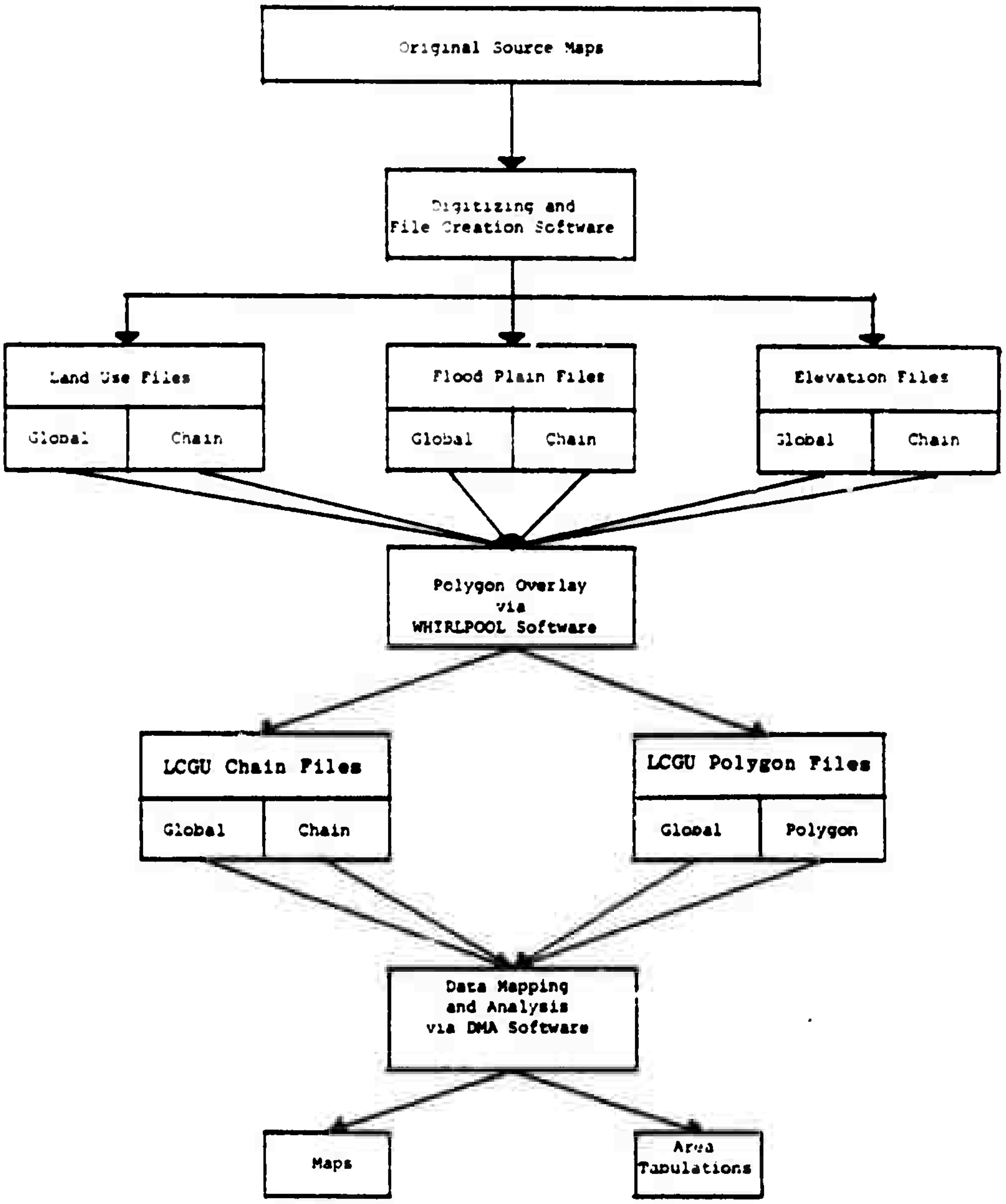


Figure 3.4-7 ETL/ODYSSEY System Overview

Module DMA

Purpose: Selection, area tabulation, and graphic display.

CDC Job Control Language Required to Execute: DMA (polygon file, polygon global file, chain file, chain global file).

Plotter Output: Appears on file named TAPE29.

Storage Location at ETL: File 1 on library software tape.

Module FRED

Purpose: Correction and update of chain files.

CDC Job Control Language Required to Execute: FRED (input chain file, input global file, output chain file, output global file).

Storage Location at ETL: File 2 on library software tape.

Module POLLY

Purpose: (1) Produces before and after plots and polygons affected by chain file modifications made with FRED.

(2) Tabulates and prints out area calculations and perimeter error estimates for polygons.

CDC Job Control Language Required to Execute: POLLY (chain file one, global file one, chain file two, global file two). If only area tabulations are desired, then POLLY (chain file one, global file one).

Storage Location at ETL: File 3 on library software tape.

Module PSYCHE

Purpose: Sorts chain files on their y-axis coordinate. Used between WHIRLPOOL steps when overlaying more than two coverage files.

CDC Job Control Language Required to Execute: PSYCHE (Input chain file, input global file, output chain file, output global file).

Storage Location at ETL: File 4 on library software tape.

Module WHIRLPOOL

Purpose: Overlays two data coverage files (red and blue), and produces a composite output file (purple).

CDC Job Control Language Required to Execute: WHIRL (red chain file, red global file, blue chain file, blue global file, purple chain file, purple chain global file, purple polygon file, purple polygon global file).

Storage Location at ETL: File 5 on library software tape.

3.4.8 Data Tape Delivered to ETL

The Data Tape consists of the digitized data generated at Harvard and delivered to ETL. This tape contains chain file data for the three original coverages in two forms: (1) original digitized data, and (2) a second set of data where all coordinates have been rounded to the nearest 10 meters in an attempt to remove trivial polygons created during the overlay process.

Each coverage on the tape is described on two files, a binary chain file in packed format plus a "global" file as shown in Table 3.4-2. The global file for a particular chain file gives characteristics of the whole chain file, such as organization, coordinate box, and counts of entities. Global files are documented in the description of the ODYSSEY file system (Ref. 21).

TABLE 3.4-2
DELIVERED DATA TAPE FILE ORDER

T-2479

FILE #	DESCRIPTION	SIZE*
1	Land Use, Original Chain	35B
2	Land Use, Original Global	5C
3	Floodplain, Original Chain	7B
4	Floodplain, Original Global	5C
5	Elevation, Original Chain	37B
6	Elevation, Original Global	5C
7	Land Use, Rounded Chain	30B
8	Land Use, Rounded Global	5C
9	Floodplain, Rounded Chain	5B
10	Floodplain, Rounded Global	5C
11	Elevation, Rounded Chain	26B
12	Elevation, Rounded Global	5C

*Where C = Card Image, B = Blocks of 510 Words Each.

3.5 ASSESSMENT OF TOPOLOGICAL DATA BASE CAPABILITY AND RECOMMENDATIONS FOR FURTHER RESEARCH

3.5.1 Capabilities and Resource Requirements

This project has demonstrated that a topological data base offers a number of highly desirable capabilities for the creation, analysis and display of polygonal geographic information. The major advantages of such an approach include:

- Automated error detection and correction during data base creation
- Integration of two or more data coverages for a given region into a single composite data base
- Selective retrieval and display of information from a composite data base
- Local versus global data processing techniques which would allow implementation of this approach in a mini-computer environment.

The equipment used to perform these tasks is described in Table 3.5-1. Table 3.5-2 describes the type and number of data elements contained within each data base. The computer resources, execution time and manual labor required are described in Tables 3.5-3 through 3.5-5, respectively. Given the size and complexity of the total data base (17,884 points), the total of 33 man-hours used to create and edit the complete data base was exceptionally low. This was possible only because of the automated error detection and correction procedures possible within a topologically-structured data base. It is particularly important to note that these manpower requirements can potentially be reduced by an order of magnitude given the use of automatic digitizing procedures rather than manual procedures such as were used in this experiment.

The composite (i.e., LCGU) data base characteristics are described in Table 3.5-2. A total of 958 LCGU polygons resulted from the overlay of the three files. Note that the intersection of the three files created additional chains, nodes and points due to the breaking of original chains at each point of intersection. The total number of additional points created amounted to approximately 20% of the three original files combined. Much of this increase reflects the difficulty

TABLE 3.5-1
EQUIPMENT USED

T-2480

Digitizer:	Tektronix 4954 with resolution of 0.01"
Computer:	DEC PDP-10 at Harvard University, KA-10 processor
	CDC 6400 at Smithsonian Astrophysical Observatory
Terminal:	Tektronix 4014 Display
	Tektronix 4061 Hardcopy Device
Pen Plotter:	CALCOMP 565

TABLE 3.5-2
DATA BASE STORAGE REQUIREMENTS

T-2481

	NUMBER OF POLYGONS	NUMBER OF CHAINS	NUMBER OF NODES	NUMBER OF POINTS
Land Use	210	602	403	5167
Elevation	70	118	70	8216
Floodplain	3	6	4	1475
LCGUs	958	2122	920	17884

Note: 1) Chain file storage required 6 or 12 words/chain and 2 words/point in chain, with no explicit storage for nodes

2) Overlaid data base (LCGU file) uses 4 words/polygon

TABLE 3.5-3

COMPUTER SYSTEM RESOURCES REQUIRED

T-2482

	MAX MAIN MEMORY	SECONDARY STORAGE	
		Average	Peak
Data Base Creation and Editing (PDP-10)	22K	100 blocks	300 blocks
Data Base Overlay (PDP-10) (CDC-6400)**	49K $137{,}000_8$	600 blocks 1200 PRU	908 blocks 1700 PRU
Display and Analysis per Plot and Tabulate Query (PDP-10)* (CDC-6400)**	35K $70{,}000_8$	543 blocks 820 PRU	543 blocks 820 PRU

*36-bit words, 1 block = 128 words

**60-bit words, 1 Program Resource Unit (PRU) = 64 words

TABLE 3.5-4

EXECUTION TIME FOR MAJOR PROCESSES

T-2475

Process	Execution Time	
	PDP-10 CPU Time	CDC-6400 CPU Time
Data Base Creation and Editing		Not Used
• Land Use	15 Min	
• Elevation	20 Min	
• Floodplain	5 Min	
Data Base Overlay	7 Min	2.6 Min
Display and Analysis	3 sec	
• Program Initialization	3 sec	1 Sec
• Tabulation Query	2 sec	0.6 Sec
• Plot Query	12 sec (average) 22 sec (peak)	4 sec (average) 7 sec (peak)
Update and Correction	20 Sec	7 Sec

TABLE 3.5-5

ESTIMATED TIMES TO DIRECT PROCESSES
AND CORRECT ERRORS (INTERACTIVE MODE)

T-2483

PROCESS	AVERAGE TIME	PEAK TIME
Data Base Creation*		
Land Use	16 hr	
Elevation	14 hr	
Floodplain	3 hr	
Data Base Overlay**	1 hr	8+ hr
Display and Analysis per Plot and Tabulate Query	1 min	2 min

*Manual Digitizing.

