Highway Coding for Route Designation And Position Description

GREG SALAS, Philco-Ford Corporation, WDL Division, Palo Alto, California

A coding system compatible with the Experimental Route Guidance System (ERGS) concept has been devised for application to over 4 million highway intersections of the contiguous United States. The derivation of this coding system involved suitability for hardware implementation, acceptability to the general driving public, and capability for the ancillary functions of highway data record files generation and processing. In addition, computer programming of an optimal route determination algorithm was accomplished.

The coding system developed utilizes a dual form of partitioning: (a) geographic so that locality can be described via the code; and (b) hierarchical so that selective transmittal of information based on highway network hierarchy reduces the total information load on the highway-vehicle communication link by orders of magnitude. Further, the code is alphabetic so that the code name is short, easily read and manipulated by the general driving public, thus inspiring confidence in its use and encouraging acceptance. The basic five-letter code substitutes a two-letter state abbreviation prefix for the first character of the code as used by the public, thereby establishing a first-order level of familiarity. Where possible, the code alphabet is made vowel-heavy so that generation of pronounceable syllables is enhanced. The resultant verbalization characteristic aids the short-term memorization process and tends to reduce human error in driver manipulation of the code.

A comparative cost analysis showed that the expected increase in hardware cost is only 2.65 percent over that of the minimum cost binary coding system. For the purpose of testing prototype models of the information exchange equipment, a FORTRAN IV computer program for solution of optimal routes on a network of 1,500 or fewer intersections was written and tested. Additionally, data processing formats for computer files of highway records utilizing the developed code were devised and applied in the construction of intersection files used for the publication of a United States Intersection Directory.

•OVER the period of years in which the present vehicular roadway system has developed from a simple roadway system to the current maze of complex interconnecting, sometimes three-dimensional network of streets, roads, and highways, the performance and quality of vehicles and roadways have improved, seemingly without corresponding improvement in highway guidance. The external (to the vehicle) visual signals that served to guide oxcarts and buggies over their routes are still with us today in a state of proliferation that matches the growth of the vehicle roadway system.

Paper sponsored by Special Committee on Electronic Research in the Highway Field and Committee on Traffic Control Devices and presented at the 48th Annual Meeting.

Our present state of the art in highway signing is unsatisfactory from a number of standpoints, particularly in regard to route guidance. First, external visual signals are network-oriented and not driver-oriented. Road signs giving guidance information tersely indicate where the road goes, not where the driver is to go, forcing the driver to have preplanned his route and have some foreknowledge of the roadside signing to expect in the vicinity of network areas where he must maneuver in order to negotiate the network maze throughout his trip. Second, only a small portion of the total roadside signing is pertinent to a given driver's trip, resulting in a high degree of superfluity, but the driver cannot indiscriminately reject all road signing without the risk of rejecting that which is pertinent, forcing almost continual surveillance of signing along the route driven. This aspect of vehicle guidance serves to create anxiety and stress in drivers, probably contributing to driver fatigue and creating incipient hazardous conditions. Third, external visual signals are impaired whenever visibility is impaired, as in night driving, bad weather, and the obscuration of thick traffic or large vehicles.

Recognizing the need for a vehicle guidance system to suit the needs of the individual driver, the Bureau of Public Roads with Philco-Ford Corporation and other contractors has engaged, over the past three years, in a research project aimed at circumventing the problems associated with present vehicle guidance systems. Within the Department of Transportation, this project is referred to as the Experimental Route Guidance Systems (ERGS).

ERGS may be considered to consist of three main constituents:

1. A system of route coding logic along with application to the U.S. Highway network. This route coding involves the development of a logically consistent technique for naming intersections and roads of a highway network. For the purposes of this paper, intersections are called nodes and their emanating paths are called branches. Quantitative observations have determined that the vast majority of nodes have eight or less branches, so that a clockwise numbering of branches from 1 to 8, starting from the north as 1, would suffice as a coding scheme for node branches. Branch and node coding can provide an efficient language for designation of any point in a highway network, and hence any point of interest to a traveler. Application to data processing of highway records is also apparent for a coding logic such as developed here.
2. Development and compilation of files of optimal routes between node pairs on the coded highway network. Several algorithms exist for solving for the shortest route between two points in a network. The size of the network and the requirement for the shortest routes between all pairs of network nodes are vital considerations in the selection of the most appropriate technique. The technique chosen for handling large networks mathematically with ease in programming, small core requirements, and rapid solution, was a modification of the Shimbel "minaddition" algorithm.
3. Development of an information exchange methodology. The methodology developed takes place in three steps. First, by use of the developed coding logic, the driver informs the system of his destination. Second, the system selects from the files of optimal routes developed by the modified minaddition algorithm, the instructions to give the motorist so that he may proceed along an optimal route. Third, the instructions must be transmitted to the driver, transformed, and usably displayed. The system element that performs this function is further divided into two subsystems—the vehicle subsystem and the node subsystem.

It is the first two constituents of ERGS to which this report is addressed.

ROUTE CODING LOGIC

Viewing a segment of the U.S. Highway network schematically (Fig. 1), each node can be uniquely identified by a name or code (for this simple case a single alphabetic designator), and each branch emanating from the node, by a single numeric digit. The scheme for choosing the digit to identify the node branch is to assume that eight equiangular sectors (45 deg each) surround the node. Each of these sectors is numbered from 1 to 8 clockwise, starting with 1 as the northernmost sector. The branch is

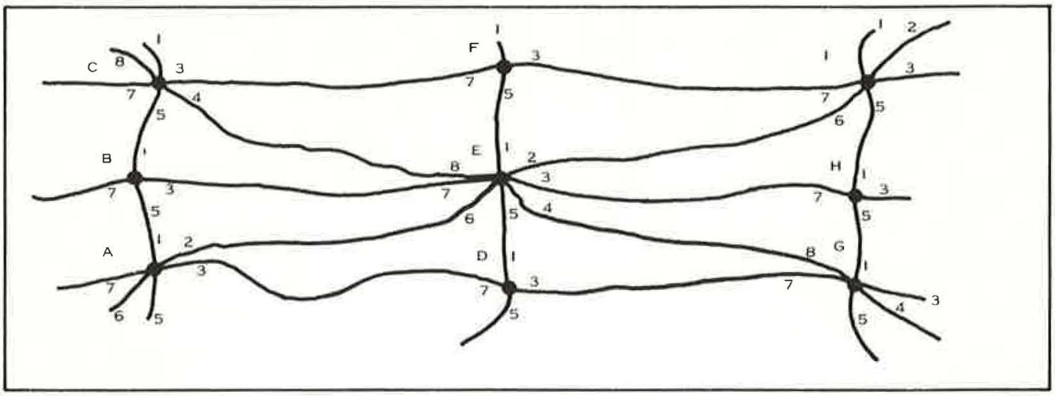

Figure 1. Schematic of segment of U.S. Highway.

identified with the number of the sector in which it lies as it first leaves the node. Complex nodes having more than eight branches can be identified as multiple nodes having separate and unique node code identifiers, each having eight or less branches.

