HAI Model

Release 5.0a

Model Description

HAI Consulting, Inc.

737 29th Street, Suite 200
Boulder, Colorado 80303

Revised: February 16, 1998

TABLE OF CONTENTS

1. Introduction 1
   1.1. Overview 1
   1.2. Evolution of the Hatfield/HAI Model 3

2. Summary of Changes Between HM 4.0 and HM 5.0 5
   2.1. User Interface 5
   2.2. Input Data 5
   2.3. Outside Plant Selection 6
   2.4. Distribution Module 6
   2.5. Feeder Module 7
   2.6. Switching and Interoffice Module 7
   2.7. Expense Modules 8
   2.8. Changes Incorporated in HM 5.0a 8
   2.8.1. Distribution Module 8
   2.8.2. Switching and Interoffice Module 9
   2.8.3. Expense Modules 9
   2.8.4. Interface Items 10
   2.8.5. Input Data Items 10

3. Fundamental Structure of Local Network 11
   3.1. Components of the Local Exchange Network 11
   3.1.1. Loop Description 11
   3.1.2. Switching and Interoffice Network Description 14

4. HM 5.0a Model Organization, Structure and Logic 18
   4.1. Overview of HM 5.0a Organization 18
   4.2. Input Data 19
   4.3. Workfiles 19
   4.4. User Interface 19
   4.5. Distribution Module 20
   4.6. Feeder Module 21
   4.7. Switching and Interoffice Module 22

4.8. Expense Modules **22**

5. Input Data **24**

5.1. Line Type Counts by Study Area **24**

5.2. Wire Center List **24**

5.3. Customer Counts by Census Block and Wire Center **25**
5.3.1. Residence Counts 25
5.3.2. Business Counts 26
5.3.3. Location and Line Counts by Wire Center 27

5.4. Customer Location **27**
5.4.1. Residence Location Data 28
5.4.2. Business Location Data 28
5.4.3. Geocoding 29
5.4.4. Gross-up 30

5.5. Customer Location Clustering **31**
5.5.1. General Criteria 31
5.5.2. Clustering Algorithm 32

5.6. PointCode Translation Processes **33**

6. Module Descriptions **34**

6.1. Input Data Files **34**
6.1.1. Demographic and Geological Parameters 34
6.1.2. Wire Center Locations and Interoffice Distances 35
6.1.3. ARMIS Data 35
6.1.4. User Inputs 36

6.2. Outside Plant Engineering **36**
6.2.1. Outside Plant Structures 38
6.2.2. Terrain and Its Impact on Placement Costs 39
6.2.3. Structure Sharing 40
6.2.4. Lines Density Considerations 40
6.2.5. Economic Adjustment of Structure Fractions 41

6.3. Distribution Module **42**
6.3.1. Treatment of Main Clusters 42
6.3.2. Treatment of Outlier Clusters 43
6.3.3. Customer Drop Arrangement 44
6.3.4. Investment Cap to Reflect Potential Wireless Technologies 44
6.3.5. Determination of Feeder Technology 45
6.3.6. "Steering" Feeder Routes 46
6.3.7. Calculation of Distribution Investments 46
6.3.8. Calculation of SAI and DLC Investments 47
6.3.9. Calculation of Drop Investments 48

6.4. Feeder Module **48**
6.4.1. Overview 48
6.4.2. Development of Feeder Investments 51

6.5. Switching and Interoffice Module **54**
6.5.1. Overview 54
6.5.2. Description of Inputs and Assumptions 55
6.5.3. Explanation of Calculations 56

6.6. Expense Modules **64**
6.6.1. Overview 64
6.6.2. Capital Carrying Costs 66
6.6.3. Operating Expenses 68
6.6.4. Expense Module Output 71

7. Summary **75**

LIST OF APPENDICES

A. **History of the Hatfield/HAI Model HM 5.0a**

B. **Inputs, Assumptions and Default Values**

C. **HM 5.0a Input Data Development Flow Charts**

D. **General Rules Governing the Creation of the HM 5.0a Distance Files**

E. **Equation Listings for the HM 5.0a Network Engineering Logic Modules: Distribution, Feeder, Switching and Interoffice**

1. Introduction

1.1. Overview

The HAI Model, Release 5.0a ("HM 5.0a") has been developed by HAI Consulting, Inc. ("HAI"), of Boulder, Colorado,[1] at the request of AT&T and MCI for the purpose of estimating the forward-looking economic costs of:

a) Basic local telephone service;

b) Unbundled network elements ("UNEs"); and

c) Carrier access to, and interconnection with, the local exchange network.

All three sets of costs are calculated based on Total Service Long Run Incremental Cost ("TSLRIC") principles, and use a consistent set of assumptions, procedures and input data.[2]

The HAI Model uses the definition of basic local telephone service adopted by the Federal-State Joint Board on Universal Service ("Joint Board") for universal service funding purposes. The Joint Board states that the following functional elements are to be considered as required components of universal service:[3]

- single-line, single-party access to the first point of switching in a local exchange network;
- usage within a local exchange area, including access to interexchange service;
- touch tone capability;
- access to 911 services, operator services, directory assistance, and telecommunications relay service for the hearing-impaired.

Excluded from this definition of universal service are many other local exchange company ("LEC") services, such as toll calling, custom calling and CLASS[SM] features,

[1] With its Release 5.0a, the model formerly known as the Hatfield Model is now named the HAI Model. Hatfield Associates, Inc., the firm that developed prior versions of the Hatfield/HAI Model no longer performs telecommunications consulting. All of the staff of Hatfield Associates who have played an active role in developing the Hatfield/HAI Model have formed a successor firm, called HAI Consulting, Inc.

[2] When applied to the costing of unbundled network elements, TSLRIC equates to Total Element Long Run Incremental Costs, or TELRIC as the term is used by the Federal Communications Commission.

[3] Federal-State Joint Board on Universal Service, CC Docket No. 96-45, Recommended Decision, November 8, 1996, ("Recommended Decision") Paragraph 45-53, 65-70.

private line services and white pages directory listings.[4] The existence of such services is taken into account in developing the cost estimates for UNEs -- to the extent that the joint provision of UNEs and other services impacts the costs of UNEs. Model users also may adjust the degree to which several specific UNEs are included in calculating universal service support requirements.

The HAI Model calculates the costs of the following UNEs:

- Network Interface Device ("NID")
- Loop Distribution
- Loop Concentrator/Multiplexer
- Loop Feeder
- End Office Switching
- Common Transport
- Dedicated Transport
- Direct Transport
- Tandem Switching
- Signaling Links
- Signal Transfer Point ("STP")
- Service Control Point ("SCP")
- Operator Systems
- Public Telephones

Finally, the model estimates the per-minute economic cost of providing local network interconnection and access. These are estimated for connection points at end office and tandem switches.

The model constructs a "bottom up" estimate of the pertinent costs based upon detailed data describing demand quantities, network component prices, operational costs, network operations costs, and other factors affecting the costs of providing local service. The model's demand data, particularly data describing customer locations, line demand, and traffic volumes, serve as the key initial drivers. From these data, the model engineers and costs a local exchange network with sufficient capacity to meet total demand, and to maintain a high level of service quality.[5] The model's inputs also include the prices of various network components, with their associated installation and placement costs, along with various capital cost parameters. These data are used to populate detailed input tables describing, for example, the cost per foot of various sizes of copper and fiber cable, cost per line of switching, cost of debt, and depreciation lives for each specific network component.

[4] Although previous versions of the Hatfield/HAI Model included the monthly cost of maintaining a white pages telephone listing for each subscriber, the Joint Board and FCC have explicitly excluded this item from the definition of supported universal service. Thus, in HM 5.0a its inclusion in cost calculations for basic service is only at the user's express direction.

[5] In general, the level of service quality engineered into the HAI Model exceeds, by a substantial margin, the customary level of basic service quality offered by the LECs over their embedded networks.

Using these data, the model calculates required network investments by detailed plant category. Next, the capital carrying cost of these investments is calculated. Operations expenses are then added to compute the total monthly cost of universal service, various unbundled network elements, stated on both a total cost and an appropriate per-unit basis, and carrier access to and interconnection with the local exchange network. Costs can then be displayed on a study area, density zone,[6] wire center, Census Block Group ("CBG"), or customer cluster basis.[7]

This document describes the structure and operation of the HM 5.0a, including a discussion of various inputs to the model. Section 1.2 describes the recent evolution of the Hatfield/HAI Model. Section 2 summarizes changes made to the model between HM 4.0 and this version. Section 3 provides a general overview of the local network being modeled. Section 4 reviews briefly the structure of the model and its data. Section 5 focuses on the method by which customer locations are determined and clustered. Section 6 describes in detail each module and its operation. Section 7 summarizes the document.

Appendix A provides a brief history of the Hatfield/HAI Model. Appendix B identifies the user inputs to the model and their default values. Appendix C provides flow charts describing the data input development process used to obtain demographic and geological information, residence and business line counts, wire center mappings and loop distances. Appendix D describes the HM 5.0a's calculation of interoffice network distances. Finally, Appendix E provides equation listings of the HM 5.0a's network engineering logic modules.

1.2. Evolution of the Hatfield/HAI Model

On May 7, 1997, the FCC released its Order implementing the mandate for universal service contained in the Telecommunications Act of 1996. In the Order, it declined, on the basis of its current record, including the Report of the State Members of the Joint Board, to endorse a model, and indicated it would issue a Further Notice of Proposed Rulemaking (FNPRM) detailing what it believed to be the appropriate requirements and guidelines that such a cost methodology should incorporate. This FNPRM was released on July 18, 1997. In this FNPRM the FCC provided a wealth of information about what the Commission believes are the appropriate properties to be incorporated into a proxy cost methodology. These include:

- A more sophisticated and precise method of locating customers;

[6] The HM 5.0a differentiates among density zones based on the number of subscriber access lines per square mile of service area.