**Representative Figures; Actual Data Not Available.

with line congruence (sliver errors) caused by overlaying files with significant differences in their line precision.

The capability provided for retrieval and display of data items from a composite polygonal data base using Boolean expressions represents a major advance in the state-of-the-art of automated cartography. This capability is particularly significant in that individual coverages may be stored separately and overlaid for retrieval purposes only if and when required.

Although the computer resources used in this demonstration experiment were large systems, not mini-computers, the concept of local processing could be implemented in a mini-computer environment. As a result these techniques have potential for use in tactical as well as production environments.

3.5.2 Recommendations for Future Research

Potential areas of further research involving the use of topological data structuring include the following:

- Data base creation using automated rather than manual digitizing hardware
- Raster data input
- Network data encoding
- United archival data base design.

The use of line following or drum scanning techniques should be evaluated to assess the potential use of topological data base techniques in a high-volume, data capture environment. LCG plans to include in their newest version of the ODYSSEY software a module (PASTA, Ref. 22) which will have the capability to accept input from digitizing scanners or automated line following devices. Benefits associated with such an approach include elimination of source map node annotation as was required in the current project (see Fig. 3.4-2 and Table 3.5-5). The only manual input required using the new software would be the assignment of an identification code to each polygon, a procedure which could be performed interactively by an operator working at a cathode ray tube terminal. Savings in operator time are expected to improve by an order of magnitude the performance of the existing software, described in Table 3.5-6.

Topological data structuring techniques should be evaluated for their ability to accept raster data bases as input for the purpose of extracting and encoding polygonal features. The ODYSSEY software has the potential for automatically extracting polygonal patterns from gridded data bases such as a digital terrain model or LANDSAT imagery which has been classified and corrected. The resulting polygonal descriptions could then be merged with other polygonal data bases for the same region such as political boundaries, demographic statistics and other social, economic and political information.

Topological data structuring techniques should be evaluated for ability to record, manipulate and display network as well as polygonal entities from a geographic data base. The ability to encode network features appears to be a natural extension of existing ODYSSEY capabilities. The use of nodes and directed line segments within ODYSSEY's data structure should allow for the eventual ability of that system to accomodate point and line information as well as polygonal elements as presently implemented.

Topological data structuring techniques should also be evaluated for their potential use in providing a means for the design of a unified archival geographic data base. The ODYSSEY design for describing and recording features of geographic space lends itself well to a general philosophy of unified data bases. This results from ODYSSEY's ability to:

- Accept as input either gridded or vector descriptions of geographic space
- Encode graphic products automatically with a minimum of manual intervention plus use of extensive automatic error detection and correction facilities
- Merge two or more separate data coverages for a given region into a single, composite digital data base consisting of least common geographic units
- Selectively retrieve for analysis (e.g., area computation) or graphic display geographic features which can be described by their occurrence on either a single data coverage or jointly on two or more coverages.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1 IMAGE ARCHIVE CONSIDERATIONS

Mapping, Charting and Geodesy (MC&G) data storage and data structuring concerns are a central issue in the organization of digital production facilities. Several different possible schemes for the arrangement of an image archive to support MC&G production were analyzed, and the meaning of "archival" storage in a cartographic production system was discussed. The different image archive formulations have different effects on data collection, normalization processing, data storage costs, and data availability.

The examination of image archive alternatives shows these implications for an MC&G production system:

- An archive of raw imagery is the recommended alternative, but it has inherent cost problems. Film storage is not likely to be attractive, due to the cost of silver and the cost of the film base, since the film is not reusable. Magnetic media offer the hope of lower long-run costs, if an appropriate storage mechanism for the required volumes of data can be identified
- A compromise between the alternatives of archiving raw imagery and relying upon the MC&G data base is forming an archive of normalized, mosaicked image data. The critical processes for this scheme are the identification of new imagery of interest and the means of incorporating it into the archive
- Storing of raw imagery in an input holding pen has no unique advantage for the production system. The need for prompt

response to the availability of new imagery will cause the holding pen to drive the remainder of the system design. Efforts to mitigate the impact of the holding pen will effectively transform it into a full archive or otherwise modify the system, creating an archive of a different sort

- Development of an archive of derived MC&G data would depend on the solution of many presently unsolved problems. Extraction algorithms for automating the updating of the MC&G data base are necessary if data base input is not to depend excessively on operator control.

Based on these considerations and the supporting analysis presented in Chapter 2, it is recommended that:

- The potential utility of a normalized, mosaicked image archive should be explored through an experimental program aimed at developing and demonstrating the feasibility of the key processes and concepts, such as information measures

- The operational definition developed herein, which identifies an MC&G archive as the lowest level of the data hierarchy which contains all of the data for generating MC&G products, should be validated by further study of present and planned DMA MC&G production methods

- The accompanying MC&G data base hierarchy ideas, especially the concept of a level of unified, product-independent data bases, should also be validated in the DMA operational context. Particular attention should be paid in those studies to the practicalities of assembling, converting and storing the mass of MC&G data already at hand in the DMA Centers, in order to fairly assess the reasonableness of the theoretical approach developed in this study.

4.2 TOPOLOGICAL DATA BASE STRUCTURES

The overall demonstration of Topological Data Base (TDB) structuring applications in a sample MC&G problem was successful and conclusive, and the principal assertions concerning the advantages of TDB structures were upheld:

- The complex problem of overlaying several disparate maps was successfully performed as an <u>automated</u>, <u>local</u> computer process, which operated within the expected low range of computer time and memory requirements*
- The ability to identify, access, and retrieve (plot) map feature sets, by querying the overlaid data base, was successfully demonstrated
- The ability to update and merge TDB data without redigitizing the basic data was demonstrated, using a preliminary approach demanded by experimental expediency. A better approach, based on extension of the overlay capability, was identified in the course of of the activity and will be applied in future work.

Some unanticipated difficulties in data base and computer program conversion to the ETL CDC 6400 computer system were encountered and overcome, as was the unexpected unavailability of the preferred digitizing equipment. These difficulties affected the experimental demonstration as follows:

*It is to be noted that no data was available on execution times, etc., for other approaches to this problem, which could have been used for comparison.

- Operator intervention time statistics, relative to TDB-related encoding, were unavoidably mixed with digitizing times, due to the mode of operation of the available digitizing tablet. This problem also carried over into data preparation time statistics, resulting in inconclusive results in this area

- Difficulties in the conversion process prevented in-depth analysis of actual data base storage requirements at all of the required levels

- All operations were performed, for this experiment, in Universal Transverse Mercator (UTM) coordinates; the transformation from digitizer coordinates to UTM coordinates is part of the LCG digitizing software, which was not modified to handle other coordinate systems.

In addition to the proposed elaboration on the topological data structuring software described in Section 3.5.2, it is suggested that future investigations in the DMA range of interest include these items:

- Demonstration of the utility of TDB structuring in a production environment, as in Digital Land Mass Simulation (DLMS) encoding and production

- Demonstration of TDB concepts applied to DMA source data, using an interactive minicomputer system such as the ETL Automated Cartography Divisions's PDP 11/45 computer

- Investigation of applications of TDBs used interactively with imagery, utilizing an interactive image processing facility.

Such studies would provide insights for the best approaches to the issue of unification of the different MC&G data bases utilized by DMA, exploring how they can be coordinated with the aid of TDB data structuring concepts. The near-term results could enhance the productivity of the DLMS production process, or other production applications where the TDB ideas are applicable.

APPENDIX A

A REVIEW OF PUBLISHED IMAGE COMPRESSION TECHNIQUES

A.1 INTRODUCTION

The transmission or storage of certain types of digital imagery required by DMA necessitates systems with large capacities. For example, a 1024 x 1024 pixel (picture element) by 64 grey shades/pixel image requires over 6 megabits for representation using standard Pulse Code Modulation (PCM) techniques. In an effort to reduce these requirements many schemes have been proposed, based either on removal of the inherent redundancy in the pictures, or on the psychovisual characteristics of the human visual system, or both. A number of heuristic techniques using dithering noise (Ref. A-1), contour coding (Ref. A-2), run length coding (Ref. A-3) and synthetic highs (Ref. A-4), to name a few, were developed, but their performance turned out to be substantially inferior to the statistical methods described below. In addition, there exist some specialized coding techniques which have application only to specific types of imagery, such as facsimile (binary black and white) photographs (Ref. A-5). These schemes are not considered here.

Of the remaining methods, the most promising can be divided into five major categories:

- Predictive
- Transform
- Hybrid

- Interpolative
- Variable Resolution.

These techniques are particularly attractive because they provide either a high level of performance, a simple implementation, or offer an effective compromise between the two factors.*

The techniques listed previously may also be further subdivided depending on whether the methods are applied intraframe (within a single frame) or interframe (across multiple frames). In Sections A.2 through A.6 each of these techniques is described in more detail. Section A.7 summarizes the methods and provides a comparative performance evaluation.

A.2 PREDICTIVE COMPRESSION TECHNIQUES

Predictive techniques have been developed to take advantage of the fact that intensity values of neighboring pixels tend to be closely related, so that information about a particular pixel can be inferred or "predicted" from the intensity values of its neighbors. Of the predictive schemes, the most widely used are Differential Pulse Code Modulation (DPCM) (Ref. A-6) and Delta Modulation (Ref. A-7). An n^{th}

*Performance can be measured in terms of the level of compression obtainable relative to a given amount of distortion. For almost all practical schemes, there is a nonlinear tradeoff between compression ratio and induced distortion. This is particularly evident at the higher compression levels (8 to 1 and greater, for 8-bit/pixel images). For moderate levels of compression ($\approx$ 4 to 1) using good techniques the compression process is nearly reversible, in that the reconstructed image is almost indistinguishable from the original when viewed by a human observer. There are some "error free" or totally reversible techniques such as entropy coding (Huffman), but these suffer from very difficult implementation problems in all but the simplest cases, where the compression ratios are low (typically about 2 to 1).

order PDCM coder works by estimating the value of the current pixel based on the previous n pixels, forming the difference between this estimate and the actual pixel value, and quantizing the difference for storage or transmission. Delta Modulation is a variant of this technique, comparing just the current and the last pixel, which can be shown to be equivalent to a first-order DPCM coder with a one-bit quantizer.