With this basic coding logic, route following can be accomplished by designating the node code and the branch to be taken out of the node as travel across the network occurs. Thus, in Figure 1, if node I is a trip destination and node A is an origin, the route from A to I may be specified as A-3, D-3, G-1, H-1, I. Another routing may be specified as A-1, B-1, C-3, F-3, I. A third routing may be specified as A-2, E-2, I. If all branches of the schematic network are measured for distance, then it becomes possible to describe minimum distance routes from each node in the network to every other node in the network. Route measurements other than distance may be used and minimized in order to determine an "optimum" routing. Trip-time on branches may be minimized. If toll charges on some roads are encountered, then costs may be minimized. Whatever the criterion, optimal routes may be determined and stored at each instrumented node, so that each suitably equipped vehicle may indicate its destination as it approaches an instrumented node, and receive from the node appropriate information regarding the correct maneuver (turn right, turn left, continue straight, etc.) to perform in order to reach its destination optimally. Figure 2 shows the ERGS system's sequential operation.

Network Partitioning

The immense size of the U.S. Highway network, which is projected to contain approximately 4 million nodes within the next score of years, requires some form of subdivision or partitioning in order to reduce the networks to a tractable level for analysis and operations. A dual form of partitioning has been chosen for coding purposes, consisting of geographical partitioning on the one hand, and hierarchical partitioning on the other. Geographic partitioning is particularly convenient since it permits locality or place information to be imported via the code; hierarchic partitioning was found to be necessary in order to reduce the total information load on the system.

For ERGS coding, three hierarchic levels were chosen. The first level called the "primary" network level consists of the principal arterial roads in the United States similar to the Federal-Aid Primary System of highways and its intersections. The second level, called the "secondary" network level, consists of secondary roads similar to the Federal-Aid Secondary System of highways, its intersections, and the primary system. The third level, called the "minor" level, consists of all other roads and intersections connected to the other two levels.

By surveying current road maps of the U.S. Highway System and counting the intersections of the principal arterial roads, approximately 7,500 primary nodes were identified. In order to afford room for expansion of the highway system, allowance was made for 10,000 primary nodes. Each primary node is unique to a given geographic

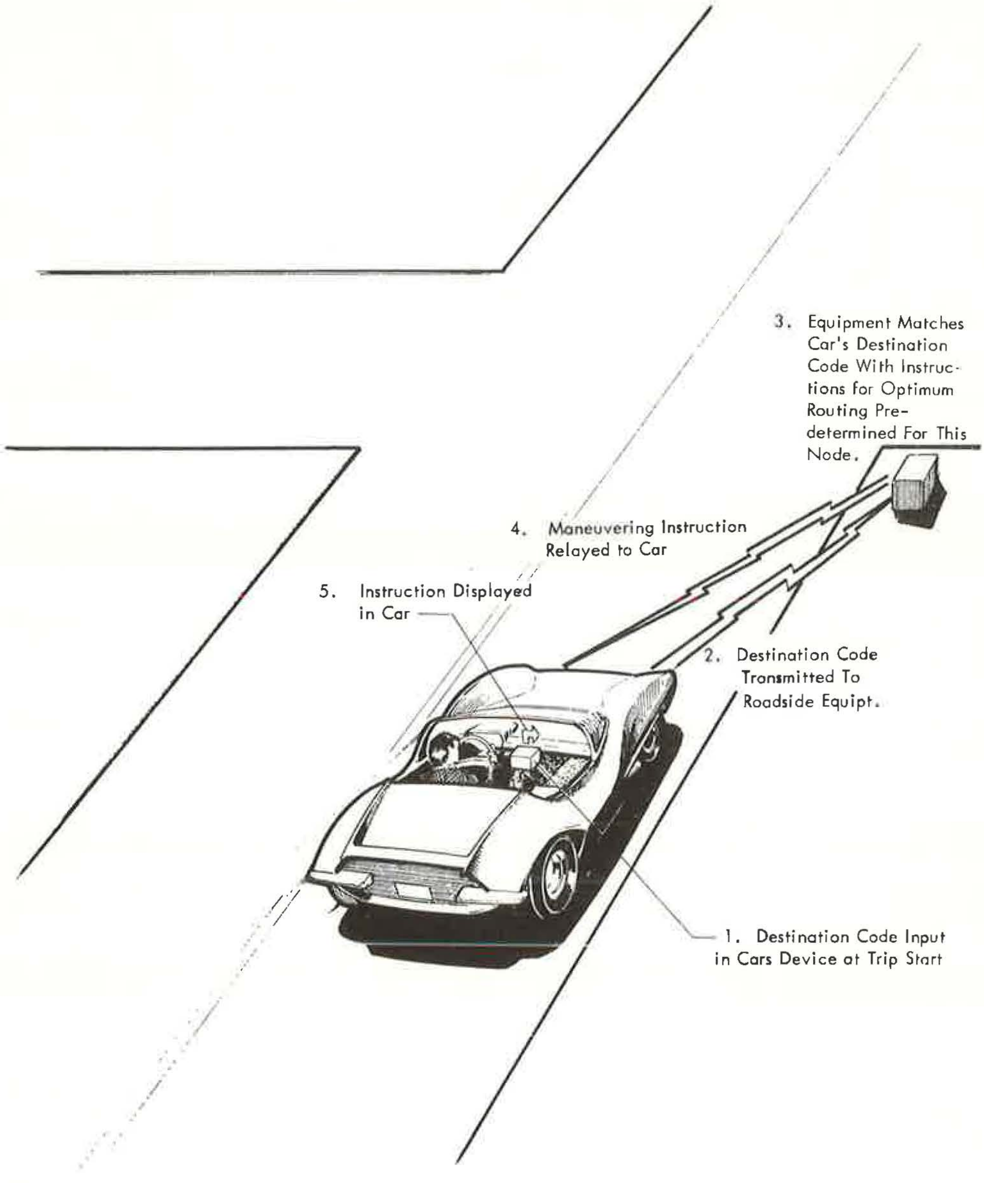

Figure 2. ERGS sequential operation.

partition called a "sector," thus defining some 10,000 geographic sectors for the United States. Within the geographic sector, allowance is made for including up to 20 secondary nodes; each secondary node being unique to a geographic partition called a "zone." Thus there is the capability of containing 200,000 secondary nodes in the U.S. Highway network (20 secondary nodes for each of 10,000 primary nodes). If, in a similar fashion, allowance is made for up to 20 minor nodes for each secondary node, or 20 minor nodes per zone, there would be the capability of containing 4,000,000 minor nodes in the U.S. Highway network (20 minor nodes for each of 200,000 secondary nodes). Three levels of network hierarchy can thus serve to identify in excess of 4,000,000 nodes.

Code Symbology

The problem next addressed in the route coding logic involves the selection of an appropriate symbology to be used in coding or naming each of the 4,000,000 nodes described by the three hierarchic levels. This aspect of the problem impinges upon two important factors in route coding: (a) equipment factors, and (b) human factors.