[7] A CBG is a unit defined by the U.S. Bureau of the Census, and nominally comprises between 400 and 600 households. Customer clusters are dynamically formed aggregations ranging from singleton isolated customer locations, up to 1800 customer locations. See, Section 5.5 below, for a description of the spatial and size criteria used by the HM 5.0a in forming customer clusters.

- A choice of outside plant technologies and structures that reflects more closely local cost conditions;

- Explicit modeling of host/remote relationships between end office switches; and

- More flexible assignments of expenses based either on lines or relative investments.

The Commission set up a series of weekly meetings, and Comment and Reply cycles to address each of these and other related issues in greater depth. The Commission also indicated its intention to select a model for determining universal service support for nonrural carriers by the end of 1997.

HM 5.0a, as here submitted, is responsive to each the Commission's requirements as presented in the Order, the requirements outlined in the FNPRM on cost modeling, and the public notice guidance provided by the Commission subsequent to its release of the FNPRM. Indeed, HM 5.0a represents a revolutionary advance in the modeling of local telephone network costs by its incorporation of:

- Actual gecoded customer locations;

- An algorithm that identifies clusters of customers that may be served efficiently together – without recourse to arbitrary geographic limitations;

- Numerous optimization routines that ensure the use of outside plant that is most technically and economically suited to particular local conditions;

- Explicit specification of host, remote and stand-alone switches;

- An optimizing algorithm for the creation of efficient interoffice SONET transport rings; and

- Opportunities to allocate flexibly expenses based on lines or relative investments.

As a result of these many changes, HM 5.0a has refined greatly the task of identifying actual customer locations, and clustering them into units logically served by telecommunications outside plant. The model has thus moved well ahead of other models that employ more geographically limited, rule-of-thumb calculation techniques.

HM 5.0 was originally submitted to the FCC on December 11, 1997. A number of small but significant changes have been made to the Model's data, logic and documentation since that time. These are incorporated into a revision referred to as HM 5.0a, released January 28, 1998. Section 2.8 summarizes the changes between HM 5.0 and HM 5.0a. To the extent those changes impact the model description, they are reflected in this document.

2. Summary of Changes Between HM 4.0 and HM 5.0

The changes between HM 5.0 and the previous release of the model, HM 4.0, are summarized in the first portions of this section. Section 2.8 summarizes the changes between HM 5.0 and HM 5.0a. All of these changes are reflected in the discussion of how HM 5.0a operates, presented in Sections 4 and 6.

2.1. User Interface

- The new features of the user interface provide the user with many additional inputs and options. Among the new inputs included are the ability to designate specific end office switches as hosts, remotes, or standalones – as well as to assign remotes to a particular host; ability to specify variable T1 repeater spacing; ability to enable the steering of feeder toward population clusters within a quadrant; the ability to invoke a wireless distribution option if its cost is less than wireline, and many more.

- The interface also now allows the user to select multiple companies from one or more states (limited only by hard drive space) to be run in automatic sequence by the model. Expense Modules and workfiles are then produced for each individual company, and their universal service calculations rolled up.

2.2. Input Data

- The HM 5.0a input data locate customers much more precisely. These data determine the actual precise locations of as many customers as possible through latitude and longitude geocoding of their addresses. The remainder are located to at least the Census Block ("CB") level of precision and are assumed to be placed along the CB's periphery.[8]

- A clustering algorithm is used to determine groupings of customers that have extremely realistic correlation to efficient distribution areas.

- The August 1997 Local Exchange Routing Guide ("LERG") is used to identify and locate LEC wire centers.

- Business Location Research ("BLR") wire center boundaries are used to associate customer locations with LEC wire centers. This ensures that all identified clusters

[8] Previous versions of the HM only located customers precisely to their Census Block Group (CBG). Within high density CBGs, customers were assumed either to be spread uniformly across the CBG. In low density CBGs, a portion of customers was assumed to be clustered in quadrants, while another portion was assumed to spread along outlying roads.

are restricted to include only customer locations that fall within the boundaries of a single wire center.

- Company line count totals are determined from the most recent available data, including that provided in the 1996 ARMIS data and NECA USF Loops filing for 1996.

- The method of estimating line counts by LEC wire center is refined, and line counts can be determined by CB.

- 1996 ARMIS data (rather than 1995 ARMIS data) are used to estimate traffic volumes and expense inputs.

2.3. Outside Plant Selection

- HM 5.0a automatically adjusts buried and aerial structure fractions to account for varying maintenance costs and placement costs occasioned by local soil conditions and bedrock. The amount of one type of structure substituted for another depends both on differences in placement costs and on a life-cycle analysis of maintenance and capital carrying costs of the two types of structure.

2.4. Distribution Module

- HM 5.0a lays its distribution plant directly over the actual identified locations of customer clusters.

- Rather than assuming that the distribution area is square, HM 5.0a engineers its distribution grid as a rectangle. The aspect ratio (height-to-width) of this rectangle is determined by the data input development process for each cluster, and distribution cable is laid out in a fashion that reflects this aspect ratio.

- HM 5.0a serves "outlier" clusters from "main clusters" on which they home, using digital T1 technology whenever the road cable length exceeds a user-adjustable maximum analog copper distance.[9] The cables carrying T1 signals to the outlier clusters are separate from the analog copper cables that extend from the T1 terminal in each outlier cluster to the customer locations within the outlier cluster.

- Assuming that the distance of a cable run is sufficiently short so that use of copper feeder is a technically acceptable option, the HM 5.0a performs an analysis of the relative life-cycle costs of copper versus fiber feeder to determine which feeder technology should be used to serve the given main cluster.

---

[9] Outlier clusters are clusters that contain fewer than five lines. Main clusters are cluster containing five or more lines. These clusters are served by feeder linking them to their serving wire center. See Section 6.3.2 for more detail.

- The HM 5.0a also incorporates an optional, user-adjustable "cap" on distribution investment. This cap is structured to reflect the potential cost structure of wireless distribution technologies.

2.5. *Feeder Module*

- HM 5.0a engineers feeder to serve actual population main clusters (and uses distribution cable to serve main clusters' subtending outlier clusters), rather than simply engineering to each CBG.

- At the user's option, the HM 5.0a "steers" feeder routes toward the preponderant location of main clusters within a given wire center quadrant.[10] When this steering is invoked, the user may also apply an adjustable route-to-airline distance multiplier to the amounts of cable placed along these "steered" feeder routes.

- Manhole placement costs are increased by a user-specified amount whenever the local water table depth is less than the user-specified threshold.

2.6. *Switching and Interoffice Module*

- At the user's discretion, HM 5.0a will both engineer and cost explicit combinations of host, remote and stand-alone end office switches. If the user does not make such a specification, the HM 5.0a defaults to computing end office switching investments using input values that provide average per-line investments for an efficient portfolio of host, remote, and stand-alone switches.[11] If the host/remote/standalone designation option is invoked, the user is required to specify whether a wire center houses switches that are hosts or remotes, as well as to assign the correspondence between host and remote switches.

- Further, when the user chooses the model to distinguish explicitly between switch types, the HM 5.0a assumes that each host and its remotes are on a Synchronous Optical Network ("SONET") fiber optics ring separate from the interoffice rings used to interconnect host, standalone and tandem switches with each other.

- The HM 5.0a calculates explicitly a set of interoffice SONET rings that interconnect host, standalone, and tandem switches with each other. Based on this explicit specification of what wire centers are on each interoffice ring, the HM 5.0a determines associated ring distances using the actual locations of the wire centers along the ring. In addition, the rings are appropriately interconnected

---

[10] The default treatment, if steering is not invoked, is for the Feeder Module to calculate feeder distances using "right angle routing" in the four cardinal compass point directions, as employed in HM 4.0.

[11] The Model defaults to an average per-line mix because accurate data on the purchase prices of a portfolio of host, remote and standalone switches of varying capacities, and on the identification of hosts, remotes, and stand-alone switches, may not be available to the user.

with each other, and tandem switches are also interconnected if they fall within the same LATA.[12].

- The HM 5.0a engineers redundant paths and associated transmission terminal equipment for the point-to-point (folded) rings that may be specified to connect small offices to the larger wire centers on which they home.[13]

2.7. Expense Modules

- A Uniform System of Accounts ("USOA") detail worksheet is included that breaks out HM 5.0a investments and expenses by Part 32 account for comparison purposes.

- The proportion of total expenses that are assigned to loop network elements (i.e., NID, distribution, concentration and feeder) can be varied based either on relative number of lines, or on the relative amount of direct expenses (direct expenses include both maintenance expenses and capital carrying costs for the specific network elements).

- Both federal and state universal service fund requirements can be calculated in the density zone USF worksheet. This separate calculation permits differing state and federal cost benchmarks to be specified, as well as different collections of local services (e.g., primary and secondary residential lines, single business lines, etc.) to receive universal service support.

- In addition to displays of costs at the lines density zone and wire centers levels of aggregation, costs can also be displayed at the CBG and individual population cluster level.

2.8. Changes Incorporated in HM 5.0a

2.8.1. Distribution Module

- HM 5.0a modifies its method of dividing clusters to more efficiently ensure that the length of cables carrying analog signals never exceeds the user-set maximum (default = 18,000 ft).