A block diagram of an n^{th}-order DPCM system is shown in Fig. A.2-1. If a set of pixels with zero mean and variance σ^2 is represented as $\{x_i\}$, then an n^{th}-order DPCM predictor for x_{n+1} can be written as

$$\hat{x}_{n+1} = \sum_{i=1}^{n} a_i \, x_i \tag{A.2-1}$$

The difference signal to be quantized is then formed as

$$e_{n+1} = x_{n+1} - \hat{x}_{n+1} \tag{A.2-2}$$

The a_i used in the estimate are calculated using a least mean square error criterion and are obtained from a solution to the equations

$$R_{n+1,i} = \sum_{j=1}^{n} a_j R_{ij} \quad i = 1,2,\ldots n \tag{A.2-3}$$

where

$$R_{ij} = E(x_i - E(x_i))(x_j - E(x_j)) \tag{A.2-4}$$

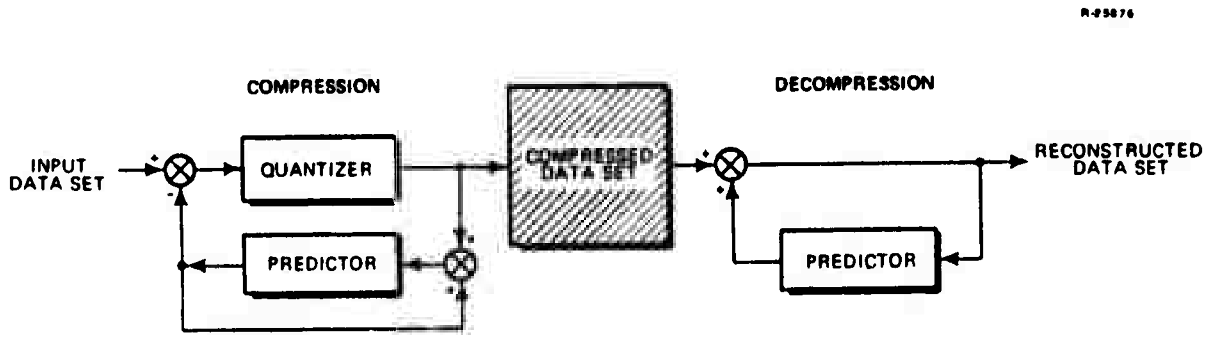


Figure A.2-1 Functional Flow Diagram of DPCM System

and E. (•) is the expected value operator. The quantizer is usually nonlinear and designed on the basis of the probability distribution for the $\{e_i\}$.

The choice of particular pixels, $\{x_i\}$ $i = 1,2,\ldots n$, to be used by the predictor is normally based on either a one- or two-dimensional selection scheme similar to those illustrated in Fig. A.2-2. Because of the two-dimensional DPCM coder's ability to remove spatial redundancy in the vertical as well as horizontal directions, it generally outperforms the one-dimensional DPCM coder. Although the accuracy of the predictor nominally increases with increasing values of n, in practical image coding experiments using real images, $n = 3$ provides performance substantially equal to the cases for $n > 3$. Often $n = 1$ is used since the simplicity of implementation of such a system outweighs the performance decrease.

Generally speaking, DPCM offers the least implementation problems of all the systems considered here, provides excellent image quality when used to compress data in the 4- to 8-bit/pixel range, but is considerably inferior to the other methods in the 1- to 3-bit/pixel range. Its use in the DMA environment would most likely be limited to the transmission of non-critical data where moderate compression ratios, high speed, and ease of implementation are the primary design factors.

Pixels Used In Prediction — Current Pixel

$$* \left[\;\cdot\; - \;\overbrace{\dot{x}_{n-3} \quad \dot{x}_{n-2} \quad \dot{x}_{n-1}}^{\text{Pixels Used In Prediction}} \qquad \overset{\odot}{x}_n \quad \cdot \;\; \cdot \;\; \cdot \;\right]$$

*Row or Column of Image Matrix (Raster)

Figure A.2-2a Example of One-Dimensional DPCM Showing Current Pixel and Those Selected for the Predictor

$$* \begin{bmatrix} & & \dot{x}_{n-2} \\ & \dot{x}_{n-4} & \dot{x}_{n-1} \\ \dot{x}_{n-5} & \dot{x}_{n-3} & \overset{\odot}{x}_n \end{bmatrix}$$

· pixels used in predictor

⊙ current pixel being estimated

*Image Matrix Partition

Figure A.2-2b Example of Two-Dimensional DPCM Showing Current Pixel and Those Selected for Predictor

A.3 TRANSFORM COMPRESSION TECHNIQUES

Transform techniques, or more properly linear transform techniques, also exploit the redundancy present in an image but in an entirely different manner from DPCM. In this case, a linear transformation is either applied one- or two-dimensionally to the image, resulting in a "coordinate transformation." Performing this coordinate transformation results in a new set of coefficients equal in number to the number of original pixels, but where the variance or "information" distribution among the coefficients is modified. For normal imagery* several unitary transforms (e.g., Fourier, Hadamard, Cosine) have the property of concentrating the preponderance of the variance in just a few of the transform coefficients and decreasing the correlation between neighboring coefficients over the original source imagery. The higher variance coefficients are then selected and quantized with more bits than are the lower variance coefficients (which are often discarded entirely). The resulting quantized coefficients are stored or transmitted in place of the original pixels. To reverse the operation, an inverse transformation is applied to the quantized coefficients which results in some distortion in the reconstructed originals, the amount of distortion depending on the number of bits retained in the transform domain. However, because of the variance compaction, this distortion is normally very much less than if the original pixels were quantized directly using the same total number of bits. Figure A.3-1 contains a block diagram of a typical transform coding system.

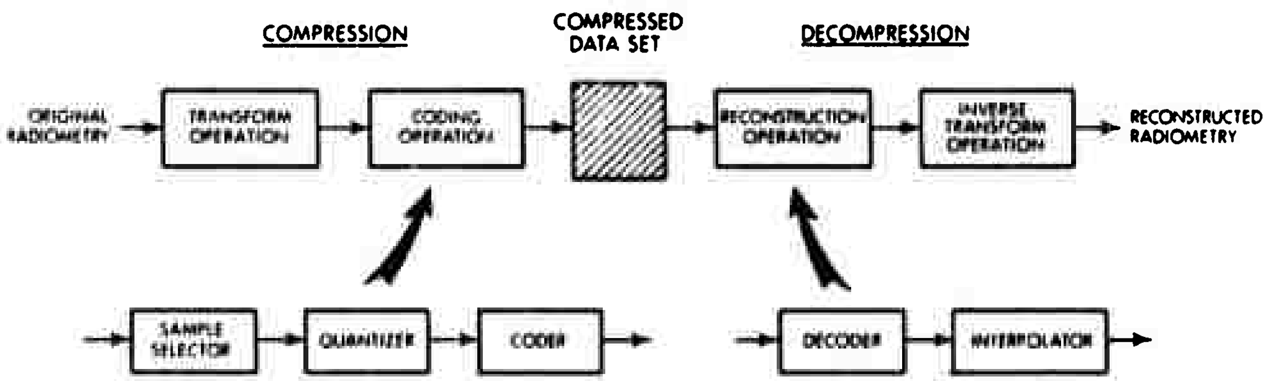


Figure A.3-1 Transform Compression-Decompression Process

---

*This excludes such pathological cases such as a "white noise image."

There are several specific concerns which must be addressed in a practical implementation of a transform coder. They are:

- Choice of transform type and mode of operation
- Selection of retained coefficients, quantization, and coding
- Implementation complexity.

Successful use of the transform techniques requires careful consideration of all these factors.

A.3.1 Transform Selection and Mode of Application

A number of possible unitary transforms have been used in transform compression techniques. Table A.3-1 lists some of the more widely used transform types along with some of their performance and implementation characteristics when used in a transform coder.

Historically, the first applications of transform coding used the Fourier transform (Ref. A-8) which was already in widespread use in other applications and could be calculated rapidly using the Cooley-Tukey Fast Fourier Transform (FFT) algorithm. This was followed by the Hadamard Transform (Ref. A-9) which offered a particularly easy implementation due to the fact that its basis vectors consist of +1 and -1 components. Although computationally simpler than the Fourier transform, its performance was not generally as good. Two additional transforms were developed specifically to exploit image characteristics, the Slant (Ref. A-10) and Discrete Linear Basis (Ref. A-11), but neither provides the performance obtainable from the Discrete Cosine Transform (DCT) (Ref. A-12). The DCT

TABLE A.3-1

CHARACTERISTICS OF SOME TRANSFORM CODING TECHNIQUES

T-2484

TRANSFORM TYPE	SUBJECTIVE QUALITY/MEAN SQUARE ERROR			ALGORITHMIC IMPLEMENTATION (ON THE ORDER OF)	CODER BUFFERING REQUIREMENT
	4:1 REDUCTION	8:1 REDUCTION	12:1 REDUCTION		
Haar Transform	Very Good MSE: 0.49%	Good MSE: 1.1%	Fair MSE: 1.9%	2(N-1) Additions	16-32 Image Lines
Hadamard Transform	Very Good MSE: 0.41%	Good MSE: 1.05%	Fair MSE 1.85%	$2N \log_2 N$ Adds	16-32 Image Lines
Fourier Transform	Excellent MSE: 0.31%	Very Good MSE: 0.75%	Fair MSE: 1.7%	$2N \log_2 N$ Complex Adds $2\left(\frac{N}{2} \log_2 N-N\right)$ Complex Mults.	32-64 Image Lines
Slant Transform	Excellent MSE: 0.29%	Very Good MSE: 0.72%	Good MSE: 1.4%	$2N \log_2 N$ Adds 2(2N-4) Mults.	16-32 Image Lines
Cosine Transform	Excellent MSE: 0.27%	Very Good MSE: 0.67%	Good MSE: 1.2%	$2(N-2) \log_2 N+2N$ Adds $8(N-\log_2 N)$ Mults.	16-32 Image Lines
Karhunen-Loeve Transform	Excellent MSE: 0.26%	Very Good MSE: 0.65%	Good MSE: 1.1%	$2N^2$ Adds & Mults. $2N^3$ Adds & Mults. (calculated only once)	Full Frame Plus 16-32 Image Lines

MSE listed are based upon the transform zonal coding of a typical reconnaissance image (transform is performed in 16 x 16 pixel blocks) is generally considered to provide the best combination of high performance with reasonable implementation complexity of all the current unitary transforms*. Except in the case where implementation complexity might dictate the choice of Hadamard or Haar transforms, selection of the DCT is usually made in current transform coders.