From the standpoint of equipment factors, the selection of a code affects the hardware items required to represent the code symbology in terms of input hardware, conversion hardware (from the code symbology to bit representation for the electromagnetic link between the vehicle and the roadside equipment), and data processing hardware, not only for the look-up process at the roadside equipment, but also for the maintenance of highway descriptive data files and the solution of optimal routes from any given highway network. Obviously, simplicity of the code will be reflected in the simplicity, reliability, and cost of the hardware representation of the code.

A simple example illustrating some of these factors is that of using the binary system for a node code, and the representation of a binary code with toggle switches. First, in order to have sufficient scope, at least 22 toggle switches would be required, since $2^{21} < 4 \times 10^6 < 2^{22}$. The binary representation of the number 4 million is

1111010000100100000000,

and 22 toggle switches would have to be set up to identify any one of 4 million nodes. To represent four million nodes decimally would require a seven-place input device, such as

4000000,

and a hardware representation of seven ten-place switches would be required.

The binary hardware implementation requires 22 "switches," each capable of representing two states, while the decimal hardware implementation requires 7 switches, each capable of representing 10 states. The binary coding system requires 22 x 2 = 44 hardware elements, while the decimal system requires 7 x 10 = 70 hardware elements. Note that the number of hardware elements is equal to the product of the number of switches, or input positions, and the number of states each position represents (the base of the number system used in the coding scheme).

Algebraically,

$$Z = B^m \tag{1}$$

where

Z = number of hardware elements required;
B = base of numbering system used in coding scheme; and
m = number of switches or positions required to represent the number in the code.

And, to satisfy the constraint that at least 4 million nodes be coded,

$$B^m \geqq 4 \times 10^6 \tag{2}$$

The question as to whether there is some combination of B and m which minimizes Z can be approached analytically by setting the derivative of Z with respect to B equal to zero and solving for B. Thus, solving for m in Eq. 2 in order to have the required scope of 4 million nodes,

$$m \ln B = \log (4 \times 10^6) \tag{2a}$$

substituting Eq. 2a in Eq. 1,

$$Z = \ln (4 \times 10^6) \frac{B}{\ln B} \tag{3}$$

Differentiating Eq. 3,

$$\frac{dZ}{dB} = \ln (4 \times 10^6) \left[\frac{\ln B\,(1) - B \frac{d(\ln B)}{dB}}{(\ln B)^2} \right]$$

$$\frac{dZ}{dB} = \ln (4 \times 10^6) \left[\frac{\ln B - 1}{(\ln B)^2} \right]$$

Setting the derivative equal to zero and solving for B,

$$\ln (4 \times 10^6) \left[\frac{\ln B - 1}{(\ln B)^2} \right] = 0$$

$$\ln B - 1 = 0$$

$$\ln B = 1$$

Therefore, B = 2.718. . . , the base of the natural logarithmic system. Substituting this value for B in Eq. 3,

$$Z = \ln (4 \times 10^6) \left[\frac{(2.718. . .)}{(\ln 2.718. . .)} \right]$$

$$= 15 \; \frac{(2.718. . .)}{1}$$

$$= 41,$$

which is the minimum number of hardware elements capable of being used in a hardware implementation of any coding scheme. However, control and display hardware is not available for use with base e numbering systems.

Base Comparison

Table 1 compares some of the base numbering systems and their associated symbology which may be considered for candidate coding schemes. The left-hand column lists the base of the numbering system, and the rows illustrate how the number "4 million" would appear in that base. The topmost row (the heading row) indicates the number of places (from right to left, the conventional way of going from low-order to

TABLE 1

RADIX-SYMBOLOGY COMPARISON FOR CANDIDATE NODE CODING SCHEMES

Base	Places 22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	No. of Elements
BASE																							
2	1	1	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	44
3									2	1	1	1	2	0	1	2	2	2	2	0	1	1	42
4												3	3	1	0	0	2	1	0	0	0	0	44
5													2	0	1	1	0	0	0	0	0	0	50
6														2	2	1	4	2	2	3	0	4	54
8															1	7	2	0	4	4	0	0	64
10																4	0	0	0	0	0	0	70
16																	3	$\overline{13}$	0	0	0	0	96
16																	3	D	0	9	0	0	96
26																		[illegible]	$\overline{19}$	$\overline{15}$	4	4	130
26																		I	T	P	E	E	130
45																			$\overline{43}$	$\overline{40}$	$\overline{13}$	$\overline{10}$	180
45																			⇹	Δ	D	Δ	180

high-order numbers) required to represent the number "4 million" in the symbology illustrated. The extreme right-hand column gives the number of hardware elements (the product of the radix and the maximum number of places for the symbology chosen) required if the given radix is chosen for a coding scheme.

Several observations are possible based on the presentation of Table 1. First, any symbology was rejected if it did not read high to low order from left to right, as is conventionally done. Second, arabic numerals and english alphabetic characters are the only serious contenders for the choice of symbology to be used in coding. Third, although it has been proven that the base e will have the minimum number of hardware elements (41), it was not considered a candidate because there is no convention for base numbering systems with other than a whole integer base, and e is irrational. From the standpoint of fewest number of elements, the choice would appear to be between the binary, ternary, and quaternary systems, with the binary system probably favored because of the preponderance of hardware currently available based upon the binary numbering system. However, there are certainly more factors to consider than just the number of hardware elements. Popular acceptance of a coding scheme requiring the recognition and manipulation of 22 digits for each node would be practically nil. Consequently, human factors considerations must certainly come to the fore in the selection of a node coding scheme.

Human Factors Considerations

There is a considerable body of research which has been conducted relative to the psychological acceptability of a symbolism used to represent information. Cited

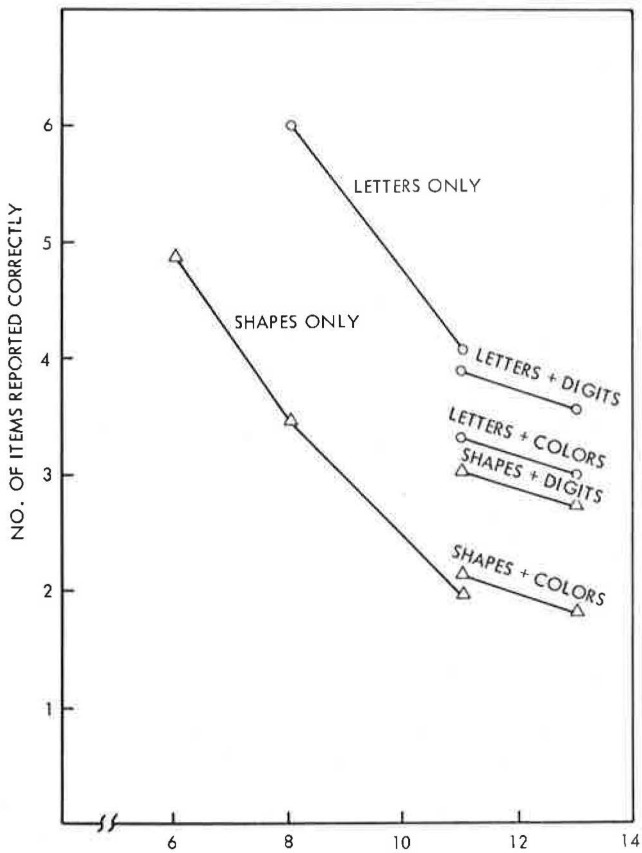

Figure 3. Recall capability.