---

[12] InterLATA links are excluded from the model because such links are not part of the local exchange network.

[13] The user may specify a minimum number of lines that a wire center must serve (default of just one) before that wire center is placed on an interoffice ring with other end office switches – rather than being interconnected directly only to its "home" wire center.

- HM 5.0a corrects minor typographical errors in equations used to calculate the portions of structure that "swing" between buried and aerial based on abnormal local life-cycle costs, and in the wireless cap equations.

- Adds columns that calculate average loop lengths.

2.8.2. Switching and Interoffice Module

- The time required to execute this module for large companies is reduced by sourcing from other portions of the workbook, rather than calculating, certain distance and DS3 count information.

- For wire centers owned by small LECs without local tandems, connectivity to a tandem is established in two pieces. First, a spur is engineered to the closest large LEC wire center that is on an interoffice ring. Second, the equivalent investments in facilities and terminal equipment associated with the required number of leased circuits on this ring that are used to connect this large LEC wire center to its tandem are calculated on a per-DS0 facilities basis. This is in contrast to the previous method of determining the cost of interoffice route between the Large LEC wire center and its tandem based on multiplying the distance between these nodes by an assumed dedicated circuit-mile charge.

- Rings now must have a minimum of four nodes, assuming there are that many wire centers, versus a prior minimum of two nodes.

- HM 5.0a provides several new "traps" to prevent certain execution problems. These include: 1) the ring-generating code is modified to expect the user-specified "host/remote enable" option as boolean type rather than a string; 2) stand-alone tandems now have an associated interoffice distance; 3) the number of allowed wire center records has been increased from 1,500 to 2,000; 4) the ring-generating code contains logic to determine whether host/remote calculations are enabled before eliminating remotes as first order ring candidates; 5) the ring-generating code uses wire center records generated from the HM5.0a database as the source of the locations associated with a particular state and operating company; 6) the ring-generating code now updates the progress bar in closer proportion to the module's degree of completion; 7) the ring-generating code writes all results into a list in the "ring io" worksheet; 8) the array dimension in the routine computing interoffice mesh distances has been increased from 25 to 100 elements; and 9) several additional "divide checks" are provided and syntax errors corrected.

2.8.3. Expense Modules

2.8.3.1. Density Zone and Wire Center Versions

- Corrects the calculation of weighted average depreciation life for non-metallic cable to include interoffice fiber facilities.

- The "Cost detail" sheet of the DZ version allows for the substitution of ICO-equivalent DS0 transport values.

- Corrects cell references for residential and business usage in the wire center USF sheet from absolute to relative.

2.8.3.2. Expense Modules – CBG and Cluster Versions

- Improves on the previous CBG expense module by associating cluster costs to the several CBGs that may overlay the cluster in proportion to the relative number of lines that each CBG displaces of the cluster's total quantity of lines.

- Adds a Cluster expense module that displays cost results on a customer cluster-by-cluster basis.

2.8.4. Interface Items

- Corrects several non-functioning items in the interface, including: 1) permitting Puerto Rico to be run through the interface; 2) fixing the OLE error that previously has occurred the initial time the newly installed HM 5.0 is run; and 3) speeding the run time of the Feeder module

2.8.5. Input Data Items

- Corrects several data discrepancies, including: 1) correcting the several "problem clusters" that previously were incorrectly sized; 2) adding the clusters that were missing from the California data; and 3) assigning correctly the lines density classification of Puerto Rico clusters; and 4) correcting the state assignment of several small LECs that operate across state borders.

- Adds CBGMulti data table that relates clusters to the several CBGs that overlay them based on relative counts of lines associated with each CBG.

- Adds data that permit easy calculation of average loop lengths by cluster and wire center.

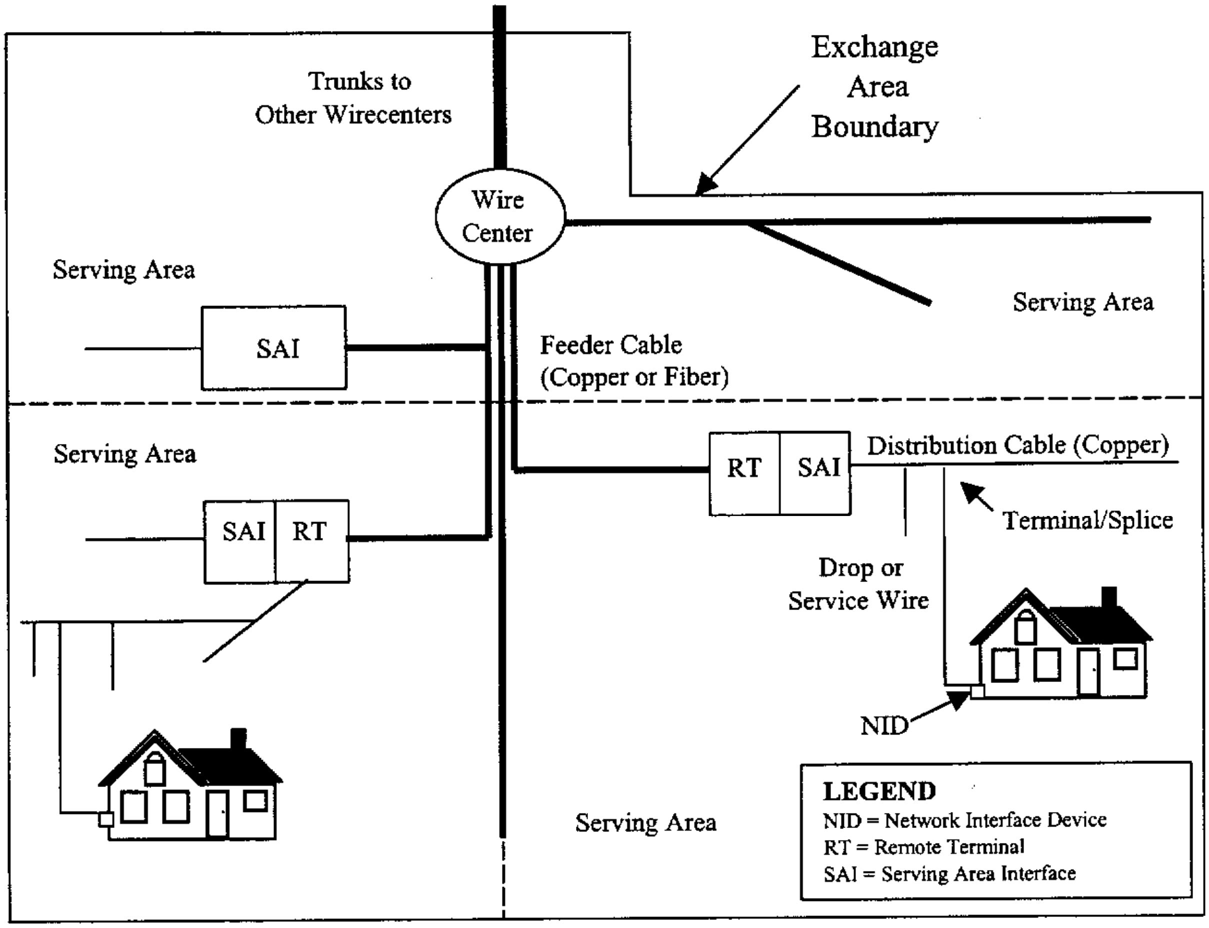
3. Fundamental Structure of Local Network

3.1. Components of the Local Exchange Network

This section describes the network configuration and components modeled in HM 5.0a. Figures 1, 2 and 3 depict the relationships among the loop, switching, interoffice, and signaling network components.

3.1.1. Loop Description

Figure 1 depicts the loop model utilized in HM 5.0a. Section 3.1.1.1 defines the serving area. Section 3.1.1.2 provides a general description of the loop, depicted in Figure 1. Section 3.1.1.3 describes the loop components in more detail.



Adapted from Engineering and Operations in the Bell System, 2nd Edition, 1983

Figure 1 Loop Components

3.1.1.1. Serving Area

The total area served by a wire center is organized into one or more serving areas, each of which contains a portion of the area and lines served by the wire center. The serving areas are delineated by dotted lines in the above figure. In HM 5.0a the serving areas equate to main customer clusters and their subtending outlier clusters, as discussed in Section 6.2.

3.1.1.2. General Loop Description

One end of the feeder portion of the loop terminates within the central office building, or "wire center." Copper cable feeder facilities terminate on the "vertical side" of the main distributing frame ("MDF") in the wire center, and fiber optic feeder cable serving integrated digital loop carrier ("IDLC") systems terminates on a fiber distribution frame in the wire center.

The other end of the feeder extends to an appropriate termination point in the serving area. Copper feeder cables terminate on one or more serving area interfaces ("SAIs") in each serving area, where they are cross-connected to copper distribution cables. Fiber feeder cables extend to a digital loop carrier ("DLC") remote terminal ("RT") in the serving area, where optical digital signals are demultiplexed and converted to analog signals. Individual circuits from the DLC are cross-connected to copper distribution cables at an adjacent SAI.

Copper distribution cable extends from the SAI along routes passing individual customer premises. At appropriate points, these cables pass through block terminals typically serving several housing units. In the terminal, individual copper pairs in the distribution cable are spliced to "drops" that extend from the terminal to the customers premises. The drop terminates at a network interface device, or NID, at the customer's premises.