The unitary transform may be applied to an image either one-dimensionally on a line-by-line basis or two-dimensionally on the entire image or block. If an image (pixel intensity) is represented as an N x N array $[X]$ and $[X_i]$ is an N x 1 vector representing the i^{th} column of $[X]$, then the one-dimensional transform of the i^{th} image line, $[\chi_i]$, is represented as

$$[\chi_i] = [A]\,[X_i] \tag{A.3-1}$$

where $[A]$ is an N x N unitary transform matrix of the type mentioned above. The two-dimensional transform is performed by sequential row and column operations† on $[X]$ and may be represented as:

$$[\chi] = [A]^T\,[X]\,[A] \tag{A.3-2}$$

*The Karhunen-Loeve Transform (KLT) is actually the optimal transform when performance is measured with a mean square error criterion. However, there is no general fast computation algorithm and it is sensitive to the image statistics. The DCT performance is almost exactly the same as the KLT without the latter's implementation drawbacks.

†Assuming a separable transform; the complexity required for non-separable transform implementation has made implementation impractical.

Both experimental and theoretical studies have shown that two-dimensional application of the transform provides considerable performance improvement over the one-dimensional case. The additional complexity of the two-dimensional approach is somewhat reduced by the fact that the transform is usually not applied over the entire image, but rather to small subsections or blocks of the image. Studies have shown that very little performance is gained, except for the Fourier transform. by using blocks much larger than 16 x 16, and often for the larger block sizes the ability to adaptively modify algorithms to exploit local image variations is lost. Figure A.3-2 shows the characteristic behavior of Mean Square Error performance with changes in block size for several important algorithms.

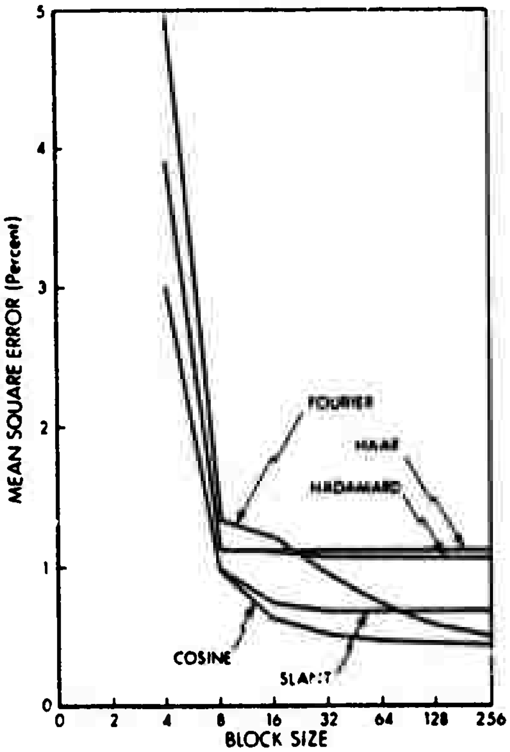


Figure A.3-2 Variation in Transform Coder Performance with Block Size (Typical)

It should be noted that the application of the transform has not resulted in any compression. In fact, the operation is totally reversible up to this point. The compression occurs when only certain of the coefficients in the transform matrix $[\chi]$ are retained.

A.3.2 Coefficient Selection, Quantization and Encoding

There are two basic strategies of coefficient selection: zonal sampling and threshold sampling. In the zonal sampling case a predetermined set of coefficients is retained (usually the ones with high variance) and the remainder are discarded. In the threshold sampling case, the coefficients retained are those which exceed a specified threshold. Both methods have advantages and disadvantages. The zonal system requires no "bookkeeping" since the locations of the retained coefficients are predetermined and are known quantities, whereas the threshold method requires transmission or storage of the locations of the coefficients exceeding the threshold. However, the threshold method is less likely than the zonal to discard coefficients with high information levels and therefore generally provides better information compaction. Once either of these methods has been used to select the retained coefficients, the next step requires that the coefficients be quantized into a certain number of levels as dictated by the desired compression ratio required. It is in this coefficient selection and quantization process that the compression actually occurs.

Because the variances of the retained coefficients vary widely the use of a single quantization scheme for all the coefficients is inefficient. Instead, more quantization levels, or bits, are reserved for those coefficients which contain the highest variances (i.e., the coefficients most necessary for an accurate reconstruction of the compressed image).

In addition, the use of a nonlinear quantizer based on the probability distribution function of the coefficient to be quantized such as the Max quantizer (Ref. A-13) can significantly reduce the mean square quantization error over what can be obtained from a uniform quantizer. The nonuniform quantizer is normally implemented in a straightforward manner by passing the coefficient through a nonlinear "table lookup" and then a linear quantizer. Methods such as the Max quantizer are strategies for optimizing the quantization of single coefficients. The use of block quantizers (Ref. A-14), which try to minimize the total mean square error of a block of coefficients, subject to a constraint on the total number of bits assigned to that block, usually provides better results ($\approx$ 1 bit/pixel reduction in the number of bits needed to represent the image). Let X_{ij} and $\hat{X}_{ij}$ represent the unquantized and quantized $(i,j)^{th}$ coefficient, respectively, σ_{ij}^2 the variance of X_{ij}, and b_{ij} the number of bits assigned to X_{ij}. Then we wish to minimize the quantization error

$$\varepsilon = E\left\{\sum_{i=1}^{N} \sum_{j=1}^{N} (X_{ij} - \hat{X}_{ij})^2\right\} \tag{A.3-3}$$

subject to the constraint that the total number of bits is a fixed number B. A number of different solutions to this problem have been investigated based on particular assumptions on the probability density functions (pdfs) for the X_{ij} and approximations of the quantization error. An example is a scheme by Ready and Wintz (Ref. A-15) which leads to a bit distribution of the form

$$b_{ij} = \frac{B}{N^2}\, 2 \log \sigma_{ij}^2 - \frac{2}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \log \sigma_{ij}^2 \tag{A.3-4}$$

Because the b_{ij} must be integers the above expression must be rounded to the nearest integer, leading to some errors typical of this approach. A further savings on the number of bits required can be obtained by using run-length or Huffman codes to further code the quantization levels. The use of the Huffman coder to optimally assign code words of unequal length to the quantization levels usually gains about 0.25 bits/pixel over using equal length codes.

Many other quantization schemes are possible including assigning bits based on various "subjective" criteria, in accordance with models of the human visual system. A number of adaptive schemes are also in use where the bit mappings are changed dynamically depending on local image statistics. This often leads to a 0.5-to-1 bit reduction in the number of bits necessary to represent an image over the non-adaptive case.

A.3.3 Implementation Complexity and Performance

Implementation of the transform coders tends to be the most complicated of the compression schemes considered, particularly when the more sophisticated transform and coefficient selection/quantization schemes are selected. However, where maximum performance is required this additional complexity can often be justified. The Haar and Hadamard transform coders, due to their simple basis vectors, have been used in real-time television compression systems with compression ratios in the 2-4 to 1 range. Even for the more complicated Discrete Cosine Transform, experimental coders have been built in the 5 mega-pixel/sec range with compression ratios of up to 8 to 1.

Table A.3-1 (on page A-8) contains some comparative mean square error performance measures and computational complexity for various transform techniques. Subjective judgements by human observers tend to confirm the same order of performance as for the mean square error. Table A.3-2 summarizes these results. In addition, the transform coders tend to be fairly tolerant of changing image statistics and are not overly vulnerable to channel errors, two strong drawbacks associated with DPCM.

TABLE A.3-2
SUBJECTIVE RATINGS OF TRANSFORM CODERS

T-2485

TRANSFORM	AVERAGE NUMBER OF BITS/PIXEL FOR ALMOST NO VISUALLY OBSERVABLE DEGRADATION
Haar	≈ 2.5
Hadamard	≈ 2.5
Fourier	≈ 2
Slant	≈ 1.5
Cosine	≈ 1

A.4 HYBRID (TRANSFORM-DPCM) CODERS

Both DPCM and Transform coding techniques have attractive characteristics, along with certain limitations. DPCM provides moderate compression with an ease of implementation for high-speed compression applications, does poorly at low bit rates, is sensitive to changing image statistics, and tends to propagate errors. Transform coding, on the other hand, provides superior coding performance at low bit rates, low sensitivity to channel errors and image statistics, but requires considerable complexity in hardware implementation. A combination of the two techniques called Hybrid coding (Ref. A-16) has been suggested which combines most of the benefits of the two systems without either's major drawbacks.

Figure A.4-1 depicts the Hybrid system consisting of a one-dimensional unitary transform followed by a bank of DPCM coders operating on the transform coefficients. The one-dimensional transform is done on a horizontal line-by-line basis and then the DPCM coders operate on each vertical row of the transform coefficients. The DPCM coders are thus matched to the statistics of the transform coefficients rather than the original pixel values. Performance at low bit rates is not quite as good as the DCT coder, but the hardware complexity required is considerably reduced. Successful application of this hybrid scheme has been made in the remotely-piloted vehicle area where weight constraints place a premium on low complexity, high-performance designs. High-quality images (almost no degradation noted by a human observer) are produced in the 1.5-2 bit/pixel range and 1 bit/pixel is possible where small amounts of degradation are acceptable.