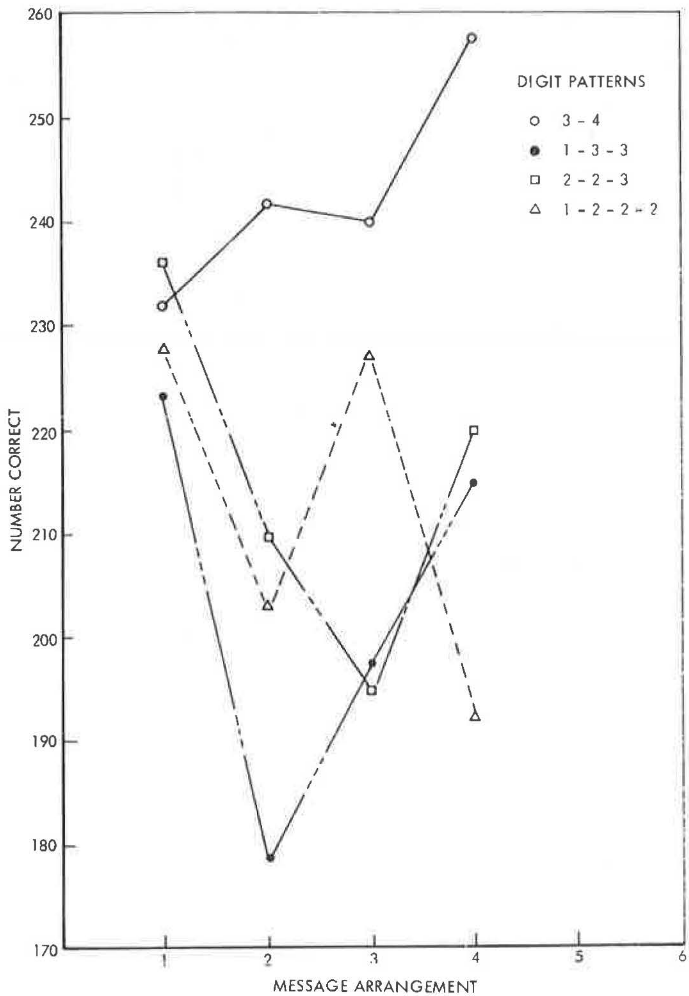

Figure 4. Effect of message format (grouping) on recall.

references (2, 3, 4, 5, 6) contain typical analyses and results of the body of data related to the parameters of code length, format, and symbolic form as these parameters influence the short-term memorization process. Figure 3 (2) shows the effect of message length and symbology upon the ability of a subject to recall a previously stated message. Figure 4 (4) illustrates results of experiments on grouping of symbols as an aid to the retentive process. The general conclusion to be reached from these data regarding preferential coding from a human factors standpoint, is that commonly recognized or used symbols, of short length, in a grouped format, would be highly desirable.

Applying these factors to the candidates (Table 1) leads one to conclude that the shorter length, higher-base symbology would be preferred from a human factors standpoint. Base 45 coding, however, contains problems in that there is no commonly used base 45 in existence today. Special symbols would have to be developed as shown on the last line of Table 1. The second from last line implements a hardware coding scheme of base 45 by doubling-up on the conventional arabic numeral digits, when necessary, at each of the four positions and uses a superscripted bar to indicate that these are special digits. The result is the equivalent of an eight-digit code requiring 180 elements, and the base 8 coding scheme in Table 1 requires only 64 elements and would therefore be more attractive.

Figure 5. Regional partitioning.

The first high-base, short code-length system in Table 1 using commonly recognized symbols is the 5-character, alphabetic-symbol base 26 system. Figure 3 indicates an alphabetic coding system is highly desirable, particularly if the code length can be held to a minimum. The question which next presents itself regards the feasibility of developing a geographic partitioning scheme which would be compatible not only with a code length of five characters, but also with the tri-level hierarchical partitioning scheme previously described.

Code Compatible Geographic Partitioning

The appellations assigned the geographic partitions derived are "region" for the most gross division of the highway network, "district" for the subdivision of the region, "sector" for subdivisions of the district, "zone" for the subdivisions of the sector, and "node" for subdivisions of the zone. Using these appellations and recalling the network hierarchy previously described, 20 intersections or nodes are grouped together to comprise a zone. These nodes are called the "minor nodes." A dominant node is chosen in the zone to be called the "secondary node." Twenty secondary nodes are grouped together to comprise a sector, so that each sector contains 20 zones. Again, a dominant node in the sector is chosen to be called the "primary node," and the 20 zones are considered to be satellites of the primary node. The primary node therefore has 400 intersections of the highway network subsequent to it. If 20 primary nodes are grouped to form a district, and 25 districts are allocated to the geographic region, then each region will have 500 primary nodes. Because of the 400-to-1 ratio of minor nodes to primary nodes, each region therefore contains 200,000 network intersections. Because it is desired to code 4 million nodes, a requirement for defining 20 geographic regions exists.

Figure 5 shows the geographic regional partitioning chosen. In order to avoid arbitrariness in boundary and area definitions for regions, the following conditions and constraints were adopted, and are reflected in Figure 5.

- Provision must be made for future expansion, hence allowance must be made for the existence of at least 10,000 primary nodes.

- The regional partitioning must provide for all existing primary nodes counted within each state. For the 48 contiguous states this amounted to 7,500 primary nodes.
- Major boundaries must follow state lines, or lie within states, but must never cross state lines. Thus, regions must lie wholly within a state or contain an integral number of states, and each state must contain an integral number of districts.
- Regions should fall into natural geographic areas for easy identification.
- Regions should contain the same number of nodes.