Feeder, distribution, and drop cables are supported by "structures." These structures may be underground conduit, poles, or trenches for buried cable and underground conduit. Underground cable is distinguished from buried cable in that underground cable is placed in conduit, while buried cable comes into direct contact with soil.[14] In more urban areas, aerial distribution cable may be attached directly to the outside of buildings, in what is called a "block cable" arrangement, or, for high-rise buildings, may consist of riser cable inside the building.

3.1.1.3. Local Loop Components

1) *Network Interface Device*

   The NID is the demarcation point between the local carrier's network and the customer's inside wiring. This device terminates the drop wire and is an access point that may be used to isolate trouble between the carrier's network and the customer's premises wiring. The NID also contains protection against externally-

[14] Although the conduit supporting underground cable is always placed in a trench, buried cable may either be placed in a trench or be directly plowed into the earth.

induced hazardous voltages, such as those associated with lightning strikes and contact between telephone and electric lines. In a multi-tenant building, the protection is located at the point at which the distribution cable enters the building.

2) *Drop*

A copper drop cable, typically containing several wire pairs, extends from the NID at the customer's premises to the block terminal at the distribution cable that runs along the street or the lot line. The drop can be aerial or buried; generally it is aerial if the distribution cable is aerial, and buried if the distribution cable is buried or underground.

3) *Block Terminal*

The "block terminal" is the interface between the drop and the distribution cable. When aerial distribution cable is used, the block terminal is attached to a pole in the subscriber's front yard at the edge of a road. A pedestal contains the block terminal when distribution cable is buried.

4) *Distribution Cable*

Distribution cable runs between the block terminals and an SAI located in the serving area. Limitations on the capacity of an SAI and/or the distribution design used in a particular serving area may lead to multiple SAIs. Distribution structure components may consist of poles, trenches and conduit.[15]

5) *Conduit and Feeder Facilities*

Feeder facilities constitute the transmission system between the SAI and the wire center. These facilities may consist of either pairs of copper wire or a DLC system that uses optical fiber cables as the transmission medium. In a DLC system, the analog signals for multiple individual lines are converted to a digital format and multiplexed into a composite digital bit stream for transmission over the feeder facilities

Feeder structure components include poles, trenches and conduit. Manholes for copper feeder or pullboxes for fiber feeder are also normally installed in conjunction with underground feeder cable. Manhole spacing is a function of population density and the type of feeder cable used. Pullboxes that are installed for underground fiber cable are normally farther apart than manholes used with

[15] Because underground distribution exists only in the highest density zones where runs are relatively short, and because in such zones underground structure is commonly shared with feeder, distribution facilities typically do not include manholes.

copper cables, because the lightness and flexibility of fiber cable permits it to be pulled over longer distances than copper cable.

Several utilities, e.g., electric utilities, LECs, IXCs and cable television ("CATV") operators, typically share structure because it is economical to do so. Manholes may be shared with low-voltage facilities. The amount of sharing of structure and manholes may differ in different density zones and between feeder and distribution cables.

3.1.2. Switching and Interoffice Network Description

This section provides a general description of the network components comprising the wire center and interoffice facilities. Figures 2 and 3 illustrate the relationships among the components described below.

3.1.2.1. Wire Centers

The wire center is a location from which feeder routes extend towards customer premises and from which interoffice circuits or "trunks" emanate toward other wire centers. A wire center normally contains at least one end office ("EO") switch and may also contain a tandem switch, an STP, an operator tandem, or some combination of these facilities. Wire center physical facilities include a building, power and air conditioning systems, rooms housing different switches, transmission equipment, distributing frames and entrance vaults for interoffice and loop feeder cables.

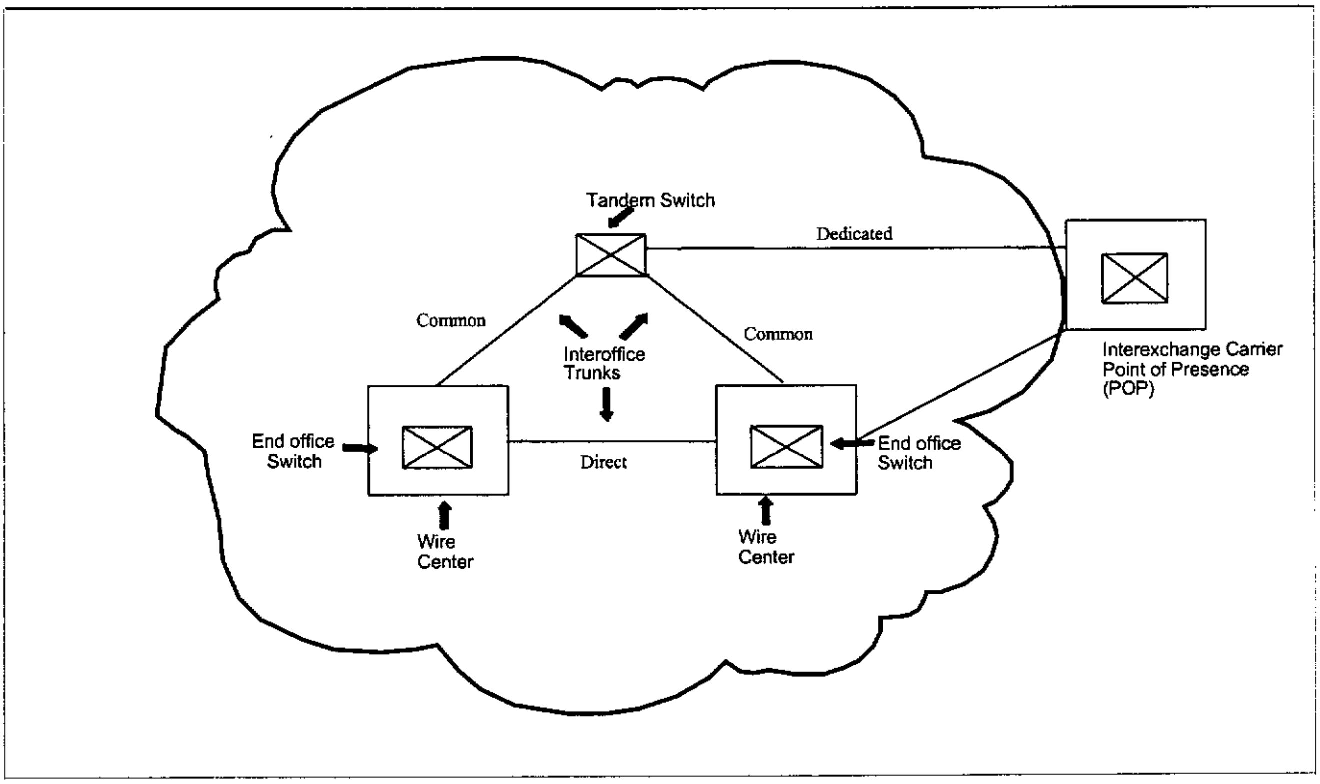


Figure 2 Interoffice Network

3.1.2.2. End Office Switches

The end office switch provides dial tone to the switched access lines it serves. It also provides on-demand connections to other end offices via direct trunks, to tandem switches via common trunks, to interexchange carrier ("IXC") points of presence ("POPs") via dedicated trunks, and to operator tandems via operator trunks.

3.1.2.3. Tandem Switches

Tandem switches interconnect end office switches via common trunks, and may also provide connections to IXC POPs via dedicated trunks. Common trunks also provide alternatives to direct routes for traffic between end offices. Tandem switching functions often are performed by switches that also perform end office functions. Tandems normally are located in wire centers that also house end office switches