A.5 INTERPOLATIVE CODERS

Interpolative coding is a technique by which information is sent for only a few of the image points and the remainder are obtained by interpolation. The efficiency of an interpolative scheme depends mostly on choice of interpolating functions, but also on the quantization and coding of the interpolating coefficients. Figure A.5-1 shows a block diagram of interpolative coding schemes. The interpolating functions have ranges from simple linear (Ref. A-17) to more sophisticated bi-cubic splines with variable knots

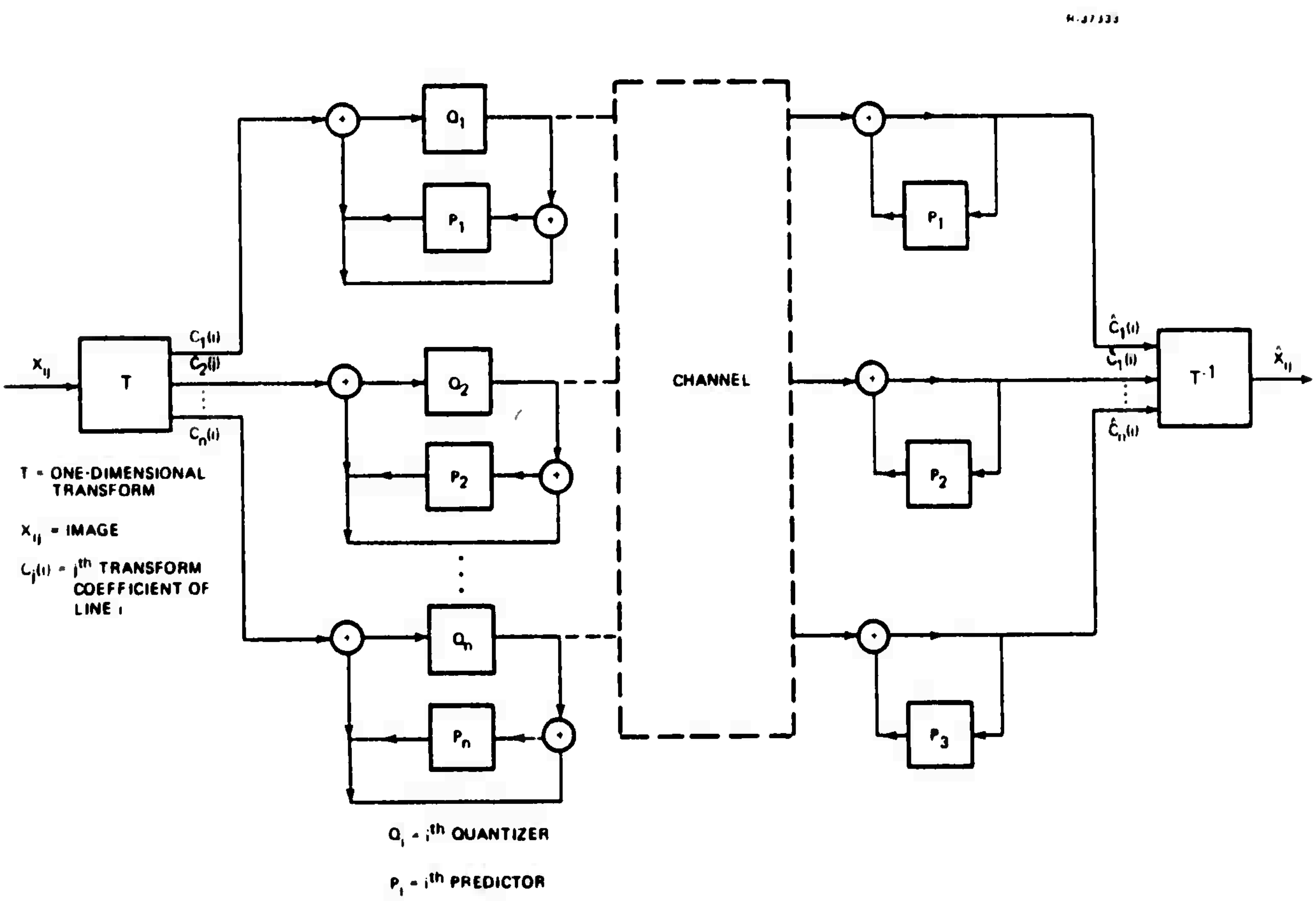


Figure A.4-1 Hybrid Transform DPCM Compression Technique

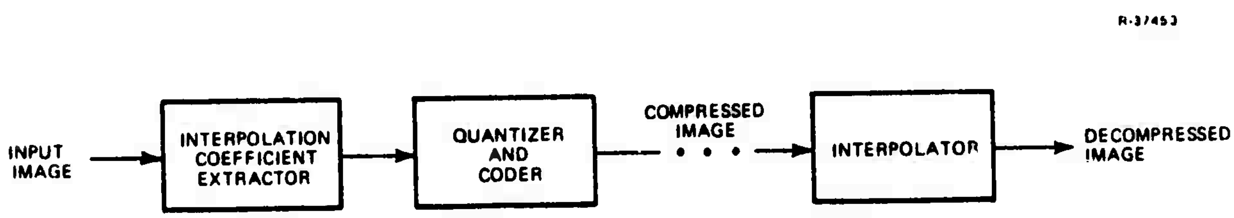


Figure A.5-1 Interpolative Coder Block Diagram

(Ref. A-18). As in the other coding techniques, the interpolation may be done either in one or two dimensions with the two-dimensional coder performing significantly better. Except in the case of special two-dimensional data such as certain types of radar returns, the performance of the simpler interpolating schemes has been inferior to the other techniques described in this report. On the other hand, preliminary work (Ref. A-19) with the variable knot bi-cubic spline functions appears to offer excellent performance in the 1 bit/pixel range or less. However, computational problems associated with obtaining the interpolating coefficients for this process appear to limit the throughput of this technique, and only preliminary results are available on the achievable compression ratios and induced distortions.

A.6 VARIABLE RESOLUTION COMPRESSION TECHNIQUES

Variable resolution coders work by preserving the full resolution or quality of the image at points judged to be important, and degrading the resolution at other points. In the earliest attempts, this amounted to using different levels of compression at various points selected by the operator. Some of the more recent attempts have used a type of activity function, typically local contrast, to automatically vary the resolution. This technique often leads to high levels of compression, but great care must be taken to ensure that each area of interest is compressed at the appropriate fidelity.

An interesting new technique called MAPS (Micro-Adaptive Picture Sequencing) has been suggested where the resolution is automatically varied by the algorithm to select the appropriate resolution locally (Ref. A-20). Implicit in the algorithm is the assumption that when an image is viewed as a whole, fine detail is noticed only when it exhibits sharp contrast. Having determined what can be represented at lower resolutions (the so-called macro fidelity criteria), the algorithm combines first individual pixels and then blocks of pixels to form new larger elements. This reduces both resolution and the number of bits required to represent the larger blocks. A procedure is presented for implementing this method which requires about 20 steps per pixel (about 1/2 to 1/3 of the amount required for a transform coder) and yields what appears to be excellent quality images at the 0.5 to 1 bit/pixel range. Although it is difficult to make definitive statements from the small number of pictures presented in the reference, A-20, this method shows potential for rivaling the transform techniques. The only noticeable artifact in the reconstructed image appears to be a slight blurring in the "background" areas.

A.7 COMPARISON OF IMAGE COMPRESSION TECHNIQUES

In terms of performance, the DCT transform coder appears to provide the best quality image for compression ratios of 4 to 1 or greater on 8 bit/pixel originals, although the MAPS variable resolution coder is a strong competitor. Its implementation is non-trival, but there exist fast DCT algorithms, and special purpose hardware has been developed to perform the operation rapidly. Other transforms such as the Hadamard may be used, with decreased performance and complexity compared to the DCT. In addition, transform coding also offers substantial freedom from propagation of errors and changing source statistics.

DPCM produces excellent results for compression ratios of 2 to 1 or so and an ease of implementation unmatched by the other techniques. DPCM performance is, however, considerably inferior to transform coding at the 4-8 to 1 compression range and it suffers from susceptibility to channel errors and changing image statistics. A considerable amount of real-time DPCM has been built for use in television image compression with moderate success.

Hybrid coding is promising because it appears to offer most of the benefits of transform and DPCM coding without their drawbacks. Its performance is nearly as good as transform coding at 1-2 bits/pixel and better at higher rates. It also has an ease of implementation which lies partly between transform and DPCM coders. For those applications where "ultimate" performance does not need to be provided, this technique appears to be one of the best choices.

Interpolative coding is not yet fully enough developed to completely evaluate its impact on image compression. The earlier, simple linear interpolation schemes provide a low level of performance and the newer variable knot spline function techniques are still undergoing testing and modification. Implementation complexity is also quite high for these methods and it is not clear whether this will preclude practical near-real-time use of the algorithm.

The variable resolution techniques, on the other hand, offer a potentially attractive alternate to the transform and hybrid techniques. They offer high compression, low distortion, and a reasonably straightforward implementation (the operations count is said to be 30-50% lower than for equivalent quality transform coders) (Ref. A-20). Image quality appears to be excellent, although background areas often have a somewhat "blurry" look.

APPENDIX A REFERENCES

A-1. Roberts, L.G., "Picture Coding Using Pseudo-Random Noise," IRE Trans. on Information Theory, Vol. IT-8, No. 2. Feb. 1962, pp. 145-152.

A-2. Schreiber, W.F., Huang, T.S. and Tretiak, O.J., "Contour Coding of Images," Conf. Record, IRE Western Electron Show and Conv., Aug. 1968, pp. 8/3(1) - 8/3(6).

A-3. Limb, J.L. and Sutherland, I.F., "Run Length Coding of Television Signals," Proceedings of the IEEE, Vol. 53, No. 2, Feb. 1965, pp. 169-170.

A-4. Schreiber, W.F., Knapp, C.F. and Kay, N.D., "Synthetic Highs, An Experimental TV Bandwidth Reduction System," J. Soc. Motion Picture T.V. Eng., 68, 1959. pp. 525-537.

A-5. Huang, T.S., "Coding of Two-Tone Images," IEEE Trans. on Comm., Vol. COM-25, No. 11, Nov. 1977, pp. 1407-1424.

A-6. O'Neal, J.B., "Predictive Quantizing Systems (Differential Pulse Code Modulation) for the Transmission of Television Signals," Bell System Technical Journal, May-June 1966, pp. 689-721.

A-7. Habibi, A., "Delta Modulation and DPCM Coding of Color Signals," Proc. of Int. Telemetry Conf., Vol. 8, Oct. 10-12, 1972, Los Angeles, pp. 333-343.