The conditions and constraints cited above were also reflected in the definition of districts, with the substitution of "county" for the "state" geopolitical entity. An example of district boundary determination is shown in Figure 6, district definitions

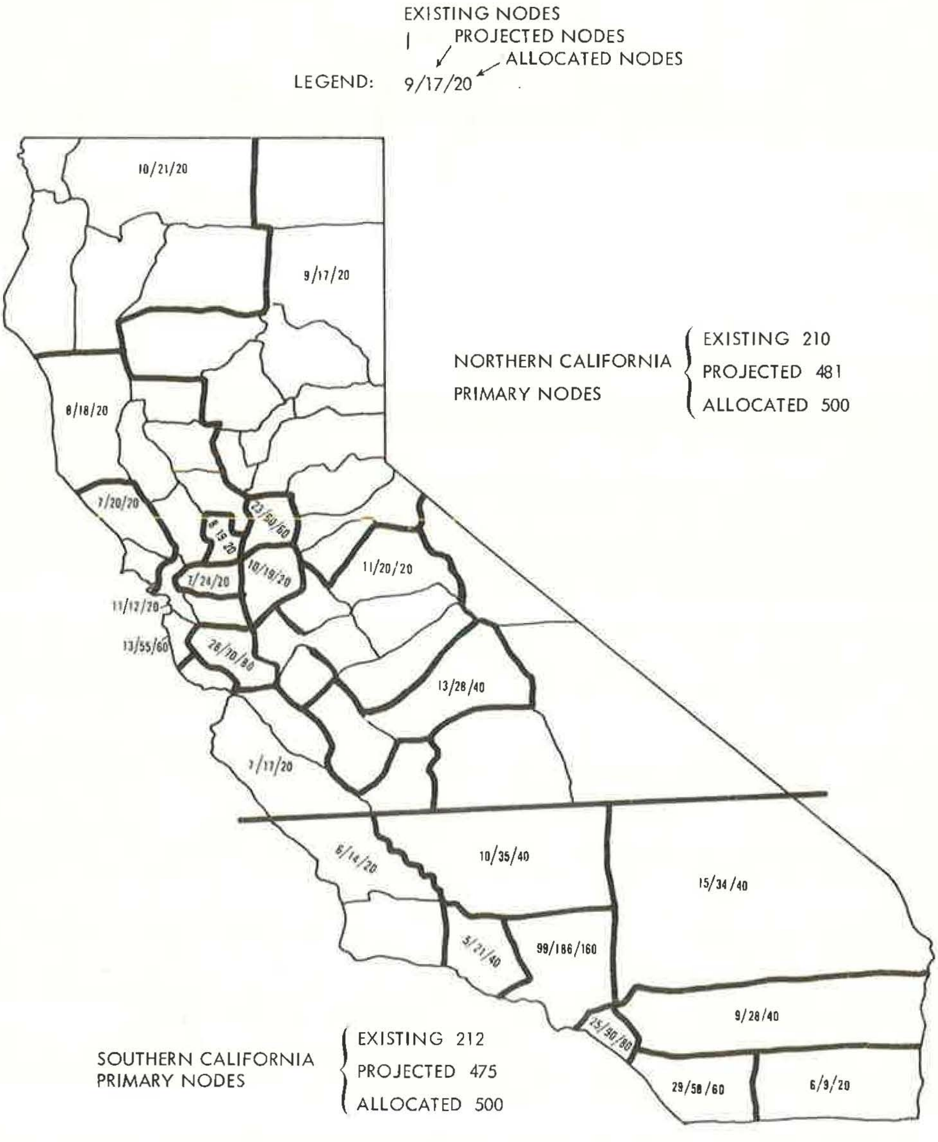

Figure 6. Interim district determination.

for California. Growth models, correlating highway network growth with population growth, were derived and applied (1) to assist in the determination of geographical partitioning boundaries. Boundary conditions for sectors and zones can be derived in a similar fashion.

Load Reduction on Communication Links

Coupling the dual geographic and hierarchic partitioning enables a reduction in information load on the system greater than two orders of magnitude. This is accomplished by recognizing that a driver's decisions in maneuvering a car through a roadway network must be made with most rapidity when the decisions concern those elements of the network closest to him. He is therefore provided with instructions regarding the finely detailed network in his immediate vicinity, which we define as the minor network in his district consisting of 8000 nodes (20 primary nodes in the district, each primary node having 20 secondary nodes, each secondary node having 20 minor nodes, and all 8000 nodes encompassed within the geographic district which is somewhat analogous in area to a county). At distances greater than his district, say in his region (equivalent somewhat in area to a state or group of states), the driver need only be furnished information about the less-detailed secondary network consisting of an additional 10,000 nodes (500 primary nodes per region, each primary node having 20 secondary nodes). For completeness, the driver need only be furnished information on a gross basis outside of his region. Thus, outside of the region in which he is receiving instructions from the roadside equipment, he is only furnished information about the primary network, which is a total of 10,000 nodes for the total contiguous United States. This hierarchic partitioning is a function of the location of the node, so that as the driver negotiates the highway network, the fine network appears to travel with him; thereby he is always furnished detailed information about the network at remote distances from him. In this way each driver is furnished optimum routing instructions for his specific destination through a network of more than 4 million nodes by only selectively referring to approximately 28,000 nodes.

Grouping destinations accessible from a given node via specific node branches enables additional reductions in roadside processing of information by the use of partial decoding of destinations common to the given node's branches. For instance, if all destinations in regions A, B, C, and D can be optimally routed out of a given node through branch 8, then an initial test on the destination's region code to see if it is the alphabetic character A, B, C, or D would suffice for optimal routing instruction transmittal to the vehicle. Further decoding of the destination's district, sector, zone, or node alphabetic characters would be superfluous and uneconomic.

Alphabetic Code Modifications

At this interim point, the coding scheme preferred is alphabetic in each of the five levels. A typical node code would be

ABCDE,

and is interpreted as the minor node E in the geographic zone of the secondary node D, which is in the geographic sector of the primary node C located in district B or region A.

Since the English alphabet can be considered a radix 26 system, almost 12 million separate nodes can be identified ($26^5 = 11.8 \times 10^6$). As only 4 million nodes need be identified, some letters may be held in reserve, and other modifications may be made to the coding scheme in order to enhance its attractiveness and acceptance by the general driving public.

One of the first modifications considered is that of state prefix coding. Region coding requires 20 alphabetic designators, so that 6 alphabetic designators can be held in reserve for growth. Since region partitioning is primarily a grouping by states, one technique which may be used is that of utilizing the common two-letter post office abbreviation for the state name included as part of the region. Thus, in the Northwest

Region of the United States, if the region designator is chosen to be A, then the constituent states of region A are Washington, Oregon, Idaho, Montana, and Wyoming, and the substituted abbreviations are WA, OR, ID, MT, and WY, respectively. If a "grouped" format is also applied to the state prefixed intersection name, then the resulting node name may appear as in the following examples:

Intersection Name		Node Code
WA-U-APE	for	AUAPE
OR-I-GON	for	AIGON
ID-A-LUU	for	AALUU
MT-M-EEK	for	AMEEK
WY-B-SAD	for	ABSAD

The mnemonic effect of the transformation illustrated is quite pronounced, and should tend to lessen reading and device input errors, and increase acceptance and confidence of the general driving public.

It has been reported that association of perceptual skills with verbalization aids the memorization process (7). This means that groups of alphabetic symbols can be recalled more readily if these groups can be made to form syllables that can be mentally verbalized. The five-level alphabetic code for 4 million nodes has the following number of divisions in each of the five levels:

Region	District	Sector	Zone	Node
20	25	20	20	20

Noting that the sector, zone, and node divisions require only 20 separate indentifiers in each, then six consonants can be eliminated from the alphabet when forming the code; resulting in a vowel-heavy alphabet and increasing the probability of forming pronounceable syllables, thus rendering the coding scheme more attractive from a human factors standpoint. Alternatively, five vowels and one consonant can be removed from each of the sector and node alphabetic designators and six consonants from the zone alphabets, so that a vowel-heavy alphabet is formed between two consonant-heavy alphabets, heightening the probability of occurrence of such syllables as DAN and BUG.