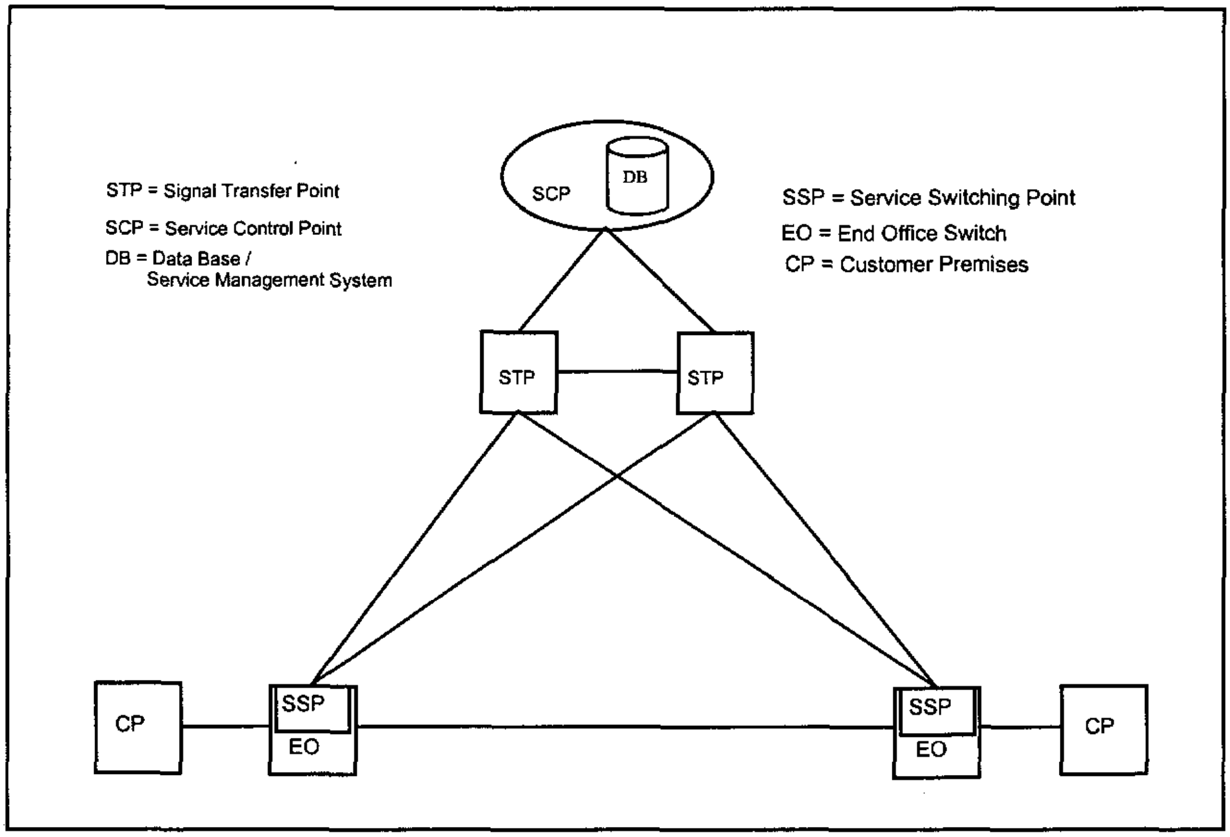


Figure 3 Interoffice Signaling Network Components

3.1.2.4. Interoffice Transmission Facilities

Interoffice transmission facilities carry the trunks that connect end offices to each other and to tandem switches. The signaling links in a Signaling System 7 ("SS7") signaling network are also normally carried over these interoffice facilities.

Interoffice transmission facilities are predominantly optical fiber systems that carry signals in SONET format. Both economic and service quality considerations increasingly prescribe the use of a fiber optic ring configuration to link switches, except for switches that serve few lines or that are too remote from other switches, where ring costs might be prohibitive. In this case, the small switches are typically connected to a nearby wire center housing another end office switch that is on a ring, or the tandem on which the small switch homes, via point-to-point links that are increasingly provided on a route-diverse (that is, redundant) basis for the sake of increasing reliability. Use of rings and redundant point-to-point links in this fashion provides an extremely secure path between any two switches, and the potential for substantial cost savings relative to the ubiquitous deployment of traditional point-to-point facilities interconnecting all switches.

3.1.2.5. Signal Transfer Points

STPs route signaling messages between switching and control entities in a SS7 network. Signaling links connect STPs and Service Switching Points ("SSPs"). STPs are equipped in mated pairs, with at least one pair in each Local Access Transport Area ("LATA").

3.1.2.6. Service Switching Points and Signaling Links

SSPs are SS7-compatible end office or tandem switches. They communicate with each other and with Service Control Points ("SCPs") through signaling links, which are 56 kbps dedicated circuits connecting SSPs with the mated STP pair serving the LATA.

3.1.2.7. Service Control Points

SCPs are databases residing in an SS7 network that contain various types of information, such as IXC identification or routing instructions for 800 numbers in regional 800 databases, or customer line information in Line Information Databases ("LIDB").

4. HM 5.0a Model Organization, Structure and Logic

4.1. Overview of HM 5.0a Organization

Figure 4 shows the relationships among the various modules contained within HM 5.0a. An overview of each component of the Model follows.

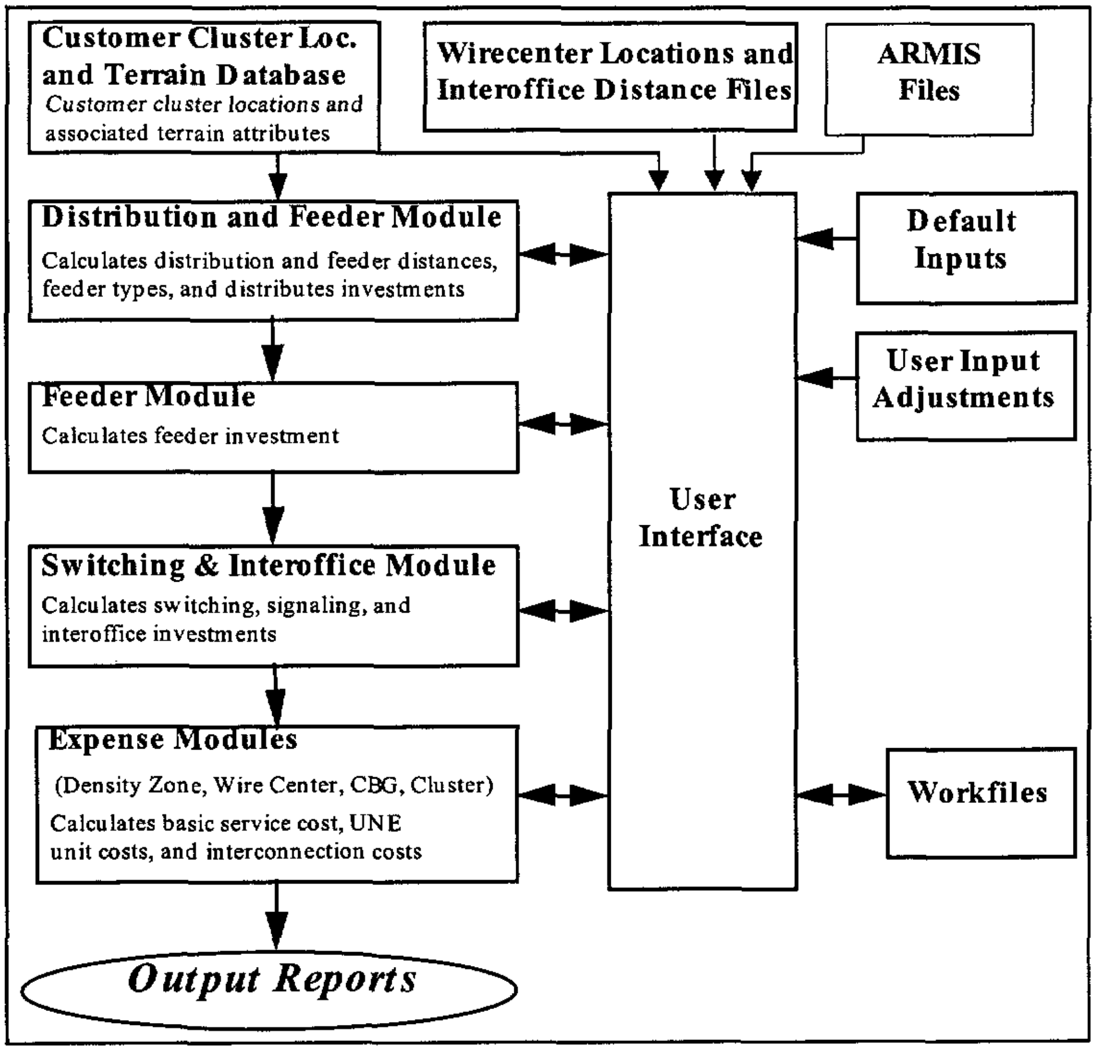


Figure 4 HM 5.0a Organization Flow Chart

4.2. Input Data

Inputs to HM 5.0a include detailed data describing the following items.

- Demographic, geographic and geological characteristics of the areas served by each telephone company. These data records are specific to individual "clusters" of customer locations (i.e., "main" clusters and their subtending "outlier" clusters). They include information on the number of customer locations and lines receiving telephone service in a cluster, the geometric configuration of the cluster's boundaries, the wire center that serves the cluster and the cluster's location relative to this wire center and to other nearby clusters, and the type of terrain that characterizes the cluster. The character and development of these data are described in greater detail in Section 5.

- Wire center locations, and interoffice distances between end offices, tandems, and STPs used to determine the route miles required for interoffice transmission and signaling facilities. These data are largely developed from Bellcore's LERG and NECA Tariff 4.

- 1996 ARMIS data reported by the Tier 1 LECs. These data provide information about current demand levels that the LEC must serve, and relationships between the LEC's embedded expenses and investments.

- Numerous user-adjustable inputs that allow users to set carrier- or locale-specific parameters, and perform sensitivity analyses. These inputs have preset default values based on the engineering experience and judgment of HAI personnel, as well as the judgements of independent subject matter expert consultants to HAI.

4.3. Workfiles

A run "workfile" is created from the input data files when a particular state/company (i.e., study area) combination is run for the first time. As a complete run of the HM 5.0a progresses, intermediate outputs from the HM 5.0a's constituent modules are stored in the run's workfile. Once the run is complete, its workfile may be examined. A great deal of information above and beyond that presented in the Expense Module spreadsheets (that contain the principal final results of the model's analysis) may be obtained from the run's workfile. One example is average loop length, by cluster and by wire center. Additionally, the user may perform separate analyses on the LEC study area by directing the model to create a new workfile for a subsequent run.

4.4. User Interface

The HAI Model includes a user interface program that facilitates model operation, including extraction of data from the input files, providing dialog boxes for users to manipulate model inputs, executing the Excel workbooks that constitute the model, saving intermediate results, and managing the flow of intermediate results between different modules.