A-8. Anderson, G.B. and Huang, T.S., "Piecewise Fourier Transformation for Picture Bandwidth Compression," IEEE Trans. on Comm. Tech., Vol. COM-19, No. 2, April 1971, pp. 133-140.

A-9. Pratt, W.K., et al, "Hadamard Transform Image Coding," Proceedings of the IEEE, Vol. 57, No. 1, Jan. 1969, pp. 58-68.

A-10. Pratt, W.K., et al, "Slant Transform Image Coding," IEEE Trans. on Comm., Vol. COM-22, No. 8, August 1974, pp. 1075-1093.

APPENDIX A REFERENCES (CONTINUED)

A-11. Hanalick, R.M. and Shanmagan, K., "Comparative Study of a Discrete Linear Basis for Image Data Compression," IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-4, No. 1, Jan. 1974, pp. 16-27.

A-12. Ahmed, N., Natarayan, T. and Rao, K.R., "Discrete Cosine Transform," IEEE Trans. on Computers, Vol. C-23, No. 1, Jan. 1974, pp. 90-93.

A-13. Max, J., "Quantizing for Minimum Distortion," IRE Trans. Info. Theory, Vol. IT-6, No.2, pp. 7-12, March 1960.

A-14. Huang, J. and Schultheiss, "Block Quantization of Correlated Gaussian Random Variables," IEEE Trans. Comm. Sys., Vol. CS-11, No. 5, pp. 289-296, Sept. 1963.

A-15. Ready, P.J. and Wintz, P.A., "Multispectral Data Compression Through Transform Coding and Block Quantization," School of Elec. Eng., Purdue Univ., Lafayette, Ind., Tech Rep. TR-EE 77-7, Jan. 1972.

A-16. Habibi, A., "Hybrid Coding of Pictoral Data," IEEE Trans. on Comm., Vol. COM-22, No. 5, May 1974, pp. 614-624.

A-17. Kortman, C.M., "Redundancy Reduction - A Practical Method of Data Compression," Proceedings of the IEEE, Vol. 55, No. 3, March 1957, pp. 253-263.

A-18. McCaughey, D., "The Degrees of Freedom of Sampled Images," Univ. of Southern Calif. Image Processing Institute Report #730, June 1977.

A-19. McCaughey, D., Private Communication, January 1977.

A-20. LaBonte, A.E., "Two-Dimensional Image Coding by Micro-Adaptive Picture Sequencing (MAPS)," Proc. of the Int. Optical Comp. Conf. 1977, SPIE, Vol. 119, Digital Image Processing, Aug. 25-26, 1977, San Diego, pp. 99-106.

APPENDIX B
DEMONSTRATION RESULTS

B.1 AREA TABULATIONS

The required area measurements were produced by the TABULATE command discussed in Section 3.4.5. Table B.1-1 contains the areas of all the land use types found on the land use map, Fig. 3.3-1a. By the SELECT command, the section of the demonstration area to be tabulated can be reduced; Table B.1-2 contains the areas of land use categories within the floodplain and below 100 ft elevation.

Areas of individual LCGUs are tabulated by invoking the FORM DETAILED INSTANCE command before doing the tabulation. Table B.1-3 lists the areas of the LCGUs in the same selected area, within the flood plain and below 100 ft. Each LCGU is identified in terms of its parent land use, elevation and floodplain polygon identifiers. The LCGUs are grouped by land use polygon identifier number. Acreage subtotals are provided for each land use category. (The land use polygon identifier numbers are the circled numbers on the annotated land use map, duplicated here as Fig. B.1-1.)

The area of an individual land use polygon is obtained by adding all entries with the same land use polygon number in the first column of the table. In Table B.1-3, land use polygons are not divided by elevation or floodplain, so each polygon number appears only once. The elevation and floodplain identifiers are constant throughout the table.

TABLE B.1-1

ACREAGE OF LAND USE BY CATEGORY
FOR THE ENTIRE STUDY AREA

ACC	361.24
ACP	1483.62
AR	63.80
AVF	1326.17
AVV	1702.54
BBR	25.79
BEQ	13.33
BES	9.13
BT	121.80
FO	813.28
LR	54.35
NEC*	10.57
R	919.88
UCB	12.00
UCC	57.84
UCR	87.46
UCW	17.20
UES	92.38
UIL	25.32
UIS	57.87
UIW	7.66
UOC	16.42
UOG	64.68
UOO	27.82
UOP	12.65
UOR	1.80
UOV	13.69
URH	10.37
URS	1532.77
UUS	15.01
UUT	102.40
WO	37.87
WS	207.42
WWP	10.75

*Not Elsewhere Classified; refers to the unlabeled area in the northeast corner of the land use map.

TABLE B.1-2

ACREAGE OF LANDUSE BY CATEGORY FOR THE SELECTED AREA WITHIN THE FLOODPLAIN AND BELOW THE 100-FT ELEVATION LINE

ACC	84.95
ACP	9.75
AVF	316.26
AVV	652.43
BBR	25.79
BEQ	13.33
BES	7.10
BT	5.03
FO	18.36
LR	51.21
R	1.47
UCC	0.18
UES	69.57
UIS	5.82
UOO	4.29
UOP	9.21
URH	0.73
URS	77.70
UUS	14.45
UUT	11.73
WO	37.87
WS	200.19
NONSELECT LANDUSE	7699.45

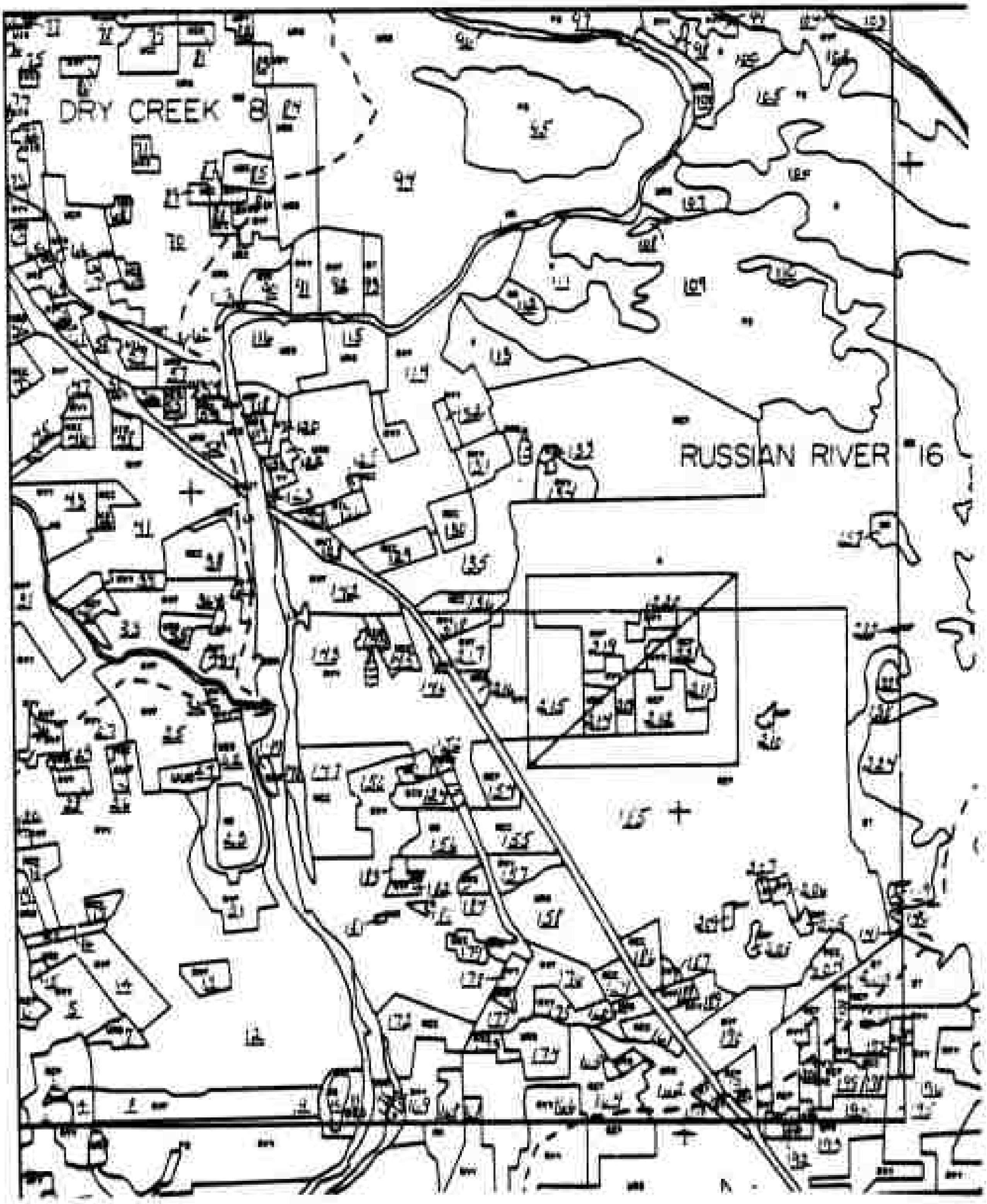


Figure B.1-1 Land Use Polygon Identification Numbers

TABLE B.1-3

ACREAGE OF LCGUs WITHIN THE FLOOD PLAIN
AND BELOW 100-FT ELEVATION

	POLYGON IDENTIFIERS LAND USE	ELEV.	FLOOD-PLAIN	
	16	8	2	2.93
	27	8	2	0.05
	147	8	2	56.02
	172	8	2	25.94
ACC				84.95
	17	8	2	6.95
	40	8	2	2.80
ACP				9.75
	8	8	2	35.22
	13	8	2	12.54
	14	8	2	43.60
	21	8	2	40.54
	25	8	2	57.92
	26	8	2	4.48
	28	8	2	0.06
	31	8	2	3.28
	33	8	2	31.38
	41	8	2	3.64
	90	8	2	15.01
	92	8	2	11.38
	114	8	2	24.27
	142	8	2	32.94
AVF				316.26
	5	8	2	6.22
	9	8	2	7.05
	12	8	2	452.93
	32	8	2	5.62
	34	8	2	1.96
	39	8	2	1.34
	43	8	2	1.32
	91	8	2	11.38
	121	8	2	6.06
	143	8	2	33.58
	150	8	2	109.98
	169	8	2	14.99
AVV				652.43
	149	8	2	4.37
	171	8	2	21.41
BBR				25.79
	11	8	2	13.33
BEQ				13.33
	119	8	2	7.10
BES				7.10
	93	8	2	5.03
BT				5.03
	97	8	2	15.77
	105	8	2	0.55
	109	8	2	2.03
FO				18.36