Table 2 lists the probabilities of various letter triplets occurring as a function of the type of letter (vowel or consonant) excluded from each coding level. The columns are arranged in such a way that the more easily verbalized triplets occur to the left. In row 6 the probability of triplets occurring in the code which can always be verbalized is maximized when vowels are excluded from the sector code, consonants from the zone code, and vowels from the node code. However, this selection also minimizes the percentage of triplets which can never be verbalized. On the other hand, if consonants are eliminated from all three levels (row 1), the possibility of no verbalization will be reduced by 0.33 while the reduction of the possibility of a "sure verbalization" is reduced by only 0.057, and the "sometimes verbalized" category is increased by 0.39. Therefore,

TABLE 2

OCCURRENCE PROBABILITIES

Letter Type Discarded (Sector, Zone, Intersection)	Probability of the Occurrence in the Code of (XXX*)							
	CVC	VCV	CVV	VVC	VCC	CCV	VVV	CCC
1. C - C - C	0.138	0.043	0.043	0.043	0.138	0.138	0.013	0.440
2. C - C - V	0.102	0.000	0.000	0.057	0.182	0.000	0.000	0.580
3. C - V - C	0.000	0.057	0.000	0.000	0.182	0.192	0.000	0.580
4. C - V - V	0.000	0.000	0.000	0.000	0.238	0.000	0.000	0.762
5. V - C - C	0.182	0.000	0.057	0.000	0.000	0.182	0.000	0.580
6. V - C - V	0.238	0.000	0.000	0.000	0.000	0.000	0.000	0.762
7. V - V - C	0.000	0.000	0.000	0.000	0.000	0.238	0.000	0.762
8. V - V - V	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	Always can be verbalized		Sometimes can be verbalized					Never can be verbalized**

*X = V; Vowel
X = C; Consonant

**With the possible exception of ZZZ and SSS.

it appears that the occurrence of vowels permits more verbalization of the code, and thus, heightens the association and memorization process.

Given that the vowels are to be retained, within a coding level (i.e., sector, zone, or intersection), there are 21!/(6! 15!) = 54,264 unique ways of removing six consonants from the alphabet. Thus, some criterion must be used to determine an appropriate set. Research aimed at measuring the association of three letter syllables has been reported (7). The six most frequently occurring consonants in syllables of low association value were observed to be H, J, Q, X, W, and Z. Consequently, when six consonants were to be eliminated from the coding, this set was chosen.

The preceding analysis has shown that hardware considerations and human factors aspects are somewhat in conflict. In order to compromise in favor of the hardware, the code should have many levels more than 5 and few elements in each level. In order to compromise in favor of human factors consideration and receive general public acceptance, a short, easily recognized and retained code is desired. The viewpoint adopted in the final selection of the code is that compromise should be made in favor of human factors and public acceptance, as long as cost penalties are not too great.

Cost Comparisons

The approach used to determine the significance of the cost differential associated with the different coding schemes is based on the following steps:

1. Estimate the equipment differences between the various coding systems, part by part, in both the vehicles and intersections.
2. Put a dollar estimate on these differences.
3. Assume a total of 70,000,000 vehicle devices costing $100 each.
4. Assume a total of 4,000,000 intersections costing $2000 each to instrument.
5. Apply the formula:

$$F_{A-B} = \frac{70 \times 10^6 (V_A - V_B) + 4 \times 10^6 (I_A - I_B)}{70 \times 10^6 \times 100 + 4 \times 10^6 \times 2000}$$

$$= \frac{70 (V_A - V_B) + (I_A - I_B)}{15,000}$$

where

F_{A-B} = fractional cost difference between the A and B coding systems;
V_A = vehicle system cost using coding system A;
V_B = vehicle system cost using code system B;
I_A = intersection system cost using coding system A; and
I_B = intersection system cost using coding system B.

TABLE 3

VEHICLE EQUIPMENT COMPARISON BY CODE TYPE

BASIC COMPONENT	BINARY	QUATERNARY	OCTAL	HEXADECIMAL	ALPHABETIC
	22 Inputs 1/2 Selec.	11 Inputs 1/4 Selec.	8 Inputs 1/8 Selec.	6 Inputs 1/16 Selec.	5 Inputs 1/25 Selec.
Switches	22 Toggle SWS	11-4 Pos. Dial SWS	8-8 Pos. Dial SWS	6-16 Pos. Dial SWS	5-25 Pos. Dial SWS
Cost Est.	$1.54	$3.35	$2.44	$2.82	$4.15
Sequencer	22 FF SR	22 FF SR	24 FF SR	24 FF SR	25 FF SR
Cost. Est.	$8.80	$8.80	$9.60	$9.60	$10.00
Total Cost Est.	$10.34	$12.15	$12.04	$12.42	$14.15
Cost Difference Compared to 5 inputs, 1/25 Sel.	-$3.81	-$2.00	-$2.11	-$1.73	0

Abbreviations: FF Flip-Flop
SR Shift Register
SWS Switches

TABLE 4

INTERSECTION EQUIPMENT COMPARISON BY CODE TYPE					
BASIC	BINARY	QUATERNARY	OCTAL	HEXADECIMAL	ALPHABETIC
COMPONENT	22 Inputs 1/2 Selec.	11 Inputs 1/4 Selec.	8 Inputs 1/8 Selec.	6 Inputs 1/16 Selec.	5 Inputs 1/25 Selec.
Destination Register	22 FF SR	22 FF SR	24 FF SR	24 FF SR	25 FF SR
Costs	$8.80	$8.80	$9.60	$9.60	$10.00
Sequencing Costs	22 6-Input Gates	22 6-Input Gates	24 6-Input Gates	24 6-Input Gates	25 6-Input Gates
Costs	$12.10	$12.10	$13.20	$13.20	$13.75
Storage Unit	13,500 Serial Mem. Bits	13,500 Serial Mem. Bits	14,500 Serial Mem. Bits	14,500 Serial Mem. Bits	15,000 Serial Mem. Bits
Costs	$270.00	$270.00	$290.00	$290.00	$300.00
Total Costs	$290.90	$290.90	$312.80	$312.80	$323.75
Cost Difference Compared to 5 Inputs, 1/25 Sel.	-$32.85	-$32.85	-$10.95	-$10.95	0

Abbreviations: FF Flip-Flop
SR Shift Register

The estimates of 70 million vehicles at $100 each and 4 million intersections at $2000 each are based upon conjecture of eventual mass production of these items in a configuration which provides for the route guidance function only.

The cost comparison for the vehicle equipment is given in Table 3. Only the switches and sequencer costs are significantly affected by the coding choice. The cost difference line, at the bottom of the chart, is the $(F_A - V_B)$ value, where B designates the five-level, 1/25 (one out of twenty-five elements) selection coding system used as a reference.