The user interface program also performs certain simple aggregation and summarization calculations in Visual Basic for Applications ("VBA"). This shortens greatly execution times and allows users examining the model's Excel workbooks to focus on the model's fundamental engineering logic.[16]

4.5. Distribution Module

The Distribution Module addresses the portion of the network extending from SAIs to the customers' premises. The module determines the lengths and sizes of distribution cable, the associated structures (poles, trenching, and conduit), the number of terminals, splices, drops, and NIDs required to provide service to the specified numbers and types of customers, and the number and type of SAIs and DLC terminals required. The module also calculates certain distances required by the Feeder Module, and, according to those calculations, determines whether to serve a given distribution area using feeder transmission facilities consisting of copper wire pairs, or using DLC running over optical fiber cable. The model selects fiber feeder if any of following five criteria are met:

a) the feeder distance exceeds a user-adjustable crossover distance (set to a default value of 9,000 feet) that limits maximum distance of any copper feeder run;

b) the total copper loop length, including feeder and distribution cable, for customer locations within a main cluster, exceeds a user-adjustable maximum analog copper distance whose default value is 18,000 feet;[17]

c) the main cluster has at least one outlier cluster subtending it;

d) an analysis of the life-cycle costs of fiber vs. copper feeder shows that fiber feeder is the more economical choice, or

e) the "wireless" investment cap is invoked.

These criteria are described in greater detail in Section 6.3.5. If, based on these criteria, copper feeder is chosen, it extends out to an SAI located at the centroid of the main cluster. Copper distribution cable then extends from the SAI to all customer premises in the cluster. If fiber feeder is chosen, it extends from the wire center out to the centroid of the main cluster. From this point, one of two configurations is used to serve the customer locations within the main cluster. If the distance to the farthest customer location in the

[16] Model versions prior to HM 3.0 used Microsoft Excel's Pivot Table feature to summarize various results at the wire center and density zone levels. Although this feature was quite flexible, applying pivot tables to the very large arrays of data required by the model led to very slow execution times.

[17] The analog copper distance refers to the distance over which signals are transmitted in analog voiceband form on copper cable.

main cluster is less than the user-adjustable maximum analog copper distance, a single DLC RT is located at the cluster centroid, and copper distribution cable extends from this DLC RT to all customer premises in the main cluster. If the distance to the farthest location in the main cluster exceeds the maximum analog copper distance, then fiber connecting cable extends vertically and/or horizontally from the centroid to two or more DLC RTs, each of which serves a portion of the main cluster and is located to ensure the longest remaining distance is less than the maximum analog copper distance. From these multiple DLC RTs, copper distribution cables extend to the customer premises in the portion of the main cluster they are responsible for serving.[18]

The HM 5.0a Distribution Module serves outlier clusters that subtend main clusters with analog copper cable if their distance from the DLC RT in the main cluster does not exceed the user-adjustable maximum analog copper distance parameter, and if this outlier cluster has no other outlier clusters either subtending it, or lying between it and the main cluster.[19] If the distance to the farthest subscriber within an outlier cluster would exceed this threshold, the Distribution Module serves the outlier cluster with digital loop carrier equipment using copper-based T1 digital transmission.[20] Once the outlier cluster has been reached, analog copper distribution cables are used to serve the customers located in the outlier cluster.

After the module has determined the quantities of all distribution elements, it calculates the investment associated with these elements, including distribution and drop cable, structures, NIDs, terminals and splices, SAIs, and DLC terminals, using as inputs the user-adjustable unit prices of each element. The numbers and types of elements engineered can be examined in the intermediate outputs of the Distribution Module as recorded in the workfile.

4.6. Feeder Module

The Feeder Module configures the portion of the network that extends from the wire center to the SAIs. Based on information it receives from the Distribution Module, it determines the size and type of cables required to reach the SAIs located in each serving area, along with supporting structures (poles, trenching, conduit, manholes, and fiber optics pullboxes). The Feeder Module then calculates the investment associated with these elements, using as inputs the unit prices of each such element. It passes these investments to the Expense Modules. The numbers and types of network elements required can be examined in the intermediate outputs of the Feeder Module as recorded in the workfile.

[18] If, however, the model invokes a wireless local distribution system, fiber feeder extends to the radio base station located in main cluster, and final distribution is made to customers in outlier clusters over-the-air.

[19] Such an outlier cluster is termed a "first order" outlier.

[20] This technology does not require the use of loading coils or coarse-gauge cable, and it also permits basic rate ISDN and other advanced narrowband services to be provided to all subscriber locations in the model.

4.7. Switching and Interoffice Module

The Switching and Interoffice Module computes investments for end office switching, tandem switching, signaling, and interoffice transmission facilities. In HM 5.0a the user can designate specific wire center locations that house host, remote, and stand-alone switches, respectively, as well as specify inputs for the per-line investments associated with each type of switch. HM 5.0a will then calculate switching investments taking into account the switch arrangements that the user designates.

The switching module determines the required line, traffic, and call processing capacity of switches based on line totals by customer type across all serving areas belonging to the wire center, and based on ARMIS-derived traffic and calling volume inputs. It also determines the required capacity and distances of interoffice transmission facilities, using the traffic data and the interoffice distances that are input to the Module. In doing so, it uses wire center locations and interoffice distances to determine an efficient mix of interoffice SONET fiber rings and redundant point-to-point fiber links. Rings are separately provided for linking host switches to their subtending remotes, and for linking host switches to each other, to stand-alone switches and to the tandem switches on which they home.[21] The numbers and types of elements involved can be examined in the intermediate outputs of the Switching and Interoffice Module as recorded in the workfile.

4.8. Expense Modules

There are four different versions of the HM 5.0a's Expense Module – one for each of the four levels of granularity at which the user can elect to have cost results displayed: by line density range (which also displays total study area costs), by wire center, by CBG, or by customer cluster.[22] Each version calculates the monthly costs for unbundled network elements, universal service, and carrier access and network interconnection. These costs include both the capital carrying costs associated with the investments, and the costs of operating the network. Capital carrying costs include depreciation, return on the debt and equity investment required to build the network, and income taxes on equity return. Network-related expenses include maintenance and network operations. Non-network related expenses include customer operations expenses, general support expenses, other taxes and variable overhead expenses.

Several sources provide information to the Expense Modules. The Distribution, Feeder, and Switching and Interoffice Modules provide network investments by specific plant

[21] At the user's option, small standalone wire centers serving fewer lines than a user-adjustable threshold (default value: one line) may be excluded from being placed on rings, and instead linked directly to its serving tandem using point-to-point links. In this case, the model attempts to physically route these point-to-point circuits through a nearby large wire center, and then over the fiber rings it has otherwise engineered for interconnecting larger offices and tandems.

[22] Although the HM 5.0a engineers no plant based on CBG granularity, the results of its engineering to individual clusters may be rolled up to display cost results at the CBG level. The association between clusters and CBGs is made based on the relative number of lines each cluster contributes to a CBG's total.

category. ARMIS and other sources are used to derive information on network operating and maintenance expense relationships.

The Expense Modules produce reports (either by density zone, wire center, CBG, or cluster) showing the key outputs of the model, including the costs of providing universal service, unbundled network elements, interconnection and IXC access. Further detail about network investments and costs is available from the workfile associated with a model run.

5. Input Data

To accommodate HM 5.0a's evolution to modeling local telephone networks based on actual clusters of customer locations, the input data used in HM 5.0a are much more granular than the CBG input data used in HM 4.0 and earlier. Flowcharts describing the development processes used to prepare these input data for HM 5.0a are attached as Appendix C to this document.

5.1. Line Type Counts by Study Area

Counts of access lines by type (i.e., residence, single line business, multiline business, public telephone and special access lines) for each distinct NECA Study Area for calendar year 1996 are developed from several data sources. These include:

ARMIS 43-08: 1996 data, released 10/01/97;
ARMIS 43-01: 1996 data, released 10/01/97;
NECA USF Loops filing: 1996 data;
USTA report: 1995 data;
RUS report: 1995 data;
USF Data Request: 1993 data;
ARMIS-based line factors.

The rules by which the best of these data are selected are as follows.

a) When NECA Study Area name matches exactly ARMIS Company name, populate line types directly from ARMIS data for business lines (43-08), single line business lines (43-01), residence lines (43-08), special access lines (43-08), and public lines (43-08) data.

b) For remaining ARMIS Companies, determine counts of line types for NECA Study Area name by applying ARMIS line type distributions to total reported NECA USF Loops.

c) For non-ARMIS Companies, match NECA Study Area name to best available data source (i.e., USTA, RUS or USF Data Request) for residence and business line splits.

d) When no company-specific line type data exist, apply average ARMIS line type distributions to reported NECA USF Loops for NECA Study Area name.

5.2. Wire Center List

The source of the wire center information used in PNR's National Access Line Model is Bellcore's LERG database, dated August 1, 1997.[23] The portions of these LERG data that are used in the HAI model are an extract of key data from the LERG called the Special LERG Extract Data ("SLED") – which has been licensed from Bellcore by the HAI model developers.

Certain switching entities (wire centers) in the SLED with Common Language Location Identifier ("CLLI™") codes not marked as end offices, hosts or remotes are then removed from this wire center database. In addition, switching entities that are inactive, or owned by wireless, long distance or competitive access providers are removed as well.

In a few instances, the SLED assigns wire centers to multiple local carriers. This may result from switch collocation. Because the HAI Model requires wire center entries to be unique, such wire centers are assigned to the local carrier having the greatest number of active NPA-NXX codes. If active NPA-NXX codes are equal among companies, assignment is to the carrier having the greatest number of residential lines.

Multiple occurrences of 8-character CLLIs may also occur in the SLED due to placements of several switches at a single wire center location. Because the HAI Model itself engineers multiple switches in a wire center if demand requires it, duplicate occurrences of 8-character CLLIs are removed from the model's wire center list.