	POLYGON IDENTIFIERS LAND USE	ELEV.	FLOOD PLAIN	
	117	8	2	4.37
	123	8	2	3.43
	148	8	2	35.55
	170	8	2	7.85
LR				51.21
	100	8	2	0.19
	111	8	2	1.28
R				1.47
	179	8	2	0.18
UCC				0.18
	22	8	2	36.81
	116	8	2	32.76
UES				69.57
	120	8	2	5.82
UIS				5.82
	127	8	2	4.29
UOO				4.29
	118	8	2	9.21
UOP				9.21
	53	8	2	0.73
URH				0.73
	7	8	2	1.69
	35	8	2	7.40
	37	8	2	3.69
	52	8	2	4.15
	57	8	2	1.88
	70	8	2	7.56
	94	8	2	38.79
	101	8	2	3.23
	107	8	2	2.57
	115	8	2	4.16
	122	8	2	2.58
URS				77.70
	24	8	2	14.26
	181	8	2	0.19
UUS				14.45
	51	8	2	1.58
	128	8	2	10.15
UUT				11.73
	10	8	2	7.41
	23	8	2	30.46
WO				37.87
	96	8	2	200.19
				200.19
NONSELECT LANDUSE				7699.45

Table B.1-4 contains the areas of all the land use polygons in the demonstration area. This table was compiled by adding the LCGU areas which comprise each land use polygon, as described above. The LCGU areas were obtained in the same manner as for the previous example

Table B.1-5 contains the areas of the land use polygons within the red triangles on the original land use map; the land use polygons are sub-divided by the diagonal line. Areas of the land use polygons resulting from the merger of the sub-divided regions are presented in Table B.1-6. The areas of the polygons affected by the land use updates are listed in Table B.1-7.

B.2 GRAPHIC PLOTS

These full size plots of the project area are provided under separate cover:

- B.2-1) Plot of Digitized Land Use
- B.2-2) Plot of Digitized Contours
- B.2-3) Plot of Land Use Areas Within the Floodplain and Below 100-ft
- B.2-4) Plot of Land Use Areas Affected By Update, Before Revision
- B.2-5) Plot of Land Use Areas Affected By Update, After Revision
- B.2-6) Plot of Red Triangle Areas
- B.2-7) Plot of Merged Red Triangle Areas.

TABLE B.1-4

LAND USE POLYGON ACREAGES FOR STUDY AREA

T-2486-1

LAND USE	LAND USE POLYGON NO.	AREA IN ACRES
ACC	16	8.26
	27	1.06
	38	33.33
	42	10.87
	46	8.57
	49	10.62
	125	1.83
	129	20.83
	130	19.88
	136	14.88
	145	15.03
	147	67.00
	155	29.96
	159	12.10
	161	12.78
	172	39.05
	186	17.41
	198	10.60
	204	27.17
ACP	1	38.62
	6	12.75
	17	8.07
	18	19.89
	40	10.76
	135	342.50
	154	19.78
	164	83.18
	183	4.98
	185	802.00
	191	16.50
	199	55.81
	202	19.82
	212	32.47
	221	16.49
AR	112	11.56
	137	7.57
	151	7.40
	156	25.23
	168	12.03

TABLE B.1-4

LAND USE POLYGON ACREAGES FOR STUDY AREA (CONTINUED)

T-2486-2

LAND USE	LAND USE POLYGON NO.	AREA IN ACRES
AVF	8	62.65
	13	12.54
	14	50.06
	20	2.49
	21	40.54
	25	92.33
	26	6.28
	28	8.25
	30	13.38
	31	49.02
	33	100.68
	41	219.28
	62	7.47
	80	6.52
	86	18.20
	90	16.14
	92	35.60
	102	60.97
	108	2.58
	114	283.01
	142	54.26
	146	1.88
	176	43.64
	182	4.27
	188	8.07
	190	57.07
	197	6.47
	207	2.39
	209	2.24
	217	17.68
	219	40.18

TABLE B.1-4

LAND USE POLYGON ACREAGES FOR STUDY AREA (CONTINUED)

T-2486-3

LAND USE	LAND USE POLYGON NO.	AREA IN ACRES
AVV	3	13.53
	4	8.04
	5	59.96
	9	10.54
	12	657.49
	32	8.40
	34	2.16
	36	2.99
	39	18.36
	43	40.10
	45	5.46
	47	5.18
	73	6.06
	87	3.84
	91	21.57
	98	2.15
	99	9.13
	121	6.14
	131	18.78
	132	6.19
	134	19.99
	143	187.50
	150	257.32
	157	16.02
	166	19.64
	167	3.89
	169	29.53
	173	9.40
	175	17.35
	178	6.51
	187	4.19
	189	18.61
	193	2.50
	195	7.09
	196	1.27
	200	13.53
	201	11.18
	206	6.82
	211	20.31
	213	18.95
	215	104.27
	218	11.84
	220	8.76

TABLE B.1-4

LAND USE POLYGON ACREAGES FOR STUDY AREA (CONTINUED)

T-2486-4

LAND USE	LAND USE POLYGON NO.	AREA IN ACRES
BBR	149	4.37
	171	21.41
BEQ	11	13.33
BES	119	9.13
BT	78	7.04
	93	15.36
	138	64.64
	163	7.38
	203	27.38
FO	95	147.86
	97	31.19
	105	150.07
	109	483.05
	180	1.10
LR	117	6.36
	123	3.43
	148	36.70
	170	7.85
NEC	103	10.57
R	100	57.20
	106	127.77
	110	581.00
	111	67.44
	113	49.93
	140	.82
	141	.34
	223	7.16
	224	28.22
UCB	69	5.20
	81	6.80

TABLE B.1-4

LAND USE POLYGON ACREAGES FOR STUDY AREA (CONTINUED)

T-2486-5

LAND USE	LAND USE POLYGON NO.	AREA IN ACRES
UCC	68	5.99
	71	7.22
	79	24.75
	88	7.03
	89	3.87
	179	8.99
UCR	55	6.14
	58	78.86
	75	2.46
UCW	61	9.97
	65	2.30
	66	4.93
UES	22	37.74
	116	54.65
UIL	59	12.38
	126	9.81
	133	3.13
UIS	48	9.23
	54	4.67
	72	16.06
	74	.48
	76	1.04
	120	7.29
	124	15.35
	165	3.76
UIW	144	7.66
UOC	85	16.42
UOG	84	64.68
UOO	56	7.92
	64	14.99
	127	4.90
UOP	67	1.61
	118	11.05
UOR	77	1.80

TABLE B.1-4
LAND USE POLYGON ACREAGES FOR STUDY AREA (CONTINUED)

T-2486-6

LAND USE	LAND USE POLYGON NO.	AREA IN ACRES
UOV	82	6.30
	83	7.38
URH	53	7.75
	152	2.62
URS	7	11.27
	15	2.27
	19	11.21
	29	3.04
	35	11.13
	37	23.09
	50	7.37
	52	22.53
	57	22.13
	63	13.96
	70	495.61
	94	591.58
	101	16.38
	107	27.75
	115	42.26
	122	3.53
	153	23.34
	158	64.69
	160	9.80
	162	49.63
	174	30.72
	177	3.26
	184	6.67
	192	3.27
	194	14.99
	214	14.81
	216	6.44
UUS	24	14.26
	181	.76

TABLE B.1-4
LAND USE POLYGON ACREAGES FOR STUDY AREA (CONTINUED)

T-2486-7

LAND USE	LAND USE POLYGON NO.	AREA IN ACRES
UUT	51	32.16
	60	10.95
	128	59.29
WO	10	7.41
	23	30.46
WS	96	203.36
	104	4.06
WWP	2	.49
	139	1.62
	205	3.24
	208	2.87
	210	2.36
	222	.17
ACP	212	32.12
	221	15.58
AVF	219	39.39
AVV	211	20.77
	213	18.62
	220	8.65
URS	214	15.07

TABLE B.1-5

AREAS OF LAND USE POLYGONS IN TRIANGULAR REGIONS TO BE MERGED

T-2517

	LAND USE	POLYGON IDENTIFIER	AREA (ACRES)
Triangle 1	ACP	221	1.96
	AVF	219	38.20
	AVV	211	1.50
	AVV	213	8.34
	AVV	220	8.65
	URS	214	4.67
Triangle 2	ACP	212	32.12
	ACP	221	13.61
	AVF	219	1.20
	AVV	211	19.24
	AVV	213	10.26
	URS	214	10.23

TABLE B.1-6

ACREAGE OF LAND USE AREAS IN MERGED TRIANGLES

T-2499

LAND USE	POLYGON IDENTIFIER	AREA (ACRES)
ACP	212	32.12
ACP	221	15.58
AVF	219	39.39
AVV	211	20.77
AVV	213	18.62
AVV	220	8.65
URS	214	15.07

TABLE B.1-7

LAND USE POLYGON AREAS AFFECTED BY UPDATES

T-2515

LAND USE	POLYGON IDENTIFIER	AREA AFTER UPDATE (ACRES)
URS	214*	151.50
AVV	213	0.0
ACP	212	0.0
AVV	211	0.0
AVF	33	89.92
AVV	221	0.0
AVF	219	0.0
ACP	40	0.0
AVV	220	0.0
AVV	39*	59.31
AVF	41	133.28
AVV	43	0.0
ACC	42*	52.10
UUT	51	32.80
UIS	48*	39.08
ACC	46	0.0
AVV	47	0.0
AVF	44**	48.99

*Polygons which exist on original land use map and are expanded to form the regions on the update map.

**New polygon formed by division of polygon 41.

APPENDIX D
CONTROL OF EXECUTION OF DEMONSTRATION SOFTWARE ON THE CDC 6400

D.1 USE OF THE SOFTWARE TAPE, VSN = E00331

To execute the delivered software, the LCG programs desired for any run must be copied off the software tape described in Section 3.4.7. The following examples show how to access the programs for separate runs.

In all cases, to identify the software tape, this card is required:

```
REQUEST,SOFT,HY,NORING,VSN=E00331.
```

(An UNLOAD should be commanded after accessing the required software). The appropriate file must be located and copied from the tape for each program:

For DMA:

```
COPYBF,SOFT,DMA.
```

For FRED:

```
SKIPF,SOFT,1,17,B.
COPYBF,SOFT,FRED.
```

For POLLY:

```
SKIPF,SOFT,2,17,B.
COPYBF,SOFT,POLLY.
```

For PSYCHE:

```
SKIPF,SOFT,5,17,B.
COPY,SOFT,PSYCHE.
```

For WHIRL:

```
SKIPF,SOFT,6,17,B.
COPYBF,SOFT,WHIRL
```

To overlay three files, both WHIRL and PSYCHE are required:

```
SKIPF,SOFT,5,17,B.
COPYBF,SOFT,PSYCHE.
COPYBF,SOFT,WHIRL.
```

Similarly POLLY is useful in association with FRED, for display and editing in the same run:

```
SKIPF,SOFT,1,17,B.
COPYBF,SOFT,FRED.
COPYBF,SOFT,POLLY.
```

D.2 USE OF THE DATA TAPE, VSN = E00332

The delivered data and the combined data base generated at ETL reside on this tape, described in Section 3.4.8. The following examples show representative runs.

Example 1 - To access the combined data base (for DMA runs to do plots and tabulations):

```
REQUEST,DATA,HY,NORING,VSN=E00332.
SKIPF,DATA,12,17,B.
COPYBF,DATA,PHB.
COPYBF,DATA,PHG.
COPYFB,DATA,CDB.
COPYBF,DATA,CDG.
REWIND,PHB,PHG,CDB,CDG.
UNLOAD,DATA.
DMA(PHB,PHG,CDB,CDG)
```

Example 2 - To access land use file for use with FRED or POLLY:

```
REQUEST,DATA,HY,NORING,VSN=E00332.
SKIPF,DATA,6,17,B.
COPYBF,DATA,LAND.
COPYBF,DATA,LANDG.
UNLOAD(DATA)
REWIND,LAND,LANDG.
FRED,LAND,LANDG,EDITED,EDITEDG.
```

D.3 SAMPLE OVERLAY RUN

The deck ETHAO performs the overlay of the delivered digitized data. The following steps are performed in order:

(1) Mount software tape and access PSYCHE and WHIRL located on files 6 and 7.

(2) Mount data tape, copy all three coverages to disk, and rewind them.

(3) Run WHIRL with first pair of files (here they are land use and floodplain).

(4) Do precautionary rewind of output.

(5) Run PSYCHE to sort output of previous WHIRL step.

(6) Do precautionary rewind of output.

(7) Run WHIRL with sorted intermediate output and third coverage.

(8) Save output files (4 of them) on tape.

This sequence could be executed with different input files such as the revised land use data.

The deck contains a control card record and one record of commands for each WHIRL step. PSYCHE requires no commands. The WHIRL commands for the first step are the minimum needed for proper operation. The PROXIMITY value of 0.01 (meters) specifies an allowable error of one centimeter on the ground for the location of line intersection. Use of a smaller value is not advisable because it may lead to excessive computation time in the overlay process.

The second control record demonstrates the ability to alter allocations of memory during the overlay process. The allocations used here are necessary to handle the elevation data in combination with the intermediate results. The statistics printed at the end of the run show the various requirements for this procedure. If run with different files (particularly if more complex), these allocations should be adjusted.

D.4 SAMPLE EXTRACTION RUN

D.4.1 Control Card Setup

The selection, extraction, and plotting module of the delivered software (the DMA program) should be executed in the following manner, as illustrated in the deck ETHT1.

(1) Request the software tape, access the program (on file 1) and unload.

(2) Request the data tape and access the output of WHIRL saved by ETHAO or a similar run and unload.

(3) Rewind the four files containing the combined data base. (The preceding steps will depend on the actual details of storage for the particular data base.)

DMA must be loaded with the system CALCOMP software if plots are produced.

The program is invoked with the local file names of the four components of the combined data base in this order: polygon data, polygon global, chain data, chain global. Note that although this is the order of files on the tape, the tape should not be used, because the files are not necessarily read in sequence. If plotting is not needed the chain data file need not be present, but the chain global is examined whether or not plots are generated. The polygon files are always read.

When plots are generated they are directed to a local file named PLOTS which must be directed to the proper deposition (on tape) to ensure plotting. Use the standard form:

```
REQUEST,TAPE29, ...
```

D.4.2 Command Language

The language of the DMA module is described in detail in Section 3.4.5. The commands in deck ETHT1 demonstrate the capabilities required to fulfill the contract deliverables. One set includes no plotting (the DRAW or "DR", command has been avoided). A duplicate set of commands performing plots is hidden after the QUIT command.

D.5 DATA BASE REVISION CAPABILITY

The deck ETHAR controls a run which uses the programs FRED and POLLY to modify TDB files and compare the results. The steps of the run are as follows:

(1) Copy FRED and POLLY from the software tape (files 2 and 3).

(2) Get the land use data from data tape. Both the rounded and unrounded versions are extracted and rewound.

(3) FRED is run to perform the updates described in the statement of work. The local file names of the input chain data and global files appear first, followed by the output chain and globals.

(4) Second run of FRED is made to demonstrate the ability to list chain topology. To perform this function with FRED, only an input pair of files is required.

(5) POLLY, the comparison module, is used to compare the original landuse with the updated version, then compare the rounded and unrounded files.

REFERENCES

1. Lillestrand, R.L. and Hoyt, R.R., "The Design of Advanced Digital Image Processing Systems," Photogrammetric Engineering, Vol. XL, No. 10, October 1974, pp. 1201-1218.

2. Bernstein, R. and Ferneyhough, D.G., Jr., "Digital Image Processing," Photogrammetric Engineering and Remote Sensing, Vol. 41, No. 12, December 1975, pp. 1465-1476.

3. Gambino, L.A. and Crombie, M.A., "Digital Mapping and Digital Image Processing," Photogrammetric Engineering, Vol. XL, No. 11, November 1974, pp. 1295-1302.

4. Gambino, L.A. and Boulis, R.L., "Defense Mapping Agency/USAETL STARAN Complex," Sagamore Computer Conference Proceedings, August 1975, pp. 132-141.

5. Enslow, P.H., Jr., Ed., Proceedings of the 1976 International Conference on Parallel Processing (Detroit), August 1976, IEEE Catalog No. 76CH1127-0C (see especially pp. 18-34).

6. Beavers, A.N., Jr., Gibbons, J.E., Jr., Sharpley, W.K., Jr., "Defense Mapping Agency Digital Data Handling System Phase 1 Study (U)," The Analytic Sciences Corporation Report No. TR-621-2 (TCS-P-545064-76), May 1976 (CLASSIFIED).

7. Gibbons, J.E., Jr., Sharpley, W.K., Jr., and Mannos, J.L., "DMA DDHS Engineering Development Phase General Test Plan," The Analytic Sciences Corporation, Report No. TR-847-2, September 1977.

8. "PAPERS/Notebook (Proceedings)," An Advanced Study Symposium on Topological Data Structures for Geographic Information Systems (Harvard University), October 1977, Four Volumes.

9. "Defense Mapping Agency Resource Objectives Plan (DROP) FY 1979-86 (U)," Defense Mapping Agency, March 1977 (SECRET).

10. Helava, U.V., "The Analytical Plotter - Its Future," Photogrammetric Engineering and Remote Sensing, Vol. 43, No. 11, November 1977, pp. 1361-1362.

REFERENCES (CONTINUED)

11. Doyle, F.J., "The Next Decade of Satellite Remote Sensing," Photogrammetric Engineering and Remote Sensing, Vol. XLIV, No. 2, Feb. 1978, pp. 155-164.

12. "Impact of Data Intensive Programs," The National Research Council Report, Geophysics Research Board (1976) (see especially page 12).

13. Price, K.E., "Change Detection and Analysis in Multi-Spectral Images," PhD Thesis, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pa., Dec. 1976.

14. Cook, H.R., "Digital Topographic Information Bank," Proceedings of the American Congress on Surveying and Mapping (Phoenix), Oct. 1975, pp. 144-166.

15. Bernstein, R. and Ferneyhough, D.G., Jr., "Digital Image Processing," Photogrammetric Engineering and Remote Sensing, Vol. XII, No. 12, Dec. 1976, pp. 1465-1476.

16. Grubbs, J.B. (Maj.), Gunther, A.C., and Orsinger, R.J., "The Army Terrain Information System Report No. 1 - Data Base Storage Requirements," U.S. Army Engineer Topographic Laboratories, Report No. ETL-0050, March 1976.

17. Jancaitis, J.R., "Elevation Data Compaction by Polynomial Modeling," Proceedings of the 1977 ASP-ACSM Joint Annual Spring Convention (Washington, D.C.), March 1977.

18. Peucker, T.D., Fowler, R.J., Little, J.J., and Mark, D.M., "Triangulated Irregular Networks for Representing Three-Dimensional Surfaces," Department of Geography, Simon Fraser University, Geographic Data Structure Project Technical Report No. 10, April 1976.

19. Dutton, G., "Navigating ODYSSEY," An Advanced Study Symposium on Topological Data Structures for Geographic Information Systems (Harvard University), Vol. 3, Oct. 1977.

20. White, D., "A New Method of Polygon Overlay," An Advanced Study Symposium on Topological Data Structures for Geographical Information Systems (Harvard University), Vol. 1, Oct. 1977.

REFERENCES (CONTINUED)

21. Morehouse, S., "ODYSSEY File System", An Advanced Study Symposium on Topological Data Structures for Geographic Information Systems (Harvard University), Vol. 4, Oct. 1977.

22. Dougenik, J., "PASTA: A Program to Process Digitized Output," Laboratory for Computer Graphics and Spatial Analysis (Harvard University), Internal Report No. 76-10, 1976.

List of Acronyms

DDHS - Digital Data Handling System

DLMS - Digital Land Mass Simulation

DMA - Defense Mapping Agency

ETL - U.S. Army Engineer Topographic Laboratories

LCG - The Laboratory for Computer Graphics and Spatial Analysis of Harvard University

LCGU - Least Common Geographical Unit

MC&G - Mapping, Charting and Geodesy

NSF - National Science Foundation

PRU - Physical Record Unit (64 60-bit words on CDC 6400)

TDB - Topological Data Base(s)

UTM - Universal Transverse Mercator