The cost comparison for the intersection equipment is given in Table 4. The destination register, sequencing gates, and storage unit costs are significantly affected by the coding choice. The cost difference line (at the bottom of Table 4) is the $(I_A - I_B)$ value, where B designates the five-level, 1/25 coding system used as a reference.

Referring to the F_{A-B} formula and the bottom lines of Tables 3 and 4, the fractional cost comparisons are evaluated below, all with reference to the five-level, 1/25 selection coding system:

<u>Twenty-two input ½ selection</u>

$$(V_A - V_B) = -\$\,3.81$$

$$(I_A - I_B) = -\$32.85$$

$$F_{A-B} = \frac{70\,(-3.81) + 4\,(-32.85)}{15{,}000} = -0.0265$$

$$= -2.65\%$$

Thus, the use of 22 toggle switches instead of five 25-position rotary switches results in a 2.65 percent overall cost saving. It may be noted that 1.21 percent of this saving is due to the low cost of small SPDT toggle switches (estimated at 7 cents per piece in large quantities) and 0.87 percent is due to the savings in memory.

<u>Eleven input, ¼ selection</u>

$$(V_A - V_B) = -\$2.00; \quad (I_A - I_B) = \$32.85$$

$$F_{A-B} = \frac{70\,(-2.00) + 4\,(-32.85)}{15{,}000} = -0.181 = -1.81\%$$

The memory savings here is also 0.87 percent.

STATE: Washington

City or County	Intersection	Node Name	Code Name
Tacoma	I 5 - RT7	WA-G-FAB	AGFAB

Figure 7. Intersection directory format, driver oriented.

Eight input, 1/8 selection

$$(V_A - V_B) = - \$2.11; \; (I_A - I_B) = -\$10.95$$

$$F_{A-B} = \frac{70\,(-2.11) + 4\,(-10.95)}{15{,}000} = -0.0127 = -1.27\%$$

Sixteen input, 1/16 selection

$$(V_A - V_B) = -\$1.73; \; (I_A - I_B) = -\$10.95$$

$$F_{A-B} = \frac{70\,(-1.73) + 4\,(-10.95)}{15{,}000} = 0.0110 = -1.10\%$$

On the basis of this cost comparison, the conclusion can be reached that on a total program basis, no coding system is capable of implementation in the foreseeable future that will realize cost savings of more than 2.65 percent over the alphabetic coding system. This additional cost is a small price to pay for convenience and higher probability of acceptance by the general public. Consequently, the coding system chosen is a five-level, alphabetical-digit code, with the region designator (the left-most letter) of the code replaced by a two-letter state abbreviation for general public use.

Intersection Directory

Assuming that the route guidance system is implemented, the problem of informing the general public of the node code presents itself. To assist resolving this problem, an intersection directory is proposed. An intersection directory of the primary nodes of the contiguous United States has been published (1). Two formats are used.

The first format is driver oriented, and assists a driver to determine the intersection (node) code closest to his destination. To use this directory, a standard road map is consulted to determine the intersection closest to the driver's destination. The route designators are then noted for this node, such as Interstate 5 (I5) and State Highway 7 (RT7). The directory is then consulted on a geographical basis by first going to the appropriate state, city or county division of the directory, and then looking up the route designators (I5-RT7) noted from the map. The directory then provides the proper node name to be input. Figure 7 illustrates this directory format.

STATE: Washington

Code Name	Node Name	City or County	Intersection
AGFAB	WA-G-FAB	Tacoma	I 5 - RT7

Figure 8. Intersection directory cross-reference format.

The second format (Fig. 8) is oriented more toward supporting the administrative functions of a highway. Here the listing in the directory is alphabetic by the five-letter code, which is the first column listed in the directory. The next column cross-references the state abbreviation prefixed code, the next column stipulates the city or county in which the node is found, and the final column gives the route markers of the roads forming the intersection.

Node Record and Data Files

All the items of information concerning an intersection (code, location, node branches, next sequential node, and branch distance measure) constitute data items which form a necessary record for the proper functioning of the route guidance concept. To this necessary record may be appended data items considered ancillary or even unrelated to the route guidance function, but of interest to the local agency maintaining files of records on highway network status. This attachment would include such data items as highway geometrics, construction details, construction costs, traffic count statistics, maintenance data, and accident and related safety statistics. Although not exhaustive, the preceding list serves to illustrate that route guidance data records could serve as a framework for construction of highway data files for local agency use.

In order to assist in the data recording function for route guidance, a format was chosen for record-keeping (Table 5). A maximum of three 80-column computer cards is allocated for entering the route guidance data for each node. If other data are desired, the data may be appended with the use of additional cards, the node code serving as the "key," for subsequent retrieval, manipulation, and processing. Using the format of Table 5, sorting of network data can be accomplished geographically (by region,

TABLE 5

CARD RECORD FORMAT

CARD #1

Column Number	Field Entry	Remarks
1	Blank	
2-6	5-level Alpha Node Name	
7	Blank	
8-9	Alpha State Code	Following U.S. Mail Standard Abbreviations
10	Blank	
11-13	County Code	
14	Blank	
15-29	City Code and/or Township-Range Location Code	Decided by City or County Officials
30	Blank	
31	Branch Designator	1-8 by Compass Octants
32	Blank	
33-43	Old Route Designator	Highway Number, Street Name, County Highway Number. Also Distance Measure could be entered.
44	Blank	
45-49	Name of next node on this branch.	
50-69	Repeat on 30-49	
70	Blank	
71	Blank or *	*After last branch to stop reader
72	Blank	
73-80	Numbering of data decks	May be numbered as desired for sorting.

CARD #2 AND CARD #3

Column Number	Field Entry	Remarks
1	X	For sorting purposes. Also reader knows of different format to follow.
2-6	Node Name	Same entry as 2-6 on Card #1
7	Blank	
8-9	Alpha. State Code	
10-29	Same as 30-49 on Card #1	
30-49	Same as 10-29	
50-69	Same as 10-29	
70	Blank	
71	Blank or *	
72	Blank	
31, 51, or 71 or not at all	*	*After last branch To stop reader. Comments may follow.

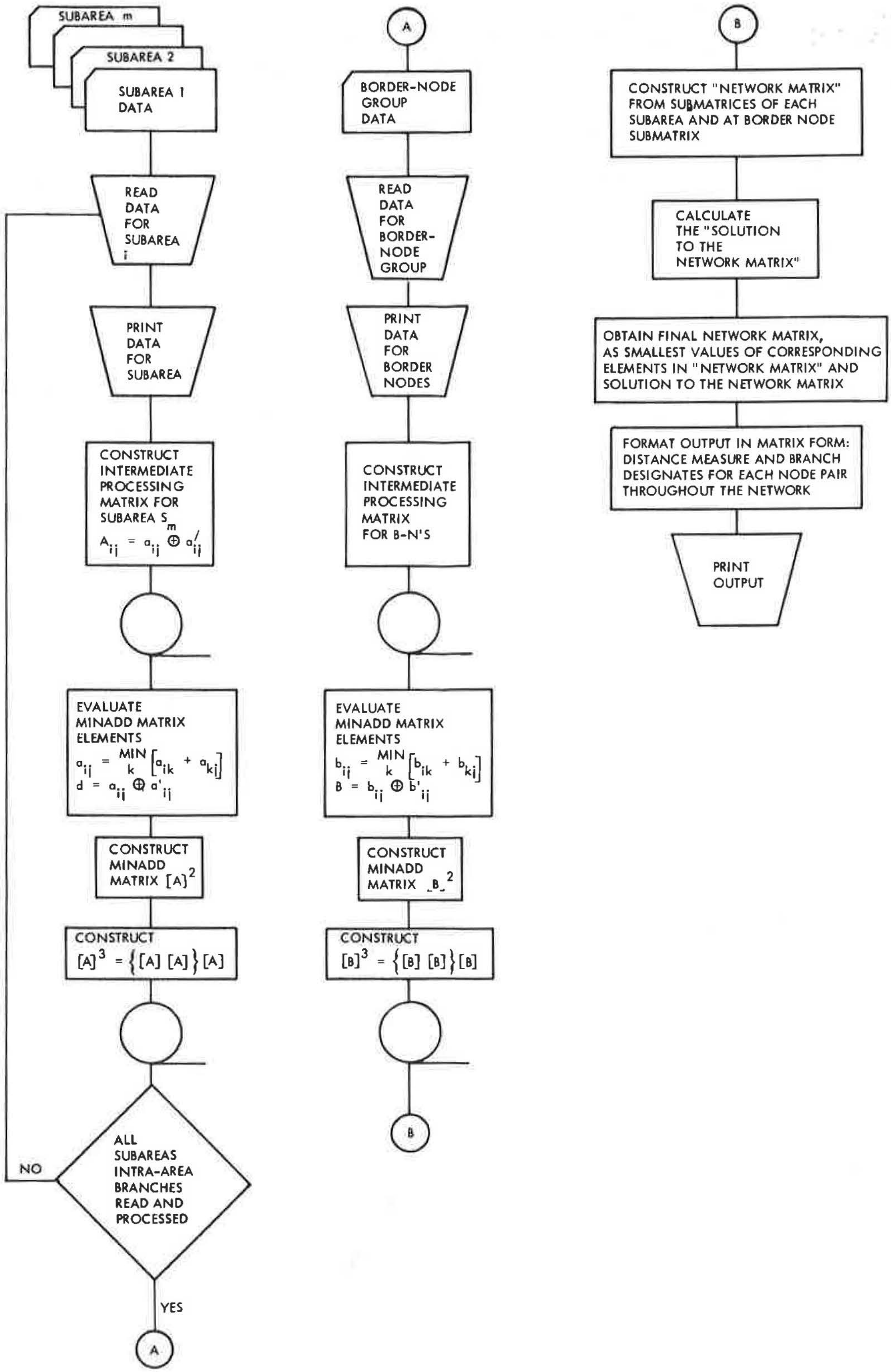

Figure 9. Shortest route program flow chart.

district, sector, zone, and node), heirarchically (by primary, secondary, and minor node), by state, by county, by city, branch designator, branch distance, route, or next node name. This type of sorting permits route following to be accomplished so that all the nodes on any given route in the network can be listed.

OPTIMAL ROUTES DETERMINATION

In anticipation of solving a convenient network of highway routes over a prototype model of the information exchange equipment, a FORTRAN IV program for solution of optimal routes has been written and compiled on Philco Model 212 computers. The flow chart for this program is shown as Figure 9.

This program determines the shortest route distance and the branch which yields this route between each pair of nodes in a matrix representation of the network. Partitioning a large network allows a considerable reduction in computation steps and processing time by decomposing the matrix into several proportionately smaller arrays, and solving each separately. The shortest routes of each subnetwork are then interstitched using the concept of "border-nodes" for each subarea. It has been found that the number of iterations of the border-node search procedure plus those for the complete search within each subarea are fewer than would be required if all of the nodes of the total network are searched without partitioning.

The computer program uses a process called minaddition first described by Shimbel (8). In ordinary matrix multiplication, the product of two matrices a and b defined as

$$a = a_1, a_2, \ldots, a_n$$

$$b = b_1, b_2, \ldots, b_n$$

is

$$a \cdot b = a_1 b_1 + a_2 b_2 + \ldots, a_n b_n$$

Minaddition of a and b is defined as

$$\text{Minaddition } (a, b) = \min_k (a_k + b_k)$$

i.e., the minimum of all sums $a_k + b_k$ taken over all k.

If the rules of matrix multiplication (and squaring) are retained except when minaddition is substituted in the formation of the product matrix, then the type of multiplication used in the computer program for determining optimal routes is defined.

Shimbel has shown that if the original network matrix is successively squared, at most r times, the resultant matrix is the solution to the original network matrix. The maximum number of times r that the matrix must be squared is given by

$$2^r \geqq n > 2^{r-1}$$

where n is the number of nodes in the network, i.e., the matrix is an nxn matrix.

Use of the Shimbel minaddition process plus the partitioning technique previously described permits the FORTRAN IV program developed at Philco-Ford to solve for optimal routes in network matrices of up to 1564 nodes with a computer run time of approximately 2 hours.

SUMMARY

A system for uniquely coding and formatting over 4 million nodes (intersections) of the contiguous United States roadway network has been devised, and a computer program for the determination of optimal routes over an instrumented network of up to 1,500 nodes has been written and tested. The code is alphabetic, short, and vowel-heavy, with notable mnemonic effects. A dual form of partitioning is used in coding the roadway network, thereby lessening the information load on the communication

link between the roadway and the vehicle. Over 7000 major intersections of the U.S. Highway network have been coded in a machine language format based on the node code as a key, and an intersection directory has been compiled and published through use of the machine language and node code.

REFERENCES

1. Highway Coding for Route Designation and Position Description, Volume I and Volume II. Philco-Ford Technical Report WDL-TR3580, Palo Alto, California, 1 July 1968.
2. Mackworth, J. F. Interference and Decay in Very Short Term Memory. Jour. Verbal Learning and Verbal Behavior, Aug. 1964.
3. Terchner, W. H., Dahlquist, E., Eddy, N., and Pesner, S. Experiments on the Effects of Input Variables on Multi Target Alphabetic Displays. Electronic Systems Division TDR 63546, July 1963 (USAF Decision Sciences Laboratory, Hanscom AFB, Massachusetts).
4. Severin, F. T., and Rigby, M. K. Influence of Digit Grouping on Memory for Telephone Numbers. Jour. Applied Psychology, April 1963.
5. Jung, J. Telephone Numbers: A Mnemonic Alternative to All-Digits. Perceptual and Motor Skills, Aug. 1963.
6. Mackworth, J. The Duration of the Visual Image. Canadian Jour. Psychology, Jan. 1963.
7. Stevens, S. S. Handbook of Experimental Psychology. John Wiley and Sons, New York, 1951.
8. Shimbel, A. Structural Parameters of Communication Networks. Bull. Mathematical Biophysics, Volume 15, 1953.