5.3. Customer Counts by Census Block and Wire Center

Customer locations must be associated both with CBs, as well as their serving wire center ("WC"). The PNR National Access Line Model, Version 2.0 ("NALM") performs both of these tasks. The PNR NALM uses PNR survey information, Bellcore's LERG, BLR wire center boundaries, Dun & Bradstreet's ("D&B") business database, Metromail's household database, Claritas' 1996 demographic database, and U.S. Census estimates to calculate both the number of residential and business locations and access lines in each CB, and in each wire center in the United States. This summary describes the methodology, data and assumptions used in developing these location and line estimates in the NALM.

5.3.1. Residence Counts

Residential customer location counts are developed by applying the following process.

a) The Metromail household database (described in section 5.4.1, below) is geocoded to the "point" level.[24] In addition to recording the precise six-decimal place latitude and longitude of this household, the CB associated with its location

[23] These LERG data are augmented by data from NECA Tariff 4.

[24] As described in more detail in Section 5.4.3, below, geocoding to the "point" level means that the geocoding software has both found the housing unit's address in its location files and determined a latitude and longitude for the location down to six decimal places of a degree.

is recorded as well. Duplicate household information is identified and eliminated. If two records appear with an identical latitude, longitude and phone number, one of the two records is eliminated.

b) Implied residential household counts are evaluated by comparing Metromail counts to Claritas' 1996 CBG-level projections of households with telephones. When Metromail households exceed Claritas households, Metromail households are used. When Claritas households exceed Metromail households, Claritas households are used, and the total differences are distributed to the constituent CBs in proportion to 1990 U.S. Census household distributions.

c) Access line counts are determined from household counts using probabilities, that is, how likely is it that a household will have a first or second telephone line installed? First line probabilities are provided by Claritas based on demographic age and income profiles by CBG. Second line probabilities are based on a logistic regression using similar demographic information and developed by PNR using its ReQuest™ III residential survey. Multiplicative probability factors are applied to the household counts defined above to derive residential line counts.

d) The above derived residential line counts by CB are then normalized to sum to Study Area wide data on total residential line counts developed in Section 5.1, above.[25]

e) This lines normalization factor is applied to the residential customer location counts in each CB, as well.

The implications of the forgoing process are as follows. Because the primary source of residential location counts is Metromail – which includes all residences that receive direct mail regardless of whether they have telephones or not – the universe of "populated" CBs that the data process captures may include CBs where telephone service is not currently offered or accepted. Thus, the "breadth" of the telephone network that these data will instruct the HM 5.0a to construct is likely greater than the embedded networks of the ILECs. However, because the counts of lines and locations in each CB are normalized to sum to given Study Area wide totals, the "depth" of the constructed network will be consistent with current levels of actual telephone demand.[26]

5.3.2. Business Counts

[25] If comprehensive LEC data on residential line counts by individual wire center are available, normalization can be done at the individual wire center level.

[26] In addition, note that these primary source residential location counts derive from precisely geocoded 1997 Metromail data. Thus, these data provide 1997 information on location counts at the CB level. As a result, the model's reliance on noncurrent or nonCB-specific location data (e.g., Claritas 1996 CBG-level projections, 1990 U.S. Census CB counts, or 1995 U.S. Census Update county-level projections) is limited to those locations that show up in such counts that are in excess of the Metromail counts.

Business location counts are developed by applying the following process.

a) The D&B national business database (described in Section 5.4.2, below) is geocoded to the "point" level. In addition to recording the precise six-decimal place latitude and longitude of this business, the CB associated with its location is recorded as well.

b) From the D&B national database, the total number of business lines, as well the probabilities of these lines being single line business lines and multiline business lines such as Centrex and PBX lines are developed. This model is based on an 800,000 firm sample.

c) Because the D&B national business database contains records for only about 11 million out an estimated total of 12 million U.S. businesses, and because the businesses that it misses are almost certainly small businesses, an additional 1 million nonD&B business locations are added to CB counts in proportion to D&B businesses located in the CB. The lines associated with these added business locations are projected by PNR based on an assumption that they employ, on average, between 1 and 4 employees, each.[27]

d) The above derived business line counts by CB are then normalized to sum to Study Area wide data on total business line counts developed in section 2.1, above.[28]

5.3.3. Location and Line Counts by Wire Center

HM 5.0a uses WC boundaries provided by BLR as its primary source to define wire center service areas. These boundaries conform to CB boundaries, and customer locations contained within all of the CBs associated with a WC are then assumed to be served by that WC. Telephone number information (NPA-NXX) continues to be used for backup and data scrubbing purposes when anomalies arise in the BLR geographical assignment process – as can occur if one wire center's boundaries fall completely within another wire center's boundaries.

5.4. Customer Location

The customer location approach used in HM 5.0a is fundamentally different from that of HM 4.0 – or any other approach that uses arbitrary geographic delineators such as CBs, CBGs or latitude and longitude grid cells. Because HM 5.0a's approach identifies the actual locations (accurate to within 50 feet) of most telephone customers, it produces the

[27] To the extent that the D&B database contains firms that are not locatable to the CB-level, these firms are assumed to be distributed across CBs in proportion to located firms within D&B. Their line counts are calculated based on the company characteristics (e.g., employees, SIC) that they report to D&B.

[28] Again, if comprehensive LEC data on business line counts by individual wire center are available, normalization can be done at the individual wire center level.

most sophisticated demographic data set of its type. The process first develops a database of about 109 million customer address records. These addresses are then geocoded (assigned latitude and longitude coordinates). These locations are then divided among wire center serving areas based on geocoded customer location and the BLR wire center boundaries.

5.4.1. Residence Location Data

Data for residence locations are provided by Metromail, Inc. The Metromail National Consumer Database© ("NCDB") is a large, nationally compiled file of U.S. household-level consumer information that includes both deliverable postal addresses (and telephone numbers, when available). The file consists of close to 100 million records – which constitute over 90% of all residential housing locations that the U.S. Bureau of the Census reported for 1995.[29]

To ensure that the data captured are the most current available, this file is updated 65 times per year, and undergoes numerous "hygiene" measures to ensure its continued high quality for direct marketing purposes. Such purposes require the data to reflect postal address standardization practices, incorporate National Change of Address ("NCOA") processing, and permit postal geocoding to street address, ZIP+4 or Carrier Route levels.

The file is compiled primarily from telephone white pages directory data, but also utilizes many other primary sources of information, such as household mover records, voter registration data, motor vehicle registration information, mail-order respondent records, realty data, and home sales and mortgage transaction information, to build a large repository of verified household-level data.

5.4.2. Business Location Data

Dun & Bradstreet collects information on more than 11 million business establishments nationwide. Information is gathered from numerous sources such as business principals, public records, industry trade tapes, associations, directories, government records, news sources, trade organizations, and financial institutions. This information is validated each night. Additionally, D&B conducts millions of annual management interviews to help improve the timeliness and accuracy of its information.

The information is organized by D-U-N-S number, a nine digit identification sequence which allows for the placement of companies within larger business entities according to corporate structures and financial relationships. A D&B family tree may be used to relate separate operating companies to each other, and to their ultimate parent company. D&B also provides "demographic" information on each of the firms in its database. Such information includes counts of employees and the SIC code of the establishment.

[29] This number is also very close to the 101 million households that the FCC finds in 1996, and exceeds the 95 million households that the FCC reports had telephones in that year. See, *Trends in Telephone Service*, FCC Common Carrier Bureau, Industry Analysis Division, March 1997, Table 1.

5.4.3. Geocoding

Geocoding is used in order to most accurately assign known customer locations to actual, physical locations. Geocoding is also known as location coding. It involves the assignment of latitude and longitude coordinates to actual street addresses. Geocoding software is sophisticated enough to provide information regarding the source and precision of the lat/long coordinates selected. This precision indicator allows PNR to select only those addresses that have been geocoded to a highly precise point location. Geocoding also allows customer location points to be assigned less granularly to the CBG level, or higher. Almost uniformly, geographical address locations are derived from enhanced versions of the USGS' TIGER database.

To perform its geocoding, PNR uses a program by Qualitative Marketing Software called Centrus™ Desktop. The enhanced data behind Centrus is provided by GDT. Premium GDT data are updated bi-monthly to ensure accuracy. These data integrate new information from US Postal Service ("USPS") databases and private sources so that new streets and additions and changes to ZIP codes, street names, and address ranges are included as soon as possible.

Centrus™ Desktop allows geocoding on two levels. The first is a match to the actual address -- which is the only type of geocoding used in HM 5.0a customer location. The second is a match to a ZIP code (ZIP, ZIP+4, ZIP+2) level. Because of the lesser accuracy in the second method, these geocodes are not used in PNR's process of assigning customer locations.

Within the geocode process, there are a number of options available to the user. Each of these options determine the quality of the matches allowed in the end-use geocode. For purposes of customer location, addresses are always matched to the "Close" setting. "Close" allows for minor misspellings in addition to incorrect or missing directionals (North, East, etc.) or street types (street, road, etc.). Although ZIP-based geocodes are generally accurate enough for most applications, they are not considered good enough for actual customer locations and are not used to develop and locate customer clusters in HM 5.0a.

Data hierarchy in address geocoding starts with the State. The hierarchy continues with City, Street Name, Street Block, and finally, House Range. Typically, a Street Block is the same as an actual physical block but it can also represent a partial block as well. The House Range displays address information from the USPS. Additionally, where there are gaps in the actual address range, the House range will account for these gaps.

Initially, the address coding module in Centrus™ Desktop compares the street addresses from the input file to the records contained in the USPS ZIP+4 directory and the enhanced street network files. If the address is located in the USPS files, the address is standardized and a ZIP+4 is also returned. If this address is also found in the street network files, Centrus™ Desktop determines a latitude and longitude for the location.

Optionally, if the address is not found in the street network files, location information may be applied from the ZIP level.[30]

Location codes generated by Centrus™ Desktop indicate the accuracy of the geocode. For purposes of customer location clustering in the HM 5.0a only those geocodes assigned at the 6-decimal place point location made directly to the street segment are used.[31]

While the software and data used allow for a much more comprehensive output of data elements, for use in HM 5.0a customer location, the following addressing elements are extracted:

Address
City
State
ZIP
ZIP+4
Latitude
Longitude
Census Block
Match Code
Location Code

5.4.4. Gross-up

The above-derived precisely geocoded locations are then counted by CB. These geocoded location counts by CB are then compared to target total line counts for that CB derived by the PNR NALM (as described in section 2.3, above). If the geocoded location counts are less than the target count, the residual number of customer location points is then computed, and geographical locations for these points are generated. This process is performed by PNR using TIGER file CB boundaries. Each of the additional number of customer location points that a CB requires to total to its target count is generated and assigned a geocode so as to place these "surrogate" points uniformly along the CB's boundary. While these boundary-assumed locations for the gross-up or surrogate points are plausible – because most CBs are bounded by roads – this is also a conservative placement of the gross-up points because it assumes they are maximally separated from one another.

[30] Note that ZIP+4 codes may be very precise. In general, they are specific to the face of single city block. While it may turn out that accuracy to the street block face is quite sufficient for accurate cost modeling of local telephone networks, in the interest of conservatism, these type of geocodes are not presently used in HM 5.0a data.

[31] Furthermore, placement of the address along the street segment is quite precise. The Centrus geocoding software and reference data also make use of USPS determinations of whether the segment contains a continuous or discontinuous range of address numbers. Thus, if the addresses on a block face run from 200 to 250 and 274 to 298 (with the range between 252 and 272 missing), an address of 250 will be geocoded, it will not simple be geocoded as at midblock.

As a result of this gross up process, the customer location file now contains records for each of the U.S.'s more than 100 million customer locations with a geocode (either calculated precisely or through the gross up process) associated with it.

5.5. Customer Location Clustering

5.5.1. General Criteria

The input development process next identifies all customer locations within a wire center's boundaries that are close enough together to be efficiently engineered as a single telephone plant serving area. This process is called clustering. While there are many available off-the-shelf clustering algorithms, efficient determination clusters of customer locations that are consistent with telephone engineering practices requires that certain engineering restrictions be imposed during the clustering process, and not afterward. Customer locations must meet the following criteria to be considered members of a particular cluster.

- No point in a cluster may be more than 18,000 feet distant (based on right angle routing) from the cluster's centroid.
- No cluster of nondegenerate area may exceed 1800 lines in size.[32]
- No point in a cluster may be farther than two miles from its nearest neighbor in the cluster.

Note that other than for the wire center boundary restriction, these criteria do not include any arbitrary geographic restrictions, such as clusters being restricted to lie within a single CBG, CB, or latitude/longitude grid cell. Thus, clusters developed pursuant to this process are likely to be the most closely representative of actual telephone distribution areas as determined by outside plant engineers. Note, however, that the last of these restrictions – no point in a cluster may be farther than 2 miles from its nearest neighbor – is not an absolute engineering limitation. It is used to ensure that customer locations that are separated by the given distance are not required to be clustered together. It is possible that efficient engineering of telephone distribution plant would suggest that a different, possibly larger value be chosen.[33]

[32] This restriction is based on the maximum unconcentrated lines capacity of an OC-3 fiber optic transmission system used to feed a DLC remote terminal (adjusted for a 90% rate of fill). This is consistent with current HM 5.0a practice that provides a separate channel for each line served by a DLC system. Because it is reasonable engineering practice to concentrate traffic on these large fiber optic DLC systems, future versions of the model may assume that traffic is concentrated on the fiber optic systems feeding DLC remote terminals. When such revisions to the HM become available, the customer location data will be reclustered with the appropriately enlarged maximum limit on cluster size. In all events, if single customer locations, such as a large office or apartment building, by themselves exceed 1800 lines, such clusters are not split. Rather, multiple DLC RTs/SAIs will be placed to serve such "oversized" clusters.

[33] Testing of different parameterizations for the maximum distance to a cluster point's nearest neighbor suggests that 2 miles is a reasonable national value.

5.5.2. Clustering Algorithm

The process used by the PNR clustering algorithm is as follows.

a) First, to provide a uniform geographic unit for clustering operations, all customer location geocodes associated with a particular wire center are "rasterized" into 150 foot cells that overlay the geographic rectangle covering the wire center's service area.[34]

b) The algorithm then inspects all neighboring cells (e.g., all cells that have their center within a 150 foot radius of the center of the initial cell) to see if any are populated with customer locations. If one of these neighboring raster cells is populated, the algorithm first checks to see whether clustering that cell (and its immediately surrounding neighbors) with the initial cell would violate any of the clustering restrictions (i.e., create a cluster that has points more than 18,000 feet from the cluster centroid, create a cluster of more than 1800 lines, or include a point more than 2 miles from its nearest neighbor).[35] If none of these conditions would be violated, the adjoining cell plus its immediate neighbors will be added to the initial cell's cluster

c) This process continues on to the next unclustered populated cell and performs the analysis described in step (b), above. This repeats until no more unclustered cells exist.

d) The process then restarts at step (b), but uses a 300 foot search for populated neighboring cells.

e) This continual enlargement of the search for neighboring populated cells continues on until no more cells can be added to the cluster without violating one or more of the engineering requirements. At this point, the algorithm is complete.

Clusters that contain five or more customer locations are classified as "main" clusters. Clusters that contain from one to four customer locations are called "outlier" clusters. Outlier clusters may be linked to their "home" main cluster via "chains" that string together other outlier clusters that home on the same main cluster. The process of determining the routing of the chain is as follows.

---

[34] Rasterization at this level is a process whereby all customer locations that are within the 150 foot square cell are counted as being placed at the center of the cell. Rasterization to 150 foot square cells ensures that the mathematical clustering process can proceed efficiently, and that nonmeaningful distinctions in customer location are ignored. Increasing the raster to 300 or 500 feet would speed up the clustering process dramatically, and still provide a very acceptable level of precision in cluster formation. Note, too, that the PNR calculation of 150 foot raster cells is precise – based on the local latitude of the cell. Thus, raster cells do not enlarge as one moves south toward the equator.

[35] Because the rasterization into 150 foot square cells may cause customer locations that actually are in the farthest corner of a cell to be considered at the cell's center, the clustering algorithm will actually check to ensure that no cells added to a cluster exceed 17,700 (= 18,000 - 2*150) feet from the cluster's centroid.

a) The clustering algorithm first calculates the distance between each cluster and every other cluster in the wire center service area.

b) The algorithm then determines the shortest distance between any two clusters, and associates these two clusters together.

c) Next, the algorithm determines the next shortest distance between any two clusters or cluster chains, and associates these together.

d) This process continues until all outliers are chained to a main cluster.

In addition to creating chains to associate outlier clusters to a main cluster; the clustering algorithm calculates the centroid location, "aspect ratio" and area of the rectangle that overlays the convex hull of each cluster; and that has the same centroid location, aspect ratio and area as this convex hull.[36]

When this process is completed, the main cluster and its subtending outliers are considered to constitute one serving area.

The description the HM 5.0a Distribution Module in Section 6.3 provides greater detail on the model's engineering of outside plant to serve main and outlier clusters.

5.6. *PointCode Translation Processes*

PointCode is a Microsoft Access '97 database process that translates between coordinate systems, computes distances and assigns additional characteristics to cluster records. Among the activities executed by PointCode are the following.

Convert the latitude and longitude coordinates provided by PNR for cluster centroids to V&H coordinates. Ensure that modeled distribution rectangles have an aspect ratio and area that reflects a minimum dimension equal to twice the default drop length for that lines density range. Calculate radial distances between main clusters and their serving wire center. Calculate radial distances between outlier clusters and main clusters.

Compute omega angles between main feeders and the clusters they serve and compute alpha angles between clusters and their subfeeders. Calculate rectilinear (right angle) distances between main clusters and their serving wire center, and between outlier clusters and main clusters.

On the basis of the characteristics of the covering CBG, assign terrain and lines density zone characteristics to the cluster.

[36] The convex hull of the cluster is the convex polygon that has as its boundary the outermost points of the cluster – as such, it covers all of the points in the cluster. The aspect ratio is the ratio of the North-South length to the East-West length of the convex hull of the cluster.

6. *Module Descriptions*

6.1. *Input Data Files*

6.1.1. *Demographic and Geological Parameters*

Demographic and geological parameters are obtained from a database developed by PNR and Associates of Jenkintown, PA. Section 5, above, explains in detail how these data are developed. The data file resulting from these processes is organized by state and telephone company (study area). Each customer location cluster (both main and outlier) identified in the wire center service areas of the LEC modeled appears as a separate record in this Microsoft Access 97 database. Each of these cluster records contains the following information:

- Identity of the LEC and wire center serving the cluster;
- Locational information about the cluster relative to its wire center, the predominant CBG in which it falls, its nearest neighboring cluster and associated distances;
- Area and dimensional measurements of the cluster and its lines density;
- Terrain and geological parameters;
- Number of telephone lines by type;
- Number of households and number and type of housing units;
- Number of business firms and employees;
- Information about the fraction of a wire center's total lines are represented by this cluster, the average loop distribution distance of its subtending outlier clusters and total number of outlier lines.

The complete list of data fields in the Cluster Input data table is as follows: