# HIGHWAY RESEARCH RECORD

**Number 293**

## Transportation Systems Planning

12 Reports

Subject Area

11 Transportation Administration
15 Transportation Economics
82 Urban Community Values
83 Urban Land Use
84 Urban Transportation Systems

HIGHWAY RESEARCH BOARD

DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D. C., 1969 Publication 1671

Price: $4.00

Available from

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

## *Department of Urban Transportation Planning*

(As of December 31, 1968)

Pyke Johnson, Chairman

HIGHWAY RESEARCH BOARD STAFF

James A. Scott

Frank N. Barker
Kurt W. Bauer
Fred J. Benson
Donald S. Berry
Siegfried M. Breuning
E. Wilson Campbell
J. Douglas Carroll, Jr.
F. Stuart Chapin, Jr.
John J. Cummings
Leo Cusick
Harmer E. Davis
John G. Duba
Alfred Eisenpreis
John R. Hamburg
Robert L. Hardin, Jr.
Thomas J. Hart
Frederick O. Hayes
Patrick Healy
Bernard F. Hillenbrand
E. H. Holmes
William L. Hooper
Edgar M. Horwood
John T. Howard
Richard Ives
J. E. Johnston
Peter A. Lewis
Mo Chih Li
Burton W. Marsh
J. O. Mattson
Robert E. McCabe
J. B. McMorran
William L. Mertz
Charles L. Miller
Robert B. Mitchell
Walter S. Rainville, Jr.
Clifford F. Rassweiler
Carlton C. Robinson
Paul W. Shuldiner
Merlin Smelker
Wilber E. Smith
Francis C. Turner
Alan M. Voorhees
Edward G. Wetzel
C. R. Wilder
F. Houston Wynn
Charles J. Zwick

COMMITTEE ON LAND USE EVALUATION

(As of December 31, 1968)

John R. Hamburg, Chairman
Creighton-Hamburg Incorporated, Bethesda, Maryland

Carl Swerdloff, Secretary
Bureau of Public Roads, Washington, D. C.

Rodolfo J. Aguilar
Brian Barber
Robert E. Barraclough
F. Stuart Chapin, Jr.
Salvatore V. Ferrera
Aaron Fleisher
William L. Garrison
Ronald S. Graybeal
David A. Grossman
Walter G. Hansen
Britton Harris
George C. Hemmens
Donald M. Hill
C. David Loeks
David D. Longmaid
Ira S. Lowry
John K. Mladinov
Kenneth J. Schlager
Morton Schneider
David R. Seidman
Joseph R. Stowers
R. J. Wheeler
Lowdon Wingo, Jr.
Richard D. Worrall

COMMITTEE ON COMMUNITY VALUES

(As of December 31, 1968)

Peter A. Lewis, Chairman
Bureau of the Budget, Washington, D.C.

Charles H. Graves, Secretary
U.S. Department of Housing and Urban Development, Washington, D.C.

Frederick T. Aschman
Leon M. Cole
Lyle C. Fitch
Bamford Frankland
Walter D. Harris, Jr.
Frederick O. Hayes
Kevin E. Heanue
John T. Howard
Jonathan B. Howes
Robert G. Howes
Richard Ives
Peter Koltnow
John B. Lansing
Morton Lustig
Robert E. McCabe
Jack Meltzer
Robert L. Morris
Walter S. Rainville, Jr.
John Rannels
Joseph L. Schofer
Margaret T. Shaffer
Allan Temko
John R. Walker
William L. C. Wheaton

## COMMITTEE ON TRANSPORTATION SYSTEM EVALUATION
(As of December 31, 1968)

William L. Mertz, Chairman
Tri-State Transportation Commission, New York, New York

Joseph W. Hess, Secretary
Bureau of Public Roads, Washington, D.C.

Frank E. Barker
Donald S. Berry
E. Farnsworth Bisbee
E. Wilson Campbell
Arthur A. Carter, Jr.
Leo Cusick
Thomas B. Deen
Stephen George, Jr.
George Haikalis
Frank W. Herring
Neal A. Irwin
Howard C. Lawson, Jr.
Warren B. Lovejoy
James J. McDonnell
Joseph C. Oppenlander
Ronald W. Pulling
Paul W. Shuldiner
Vergil G. Stover
Edward F. Sullivan
Edwin N. Thomas
Robert S. Vogt
Martin Wohl
F. Houston Wynn
Charles J. Zwick

## COMMITTEE ON FUTURE CONCEPTS
(As of December 31, 1968)

Siegfried M. Breuning, Chairman
Massachusetts Institute of Technology, Cambridge

Richard D. Desrosiers, Secretary
University of Maryland, College Park

Albert A. Atwell
Edith Bairdain
Wesley R. Bellis
Hermann S. Botzow
Daniel Brand
Vincent R. Desimone
Alfred Eisenpreis
Thomas H. Floyd
David Q. Gaul
Edmund Hodgkins
George A. Hoffman
William L. Hooper
M. Peter Jurkat
Alan F. Kay
James H. Kell
Robert K. Lyons
Marvin L. Manheim
Lawson M. McKenzie
Theodore F. Morf
Edward K. Morlok, Jr.
Wayne M. Pecknold
Louis Pignataro
Clifford F. Rassweiler
Paul W. Shuldiner
Morton I. Weinberg
Edward G. Wetzel
Martin Wohl
Vladimir K. Zworykin

## *Department of Economics, Finance and Administration*

R. C. Blensly, Chairman
Oregon State University, Corvallis

HIGHWAY RESEARCH BOARD STAFF

Kenneth E. Cook

COMMITTEE ON HIGHWAY COSTS AND PROGRAMMING
(As of December 31, 1968)

Donald O. Covault, Chairman
Georgia Institute of Technology, Atlanta

Harold W. Hansen, Secretary
Portland Cement Association, Skokie, Illinois

Wayne C. Allison
Alvin F. Bastron
Orhan Baycu
Ralph D. Brown, Jr.
J. P. Buckley
B. G. Bullard
S. F. Coffman
Benjamin M. Givens, Jr.
A. A. Grant
James O. Granum
Stearns R. Jenkins
Ralph D. Johnson
Samuel L. Lipka
Edwin C. Lokken
C. F. McCormack
Charles H. Moorefield
Willa W. Mylroie
G. A. Riedesel
Robert S. Scott
James R. Turner
Walter G. Wigle
H. S. Wiley
Neil H. Wilson
Reed H. Winslow

COMMITTEE ON SOCIO-ECONOMIC ASPECTS OF HIGHWAYS
(As of December 31, 1968)

Bamford Frankland, Chairman
California Division of Highways, Sacramento

William G. Adkins
V. Lewis Bassie
Henry W. Bruck
R. Kirk Dansereau
Hays B. Gamble
Edmond L. Kanwit
Paul R. Lowry
Walter C. McKain, Jr.
Charles Thomas Moore
James R. Nelson
Robinson Newcomb
Robert W. Paterson
Norman Pearson
James M. Smith
C. A. Steele
Floyd I. Thiel
Christopher Tunnard
Paul Wendt

# Foreword

The various papers and abridgments contained in this RECORD cover a rather wide range of component elements of the transportation planning process. Various topics covered include concepts for determining urban subarea transportation requirements, cost comparisons of bus and rail systems, techniques for determining alternative transportation systems and for planning and developing transportation corridors, computer simulation techniques to simulate a demand-scheduled bus system, and an analytical technique to estimate highway impedances in urban areas.

Two papers discuss the techniques involved in small subarea planning of transportation requirements. Hoel discusses central-city distribution and states that, while the final choice of systems depends on a variety of economic, political, and environmental factors, it would appear that small-vehicle systems for central-city distribution present a viable alternative for system design. Bellomo and Provost emphasize the need for subarea planning to be carried on as an integral part of regional system planning. The authors discuss the application of an approach for the evaluation of alternative transportation systems of subareas.

Deen and James compare rail and bus costs by comparing relative service costs associated with four separate operating patterns. Sensitivity of the conclusions were tested to determine the effect of line length, passenger-traffic patterns, feeder-bus route length, station spacing, and other variables. The authors state that no particular operating method is inherently superior to any other and that any one method may be superior in a particular instance depending on the operating environments and conditions that must be met.

Manheim postulates in his paper that transportation is not a question of hardware alone, but of social and political choice. The real issues, according to the author, are not alternative technologies, but their ramifications—how each alternative transportation system affects differently the various segments of a society. The author focuses on the process of search and choice—how alternatives are generated and what the interactions are between search and choice.

The paper by Leisch and the abridgments of papers by Moore and Mason and by Stuart emphasize various aspects of transportation-corridor planning and development. Leisch develops a concept of permanent transportation corridors and network configurations that allow for long-range planning. The author states that an optimum network of major transportation facilities, initially spaced at 4 to 6 miles and eventually spaced at 2 to 3 miles in built-up areas, can function at reasonable levels of service for any predicted rates of population growth, urban expansion, and travel demand. The Moore and Mason abridgment is a summary of their research work to ascertain whether there is justification for the concept of a household-travel carrier corridor. Stuart discusses freeway corridor planning and states that it is a means for achieving integrated transportation and land use development, which can strengthen the process of specific route location for both urban and interurban freeways.

Morris discusses the opportunities for transportation planners to build on the basic research carried out in urban areas in the designing of self-contained communities of new towns. The varying concepts of traffic and transportation planning as applied to new towns are discussed by the author.

Bruggeman and Heathington discuss a computer simulation model that simulates a demand-scheduled bus system offering door-to-door service. Sensitivity analysis was performed on some of the parameters in the model. Parameters studied were link travel time, maximum pickup time, shape of serviced area, frequency of calls, bus capacity, and length of trips. Outputs considered sensitive to change in these parameters were cost of operation per passenger mile, waiting time to be picked up, passenger travel time on the system, and the total time required to make a trip.

Wegmann and Berry discuss a link-analysis technique that relies on multivariate statistical procedures as a possible guide for designating major rural highways as freeways, expressways, and other major rural road systems.

Rassam and Ellis propose an Urban Network Impedance Model for analytically estimating point-to-point highway impedances within urban areas without employing current network-analysis procedures. The model was tested by comparing the impedance estimates of the model to those provided by conventional network-analysis procedures. Test results suggest that the model provides a reliable and efficient procedure for estimating point-to-point impedances within urban areas.

Schneider's paper proposes a theoretical model by computer application that will show the relationships among travel, land development, and the time and cost characteristics of transportation.

# Contents

# Evaluating Alternative Strategies for Central-City Distribution

LESTER A. HOEL, Transportation Research Institute, Carnegie-Mellon University

Central-city distribution is an essential element in the hierarchy of a regional transportation network. It is the total system that the user views, and consequently both collection and distribution systems require separate evaluation.

A central-city distributor system can be composed of pedestrian ways, bus systems on existing streets, or mechanized loops on exclusive rights-of-way. The distributor system technology is influenced by the nature and configuration of the regional system. If the regional system penetrates the core area, then both delivery and distribution must be accomplished within one set of criteria. If two or more technologies are considered, then a greater variety of alternatives is possible because the regional system is freed of its role as a CBD distributor and the distributor system is freed of the restraints dictated by the location of the regional lines. The selection of alternative strategies is based on a variety of factors including facility cost, level of service, environment, and community impact.

Travel time, including waiting, walking, station dwell, transfer, and riding, is utilized in this study to evaluate a variety of single-technology and small-vehicle systems developed for the Pittsburgh central business district. The model developed calculates all appropriate paths between each cordon point and the zone centroid for a convenience (minimum time) rule and for an accessibility (minimum walk) rule.

•CENTRAL-CITY distribution is one element in the hierarchy of a regional transportation network. For corridor-oriented systems, transportation service is normally between low-density residential areas and high-density commercial activities in the central business district or other urban cores.

A regional network comprises facilities consisting of three basic elements: (a) the collection of persons located in relatively scattered residential areas by the use of facilities such as feeder bus systems or park-and-ride; (b) the delivery between residential areas and the downtown core by the use of line-haul rapid transit systems, either bus or rail; and (c) the distribution to ultimate destinations by the use of a variety of distribution systems, including walking. It is the total system that the user views when he compares alternative modes, and consequently both collection and distribution systems are essential elements of a regional network and require separate evaluation.

The technologies that are appropriate to the function of delivery and central-city distribution can influence the resulting system configuration. For example, if a single technology is maintained throughout the system, then both delivery of high volumes to the core area and distribution within the core must be accomplished within one set of criteria (1). If, however, two or more technologies are incorporated, then each can be tailored to serve the specific requirements set by travel demands.

Paper sponsored by Committee on Transportation System Evaluation and presented at the 48th Annual Meeting.

## ALTERNATIVE DISTRIBUTOR SYSTEMS

If the regional system, which may be either rail transit line or bus on a separate right-of-way, is to act also as the central-city distributor, then the regional lines must penetrate the core area. This conventional approach, illustrated in Figure 1, requires that stations be located in the core area along the regional lines.

On the other hand, if two or more technologies are considered, then a greater variety of alternatives are possible because the regional system is freed of its role as a CBD distributor and the distributor system is freed of the restraints dictated by the location of the regional lines. Consequently, each can be designed to its individual capacity requirements rather than to that of the entire system. This more novel approach, illustrated in Figure 2, relies upon the development of small-vehicle systems tailored specifically for the purpose of central-city distribution.

If the regional system does not penetrate the core, a transfer to the distributor system will be required to a greater extent than that for the single technology system, although transfers are unavoidable in either system. The transfer is generally considered to be highly undesirable under most present-day systems, as it represents an inconvenience in both time, energy, and comfort, quite intolerable in our affluent motorized society. Careful design of the transfer point from the viewpoint of a pleasurable environmental experience and coordinated operation between both systems is required to minimize the delay at the interface and thus enhance the public's acceptance.

## REQUIREMENTS OF SMALL-VEHICLE SYSTEMS

The central-city distributor system can take a variety of forms, ranging from simple pedestrian paths or bus systems on existing streets to mechanized loops on exclusive

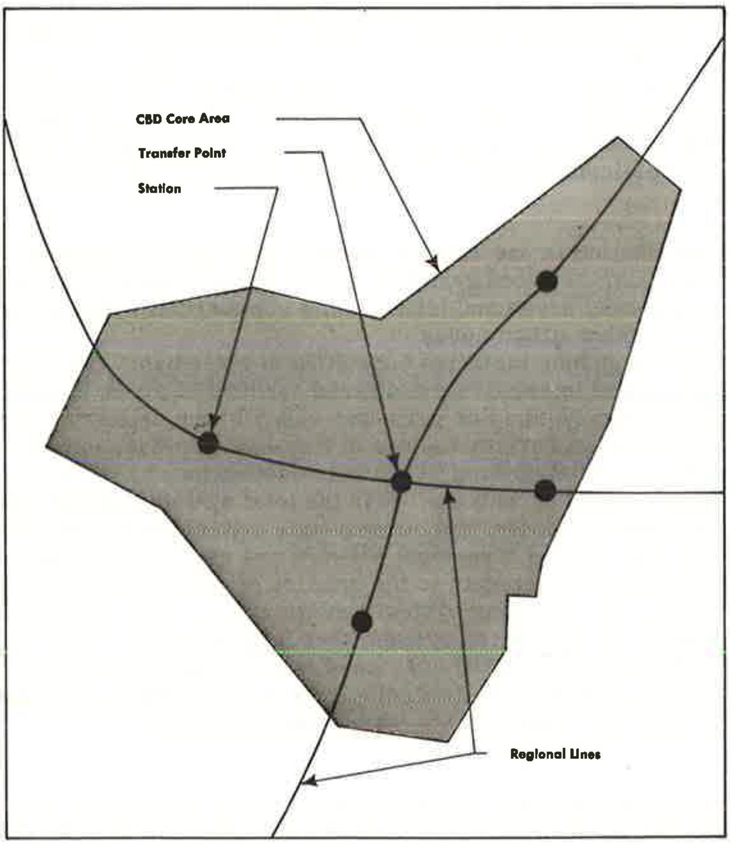



Figure 1. Regional system acting as central-city distributor.

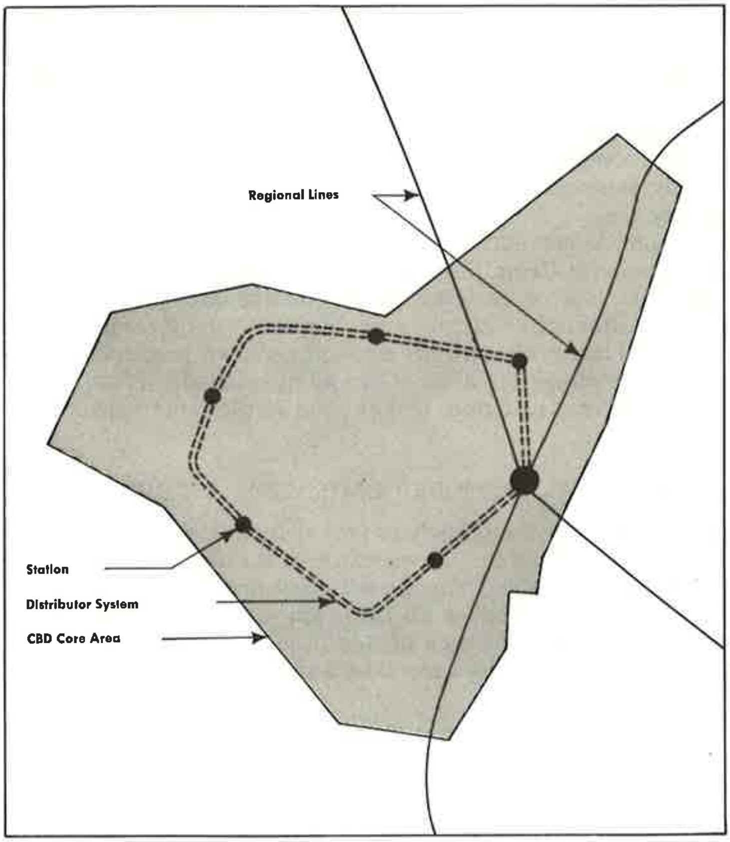



Figure 2. Small-vehicle systems for central-city distribution.

rights-of-way. Although small-scale systems have been proposed for central-city distribution, none has been tested in a full-scale urban setting (2). Basic requirements of such a system include ability to be integrated within an existing environment; easy accessibility and availability throughout the day and evening; minimum waiting; minimum fare; silent operation, free of fumes and noise; flexibility and capability of modification including complete removal; ability to adjust to changing capacity requirements; pleasurable riding experience; and moderate average speeds (3).

## FACTORS AFFECTING SELECTION

This paper presents an evaluation of alternative central-city distribution systems using travel time as the measure of effectiveness. The model used compares both single-technology systems and small-vehicle systems, and results are presented based on data for the Pittsburgh central business district. The model tests all reasonable combinations of elements that form a complete trip between each origin and each destination and measures the relative service ability of each alternative. [For an evaluation of environmental and cost factors see Netsch (4) and the Allegheny Study (5).]

The selection of alternative strategies for central-city distribution is based upon a variety of factors, including facility cost, travel costs and benefits to the user (i.e., level of service), compatibility with the environment, impact upon the community, and other intangibles. This paper is primarily concerned with service-level aspects, but a brief look at cost factors and aesthetics is appropriate.

Transportation-facility costs are conveniently categorized as the capital cost for construction and right-of-way acquisition, and annual costs for maintenance, operation, and administration. Calculation of these costs is essential for the evaluation of each strategy.

Compatibility with the environment and community impact require an evaluation of the technology of each strategy with respect to items such as architectural effect, noise levels, effect upon abutting land uses, disruption of normal activities during construction, and the interrelationship with other community objectives. These items may be quantifiable or highly subjective, but information concerning these factors is essential in the evaluation process.

Level of service can be measured by several characteristics including user cost, safety, reliability, aesthetic (from the rider point of view), and time.

For central-city distribution systems, it may be assumed that the fare structure and accident cost for each alternative strategy are equal. Furthermore, factors related to rider identification with the system, while recognized as a positive attribute for aerial or surface systems, are subjective and not easily quantified. Thus, for evaluating the service level of alternative strategies, travel time is the most appropriate measure of effectiveness.

## ELEMENTAL ACTIONS COMPRISING A CBD TRIP

The elements of a trip may be defined as travel from a cordon point to a zone centroid. (The cordon is defined as the intersection of the regional system with the core-area boundary. This definition provides an efficient and equitable means of comparison between alternative strategies because all trips are considered to originate at a common boundary.) A trip can be described by a series of elemental actions, such as those illustrated in Figures 3 and 4. The model considers the time involved in each of these

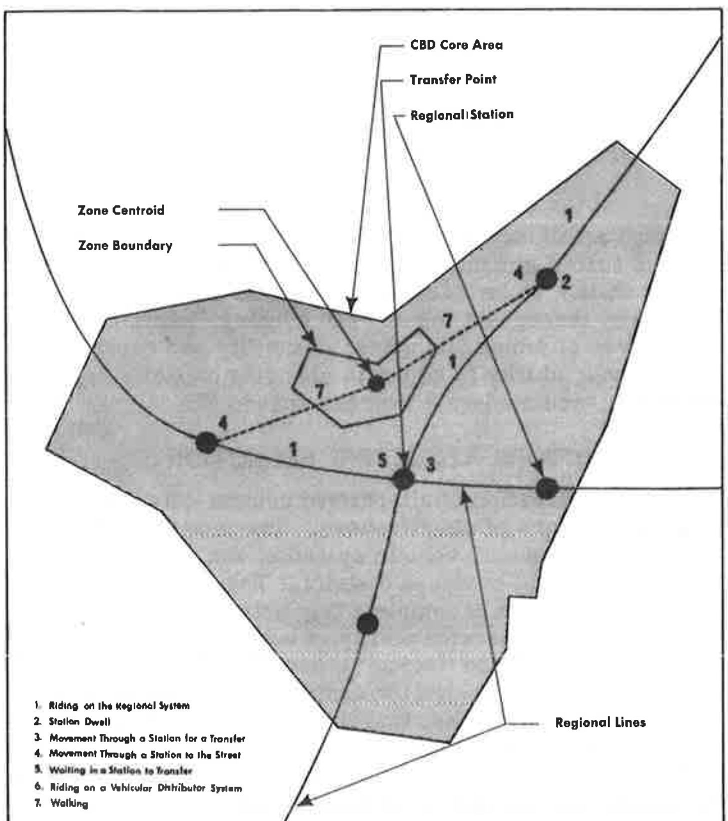



Figure 3. Elemental actions comprising total trip on regional system distributor line.

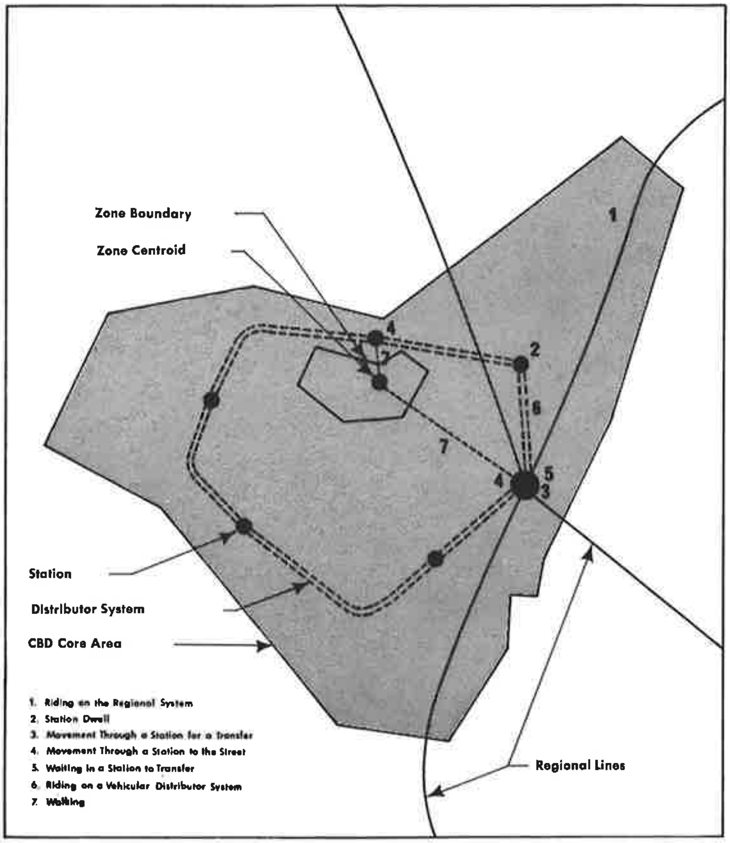



Figure 4. Elemental actions comprising total trip on small-vehicle system distributor line.

actions when measuring the total time through the system. These elemental actions are:

1. Riding on the regional system;
2. Dwelling in the station (i.e., time spent on a vehicle in the station when the passenger is neither embarking nor debarking);
3. Moving through a station for a transfer to another vehicular mode;
4. Moving through a station to the street;
5. Waiting in a station to transfer;
6. Riding on a vehicular distributor system other than the regional system; and
7. Walking.

For example, in Figure 3, a trip originating at the northern cordon point has two possible paths to the zone centroid, one composed of elemental actions 1, 4, 7 and the other of 1, 2, 1, 3, 5, 1, 4, 7. In Figure 4, two possible paths are 1, 4, 7 or 1, 3, 5, 6, 2, 6, 4, 7. One of these paths involves less time than the other. However, the user might select the longer time path if it involves less walking.

Two sets of information are required to reduce the ride on a regional or distributor line to a time interval. First is the spatial arrangement of the guideway and the stations, and second is the equations relating the motion of the vehicular system to space and time. Solving the equations of motion for each particular alignment yields the ride time between two stations.

The station dwell (i.e., in-station time for a vehicle) can be fixed for a particular system, or it can be determined for each station but varied from station to station depending on the time to load or unload a vehicle.

Movement through a fixed-guideway system station either to another vehicle or to the street requires knowledge of the spatial arrangement of the station interior and the various modes employed to move passengers. Time intervals can be derived by knowing walking distances, pedestrian speeds, and travel rates of escalators, elevators, and moving walkways. The level of sophistication can vary from crude approximations of in-station movements to simulated transfer analysis that accounts for queuing at escalator points and for other factors. For surface bus systems, the in-station time is usually negligible as the passenger debarks directly onto the street.

The wait for a transfer is dependent upon the regularity of schedules and system capacity. Expected waiting time is proportional to the headway between successive vehicles, but could be increased if capacity is exceeded.

## COMPUTING TRIP TIMES: CONVENIENCE VS ACCESSIBILITY

The rule for computing trip times involves knowledge of the behavioral characteristics of the system user in order that the correct route choice and combination of links be determined for each origin and destination pair. To date, there are no definitive answers to this behavioral question and, consequently, two route-choice rules are assumed, one based on convenience and the other on accessibility.

The convenience rule is based on the hypothesis that all users will choose the absolute minimum-time path from origin to destination. This route-choice rule assumes that the user places equal weight on all elements of the trip (i.e., it is immaterial if the user is riding, waiting, or walking provided that the total trip time is minimized).

The accessibility rule is based on the hypothesis that all users will choose the minimum-time path that provides delivery nearest to the ultimate destination. This rule assumes that the user places greatest weight upon the walking element of the trip, (i.e., it is immaterial that the user rides or transfers provided that the walk time is minimized).

The weighted averages of all travel times obtained by applying these rules are the measures of effectiveness and are termed the convenience measure or the accessibility measure.

These rules represent extremes, but the model is flexible and allows for computing values between these measures. In operation, the accessibility measure is computed by heavily weighting the street walk time. To obtain results between these extremes, it is only necessary to change the weight assigned to the walk element.

The convenience and accessibility measures are obtained by computing the weighted average total trip time for the system. The expression for the weighted average total trip time for the system is:

$$t^{T}_{AVE} = \sum_{j=1}^{m} \sum_{i=1}^{n} \rho_{ij}\, t_{ij} \qquad (1)$$

where

$t^{T}_{AVE}$ = weighted average system trip time,
$i$ = index of origins,
$n$ = total number of origins,
$j$ = index of destinations,
$m$ = total number of destinations,
$\rho_{ij}$ = weight placed on trips from i to j,
$t_{ij}$ = total minimum trip time from i to j under a measure,

and

$$\sum_{j=1}^{m} \sum_{i=1}^{n} \rho_{ij} = 1 \qquad (2)$$

The weight, $\rho_{ij}$, is that proportion of the total passengers going from i to j. This relates to the distribution of originations and terminations by

$$D_j = \sum_{i=1}^{n} O_i \rho_{ij} \tag{3}$$

or

$$O_i = \sum_{j=1}^{m} \rho_{ij} D_j \tag{4}$$

where $D_j$ is the proportion of total passengers traveling to destination j, and $O_i$ is the proportion of total passengers originating at i.

As an indicator of dispersion, the standard deviation can be calculated using the following expression:

$$\sigma_T = \sqrt{\sum_{j=1}^{m} \sum_{i=1}^{n} \rho_{ij} t_{ij}^2 - \left(t_{AVE}^T\right)^2} \tag{5}$$

These formulas consider only the total trip time. Under each hypothesis the mean and standard deviation of each primary trip element can also be similarly computed.

Other service statistics are available from the model that can provide more detailed information about all elements of the trip. Furthermore, as a result of constructing a general model for testing the level of service, other relevant statistics, such as the number of passengers that transfer or terminate at a given station, can be generated. These are useful for comparing alternative strategies, and with the model as posed this is simply a matter of bookkeeping.

## THE PITTSBURGH CBD: A CASE STUDY

The present and future land use plan for the central city is a primary determinant in selecting a transportation alternative. The density configuration of trip destinations based on the land use plan is a reflection of the alignment and location of the core area distribution system. Ideally, station locations are selected to be coincident with the centers of trip density but also to be consistent with minimum station spacing requirements, geometrical constraints on route alignments, and site availability.

The Pittsburgh central business district, for example, illustrates a strong focus of employment and commercial activity. Although eighth in population (as of the 1960 census), the city ranks third in the location of the largest industrial corporations and serves as a regional hub for the tri-state area. Constrained by two rivers and unusually rugged topography, this compact area of approximately two-thirds of a square mile is bounded on the east by an elevated highway and railroad tracks. This area, known as the Golden Triangle, is the focal point of almost 30 million square feet of office space and has a daily influx of 100,000 commuters.

The Golden Triangle defines the central-city cordon and is shown in Figure 5. The proposed regional lines, designated as north, south, east, and west, are depicted as they would enter the central area. The alternative strategies for central-city distribution are considered as the distribution systems that serve trips between these four regional lines at the cordon points and the land use activity in the CBD.

The points of destination within the Pittsburgh CBD are defined by partitioning this area into zones and placing centroids within each of the zones. The centroids are then considered, for analytical purposes, as destinations. Figure 6 illustrates the selection of zones and the placing of centroids for this analysis. The zones generally are city blocks or multiple blocks in areas where streets are presently developed. In other more irregular areas, zones are selected such that available land space is covered.

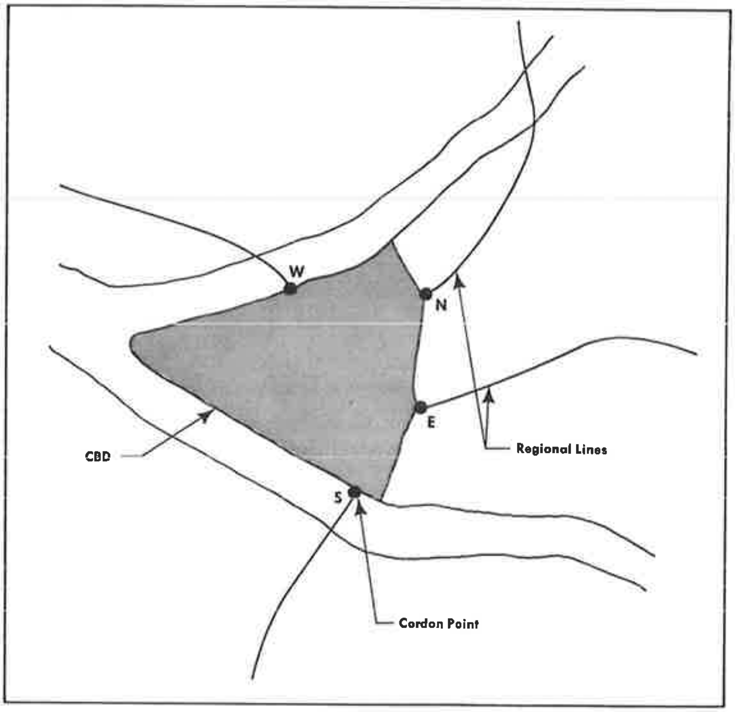



Figure 5. Pittsburgh's central business district, the Golden Triangle, and proposed regional lines.

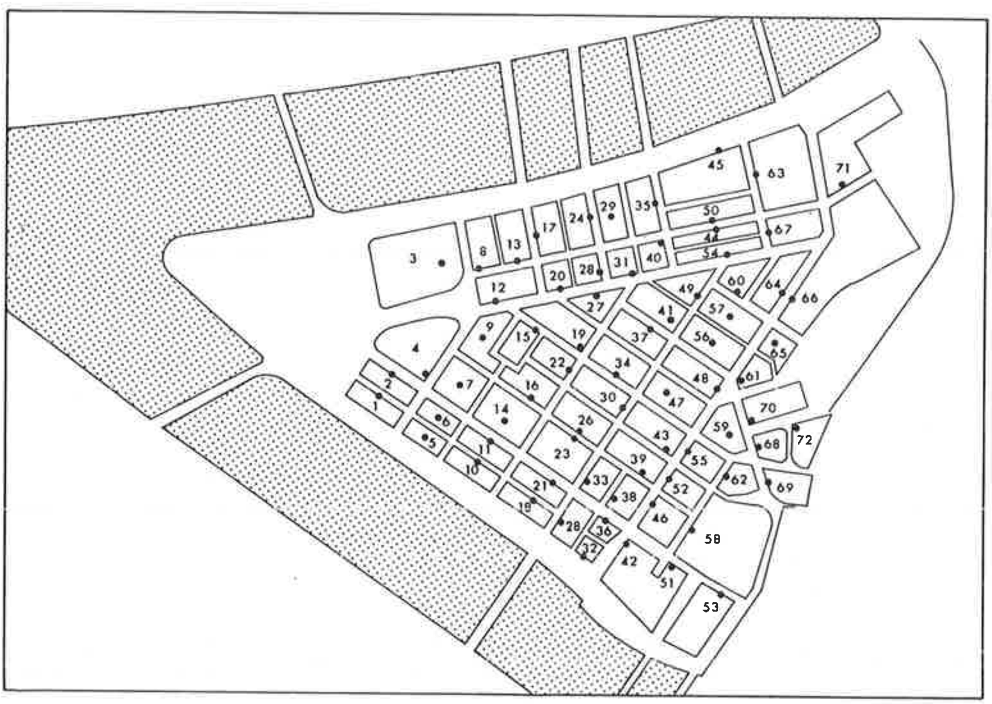



Figure 6. Analysis zones and centroids in Pittsburgh's Golden Triangle.

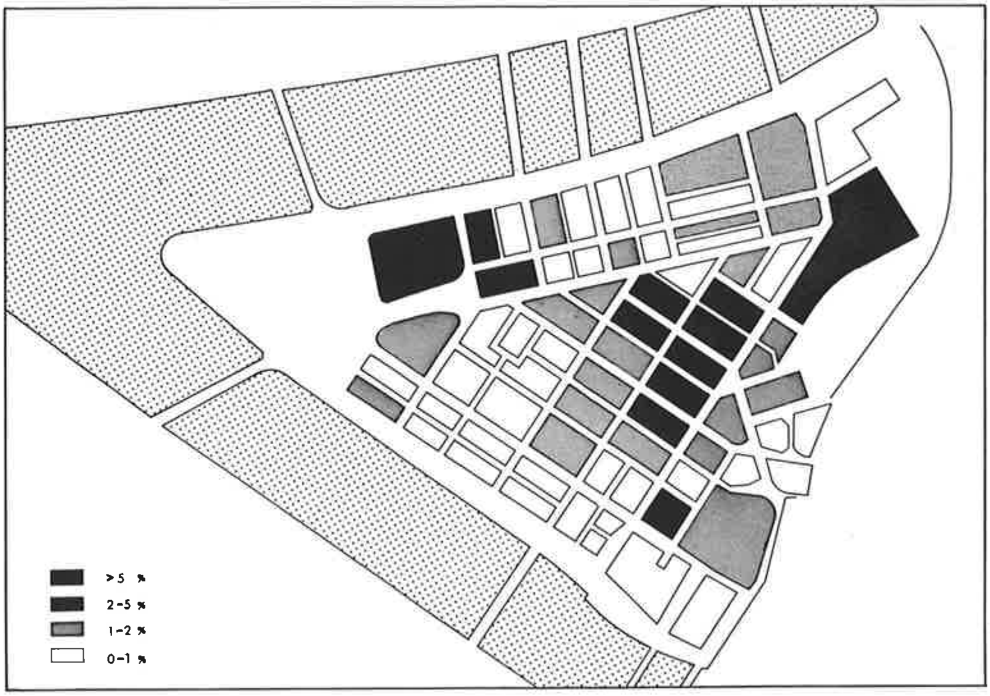



Figure 7. Density of trip destinations in Pittsburgh's Golden Triangle.

The centroids are located in logical positions in the zone depending on information available for the current and projected land use development of the area. The centroid of zone 3 illustrates a case where there are several large buildings in the zone, and the centroid is located at a common access point to these buildings. Zone 8 is a single building, and the centroid is placed at the main entrance. Placement of centroids is primarily based on present or projected knowledge of development in the core area. In many cases placement is quite realistic, and in other cases it represents an aggregation of known or projected destinations.

The employment density pattern for the Golden Triangle indicates that major concentrations of trip destinations occur along the base of the Triangle and in the vicinity of the Point. Analyses of these and similar trip-density patterns suggest several likely sites for downtown stations along the base of the Triangle and at its vertex. The distribution of destinations is illustrated in Figure 7.

## CANDIDATE DISTRIBUTOR SYSTEMS

Figure 8 illustrates six candidate distribution systems that were tested. Systems I, II, and III are single-technology systems and involve penetration of the core area by the four regional lines. Systems IV, V, and VI are small-vehicle, closed-loop aerial distributor systems, contiguous with the regional lines at one or two key transfer points.

The single-technology systems differ with respect to both the location and the number of stations. Systems I and II have three stations located in the core area but on different alignments, whereas System III is along the same alignment as System I but has five downtown stations instead of three.

The small-vehicle systems also differ with respect to the location and number of stations. System IV utilizes one major interchange point between the regional lines and the distribution system. The location of this "transportation center" is at a site that is within a 5-min walk of approximately 50 percent of all destinations. System V utilizes two major interchange points between the regional lines and the distributor system.

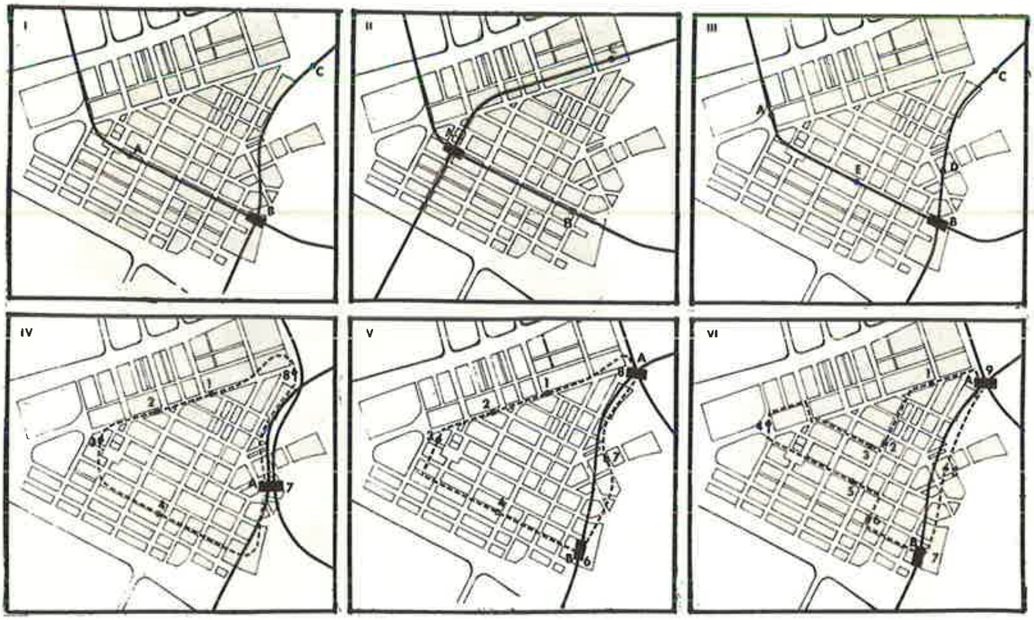


Figure 8. Candidate systems for Pittsburgh's central-city distribution.

This alternative has the advantage of reducing the volume on the most heavily used link and, consequently, requires shorter stations on the distributor system. System VI differs from System IV and V in the alignment of the distributor system. Rather than circling the Golden Triangle in a loop arrangement, the system follows a T-pattern, with the consequent advantages of improved environmental and architectural effects.

## TRAVEL-TIME ELEMENTS

The computer program developed for this analysis is capable of calculating all appropriate paths between each cordon point and each centroid. In addition to a minimum-path algorithm for computing weighted-travel times for each trip pair, the program collects relevant statistics for comparison of alternative strategies. Computational efficiency is obtained through a dynamic-programming algorithm using only a minimal number of calculations to obtain the shortest time path. For each system analyzed, execution time is less than seven seconds using an ALGOL program run on the UNIVAC 1108.

Travel times between stations were determined with the aid of train performance simulators or equations of motion that assume constant acceleration and deceleration between the station midpoints. The average wait for a transfer is considered as one-half the headway, which is 120 sec on the east-west line, 90 sec on the north-south line, and 60 sec on the small-vehicle distributor system. The dwell time at regional stations is 20 sec and 10 sec on the small-vehicle distributor system. The travel time for walking from a station to a zone centroid was determined by scaling the actual walking distance and by computing the walking time based on an average pedestrian speed of 4 ft per sec.

Time spent moving through a station, either to transfer to another vehicle or to reach the street, was determined by a careful analysis of probable station designs including escalator locations, mezzanines, and platform designs. Based on these studies, the travel times given in Tables 1 and 2 were established for movement within each station.

The proportion of travelers arriving in the downtown area on each line was established from the results of home-interview origin-destination data and travel forecasts using standard modal-split and assignment techniques. These results indicated the following

TABLE 1

STATION-TO-STREET TIMES[a]

| System | Line | Station A | Station B | Station C | Station D | Station E |
|---|---|---|---|---|---|---|
| I | E-W | 110 | 136 | | | |
| | N-S | | 103 | 122 | | |
| II | E-W | 136 | 110 | | | |
| | N-S | 103 | | 122 | | |
| III | E-W | 110 | 136 | | | 110 |
| | N-S | | 103 | 122 | 110 | |
| IV | E-W | 75 | | | | |
| | N-S | 39 | | | | |
| V | E-W | 61 | | | | |
| | N-S | 107 | 108 | | | |
| VI | E-W | 61 | | | | |
| | N-S | 107 | 108 | | | |

[a]In seconds. Station-to-street times for all small-vehicle systems are 50 sec except that at points contiguous with regional stations times are 107 sec.

TABLE 2

IN-STATION TRANSFER TIMES[a]

| System | From | To | A | B |
|---|---|---|---|---|
| I | E-W | N-S | | 107 |
| II | E-W | N-S | | 107 |
| III | E-W | N-S | | 107 |
| IV | E-W | Loop | 39 | |
| | N-S | Loop | 75 | |
| V | E-W | N-S | 126 | |
| | E-W | Loop | 82 | |
| | N-S | Loop | 146 | |
| VI | E-W | N-S | 126 | |
| | E-W | Loop | 82 | |
| | | | 146 | |

[a]In seconds. Does not include waiting time for transfer to another vehicle.

amounts from each direction: north, 38.6 percent; south, 30.8 percent; east, 14.2 percent; and west, 16.4 percent.

## RESULTS

Figure 9 illustrates the results for System I and System IV, considering both the convenience and accessibility rule. Shown in this figure are the frequency distribution of total trips times, the sensitivity between measures, and the mean and standard deviation of the total time, ride time, station time, and walk time.

The single-technology system shows little difference between travel times based on the convenience or accessibility criteria, whereas the values computed for small-vehicle systems change considerably. This result indicates that a specialized-distribution system can afford a wide range of user choice, whereas a single-technology system of few stations is quite insensitive to varying user demands. The total trip time

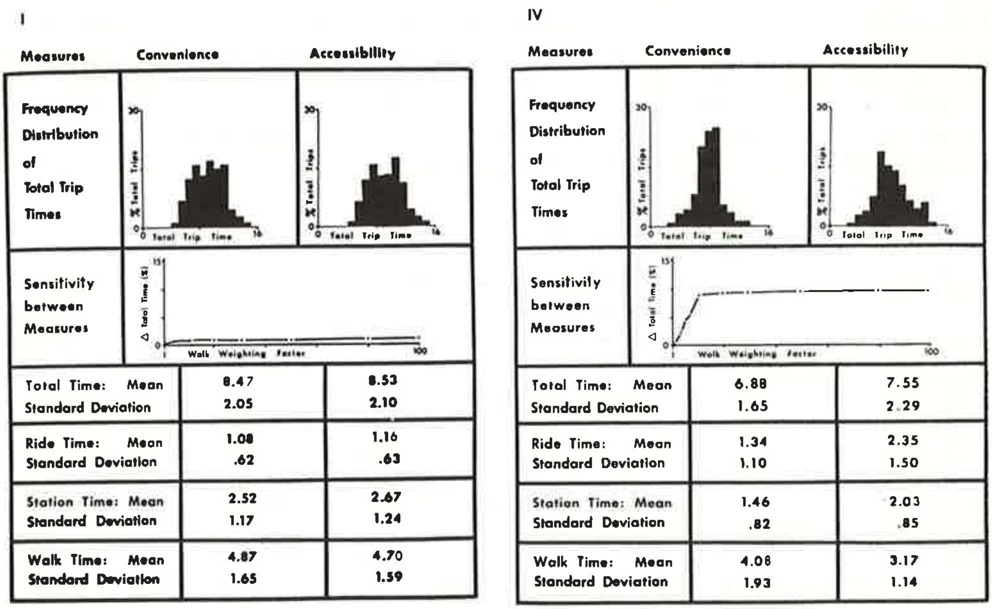
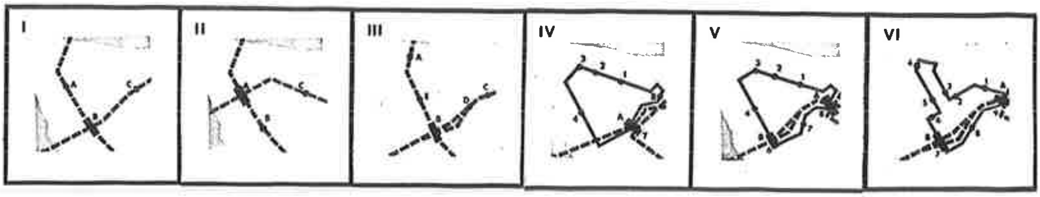



Figure 9. Comparative data: regional system and small-vehicle systems for central-city distribution.




| | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| CONVENIENCE | | | | | | |
| TOTAL | 8.47 | 7.10 | 8.07 | 6.88 | 8.10 | 8.20 |
| RIDE | 1.08 | 1.13 | 1.38 | 1.34 | 1.74 | 1.92 |
| STATION | 2.52 | 1.97 | 2.45 | 1.46 | 2.50 | 2.48 |
| WALK | 4.87 | 4.00 | 4.24 | 4.08 | 3.86 | 3.80 |
| ACCESSIBILITY | | | | | | |
| TOTAL | 8.53 | 7.15 | 8.19 | 7.55 | 9.00 | 9.53 |
| RIDE | 1.16 | 1.21 | 1.64 | 2.35 | 3.09 | 3.86 |
| STATION | 2.67 | 2.06 | 2.75 | 2.03 | 2.82 | 3.03 |
| WALK | 4.70 | 3.88 | 3.80 | 3.17 | 3.09 | 2.64 |

Figure 10. Summary comparison of mean times (in minutes) for total trip and primary trip elements.

was 8.46 min for the single-technology distributor system and 6.88 min for the small-vehicle system. This saving of 1.58 min in average time results partly from the lower walking time and shorter time spent in moving through stations. This time saving could be reduced if station design were improved for System I or if a longer in-station time were required for System IV.

The frequency distribution of travel times indicated a wider spread for System I than for System IV. The average time for System I is 8.46 min, yet 63.9 percent of users are required to travel 9 min or longer; the comparable figure is 37.7 percent on System IV.

Figure 10 shows a comparison of total, ride, station, and walk times based on the convenience and accessibility rules.

The percent of users that are required to make a transfer, under the convenience and accessibility rule, was determined to range between approximately 21 and 33 percent for the single-technology system and between 42 and 88 percent for the small-vehicle system, depending on the route-choice rule and the particular system configuration. Station usage was also determined. For example, when the convenience rule was used approximately 58.4 percent debarked at the transportation center of System IV and 46 percent at the two regional stations of System V and VI.

## SUMMARY

This paper has discussed alternative strategies for central-city distribution and their evaluation based on total travel time as the measure of effectiveness. Other relevant parameters useful in design and evaluation, such as percentage of transfers and station usage, were also developed.

Single-technology systems show little difference in travel time for either criterion, whereas values computed for small-vehicle systems change considerably. Total trip time for the single-technology system was 8.46 minutes, with 63.9 percent traveling 9 minutes or longer, and 6.88 minutes for the small-vehicle system, with 37.7 percent traveling 9 minutes or longer. However, only 21 to 33 percent must transfer on the single-technology system, whereas from 42 to 88 percent transfer on the small-vehicle system.

The final choice of either type system depends on a variety of economic, political, and environmental factors. From a level-of-service point of view, it would appear that small-vehicle systems for central city distribution present a viable alternative for system design. Aside from their inherent flexibility, these systems, which should be considered as adjuncts to larger networks and thus free of charge to users, have other advantages, including circulation for shoppers and workers throughout the central city at all times and at frequent intervals, and service to other regional bus lines, which are relieved of the necessity of circulating through the central district to distribute passengers.

The analysis presented was based upon results from a single city and a limited set of alternatives. Further study is required together with carefully designed test programs in order to develop improved circulation systems for areas of high concentration.

## ACKNOWLEDGMENTS

The data for the central city analysis described in this paper were based on forecasts prepared by the Pittsburgh Area Transportation Study, and modal-split demand estimates were developed by the consulting firm of Parsons, Brinckerhoff, Quade and Douglas as part of a study of regional rapid transit for Allegheny County, Pennsylvania. The author also wishes to acknowledge the contribution of Frank DiCesare and James W. Hoag to the development of the model used in the analysis. Illustrations were prepared by Robert K. Moorhead.

## REFERENCES

1. Meyer, J. R., Kain, J. F., and Wohl, M. The Urban Transportation Problem. Harvard Univ. Press, Cambridge, Mass., 1966, Ch. 11.
2. Future Urban Transportation Systems: Impacts on Urban Life and Form. Stanford Research Institute. Final Report II, Mar. 1968, p. 38.
3. Richards, B. Minisystems in the City. Architectural Forum, Jan.-Feb. 1968, p. 99.
4. Netsch, W. Center City Aerial Distribution for Pittsburgh. Jour. of the Franklin Institute, Vol. 286, No. 5, Nov. 1968.
5. Allegheny County Rapid Transit Study. Parsons, Brinckerhoff, Quade and Douglas. Dec. 1967.

# Toward an Evaluation of Subarea Transportation Systems

SALVATORE J. BELLOMO and STEVEN C. PROVOST,
Alan M. Voorhees and Associates, Inc.

In the traditional approach to transportation plan development, subarea plans are prepared after the regional plan has been completed. Often this approach results in subarea alternatives that are in direct conflict with the regional system elements. It is the viewpoint of the authors that subarea planning, which is oriented to short-range decisions, must be carried out as an integral part of regional planning. This will allow subarea and regional goals to be achieved concurrently, thereby permitting decisions to be made on projects that are useful to the local community and the region.

The Bethesda-Chevy Chase area in Maryland is used to illustrate the approach and the measures for the evaluation of alternative multimode transportation systems for well-established and intensely developed subareas within a metropolitan region. Of concern to the subarea are measures that reflect community impact (both positive and negative), land service, level of transport service, and economic considerations. The number of structures displaced and system costs were evaluated as were also (a) the impact of transit stations on developable land and the service they afford to the transit users; (b) the accessibility that the transit system affords the residential interests through reduction of through traffic, access to major employment concentrations, and access of emergency service vehicles; and (c) the accessibility of major employment concentrations to the areas that these concentrations attempt to serve. This paper also discusses the difficulty of using a weighted index as an input to the decision-making process.

•ONE OF THE most challenging problems faced by planners and decision-makers is the development of a comprehensive plan for subareas within a metropolitan region. These subareas can be many types. At one scale are counties that may comprise 20 percent of the regional land area. At another scale are new towns, such as Columbia and Reston; and at even a smaller scale are activity centers such as Fort Lincoln in Washington, D.C.

One common problem faced for each of these areas is how to relate the subarea under study to various elements of regional systems such as sewers, water, transportation, and public services. The approach taken in a particular case depends on the type of subarea, the regional system elements that will have a major impact on the planning, and the specific community objectives to be achieved.

This paper is concerned with the evaluation of alternative transportation systems for one type of subarea—an area located between a center city and its rapidly growing outer suburbs. Often the type of subarea under consideration contains long-established communities that have become engulfed by the expansion of the metropolitan region. They tend to have little vacant land so that the continued growth pressures can

Paper sponsored by Committee on Transportation System Evaluation and presented at the 48th Annual Meeting.

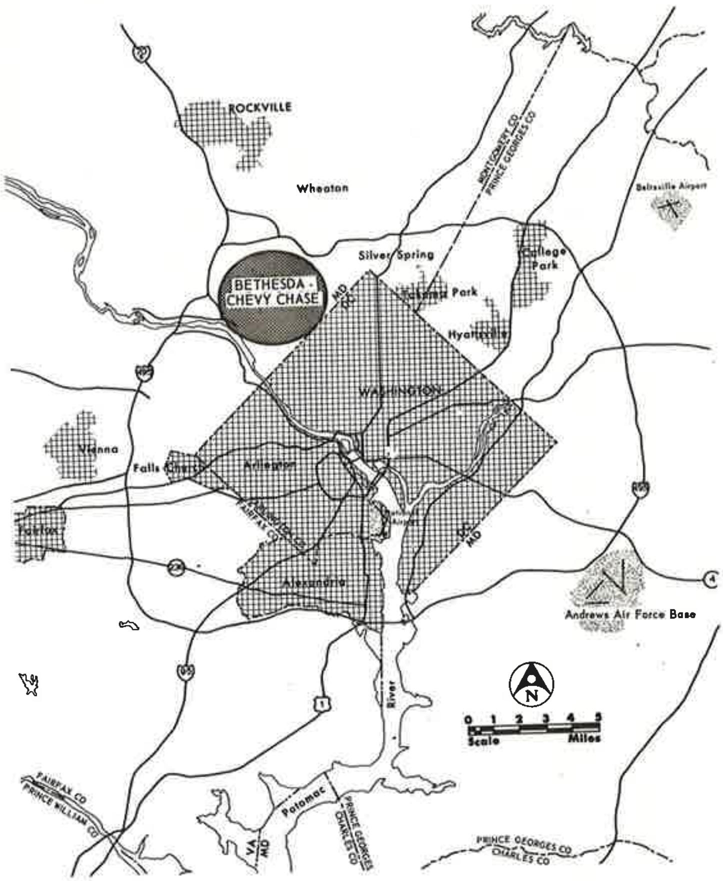



Figure 1. Location of Bethesda-Chevy Chase subarea.

only be satisfied by increasing the intensity of the existing development. In addition, many of the regional system elements, such as hospitals, colleges, water mains, and major highways, have been established for some time.

A typical subarea of this type is the Bethesda-Chevy Chase area in Montgomery County, Maryland, located north of Washington, D.C., between I-495 (Washington Beltway) and the District (Fig. 1). Table 1 summarizes the growth in land activity forecast for this subarea.

Due to the development constraints imposed by the lack of vacant land, these subareas will grow much slower than the region and may generate little need for additional capacity in the regional system elements. However, as the surrounding region grows, additional demands may be placed on the regional system elements located within the subarea. This is most true of the transportation system where the residential growth in the other suburbs and the increase in center-city jobs create additional travel

TABLE 1

SUMMARY OF LAND ACTIVITY GROWTH FOR THE BETHESDA-CHEVY CHASE SUBAREA

| Area | 1966 | 1975 | 1975/1966 (percent) | 1990 | 1990/1966 (percent) |
|---|---|---|---|---|---|
| | | | Population | | |
| Bethesda | 9,400 | 10,500 | 112 | 11,600 | 123 |
| Chevy Chase Lake | 4,100 | 6,000 | 146 | 12,000 | 293 |
| Chevy Chase-Friendship Heights | 7,400 | 9,500 | 128 | 11,200 | 138 |
| Remainder of subarea | 70,800 | 82,000 | 116 | 95,200 | 134 |
| Total | 91,700 | 108,000 | 118 | 130,000 | 142 |
| | | | Employment | | |
| Bethesda | 11,600 | 14,000 | 121 | 19,500 | 168 |
| Chevy Chase Lake | 200 | 1,200 | 600 | 1,500 | 750 |
| Chevy Chase-Friendship Heights | 3,900 | 5,400 | 138 | 8,000 | 205 |
| Remainder of subarea | 22,600 | 29,400 | 130 | 31,000 | 137 |
| Total | 36,300 | 50,000 | 130 | 60,000 | 157 |

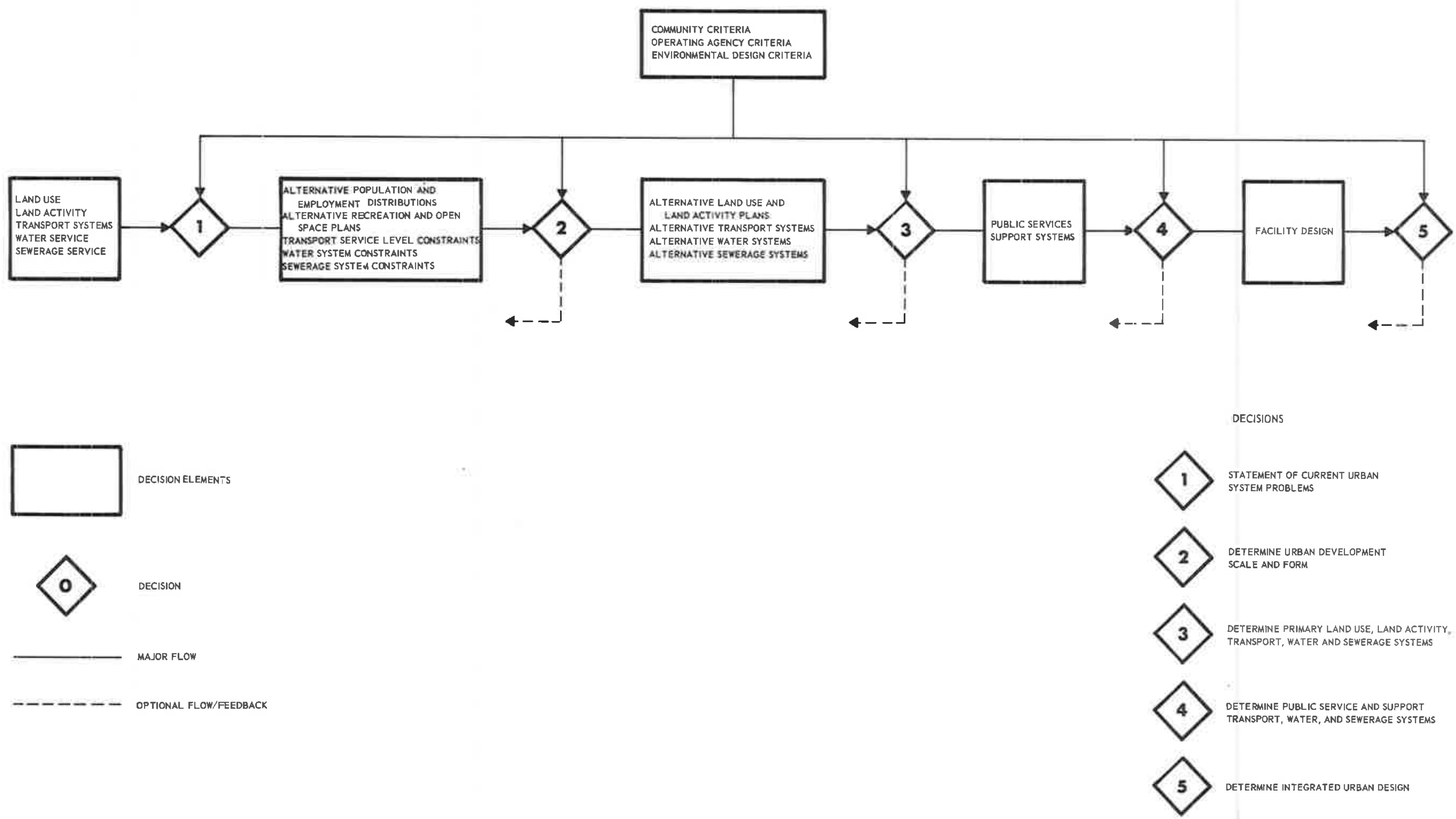



**Figure 2. Information flow for major policy decisions.**

within major corridors. To meet the increased regional demand requires that additional capacity be added to the regional transportation system.

Providing added transportation system capacity in these areas, whether it be through the addition of new major highways or transit lines, represents a complex task, the successful completion of which depends on the decision-makers' full understanding of the implications of each alternative to the region and subarea. This understanding is usually satisfied through the provision of information at critical points in the planning process. At these points, performance criteria and related measures from the viewpoints of the community and the operating agency should be considered. Figure 2 shows in a general fashion this process for very broad decisions on urban form to detailed decisions on integrated urban design. The approach presented in this paper considers some of the informational requirements for transportation decisions (3, 4). The authors feel that the approach and evaluation measures presented are steps toward obtaining an understanding between decision-makers and planners regarding the subarea implications of various transportation alternatives.

## APPROACH PHILOSOPHY

Implementation of a future regional transportation system depends on satisfying, to the greatest extent possible, both regional and subarea goals. In the studies for system evaluation, the decision-maker should be shown how well each alternative achieves these goals. Subarea goals usually relate to the achievement of aggregate goals of the property owners while the regional goals are usually oriented toward broader development patterns and operations of the regional systems, such as transportation, water, sewer, recreation, and open space.

A basic premise of the recommended approach is that subarea planning must be carried out as an integral part of the regional system planning if subarea goals are to be effectively considered. In this way, the local planning, which is concerned with the detailing of physical and service facilities, can interact with and help shape the regional system within the subarea.

Figure 3 illustrates the major steps and decision points in the subarea transportation system evaluation. Each of the steps requires interaction between the region and subarea before a major decision point is approached. The five major points at which there should be agreement by both policy and technical decision-makers at the regional and subarea level are, in sequence, (a) subarea land use and land activity forecasts, (b) corridor location for major highway system components, (c) corridor location for transit lines, (d) location of transit stations, and (e) locations of primary and secondary highways.

To be able to make these technical decisions requires an understanding of subarea and regional implications for the following: (a) differences in scale of economic development, (b) expansion of the capacity of the existing arterial street system, (c) alternative interchange locations, (d) alternative station locations and functions (automobile-oriented vs pedestrian-oriented), and (e) alternative central area parking arrangements.

Each of these items is viewed differently by the regional operating agency and the community. Therefore, analyses must be carried out so that both viewpoints are examined. It is interesting to note that in many European countries, presentation of information at public hearings includes integrated urban design alternatives that highlight the pros and cons from each of the major viewpoints discussed. This differs from the typical U.S. approach of separate hearings for each element of the plan.

The next section describes the application of the integrated plan evaluation approach for the Bethesda-Chevy Chase subarea. In this case example, local viewpoint evaluation measures and criteria were developed to reflect community impact, land service, transport service, and economic cost. Community impact factors are concerned with measures and criteria that reflect the effects of the transport system on the physical and social environment, such as existing parks, vacant land, developed areas, and various community programs and services such as police, fire, and schools. Measures considered under land service illustrate how the various alternatives promote

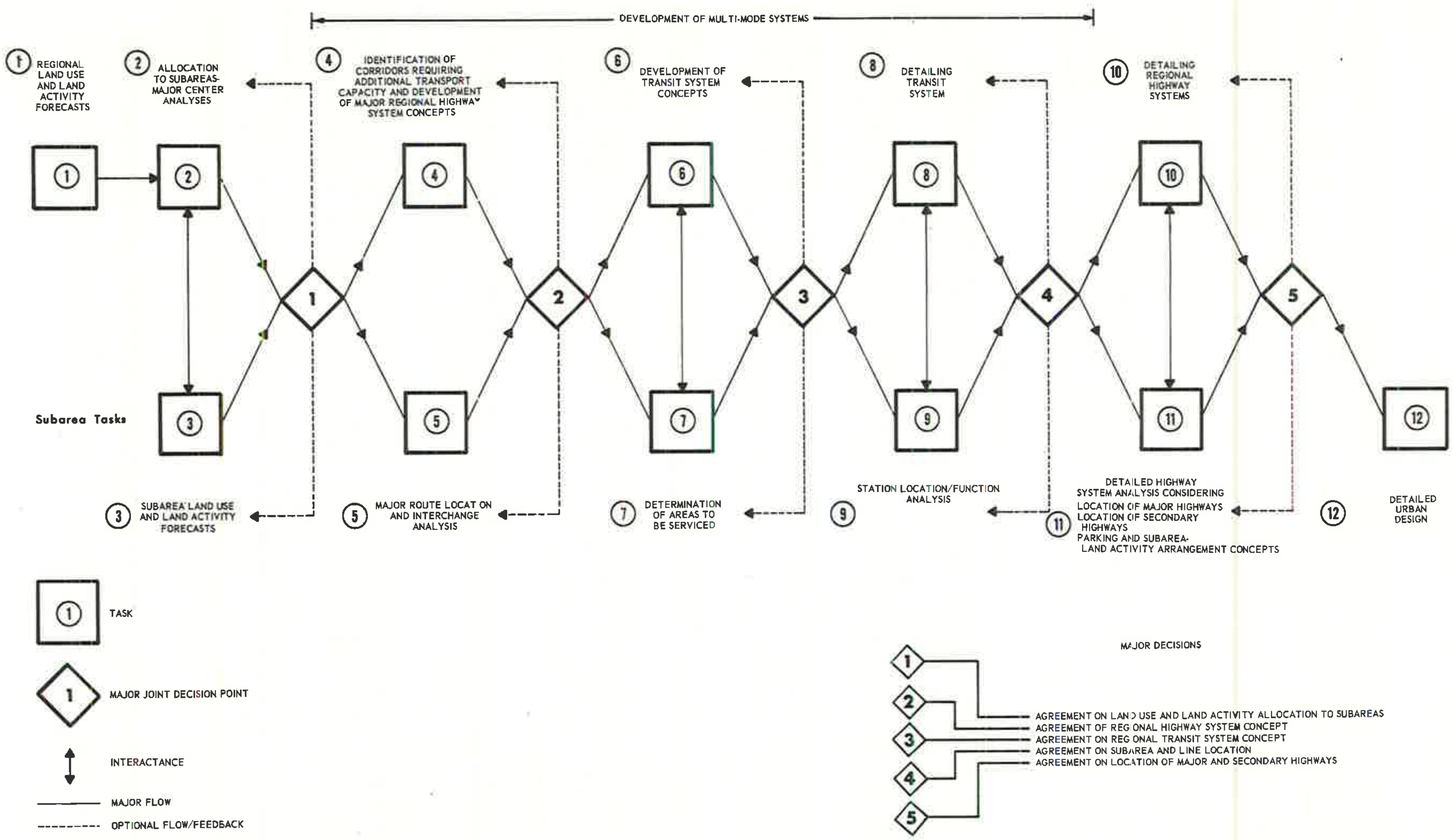



**Figure 3. Transportation decision-making process for subareas.**

the development and implementation of a particular land development concept. The measures that a community considers with respect to the level of transport service usually are concerned with accessibility and system efficiency. The economic measures, on the other hand, deal with such factors as construction, operating, and user costs within the subarea or financial programs that have been well documented (9, 12).

## BETHESDA-CHEVY CHASE EVALUATION

The comprehensive planning analysis undertaken for the Bethesda-Chevy Chase area was conducted in parallel with the evaluation of alternative regional rail systems by the Washington Metropolitan Area Transit Authority. Several regional agencies disagree with the decision that has been reached in the District of Columbia on the regional freeway system; however, results of the subarea analyses provided useful information for certain major decisions concerning the location of the rail rapid transit stations within the northwest corridor of the metropolitan region and the Bethesda subarea.

Figure 3 does not indicate the many technical decisions that were required at both the regional and subarea level, but it does show the sequence of major decisions that were made during the study. The location of the transit stations (and consequently the corridor line) has been agreed on by both the region and subarea. The study has advanced to the point of preparing alternative location plans for the major and secondary highway system.

### Data Sources

The 1990 morning peak-hour automobile and transit traffic estimates generated by the regional study were used to analyze transporation needs, costs, and priorities. The methodology followed to obtain the information at a more detailed level for subarea planning is summarized in the Appendix (3, 4, 5). With the completion of this initial data breakdown, it was possible to identify the future transportation problems that will exist at the subarea level and to evaluate the various alternatives considering the viewpoint of the community as well as regional system implications.

### Identification of the Problem

As a pointof beginning, the future (1990) transportation problems within the subarea were determined by analyzing the key points of capacity deficiency. The results were illustrated by a flow map showing the location and magnitude of the morning peak-hour deficiencies. This was done on a short-range (1975) as well as on a long-range (1990) basis. The nature of the traffic at the congested points was then determined so that alternative solutions could be developed. Figure 4, an example of one of the steps undertaken, shows the distribution of 1990 morning peak-hour trips that pass through certain selected links on Wisconsin and Connecticut Avenues within the study area. The significant characteristics of these trips are summarized in Table 2.

The results of these analyses precipitated interaction with the region to determine whether additional capacity would be available on the assumed rail or highway system. Also, the population and employment allocations were reevaluated, and a determination made of the additional capacity obtainable on the arterial system from operations techniques such as reversible lanes, progressive signal timing, and intersection widenings.

The interactions with the regional system planners resulted in a revision of the estimated rail patronage caused by diverting a percentage of the longer work trips (greater than 30 min) to the rail system. (It is not the purpose of this paper to present a documentation of the procedure developed to divert highway traffic to the transit system; the approach used was based on experience gained in other large urban area transit studies. Inasmuch as the rail transit system does not exist in Washington, there was, of course, no way to "calibrate" the diversion procedures to the local system.) The results of this diversion (shown in Table 3) had a significant impact on the transit passenger volume handled at certain stations in lower Montgomery County. However,

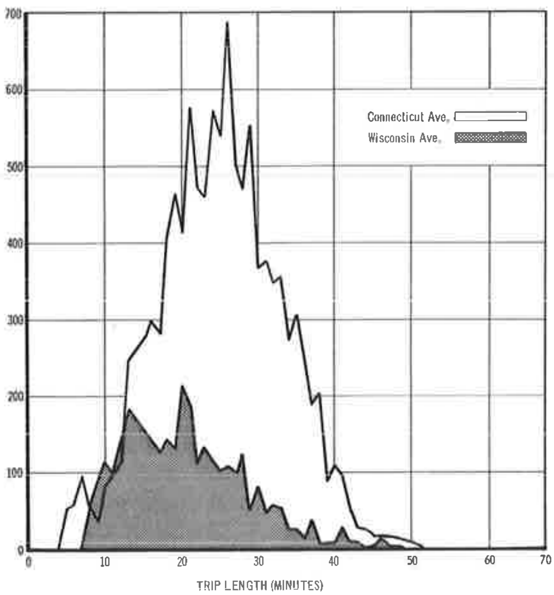



Figure 4. Trip-length distribution on selected links in the Bethesda-Chevy Chase subarea.

TABLE 2

SUMMARY OF 1990 MORNING PEAK-HOUR TRIPS ON SELECTED LINKS

| Trip Characteristics | Wisconsin Avenue | Connecticut Avenue |
|---|---|---|
| Number of trips | 3,400 | 11,000 |
| Vehicle-miles | | |
| Arterials | 24,800 | 107,200 |
| Freeways | 8,700 | 55,000 |
| Total | 33,500 | 162,200 |
| Vehicle-hours | | |
| Arterials | 900 | 3,500 |
| Freeways | 200 | 1,200 |
| Total | 1,100 | 4,700 |
| Avg. trip length (min) | 17.7 | 25.3 |

it still left unused over 50 percent of the transit line capacity within the subarea. With the revised patronage estimate and an understanding of the additional capacity available in the existing system, it was then possible to formulate and evaluate alternative ways of reducing the remaining deficiencies.

## The Alternatives Studied

The determination of alternative station locations and the development of alternative highway networks evolved after discussions with regional and local decision-makers. For the most part, the alternatives were developed in series from preliminary evaluations of preceding systems. In this way, each succeeding alternative would bring a consensus on a system closer to hand. In some cases, policy-makers deviated from a testing of series alternatives and requested that certain highway system links be incorporated into the alternative for testing purposes. The reason for these departures was to ascertain transportation impacts so that strategies could be developed for links that were critical to the subarea plan. Certain of these departures took the form of short-range (1975) traffic assignments and analyses. The key transit delineations tested are shown in Figure 5; the major highway alternatives are illustrated in Figure 6.

TABLE 3

DIVERSON OF AUTOMOBILE TRIPS TO RAPID TRANSIT BY MODE OF ARRIVAL IN 1990, WISCONSIN AVENUE LINE

| Station | A (Without Diversion) | | | | $A_1$ (Includes Diversion) | | | |
|---|---|---|---|---|---|---|---|---|
| | Total Persons | Walk | Bus | Auto-mobile | Total[a] Persons | Walk | Bus | Auto-mobile |
| Rockville | 920 | 230 | 456 | 234 | 1,960 | 230 | 456 | 1,274 |
| Halpine Road | 1,540 | 200 | 714 | 626 | 2,139 | 200 | 714 | 1,225 |
| Outer Beltway | – | – | – | – | – | – | – | – |
| Nicholson Lane | 1,140 | 133 | 480 | 527 | 1,423 | 133 | 480 | 810 |
| Parkside | 480 | 90 | 167 | 223 | 665 | 90 | 167 | 408 |
| Grosvenor Lane | – | – | – | – | – | – | – | – |
| Pooks Hill | 1,040 | 130 | 740 | 170 | 1,060 | 130 | 740 | 190 |
| Nat'l. Inst. of Health | – | – | – | – | – | – | – | – |
| Bethesda | 3,950 | 572 | 3,108 | 270 | 3,950 | 572 | 3,108 | 270 |
| Friendship Heights | 1,340 | 412 | 883 | 45 | 1,340 | 412 | 883 | 45 |
| Total | 10,410 | 1,767 | 6,548 | 2,095 | 12,537 | 1,767 | 6,548 | 4,222 |

[a]This diversion includes approximately 2,000 person trips diverted from the congested Connecticut Avenue corridor.

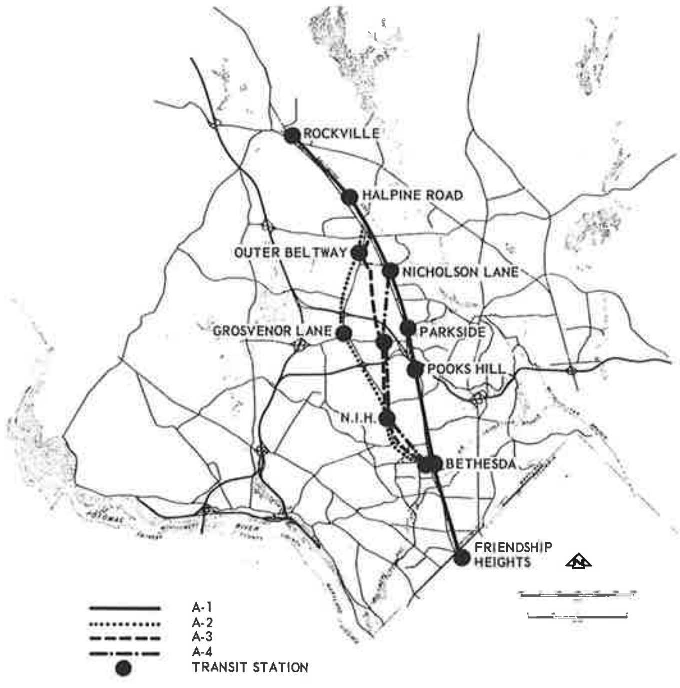



Figure 5. Alternative transit lines tested for Bethesda-Chevy Chase subarea.

## Measures and Criteria Considered

To evaluate the alternative rapid transit station locations and highway alternatives, measures reflecting community impact, land service, transport service levels, and economic costs were developed.

Community Impact—The need for a methodology that will consider the impact of transportation systems on the communities through which they pass or terminate are obvious to those concerned with the planning of urban transportation systems. Numerous freeway controversies throughout the country and in this planning area have developed with regard to the location and design of major transportation facilities. The viewpoint of this community regarding new transportation facilities was expressed at various meetings that were held with key citizen groups. What evolved from these meetings was concern over the following: (a) the number of residential homes and businesses that would be displaced by a transportation facility; (b) the land area that would be required for transportation use that would reduce the tax base; (c) the heavy through traffic on commercial distributor roads in the existing business district; (d) the transportation system by-products, such as pollution and noise; (e) the flexibility of the short-range transportation system to expand to higher capacity levels or to be integrated into alternative future regional freeway system configurations; (f) access of key generators to the system in the peak-hour; and (g) the maintenance of basic neighborhood integrity.

The first two measures, building displacement and land area for transportation use, were obtained from aerial photos and land area measures. The percentage of through traffic removed from existing streets in the business district was derived from analyses of peak-hour traffic flows, deficiencies, and selected link assignments.

Transportation by-products (pollution and noise) were difficult to quantify because little has been done to model the production of noise or pollution by various facility and

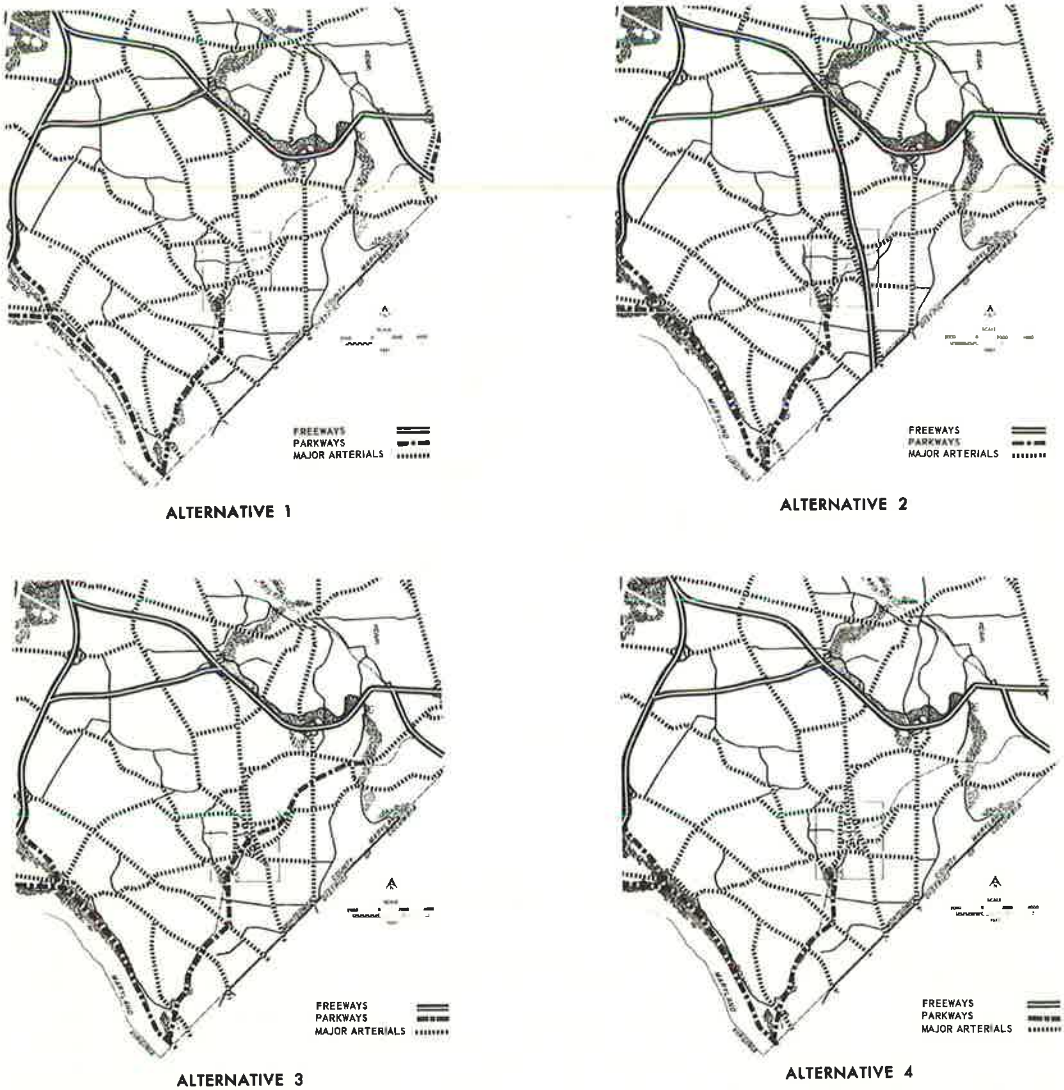



Figure 6. Alternative highway systems tested for Bethesda-Chevy Chase subarea.

traffic combinations. Research conducted by Michigan State University (6) has shown, however, that the average fuel economy for passenger vehicles is related to the number of stops and the average speed for various types of urban highways as indicated in Table 4. Therefore, the alternative highway system with the lowest average speed and greatest number of stops was considered to emanate the greatest amount of pollution. Each alternative was, therefore, qualitatively ranked by these two measures.

Research conducted by the California Division of Highways (7) indicates that, for a given volume of traffic, the noise level in the vicinity of the facility varies with facility design. In particular, depressed facilities generate less noise than do at-grade or elevated facilities. Inasmuch as certain alternatives contained depressed facilities, this criterion was used to qualitatively rank the alternatives.

The flexibility of the new facilities in the system to expand to higher capacity levels was also considered by the community. This was brought about by the community's concern for delays in the construction schedule or uncertainties in the long-range (1990)

TABLE 4

AVERAGE FUEL ECONOMY OF PASSENGER VEHICLES ON MAJOR URBAN HIGHWAYS

| Type of Urban Highway | Average Fuel Economy (mpg) | Average Speed (mph) |
|---|---|---|
| Freeway | 17.4 | 46.0 |
| Nonsignalized urban arterial | 20.0 | 36.9 |
| Signalized urban arterial with median | | |
| 1 or 2 signals per mile | 18.7 | 30.6 |
| 3 or more signals per mile | 16.1 | 25.0 |
| Signalized urban arterial without median | | |
| 1 or 2 signals per mile | 16.6 | 26.1 |
| 3 or more signals per mile | 16.1 | 23.0 |
| Signalized downtown arterial | 9.1 | 9.5 |

Source: Highway Traffic Safety Center, Michigan State Univ., East Lansing, 1957-1958.

TABLE 5

ACCESSIBILITY OF MAJOR EMPLOYMENT CENTERS TO FUTURE POPULATION

| Major Employment Center | Mean Opportunity Times (min) for Highway Alternatives | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Bethesda | 6.17 | 6.34 | 6.33 | 5.85 |
| Chevy Chase Lake | 8.10 | 8.10 | 8.05 | 7.19 |
| Friendship Heights | 6.51 | 8.00 | 7.65 | 7.50 |
| River Road | 7.52 | 7.52 | 7.52 | 6.73 |
| Army Map Service | 9.45 | 9.45 | 9.45 | 8.06 |
| Total study area | 6.51 | 6.58 | 6.52 | 6.28 |

traffic forecasts that are predicted on assumed freeway networks and development scales. Therefore, a unit to express this became the miles of actual street capable of expanding to higher capacities within the right-of-way. Consideration was also given to the flexibility of the subarea system to tie into future possible alternative freeway configurations.

Land Service—The business and industrial interests in the community were concerned with the secondary effects that the systems afforded them in terms of (a) accessibility of major employment concentrations to future development; (b) availability of development opportunities around rapid transit stations; and (c) development scale possible over and above the economic forecasts. To date, there are no satisfactory measures of these effects that are generally recognized for use in subarea evaluation. Therefore, certain measures were calculated to reflect these land service objectives. The accessibility of major employment concentrations to future development was developed by analyzing the trip length distribution generated by use of the gravity model formula:

$$T_{ij} = \frac{E_i (P_j + E_j) t_{ij}}{\Sigma (P_j + E_j) t_{ij}} \qquad (1)$$

where

$T_{ij}$ = interaction between zone i and j,
$E_i$ = employment forecast for zone i,
$P_j$ = population forecast for zone j,
$E_j$ = employment forecast for zone j, and
$t_{ij}$ = future morning peak-hour travel time from zone i to j.

A morning peak-hour, skim-tree matrix reflecting each highway alternative and its congestion was developed, and the average trip length from the major employment concentrations was generated. Table 5 indicates how the mean trip times generated from this distribution (opportunity trip lengths) varied for the major employment concentration for each of the highway alternatives. Alternative 4 was the most accessible of those examined because it had the lowest mean opportunity trip length.

Figure 7 indicates the cumulative distribution of opportunities around the business district for each highway alternative. There were three times as many opportunities within 4 minutes with alternative 4 than there were with the other alternatives. Though seemingly theoretical, the measures provided not only an evaluation metric but also a tool to generate alternative highway networks to provide better access to major employment concentrations—a goal of the study.

The ability of the transit station to generate development opportunities has been observed in major cities such as Toronto, Canada, that contain rail rapid transit systems. The criterion for the placement of the rapid transit station within the planning area was postulated to be related to the amount of developable land that would be adjacent to a new station. There were several sites where these new stations could be placed to service the patronage that would be generated. A study by the land-use planners was conducted to identify vacant land and areas containing one-story structures that could be

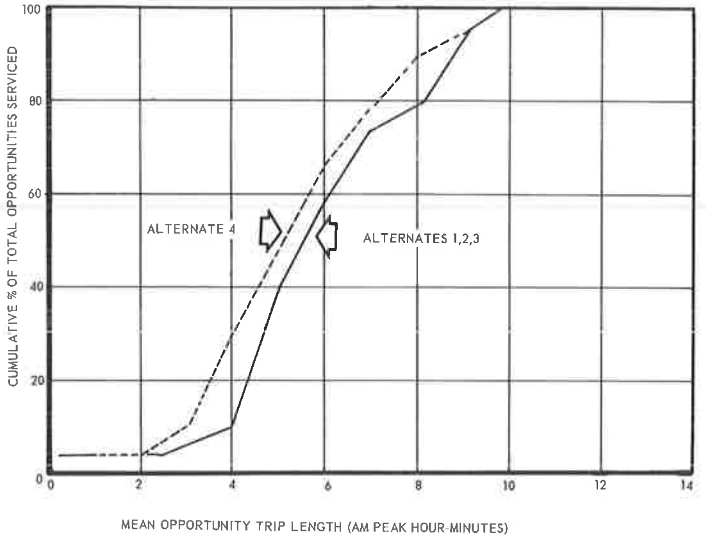



Figure 7. Accessibility of the business center to population and other employment.

economically replaced. Each transit line (Fig. 5) was then evaluated with respect to the cumulative percentage of developable land that surrounded it (Fig. 8). Of particular concern was the amount of developable land within 1,500 to 2,000 ft of the station. Alternative transit lines A2 and A3 had the maximum potential in this regard.

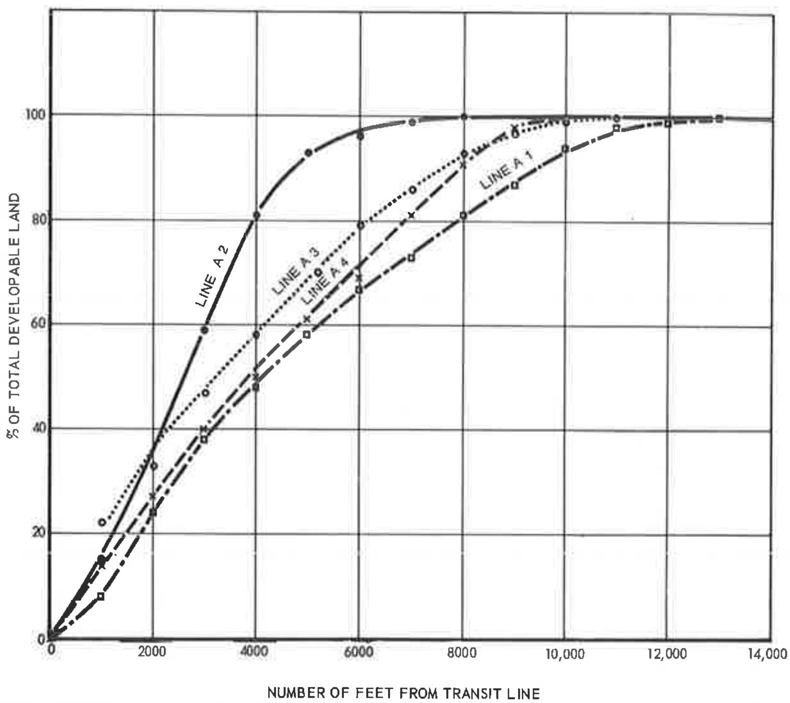



Figure 8. Percentage of developable land in proximity to rapid transit stations in Bethesda-Chevy Chase subarea.

A third metric to reflect the ability of the transport system to service land was its capacity to contain a larger scale of development activity because of additional road capacity. This development scale was estimated by formatting the peak-hour trip table to a major center, calculating the through traffic, and estimating the amount of additional traffic that could be carried in the peak hour. This additional traffic was then converted (through estimates of automobile occupancy, directional split, and percentage by transit) into economic development that could be accommodated in the area.

Transport Service Levels—In addition to concern for standard measures of future traffic congestion (such as vehicle-miles of peak-hour capacity deficiency on the highway network, highway peak-hour travel times from key traffic generators, and plots of congestion surrounding rapid transit stations), there was also concern for mobility levels and access opportunities. The measures to reflect these were (a) accessibility of the subarea population to job opportunities in the planning area; (b) accessibility of emergency vehicles to population and employment; (c) person-miles of travel to and from the rapid transit station; (d) miles of streets carrying transit-station traffic; (e) average trip length of person trips to the rapid transit stations; and (f) uniform distribution of person arrivals at any one station on a given transit line.

The accessibility of the subarea population to job opportunities within the planning area was generated by means of the opportunity distribution described previously. In this case, however, the population in the zone was used as the production index, and the attraction was the total employment within the zones. No discernible differences were found for this measure. Because of the high median family income and mobility within the planning area, this issue was not of major concern. However, the metric may prove useful in studying low-income areas where these concerns are more pronounced.

The access that the transportation system affords emergency vehicles such as fire, police, and ambulances is of major concern to the community that will be serviced by such vehicles. Peak-hour congestion can limit severely the ability of these vehicles to respond to an emergency situation. The comprehensive planning study pinpointed the locations where these major services would be generated. By means of the gravity model distribution formula, the accessibility that these service centers provided the subarea population in the peak hours was developed using the following distribution formulas for police and fire services:

Police

$$T_{ij}^{P} = \frac{P_i\,(P_j + E_j)\,t_{ij}}{\Sigma\,(P_j + E_j)\,t_{ij}} \qquad (2)$$

Fire

$$T_{ij}^{F} = \frac{F_i\,(P_j + E_j)\,t_{ij}}{\Sigma\,(P_j + E_j)\,t_{ij}} \qquad (3)$$

where

$T_{ij}^{P}$ = interchange of police service from zone i to zone j,

$P_i$ = population in zone i (police cars are dispatched to cover certain areas; number of cars dispatched was assumed to be related to the zonal population),

$P_j$ = population in the attraction zone,

$E_j$ = employment in the attraction zone,

$t_{ij}$ = zone i to zone j peak-hour travel time,

$T_{ij}^{F}$ = interchange of fire services from zone i to zone j, and

$F_i$ = number of firemen available in the firehouse of a given zone.

The running of these distributions permitted a comparison of the cumulative travel time required for such services for each of the alternatives. These distributions are summarized in Figures 9 and 10. The comparison indicated the superiority of alter-

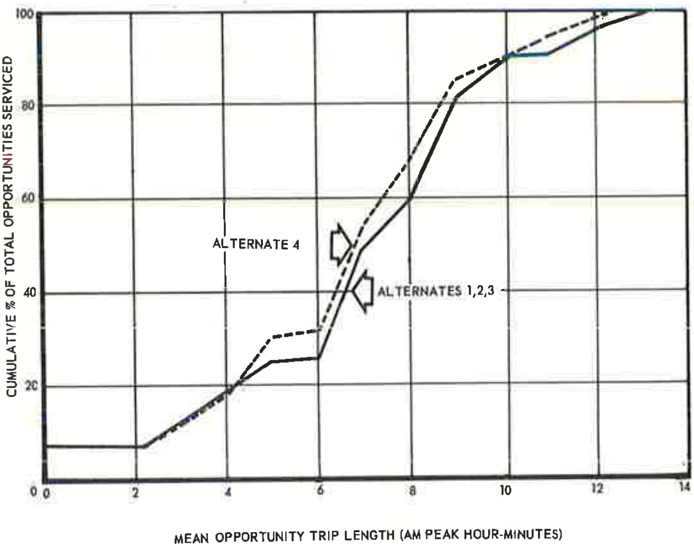



Figure 9. Accessibility of firemen to population and employment.

native 4 over the other highway alternatives. It was also found that the mean opportunity times for police and fire services were 5.5 and 6.9 minutes respectively.

The service that the transit stations provided the users of the planning area was also of concern. This service was measured by analyzing the traffic patterns of the various arrivals and departures estimated for each transit station. These patterns were determined for each transit delineation based on the patronage forecasts shown in Table 6. It should also be noted that the submodal split (bus vs automobile vs walk), which is an input to the regional study, was made after inputs of parking availability and walking generated by surrounding development were developed by the subarea studies. The regional study could not proceed with its traffic estimates until these inputs were furnished. This fortifies, in part, the necessity to conduct the regional and

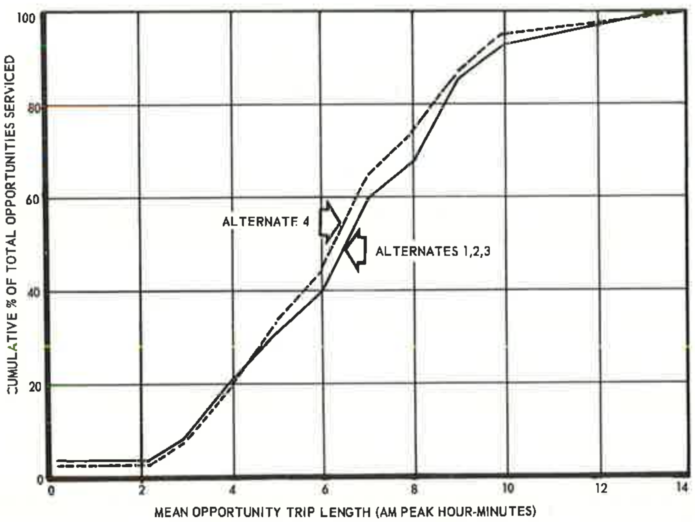



Figure 10. Accessibility of police to population and employment.

TABLE 6

MODE OF ARRIVAL AT RAPID TRANSIT STATIONS

| Station | Transit Line Alternatives | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Total Persons | Walk | Bus | Auto-mobile | Total Persons | Walk | Bus | Auto-mobile |
| | A1 | | | | A2 | | | |
| Rockville | 1,960 | 230 | 456 | 1,274 | 1,960 | 230 | 456 | 1,274 |
| Halpine Road | 2,139 | 200 | 714 | 1,225 | 1,489 | 200 | 624 | 665 |
| Outer Beltway | – | – | – | – | 2,290 | 350 | 570 | 1,370 |
| Nicholson Lane | 1,423 | 133 | 480 | 810 | – | – | – | – |
| Parkside | 665 | 90 | 167 | 408 | – | – | – | – |
| Grosvenor Lane | – | – | – | – | 995 | 420 | 167 | 408 |
| Pooks Hill | 1,060 | 130 | 740 | 190 | – | – | – | – |
| Nat'l. Inst. of Health | – | – | – | – | 1,660 | 130 | 1,280 | 250 |
| Bethesda | 3,950 | 572 | 3,108 | 270 | 3,350 | 572 | 2,568 | 210 |
| Friendship Heights | 1,340 | 412 | 883 | 45 | 1,340 | 412 | 883 | 45 |
| Total | 12,537 | 1,767 | 6,548 | 4,222 | 13,084 | 2,314 | 6,548 | 4,222 |
| | A3 | | | | A4 | | | |
| Rockville | 1,960 | 230 | 456 | 1,274 | 1,960 | 230 | 456 | 1,274 |
| Halpine Road | 1,489 | 200 | 624 | 665 | 2,139 | 200 | 714 | 1,225 |
| Outer Beltway | 2,290 | 350 | 570 | 1,370 | – | – | – | – |
| Nicholson Lane | – | – | – | – | 1,423 | 133 | 480 | 810 |
| Parkside | – | – | – | – | – | – | – | – |
| Grosvenor Lane | 1,095 | 520 | 167 | 408 | 1,095 | 520 | 167 | 408 |
| Pooks Hill | – | – | – | – | – | – | – | – |
| Nat'l. Inst. of Health | 1,660 | 130 | 1,280 | 250 | 1,660 | 130 | 1,280 | 250 |
| Bethesda | 3,350 | 572 | 2,568 | 210 | 3,350 | 572 | 2,568 | 210 |
| Friendship Heights | 1,340 | 412 | 883 | 45 | 1,340 | 412 | 412 | 45 |
| Total | 13,184 | 2,414 | 6,548 | 4,222 | 12,967 | 2,197 | 6,548 | 4,222 |

subarea studies concurrently. The person-miles, miles of street carrying transit station traffic, and average person-trip length to each station were straightforward measures after the pattern and trip generation were determined. The uniformity of station arrivals on a given line required a new measure, however. This measure was defined as

$$U = |X - \overline{X}|$$

where

$U$ = uniformity of station arrivals on a given line,

$X$ = ratio of individual station patronage to the sum of the patronage for the five stations (percentage) along the line, and

$\overline{X}$ = average station patronage (percentage).

In simple terms, this measure is a standard deviation that indicates disproportionate station arrivals. The larger the standard deviation, the more uneven would be the number of arrivals at a station and the waiting time within a station—both of which would cause additional congestion at certain over-crowded stations.

Economic Cost—This part of the evaluation consisted of a straightforward comparison of construction and user costs for both systems. User costs included the standard estimations of operating, accident, and time costs. The standard methodology and basic considerations of an economic evaluation are discussed elsewhere (15, 16). Also, the ability of the various jurisdictions (federal, state, and county) to finance the recommended improvements was considered a constraint.

## Evaluation Summary

With an understanding of the quantifiable and nonquantifiable measures that were of concern to the residents and employers of the subarea, it was then possible to summarize these measures and criteria as indicated in Table 7. This type of information display accompanied with the appropriate graphics proved to be the most effective device for the evaluation of the alternatives. The criteria required a maximization or minimization of the quantitative or qualitative rankings indicated.

TABLE 7

SUMMARY OF EVALUATION MEASURES AND CRITERIA[a]

| Measure | Dominant Mode (S) | Criteria | Transit System Alternations | | | | Highway System Alternatives | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A1 | A2 | A3 | A4 | 1 | 2 | 3 | 4 |
| Community Impact | | | | | | | | | | |
| 1. Number of structures displaced | highway/transit | minimize | | (tunnel) | | | 0 | 250 | 30 | 50 |
| 2. Land area taken for transportation use (acres) | highway/transit | minimize | 3 | 3 | 3 | 3 | 33 | 157 | 79 | 126 |
| 3. Percentage of through traffic removed from existing streets entering Bethesda CBD | highway | minimize | | | | | 0 | 1.3 | 11.5 | 31.4 |
| 4. Estimation of transportation system by-products, quality of atmosphere and sound level | highway | minimize | | | | | 3 | 3 | 2 | 1 |
| 5. Flexibility of system to expand to higher capacities (miles of street capable of expanding to higher capacities) | highway/transit | maximize | | | | | 0 | 0 | 2.4 | 2.4 |
| Land Service | | | | | | | | | | |
| 1. Accessibility of major employment centers to future population [mean opportunity times (min)] | highway | maximize | | | | | 6.17 | 6.34 | 6.33 | 5.85 |
| 2. Percentage of developable land within 2,000 feet of rapid transit stations | transit | maximize | 23 | 35 | 35 | 25 | | | | |
| 3. Scale of development possible under the capacity of a particular scheme (transit, population possible within 2,000 feet of stations; highway increase or decrease of employment within Bethesda) | highway/transit | maximize | 29,000 | 38,000 | 39,000 | 36,000 | 960 | NA | -360 | 2,000 |
| Transport Service Levels | | | | | | | | | | |
| 1. Peak-hour (a.m.) congestion (miles of capacity deficiency) | highway | minimize | | | | | 19 | 20 | 18 | 17 |
| 2. Travel time (a.m. peak hour) from major traffic generators | highway | minimize | | | | | 4 | 2 | 3 | 1 |
| 3. Congestion surrounding transit stations (with existing street system) | transit | minimize | 3 | 2 | 2 | 1 | | | | |
| 4. Accessibility of fire and police vehicles to population and employment | highway | maximize | | | | | 2 | 2 | 2 | 1 |
| 5. Person-miles to and from rapid transit stations | transit | minimize | (base) | -930 | -674 | -850 | | | | |
| 6. Miles of street carrying transit station traffic | transit | minimize | 12.4 | 10.6 | 10.6 | 10.6 | | | | |
| 7. Average trip length of automobiles to and from the station (miles) | transit | minimize | 4.5 | 3.8 | 3.8 | 3.8 | | | | |
| 8. Standard deviation of person arrivals for rapid transit stations on a given line | transit | minimize | 51.8 | 35.2 | 33.6 | 34.2 | | | | |
| Economics | | | | | | | | | | |
| 1. Cost of transit | transit | minimize | 3 | 4 | 1 | 2 | | | | |
| 2. Cost of highways | highway | minimize | | | | | 1 | 4 | 2 | 3 |
| 3. User costs | highway | minimize | | | | | NA | NA | NA | NA |

[a]For measures ranked qualitatively, 1 indicates the alternative that best satisfies criteria.

## Decision-Making Technique

Once the measures and criteria are established for the alternative subarea transportation system, the difficult task, as always, is that of making the decisions. The professional in this role usually does one of two things: (a) makes a recommendation based on a weighting or rating calculation of the measures or (b) lets the policy-makers do the deciding.

In the case of the Bethesda-Chevy Chase study, the first approach was attempted after a review of the literature (8 through 14). In this approach, an overall weighted index was used to evaluate the alternatives. This proved ineffective because the decision-makers had different weighting values and hence could not agree to or accept a common weighting scheme. Furthermore, because transportation planning is a continuous process, each alternative that is generated should contain the accepted elements of the plan that preceded it and make improvements in reducing negative community impact and in increasing land and transport service levels or economics with each succeeding alternative. Unlike regional plan alternatives, the alternative systems are not as broad because of the constraints (objectives) imposed by the community. A review of the work done by the Harvard Transport Research Program (14) indicated that usage of such a weighted index, representing the overall value of a particular alternative, "...implies a rather strict set of conditions on both the value set and the performance measures. The set of performance measures must be an exhaustive set containing all of the relevant consequences without any repetitions. It must, therefore, be mutually exclusive and collectively exhaustive."

It is further stated that "...in practice, obtaining a final objective measure may be a monumental task although ...it is conceptually straightforward." This approach sums up the difficulty in trying to obtain a decision from people who have different weighting schemes and who often consider measures that are interrelated and, therefore, are not mutually exclusive. Furthermore, because planning is an inexact science where variations in the estimates can occur, the sensitivity of the weighted index becomes of concern. The tools to evaluate the variations that can occur in the weighted index due to the variations in the input measures do not exist.

The second procedure of providing the information to the decision-maker and letting him decide is also ineffective. The planner must communicate the measures accompanied with the margin of error and an estimation of interdependency with other measures if he is to relate the pros and cons of one alternative over another.

The technique that appeared to be most workable in the Bethesda-Chevy Chase example was a conveyance of the measures with their accompanying criteria and an evaluation of how each succeeding alternative could improve on the one that was previously developed. In addition, the measures were kept uniform throughout the analysis and contain the data sets that are, for the most part, mutually exclusive. In this way the needs were stated, and the systems to meet those requirements were developed with the decision-makers in series rather than in parallel. Therefore, the evaluation technique, regardless of its technical efficacy, is understood and acted on by those whose responsibility it is to choose.

## SUMMARY

This paper has presented an approach and measures for the evaluation of alternative transportation systems for established and well-developed subareas of the metropolitan region. It is a starting point toward the determination of the factors that represent the viewpoints of the community that must be considered at key points in the technical decision-making process.

The approach for evaluating subarea plans suggests that subarea and regional planning be done concurrently, interfacing at key decision points in the process. Furthermore, the approach suggests usage of information generated by the region and subarea derived from a common data base for land use and activity scale and traffic forecasting procedures.

The measures and criteria presented relate to the local viewpoint regarding negative community impact, land service, levels of transport service, and economic considerations. Certain new measures are presented that indicate how rapid transit, highway systems, and intermodal transfer points can be measured in the interest of the community as well as the agencies that must own and operate the regional systems. The measures developed included (a) impact of transportation facilities on community objectives; (b) impact of transit stations on developable land and the service they afford to the transit users; (c) accessibility that the system affords the residential interests through the quantification of through traffic reduction, access to major employment concentrations, and access of emergency service vehicles; and (d) accessibility of major employment concentrations to the areas that they are attempting to service. The paper also highlighted the difficulties in using utility functions or weighted matrixes in effectively arriving at a transportation plan or generating new systems for testing in the evaluation process.

As a start for further work in this area, it is suggested that new research focus on (a) investigation of a simple set of mutually exclusive performance criteria that could be input to community decisions; (b) measurements of the uncertainties present in the performance criteria noted above; (c) development of new measures for the effect of noise and air pollution for each new facility or the community; and (d) development of measures and procedures to evaluate the effectiveness of testing interrelated parking and traffic control solutions to downtown core area problems.

## ACKNOWLEDGMENTS

The authors are indebted to the staff of Alan M. Voorhees and Associates, Inc., for their comments that helped shape this paper. Particular acknowledgment goes to

Grant Malinsky, who was responsible for preparing many of the computer runs and analyses for the Bethesda-Chevy Chase planning study.

## REFERENCES

1. Garrison, William L. Identifying Alternative Urban Environments. Science Engineering and the City.
2. Traffic Forecast. Alan M. Voorhees and Associates, Inc., Tech. Rept. 3, Dec. 1967.
3. Technical Procedures Employed in Traffic Forecasts for the Bethesda-Chevy Chase Planning Study. Alan M. Voorhees and Associates, Inc., Dec. 1967.
4. Transportation Projections—Lower Montgomery County (1975 and 1990). Alan M. Voorhees and Associates, Inc., Tech. Memorandum, Jan. 17, 1968.
5. Bethesda-Chevy Chase Planning Study Technical Report—A 1990 Recommended Transportation System. June 1968.
6. Voorhees, Alan M. Variables Affecting Traffic and Vehicular Operating Conditions in Urban Areas. Systems Workshop on Vehicular Contamination Control, Jan. 24-26, 1968.
7. Beaton, John L., and Bourget, Louis. Can Noise Radiation From Highways Be Reduced by Design? Highway Research Record 232, 1968, pp. 1-8.
8. Haikalis, George, and Cambell, E. Wilson. Evaluation of Urban Transportation Systems. Jour. City Planning Div., Vol. 89, No. CP-1, Sept. 1963.
9. Gendell, Davis S. Evaluation of Alternative Networks, Urban Transportation Planning Short Course. U.S. Bureau of Public Roads.
10. Jessiman, William, Brand, Daniel, Tumminia, Alfred, and Brussee, C. R. A Rational Decision-Making Technique for Transportation Planning. Highway Research Record 180, 1967, pp. 71-80.
11. Hill, Morris. A Method for the Evaluation of Transportation Plans. Highway Research Record 180, 1967, pp. 21-34.
12. Perazich, George, and Fischman, Leonard L. Methodology for Evaluating Costs and Benefits of Alternative Urban Transportation Systems. Highway Research Record 148, 1966, pp. 59-71.
13. Irwin, Neal A. Criteria for Evaluating Alternative Transportation Systems. Highway Research Record 158, 1966, pp. 9-19.
14. Committee on System Evaluation, Institute of Traffic Engineers. Unpublished research, Aug. 1968.
15. Road User Benefit Analyses for Highway Improvements. American Association of State Highway Officials, Washington, D.C., 1960.
16. Evaluation of Alternatives. Niagara Frontier Transportation Study, Albany, N.Y., Dec. 1966.

# *Appendix*

## A SUMMARY OF THE TRANSPORTATION PLANNING METHODOLOGY FOR THE BETHESDA-CHEVY CHASE SUBAREA ANALYSIS

The Bethesda-Chevy Chase subarea analysis was carried out using relevant regional system and travel data developed in studies conducted for the Washington Metropolitan Area Transit Authority (WMATA). This Appendix discusses the data used and the general procedures followed to make the data usable for the subarea analysis. A more detailed discussion of the procedures used in the regional studies is given elsewhere (2, 3).

### Traffic Zones

Of the 552 traffic zones established within the Washington metropolitan region, the 23 representing the Bethesda-Chevy Chase planning area were subdivided into 52 zones

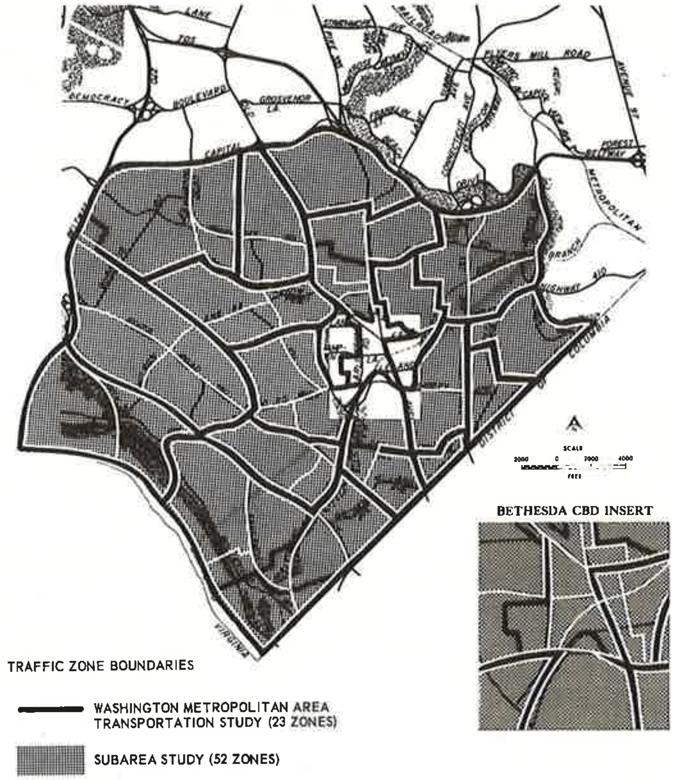



Figure 11. Traffic zones for the Bethesda-Chevy Chase subarea.

(Fig. 11). The 45 zones surrounding the area were split into 64 zones, and the remaining 484 traffic zones plus the 45 external stations were aggregated to 55 super-districts. In subsequent analyses, the zones representing the Bethesda business district were subdivided so that traffic could be assigned as necessary to parking location for alternative circulation systems in the Bethesda business district.

## Assignment Networks

The assignment networks developed during the regional study consisted primarily of the freeways and major arterials. As a starting point for the subarea analysis, additional lower order streets were added to the network consistent with the revised zone system. Outside the planning area, the regional networks were simplified by removing streets that were not oriented to intercounty movements.

## Trip Tables

The WMATA tables of morning peak-hour automobile-driver trips generated by the regional study were compressed to the zone and super-districts system described above. The morning trips (bus and automobile) to the rapid transit stations were formulated and manually analyzed as an increment over the peak-hour automobile-driver traffic. These trip tables represented the base peak-hour travel patterns used in alternative tests.

## Assignment

The assignment of the modified WMATA morning peak-hour trip table to the revised network was accomplished using minimum-path, "all-or-nothing" computer assignment techniques.

Peak-hour travel times were calculated to account for the slower speeds in the peak hour due to increased traffic congestion. Because the volume/capacity (v/c) ratio is a measure of congestion, it was calculated for each network link and used to determine the link speed. If the v/c ratio was less than 0.6, it was assumed that congestion was not severe enough to lower the off-peak operating speed. If the v/c ratio was greater than 1.5, it was assumed that extreme congestion was present and the peak speed was lowered to 5 mph. For v/c ratios between these two values, the peak speed was obtained from a series of curves relating the v/c ratio to speed.

## *Discussion*

GERALD D. MACKIN, Texas Instruments, Inc.—The authors have addressed a problem area that is becoming increasingly apparent in the urban environment: how does one make decisions when there are numerous subjective parameters, low confidence in forecasts, and many people involved, all of whom would probably react differently to different planning decisions? As pointed out in the paper, the best transit or highway system is readily determined from a pure performance and cost standpoint, but the impact on the community has been traditionally assumed nonquantifiable.

It is agreed that the decision-making process is difficult but that it must be accomplished to satisfy future demands. Therefore, the following alternatives are apparent: (a) utilization of an urban planning expert to make the decision, (b) a public relations campaign to help realize the benefits of a new system, although it will cause hardship on some, (c) a model that would quantify the unquantifiable, or (d) some combination of these.

A model could be constructed to evaluate various systems, the need for which has been determined, for whenever a decision is made, someone will either benefit or lose, including the decision-maker, and it is possible to minimize the loss. The model would have to be exceedingly flexible to factor in many variables, such as aesthetics, displacement of people and business, social, economic, and political impact, cost, performance, and utilization. The units could be dollars, or some nondimensional units of utility or value.

Although the model could be easily constructed, its inputs admittedly would require creative research and continual reiteration. The subarea problem demonstrates the importance of geographical and jurisdictional partitioning, in addition to the conventional functional (user, supplier, operator), organizational, and socioeconomic groupings.

As thorny as the subarea problem is today, it promises to get worse at an increasing rate as our urban explosion accelerates. The health of our cities, particularly the central core, depends on regional planning that appropriately and effectively integrates subarea planning, although this is not unique to transportation, and it does argue strongly for increased research along the lines presented in this timely paper.

# Relative Costs of Bus and Rail Transit Systems

THOMAS B. DEEN and DONALD H. JAMES, Alan M. Voorhees and Associates, Inc.

There is a growing need for information on the relative costs of bus and rail systems for providing rapid transit service. Rapid transit is here defined as transit operating on its own exclusive right-of-way. The first step in comparing bus and rail costs is to analyze the various methods suggested for operating busway systems. This was done by comparing the relative service and costs associated with four separate operating patterns. The sensitivity of the conclusions was tested to determine the effect of line length, passenger traffic patterns, feeder bus route length, station spacing, and other variables. It appears that no particular operating method is inherently superior to any of the others. Any one method may be superior in a particular instance, depending on the operating environments and conditions that must be met. The bus and rail costs can then be compared with assurance that the analysis is not distorted by the method of bus operation selected or, at least, with insight into the direction and magnitude of the bias.

Hypothetical bus and rail systems were described so that each provided identical services. Relative costs for providing the service vary depending on line length, proportion of the line requiring subways, and passenger loadings. Sensitivity of costs to rising wage rates and variable interest rates was also examined.

Rail systems can demonstrate cost superiority where peak-hour passenger volumes exceed 12,000 and/or where more than 20 percent of the system requires subways. At volumes of 4,000 peak-hour passengers, and where no subways are required, buses show cost superiority.

•MANY CITIES in the United States are studying the desirability and feasibility of improving their public transit systems. Recognition of the importance of the transportation needs of those who cannot drive as well as a growing disillusionment with the concept of providing for all future travel demands with additional freeways has spurred interest in transit development. Several large cities have made commitments toward construction of high-capacity rail rapid transit systems. A number of others are in planning stages for rapid transit using either bus or rail vehicles.

Perhaps the single characteristic that distinguishes rapid from conventional transit is the use of an exclusive right-of-way, under the control of the transit system, that permits high-speed service unhampered by other traffic. The type of vehicle used may be rubber tired or steel wheeled, run in trains or as single vehicles, run on tracks or on pavement, and be guided by drivers or guideways. Each vehicle system has certain inherent characteristics as to speed, acceleration, construction costs, operating costs, noise, visual amenities, and contribution to air pollution that favor its use in particular circumstances.

For all but the largest cities, the cost of construction, when compared to the number of potential patrons, makes the decision to construct rapid transit difficult. Reviews of

Paper sponsored by Committee on Transportation System Evaluation and presented at the 48th Annual Meeting.

recent transit feasibility studies performed for several cities reveal potential peak-hour line volumes in the range of 3,000 to 15,000 per hour—from one-tenth to one-half the capacity of rail transit. A fundamental question in cases where full rail capacity is not required is whether some other vehicle type might have a cost advantage.

Buses on exclusive roadways have been suggested as a cost saver compared to rail, but there is disagreement as to whether and in what circumstances such savings might result. The present study was conducted in order to develop answers to these questions. Analytical work was done as part of studies conducted for the Atlanta Area Transportation Study (AATS). Alan M. Voorhees and Associates was commissioned in early 1968 to develop a preliminary transportation plan for highways and transit in the Atlanta area. AATS includes representation from the Georgia Highway Department, the Metropolitan Atlanta Rapid Transit Authority (MARTA), the Atlanta Regional Metropolitan Planning Council (ARMPC), and the Atlanta Transit System (ATS), as well as local jurisdictions. MARTA had proposed a regional network of rail rapid transit; the Atlanta Transit System, in late 1967, proposed a system of busways to be used as an interim solution to the transit problems. Some local officials felt buses operating on freeways instead of exclusive roadways would provide adequate transit. Much of the debate centered around relative cost, and it was essential that information be developed to provide insight into this problem.

## BUS OPERATIONAL CONCEPTS

Rail rapid transit systems are in operation in a number of cities around the world, and a completely new system that takes advantage of the latest rail transit technology is under construction in the San Francisco Bay area. Operating procedures, performance, and costs of rail systems can therefore be reasonably approximated. Express transit utilizing buses, however, is an unknown in experience. There is no agreed-upon method of operating buses on exclusive roadways. Because buses are individually steered and can be driven on the street system, they can be operated in a variety of ways. Each method has its adherents, and consensus must be gained on the relative merits of various bus operational procedures before any acceptable comparisons to other modes can be made.

### Type of Operation

Figure 1 illustrates four of the principal methods suggested for bus operations.

Type 1 uses buses in a fashion similar to most rail transit operations. Trunk-line buses run only on the exclusive busway and stop at all stations, including the station in the central business district (CBD), so that very frequent service is available. Passengers have access to busway stations either by the use of separate feeder buses (as illustrated by lines AB, CD, and EF) or by walking or driving to stations. Because buses need not operate on the streets, they can be specially designed with high platforms and extra-large doors; fare collection can take place outside the vehicle. This allows rapid boarding and alighting of passengers and reduces dwell times at stations to a minimum. Scheduled speeds are restricted because of the necessity to stop at all stations. Passengers going from, say, point E to the CBD are required to make a transfer from a feeder vehicle to the trunk-line vehicle at the station. On the other hand, crosstown movements, say from C to D, can be made without a transfer.

Type 2 service also utilizes the trunk-line vehicles on the exclusive right-of-way and separate feeder vehicles to serve the stations. Type 2 operation differs from Type 1, however, in that once a bus loads at a given station it moves to the CBD without stopping at intervening stations. This allows higher speed service but reduces the frequency of service at each station because fewer stops are made on any one trip. Another disadvantage of Type 2 service is the inability to move between outlying stations, say from point X to point Y or from C to F, without moving all the way into the CBD and back out again after transferring. Stations may require more right-of-way since turnarounds are needed at each station.

Type 3 operation has often been suggested for exclusive busways because it takes advantage of the capability of buses to move on local streets as well as over an exclusive

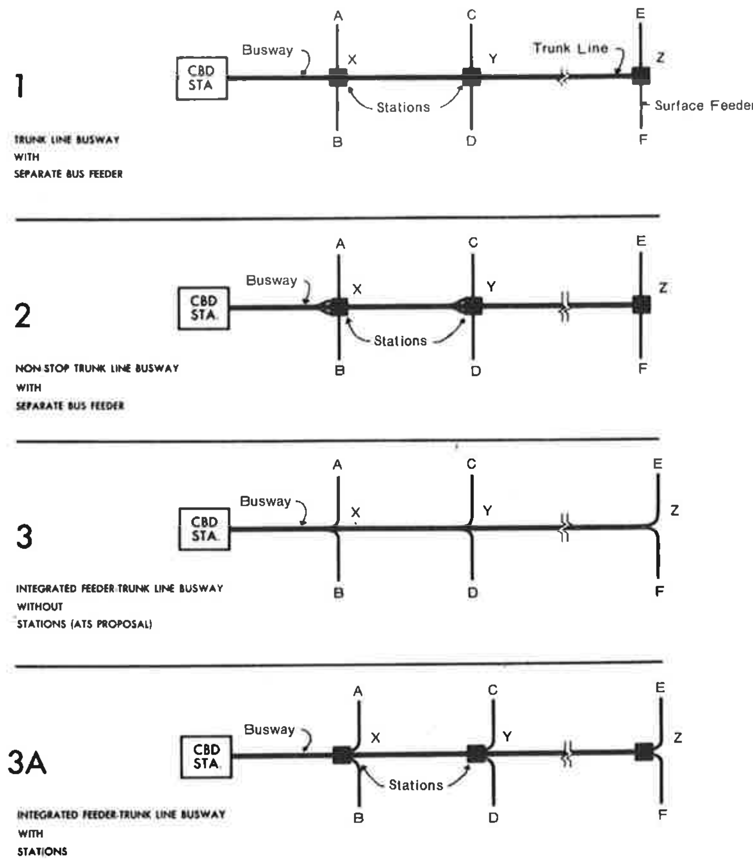



Figure 1. Alternative busway operating concepts.

busway. The same bus, therefore, feeds through local neighborhoods, enters the busway through the use of a ramp, and moves nonstop into the CBD. Such service eliminates transfers for CBD-bound patrons and also provides for nonstop service once the busway is reached. However, crosstown travelers, say moving from point E to point F, must transfer, and there are no points of high-frequency service such as that at Type 1 stations where all buses stop. Type 3 service has the same disadvantage as Type 2 in that movements between outlying points, such as from point X to point Y or from C to E, require a passenger to go all the way into the CBD station, transfer, and go back out again. Some capital cost savings are possible because stations are not required.

Type 3A service is a modification of Type 3 to add stations at the points where ramps feed into the busway. Such stations provide for interchange of passengers going crosstown as well as allow for movements between outlying stations, say from X to Y, without requiring a trip all the way into the CBD and back out again. Extra costs, of course, are incurred with Type 3A as a result of having to provide both ramps and stations, and the patrons are not provided with a nonstop ride into the CBD once gaining the busway because intervening stops must be made.

Doubtless other operating techniques could be developed, but these seem to be the principal ones advocated. There is nothing to prohibit an operation from using a combination of these procedures either simultaneously or at different times of the day in response to changes in volumes or travel patterns. However, there are limits to such changes, e. g., one cannot switch from Type 1 service to Type 3 without providing for ramps instead of stations. For purposes of relative costing and service comparisons, this study made estimates only for the four types described above.

## Evaluation of Bus Operation

In order to determine whether one method of operation has a clear advantage, a series of hypothetical tests was made for each type. Evaluation was made of the relative service and cost of operation. Because all of the systems require right-of-way, grading, drainage, roadways, vehicles, communication systems, yards and shops, and stations (or ramps), they were assumed to have equal capital costs and these were not estimated. However, Type 3A requires both ramps and stations and, in fact, would have a somewhat higher capital cost than the others.

Service was measured in terms of total equivalent passenger time. Total equivalent time is defined as total movement time on the vehicle plus twice time spent waiting or transferring. Waiting time or transfer time is assumed to be one-half the headway of the line being waited for.

Costs were measured by estimating the number of vehicle hours required for one peak hour of operation. Vehicle hours were considered to be the best measure of costs because most bus operating costs are elastic and proportional to bus hours.

## Assumptions

Line Lengths—Evaluations were made for two line lengths, 6 and 12 miles. The 6-mile line has 7 stations and/or ramps spaced 1 mile apart, including 6 outlying and 1 in the CBD. Additionally, there are 12 other loading points (shown as points A, B, C, D, E, and F in Fig. 1) that represent loading locations for feeder bus passengers. Similarly, the 12-mile line had a CBD station, 12 line stations (or ramp entrances), and feeder bus load points.

Traffic Patterns—In certain of the tests it was assumed that 67 percent of the passengers boarding at each point would have final destinations in the CBD. The remaining 33 percent would have final destinations divided evenly among the remaining origin points. In other words, 67 percent of all the persons boarding at point A in Figure 1 would have destinations at the CBD station, and the remaining 33 percent would be divided equally among B, C, D, E, F, X, Y, and Z. In other tests the traffic pattern was changed so that 80 percent of the patrons went to the CBD and 20 percent were divided equally among the remaining origin points.

Volumes—In most of the tests, volumes were assumed to be the same from each station and access point. That is, each point A through F and X, Y, and Z in Figure 1 had equal loadings. In one test, passenger volumes were varied in that the number of passengers boarding at each point became progressively less as the distance from the CBD station increased. This was done to see if non-uniform loading volumes would influence the relative cost-service factor of the different operational types. The four stations nearest the CBD were assumed to board 50 percent more passengers than the four middle stations, and the four outermost stations were assumed to board 50 percent less than the middle ones.

Acceleration and Deceleration Rates—All tests assumed identical acceleration and deceleration rates; however, speeds will vary for different operating types depending on the amount of nonstop operation and lengths of dwell times at stations. Acceleration and deceleration were set at 2 miles per hour per second (mphps).

Maximum Speed—Maximum speed of the busways was set at 50 mph; average speed on the feeder links, including acceleration, deceleration, and layover, was set at 10 mph.

Dwell Time and Layover—Layover was set at a minimum of 10 percent of round-trip running time; dwell times were dependent on passenger loading at particular stations. Dwell times were considered only at the stations and/or access points denoted by points X, Y, and Z in Figure 1. These dwell times vary with the number of passengers getting on and off each vehicle at each point. Dwell times were set at 2 seconds per boarding passenger plus 10 seconds for door open/close and clear time. Therefore, because one bus could board as many as 50 passengers at a point (as in Type 2), the dwell time could be as much as 110 seconds. A minimum dwell time of 20 seconds could occur in some cases where a large number of buses are servicing a particular point.

Headways—In all cases, buses were assumed to have 50 seats and to carry a maximum of one passenger per seat. Headways were set for each line by the number of buses required to service the passengers at the maximum load points. If 1,000 passengers boarded the busway system at each point and each bus carried 50 passengers, feeder bus headways for Types 1 and 2 (in these types, feeder buses never enter the busway) would be 3 minutes (20 buses per hour). In Types 3 and 3A feeder buses enter the busway and have headways somewhat less (more frequent) because each feeder is required to take care of the total volume of its own feeder area plus half the volume of the station (e.g., points X, Y, or Z in Fig. 1) Therefore, the basic feeder headway in such circumstances is 2 minutes on each leg. Headways on the busway portions of each of the alternative concepts would be as follows:

On Type 1 the headway on the trunk line at the terminal station, Z in Figure 1, would have to be sufficient to carry the number of passengers originating at E, F, and Z; that is, 3,000 passengers (assuming that 1,000 loaded at each point) minus a small number bound from E to Z and F, from F to E and Z, and from Z to E and F. This small number is disregarded, and the headway to be scheduled from Z to Y would be 1 minute (60 buses per hour). Accordingly, that from Y to X would be approximately one-half minute and that from X to the CBD, approximately one-third minute. These headways would hold if turnbacks are considered at each station. However, if turnbacks are not considered, then the headway from X to the CBD (or one-third minute) would be controlling throughout the length of the busway portion of the system.

The headway is set individually at each station in the Type 2 system because the concept is that of a loop wherein each station has its own nonstop service to and from the CBD. In this case there would be approximately 3,000 persons at each station, and a 1-minute headway from each station would be necessary. Type 3 has 1-minute headways for boarding passengers at each ramp access point because this is controlled entirely by the feeders. (There would be additional buses coming from upstream stations, but these would not stop at intermediate stations.) However, in Type 3A buses entering at each previous access point stop at all successive points to the CBD, and they would have headways of 1 minute at point Z, one-half minute at point Y, and one-third minute at point X, a situation similar to that in Type 1 using turnbacks.

Feeder Bus Route Length—Feeder buses were assumed to run on routes perpendicular to the busway for a distance of one mile either side of the busway. In one test feeder routes were lengthened to two miles.

## Results

Tests were made for the four-system operating types with variable line lengths, traffic patterns, feeder lengths, station spacing, and loading distribution. Conditions could be found that favor the use of any of the four types of operation, depending on the combination of variables chosen.

For purposes of illustrating the sensitivity of the variables, a system was selected that could be used for comparisons to other combinations. The basic system selected was 12 miles in length, had stations (or ramps) spaced a mile apart, had feeder bus routes of a mile, had uniform passenger loading with 67 percent of all passengers bound for the CBD, and allowed no bus turnback points for the Type 1 operation (i.e., all buses in the Type 1 system ran from terminal to CBD and back, even though most were relatively empty as they traveled the outer portions of the line).

A single cost-service measure was computed as:

$$I_n = BH_n \times EPT_n$$

where

$I_n$ = Cost-service measure for Type N operation,
$BH_n$ = Bus hours required to serve the trunk line and feeders for one hour using Type N operation, and
$EPT_n$ = Equivalent passenger time in minutes $\times 10^{-3}$.

A system is better as $I_n$ decreases because the objective of a good system is to reduce costs and passenger time to a minimum.

A comparison of the basic system to several variations is shown in Figure 2. For each test the type operation that gave the lowest cost-service measure was set at 100, and a cost-service index was computed for the other types relative to 100. For example, Figure 2 shows that Type 3A had a cost-service measure of 100 in Test A, and Type 3 had a measure 1.06 times this. One may say that Type 3 was 6 percent poorer than Type 3A from a composite cost-service standpoint.

Figure 2 shows that Types 2, 3, and 3A perform similarly under varying conditions. Type 1 appears to show an inferior cost-service performance. This may result partially from the no-turnback restriction. Types 2, 3, and 3A are designed to be inherently sensitive to passenger travel demands. Service on the outer ends of the busway is reduced, whereas Type 1 buses all run the full length of the busway. When two intermediate turnbacks are introduced, as in Test D, Type 1 narrows the differentials with the other types significantly.

The length of feeder route is also an important variable because Types 3 and 3A use the same vehicles for feeder and trunk service. If these feeder routes are longer, they affect a higher proportion of the total buses. Test E shows that the use of 2-mile feeder routes works to the relative advantage of Types 1 and 2. Though no test was run that combined longer feeder routes with cutback service for Type 1, it is possible that Type 1 could be superior to Types 3 and 3A under such conditions.

| TEST | DESCRIPTION | Passenger Service Cost Index for Types 1 | 2 | 3 | 3A |
|---|---|---|---|---|---|
| A | BASIC SYSTEM<br>Line Length: 12 miles<br>% Passengers to CBD: 67%<br>Type 1 Turnbacks: None<br>Feeder Bus Route Lengths: 1 mile<br>Station Spacing: 1 mile<br>Passenger Loading: Uniform | 124 | 104 | 106 | 100 |
| B | CHANGE LINE LENGTH<br>Line Length: 6 miles<br>All other parameters same as basic system | 110 | 108 | 114 | 100 |
| C | CHANGE PASSENGER DISTRIBUTION<br>% Passengers to CBD: 80%<br>All other parameters same as basic system | 127 | 104 | 101 | 100 |
| D | USE TURNBACKS IN TYPE 1 OPERATION<br>No. of Turnback Points: 2<br>All other parameters same as basic system | 114 | 104 | 101 | 100 |
| E | USE LONGER FEEDER ROUTES<br>Feeder Bus Route Length: 2 miles<br>All other parameters same as basic system | 118 | 100 | 121 | 115 |
| F | USE LONGER STATION SPACING<br>Station Spacing: 2 miles<br>All other parameters same as basic system | 123 | 101 | 109 | 100 |
| G | VARY LOADINGS WITH DISTANCE FROM CBD<br>Passenger Loadings:<br>First four stations from CBD load 50% more passengers than the second four stations. The third four stations from CBD load 50% less passengers than second four stations. | 166 | 100 | 100 | 100 |

Figure 2. Summary comparison of types of bus operation.

Test G illustrates again the cost effectiveness of tailoring service to passenger demands. Type 1 service, without cutback, is clearly inferior when passenger demand diminishes with distance from the CBD. Types 2, 3, and 3A automatically tailor service, and they are not so adversely affected. Under such loading conditions, one would use cutbacks frequently if Type 1 service were in use.

Type 3A appears to have some advantages for the range of variables tested here. However, its operation is more complex, and its ramps and stations require somewhat more capital cost and would be more difficult to superimpose on an urban environment.

Figure 2 seems to suggest that Type 1 operation is inherently inferior to the others. However, this conclusion is not warranted because different results could be obtained by running the tests with multiple turnbacks on Type 1 operation and increasing the length of feeder routes. A more appropriate conclusion is that any particular operating type could be superior in a given instance, depending on the environment for which service is required. Additional study to specify boundary conditions where each operating type is superior could be helpful.

In the following section, Type 1 bus operation (with turnbacks) is used for comparison to rail systems. We can now proceed with this analysis with some assurance that the selection of Type 1 operation will result in an unbiased comparison or, at least, with some insight into the direction and magnitude of the bias.

## BUS-RAIL COMPARISONS

Cost comparisons of bus and rail are often clouded by difficulties in ensuring that equal service is provided by each mode. Buses typically have on-board fare collection, confined seating, narrow doors, entrance steps, and low acceleration rates. These characteristics are generally not found in the newer rail systems. Special attention was given in these analyses to the task of ensuring that the service level was equal for each mode so that any differentials found in cost would be meaningful.

Because one bus operating method is reasonably equivalent to another from a cost-service standpoint and because it was essential to ensure equal service for bus and rail systems being compared, bus and rail vehicles were assumed to operate in identical patterns equivalent to those in the Type 1 system. All express transit vehicles would operate only on exclusive roadways, pick up or discharge passengers only at stations, and require either a walk or a feeder conveyance to gain access to the system. Station spacing was assumed to be 1.5 miles, and line lengths of 6 and 12 miles were tested.

### Bus Performance and Comfort

Ordinary buses operating in such a system could not compete with modern rail vehicles because of longer passenger loading and discharge time and slower acceleration. To ensure that dwell times of bus and rail would be equal, it was assumed that the buses were provided with extra-wide doors so that the linear feet of door openings per square foot of passenger space would be approximately the same on both rail and bus. This meant that a 40-ft bus required one 30-in. door and one 50-in. door in order to be equivalent to the two 54-in. doors found on the 70-ft rail car proposed by MARTA. The MARTA rail car proposed for Atlanta was used as the standard for all comparisons. Seventy ft long, and 10.5 ft wide, it has 72 seats and two 54-in. doors. The seats are 41 in. wide, and there is a 34-in. longitudinal distance between seats (seat back to seat back). For the buses to operate in and out of high-platform stations without delay, it was assumed that a device would be installed at each door sill that would extend outward to bridge the gap between bus and platform.

Fare collection was assumed to take place in stations (off the vehicle) for all systems. Bus seats were widened and foot room increased by providing two seats on one side and one on the other. This was required in order to provide the same seating comfort (number of seats per square foot of floor space) as that provided on the MARTA rail vehicle. Air conditioning was assumed on both systems.

Equipped with the latest V-8 diesel engines, high-torque transmissions, and separate engines to power air conditioning equipment, a bus could be made to accelerate at 2 mphps up to 30 mph, at which point tapering begins. In contrast, the MARTA rail vehicle is designed to accelerate at 3 mphps up to 30 mph. There appeared to be no way to overcome this discrepancy in service. On the 12-mile line this differential in acceleration caused the bus to take 4 minutes longer than rail from terminal to terminal. The difference for most passengers, of course, would be much less than this. On the other hand, at most volumes buses must operate at closer headways to carry the same load, and this is a plus service factor for buses. For this analysis it was assumed that these factors traded off equally. Slower speeds, however, did reflect adversely on bus operating costs because relatively more vehicles were needed to provide the same service. All modifications to buses described above were included in vehicle cost estimates. Vehicle manufacturers were contacted to confirm the feasibility of some of the modifications.

Articulated Buses

One of the disadvantages of buses, particularly as passenger volumes increase, is that added passengers mean more drivers and higher cost, whereas longer trains can be operated on rail with the same number of personnel. Clearly, bigger buses that would increase the passenger/driver ratio could potentially reduce bus costs. Large articulated buses have been used in Europe for urban service for years. The Alameda-Contra Costa Transit District in the San Francisco Bay area has been experimenting with a similar vehicle. Such a bus, 60 ft long, could carry over 40 percent more passengers than a 40-ft bus using the same space standards. Although the larger bus would require somewhat large platforms at stations, its potential advantages were sufficient to justify separate analyses. The articulated bus is hereinafter referred to as the 60-ft bus.

Underground Systems

Ventilation and driving clearance requirements usually work to the disadvantage of buses used in tunnels. Thus, separate analyses were made for lines requiring no subway and those requiring 20 percent subway. Several lines in Atlanta could be built with virtually no subway construction because of the availability of surface right-of-way along railroad lines that continue through the center of the central business district.

Lower ventilation costs would be incurred if electric trolley buses were used instead of diesel coaches. However, there would be added costs for a larger tunnel to accommodate overhead wires and collection devices. The cost of the power distribution system that would be required over the entire private right-of-way line would tend to offset ventilation savings. Thus, diesel buses were assumed for all tests.

Wage Rates

All comparisons were made using 1967 unit costs and wage rates. Buses are more labor intensive than rail, and future increases in operator wages could work to the relative disadvantage of bus systems. An analysis was prepared of trends in transit personnel wage rates (after correcting for inflation) in recent years, and from this a 15-year projection was made. The future rates are expected to be 67 percent above today's rates, or the equivalent of a 3.5 percent annual increase. One could well project wage levels 20 or 30 years into the future, or use a different rate of increase, but the values used here are believed reasonable and were selected to show the relative effect of increased wages.

Cost of Capital

Rail systems cost more to construct than bus systems, and thus are more capital intensive. Two different costs of capital were assumed: (a) 4 percent, which was selected to represent a minimum borrowing cost for public agencies, and (b) 7 percent, which was selected to represent the opportunity cost of using capital that could otherwise be used in the commercial or private market. Presumably commercial risk capital would be worth 7 percent or more, and one may reasonably consider a denial of its use in the private market as a cost that should be attributed to the transit system. Some economists might suggest a still higher opportunity cost, but these analyses will indicate the relative sensitivity of rail and bus cost to this variable.

## Physical Standards

Physical standards were prepared for each system so that equivalent levels of service would be provided. Figures 3A and 3B illustrate the various standards used.

Roadway widths for bus and rail are quite similar (12.5 ft for rail and 12.0 ft for bus) but the right-of-way widths on grade, cut, and fill sections are somewhat wider for rail because a 2.5-ft safety walk is provided away from the third rail. The difference in subway widths (2.5 ft more for bus) is to provide buses with additional maneuvering room. Buses, which are 8.5 ft wide, are provided with standard 12-ft freeway lanes, plus 2-ft curbs on either side for safety. This gives actual wall-to-wall clearance of 16 ft. In aerial configuration, there is no difference between systems because both use identical structures. The safety walk for the rail system is in the center of the structure away from the third rail. (There is a 3-ft clearance between cars of opposite-direction trains, and the safety walk is elevated above the third-rail level.)

Rail stations are long enough to handle the peak number of cars per train indicated for each volume—140 ft (two 70-ft cars) for the 4,000-passenger tests and 420 ft (six 70-ft cars) for the 12,000-passenger tests. Two escalators were assumed at each station platform (4 per station); six escalators were assumed at the CBD station for the 12,000 volume test. Busway stations are shown in Figures 4A and 4B. The tests with 4,000 passengers maximum volume for the 40-ft bus assumed the type A1 station with two

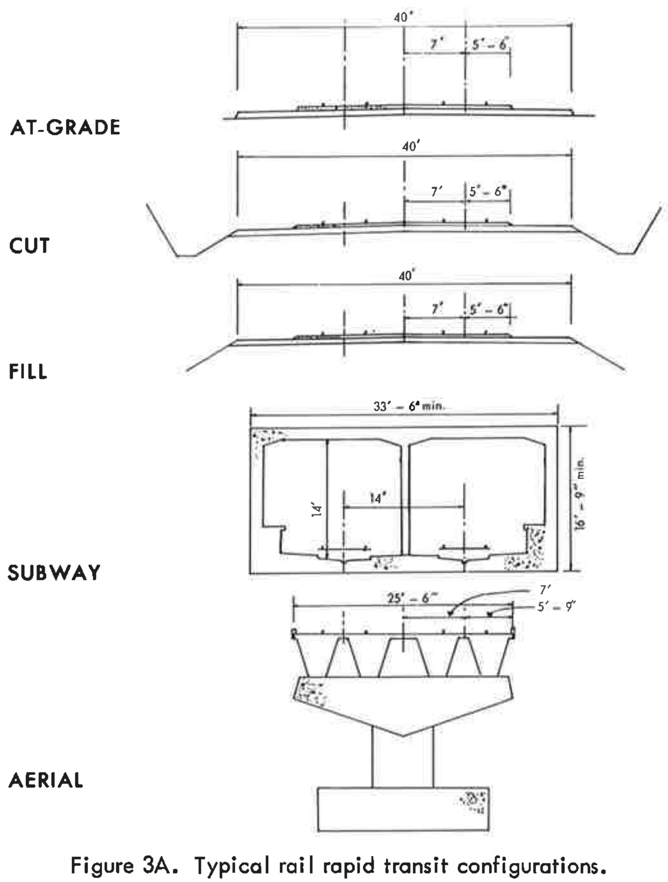



Figure 3A. Typical rail rapid transit configurations.

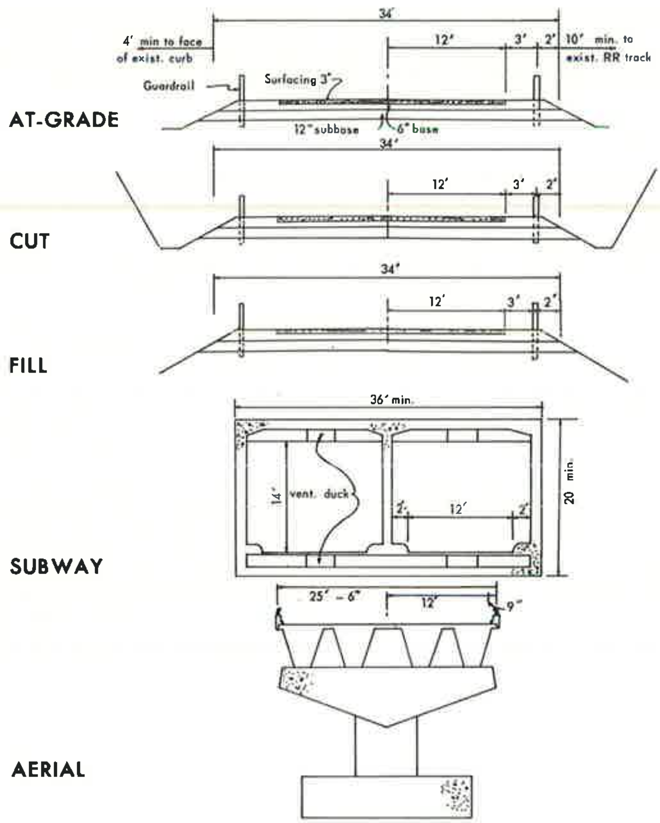



Figure 3B. Typical busway configurations.

escalators per platform (4 per station). The 60-ft bus tests, at the same volume, used the A2 station with the same number of escalators. At the 12,000-passenger level, the 40-ft bus system assumed type B2 stations up to and including the turnback point. Beyond the turnback point (3 stations), the A1 type station was assumed. The 60-ft bus used the B3 type station for the initial 6 stations, and the A2 station for points beyond the turnback to the end of the line. The reason, of course, for the increased number of platforms for the busway stations as volume increased is that the headways decreased as follows:

| Passenger Volumes | 40-ft Bus | 60-ft Bus |
|---|---|---|
| 4,000 (max. load) | 35 sec | 51 sec |
| 12,000 (max. load) | 12 sec | 17 sec |
| 12,000 (beyond turnback) | 24 sec | 34 sec |

Construction costs were estimated using unit costs for labor and materials in the Atlanta area. These costs were estimated by Daniel, Mann, Johnson and Mendenhall, consulting engineers.

## Operating Costs

Operating costs for each system were developed by using standard operating accounts and by staffing both systems at Atlanta wage rates for comparable positions. The rail accounts were comparable to MARTA estimates, and the bus costs reflect current ATS

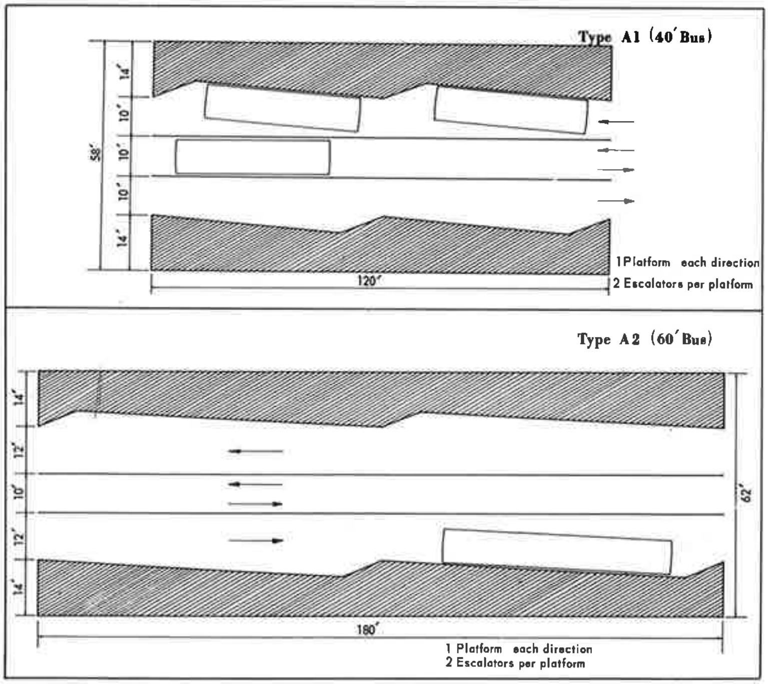



Figure 4A. Typical busway station layouts.

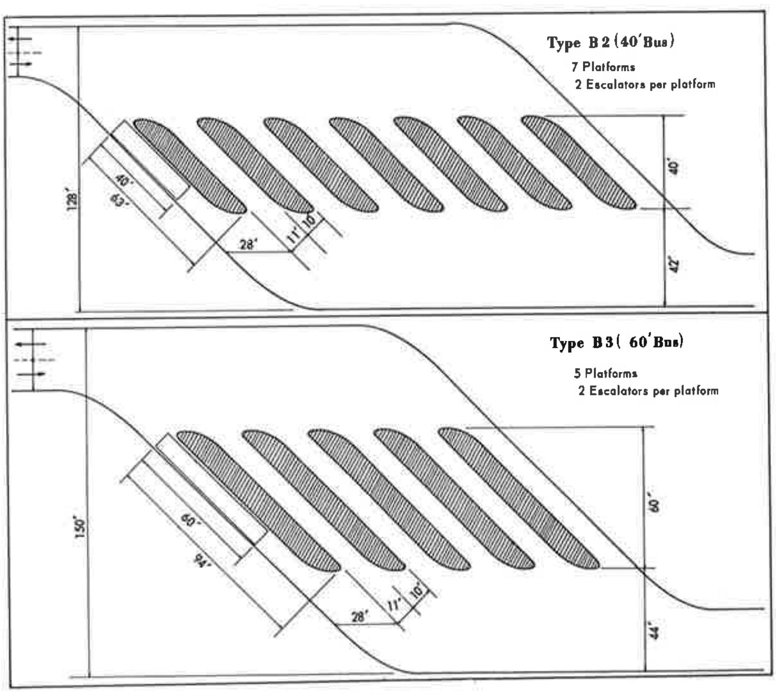



Figure 4B. Typical busway station layouts (suggested by Mayer, Kain, and Wohl in The Urban Transportation Problem).

costs plus necessary provisions for station agents, communications and control systems, and maintenance-of-way and structures. Where staffing of a similar nature was required for both modes, e.g., station agents, equal costs were used.

Rail vehicles were assumed to cost $185,000 each and were amortized at 6.5 percent annually for 30 years. The 40-ft bus was costed at $45,000, and the 60-ft bus at $70,000. Both types of buses were amortized at 6.5 percent for 15 years.

### Scheduling

Annual miles and hours for costing purposes were determined for rail as follows: The number of peak-hour passengers at the maximum load point was divided by the assumed maximum train frequency (30 per hour) to produce the requisite number of seats per train. This number of seats was divided by seats per car (72), and the number of cars per train was determined. To obtain off-peak values, the peak-hour volume was divided by 6.2 (the average peak hour was assumed to have 6.2 times more passengers than the average off-peak hour), then divided by minimum frequency (12 trains per hour) to determine the number of seats required per train. Because the off-peak load factor was assumed to be 75 percent of seats, a maximum of 54 passengers per car was assumed ($0.75 \times 72 = 54$). The number of seats required was divided by 54 to produce cars per train. When these train consists had been determined (never fewer than two cars per train), the requisite trains were scheduled over the individual systems to determine train hours and car miles for one peak hour and one off-peak hour.

A similar procedure was followed for buses except that bus frequency was determined by dividing the proper number of passengers per hour by the given bus capacity.

For both rail and bus systems, turnbacks were provided at the 9-mile station from the CBD, and turnbacks were permitted at that point as long as headways did not exceed 5 minutes at any point along the system. Turnbacks were not contemplated on the 6-mile systems.

These parameters were used to construct operating schedules and to derive vehicle miles and train or bus hours.

### Annualization

Annual vehicle miles and train or bus hours were determined by the following means: A weekday was assumed to consist of 4 peak and 17 off-peak hours, a Saturday of 21 off-peak hours, and a Sunday and a holiday of 10.5 off-peak hours (Saturday divided by 2). A year contains (normally) 255 weekdays, 52 Saturdays, and 58 Sundays and holidays. Therefore, the following formula was derived:

$$\text{Annual Miles} = 255(4P + 17B) + 52(21B) + 58(10.5B)$$

where

$P$ = peak-hour miles, and
$B$ = off-peak-hour miles.

Substituting hours for miles in this formula will give annual hours. When the appropriate costs per mile and hour were applied, annual costs for each system were determined.

## SUMMARY OF ASSUMPTIONS USED IN BUS-RAIL COMPARISONS

Assumptions used in this analysis can be classified into three main groups: (a) constant assumptions, or those applied equally to all modes for all cases; (b) variable assumptions, or those applied equally to all modes but varying for different cases; and (c) special assumptions, or those applying uniquely to each vehicle type tested.

### Constant Assumptions

Operating Methods—Rail cars and buses were operated identically: all vehicles stop at all stations, run from terminal to terminal on grade-separated roadways, and restrict their movements to within the closed system, i.e., busway buses did not at any point operate on local streets. One exception was made to this: one-half of all buses on the 12-mile tests were turned back at the 9-mile station. This provides operating economies to the bus. However, the 5-minute maximum headway standard prohibited the turnback for the 60-ft bus in some instances. It is this same standard that keeps the rail vehicle from being turned back except in the case of higher volumes and then only in the peak hours. All fares were collected off the vehicles. Each test was for a single line with one CBD station and a variable number of outlying stations.

Station Spacing—Stations were assumed at 1.5-mile intervals.

Passenger Traffic Patterns—All line stations (outside the CBD) were assumed to handle equal volumes of passengers. Seventy percent of all passengers trips were assumed to begin or end in the CBD. The other 30 percent were distributed equally to non-CBD stations. Travel in the 4 peak hours was assumed to be approximately 60 percent of the total day.

Passenger Access to the System—Passengers were assumed to reach the stations of either rail or bus systems by walking, driving automobiles, or riding in feeder buses. No analysis was performed for costs of feeder service. Comparisons are for trunk-line service only.

Equal Comfort—So that similarity of service between modes was ensured, seats in all vehicles were assumed to be the same size (41 inches wide by 34 inches, back to back). This dictated 72, 40, and 57 seats in the rail, 40-ft bus, and 60-ft bus, respectively. All vehicles were assumed to be air conditioned. Escalators were at all stations. Maximum passenger loadings were 100 percent of seats during the peak hours and 75 percent of seats during the off-peak periods.

Headways—Maximum headways were assumed to be 5 minutes for all modes. This is a critical assumption. If a 10-minute maximum had been assumed, it could have made the rail system relatively more favorable. During low-volume conditions the larger rail units could have carried loads adequately at 10-minute headways, but the 40-ft bus would have required 3-minute service. Clearly the service differential between 3- and 10-minute headways is not negligible. Thus, a 5-minute maximum headway was assumed.

## Variable Assumptions

Line Lengths—Analyses were made for lines 6 miles (5 stations) and 12 miles (9 stations) in length.

Passenger Volumes—It was assumed that there were passenger loadings of 12,000 and 4,000 at the maximum load point of the line for the peak hour, peak direction.

Station Stop Time—Both bus and rail vehicles were assumed to operate with equal dwell times for equal line loadings and line lengths. This was possible by assuming (and including in the cost estimates) that there were high platforms, equal door widths (per passenger carried), equal space (per passenger), and off-vehicle fare collection.

Wage Levels—Separate estimates of operating costs were prepared using today's wage levels and those expected to prevail 15 years hence (1.67 times today's levels).

Costs of Capital—Separate estimates of annual capital cost were prepared using 4 percent and 7 percent annualization rates.

Proportion of Line in Subway—Separate analyses were prepared for lines having no subway (90 percent at grade and 10 percent aerial), and those having 20 percent subway (70 percent at grade and 10 percent aerial).

## Special Assumptions

Vehicle Specifications—The specifications for each vehicle tested are given in Table 1; note in the table that even during off-peak periods at least two 70-ft car units were assumed to be coupled to form a 140-ft train. One-car units could theoretically be used but were not for this analysis nor are they generally used in practice for the following reasons:

1. There generally is an operating console only in one end of a car. To operate in either direction this requires either a loop track at the ends of each line or two cars coupled with a console at opposite ends.

TABLE 1

VEHICLE SPECIFICATIONS

| Item | Rail | 40-Ft Bus | 60-Ft Bus |
|---|---|---|---|
| Maximum speed | 70 mph | 50 mph | 50 mph |
| Acceleration rate (initial, before taper) | 3.0 mphps to 30 mph | 2.0 mphps to 30 mph | 2.0 mphps to 30 mph |
| Length, vehicle | 70 ft | 40 ft | 60 ft |
| Width, vehicle | 10.5 ft | 8.5 ft | 8.5 ft |
| Width, door | 2, 54 in. | 1, 30 in. and 1, 50 in. | 1, 30 in. and 2, 50 in. |
| Seat spacing (back of seat to back of seat) | 34 in. | 34 in. | 34 in. |
| Double-seat width | 41 in. | 41 in. | 41 in. |
| Number of seats | 72 | 40 | 57[a] |
| Passenger loading: | | | |
| Peak hour | 100% seats | 100% seats | 100% seats |
| Off-peak hour | 75% seats | 75% seats | 75% seats |
| Headways | | | |
| Maximum | 300 sec | 300 sec | 300 sec |
| Minimum | 120 sec | — | — |

[a]Articulation unit occupies equivalent of three seats.

2. To supply an operating console at each end of a car costs more and requires some space otherwise available for passengers.

3. There are few hours per day where volumes are low enough to allow one car operations if the seating and headway standards are met.

4. Savings would be nominal because operating two cars in light periods requires the same number of operators, station personnel, and administrative costs. Power savings would be nominal because the so-called "demand" or peak-hour charge would not change. Some maintenance would be saved, but additional hostlers would be required to uncouple and couple cars.

Speed of Operation—Acceleration rates and maximum speed limitations caused the buses to make the trip from terminal to terminal on the 12-mile line 4 minutes slower than rail, although the differential for most riders would be less than half that. Buses, being smaller, operated at shorter headways for most volumes, and this tends to offset the speed disadvantage. Also, although it was not assumed in this analysis, buses are capable of operating skip-stop with little loss of service because of their very frequent headways at most volumes.

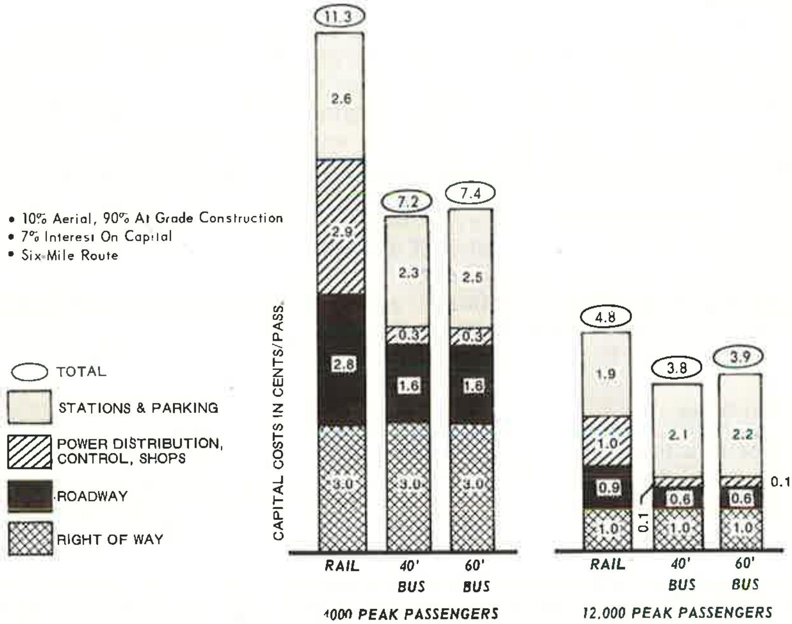



Figure 5. Comparison of rail and busway capital cost (cents per passenger).

## CAPITAL COST ESTIMATES

Capital costs for either a bus or a rail system may be broadly classified into four main categories: (a) right-of-way; (b) roadway, including grading, drainage, structures, ballast, tunneling, track or paving, guardrails, yards, and shops; (c) power distribution, control, and communications systems; and (d) stations and parking lots. Vehicle costs are included with operating costs for these analyses.

Figure 5 shows that costs of stations and rights-of-way are about equal for bus and rail for either high or low volumes. (It should be noted that this so-called high volume, 12,000 peak direction, is less than one-half of the volume carried on a number of rail transit lines in North America. Rail is clearly superior for these higher loads.) However, the costs of power distribution and control systems are more than five times as great for rails. Also, costs for track and ballast for rails are greater than are those for paving for buses. The overall capital costs are 50 percent greater for rail than for buses at the 4,000-volume level and 25 percent greater at the 12,000-volume level.

The cost per passenger for almost all cost items is substantially lower for the high-volume level, illustrating the easier task of justifying express transit systems where significant traffic can be generated. However, the cost per passenger for bus stations decreases very little for the high-volume levels. High-volume bus stations require multiple platforms, extra escalators, and other added costs, almost in direct proportion to added passengers.

Capital costs for the 12-mile line are almost all twice the cost of the 6-mile line, and thus are not graphically illustrated. Capital costs for the route with 20 percent subway are over 50 percent greater than those for the route without subway, illustrating the high cost of underground construction. Figure 6 indicates that roadway costs (including tunneling) are 12 to 18 percent higher for buses than for rail. Clearly, as the proportion of the line involving subways increases, the more roadway costs will transcend other

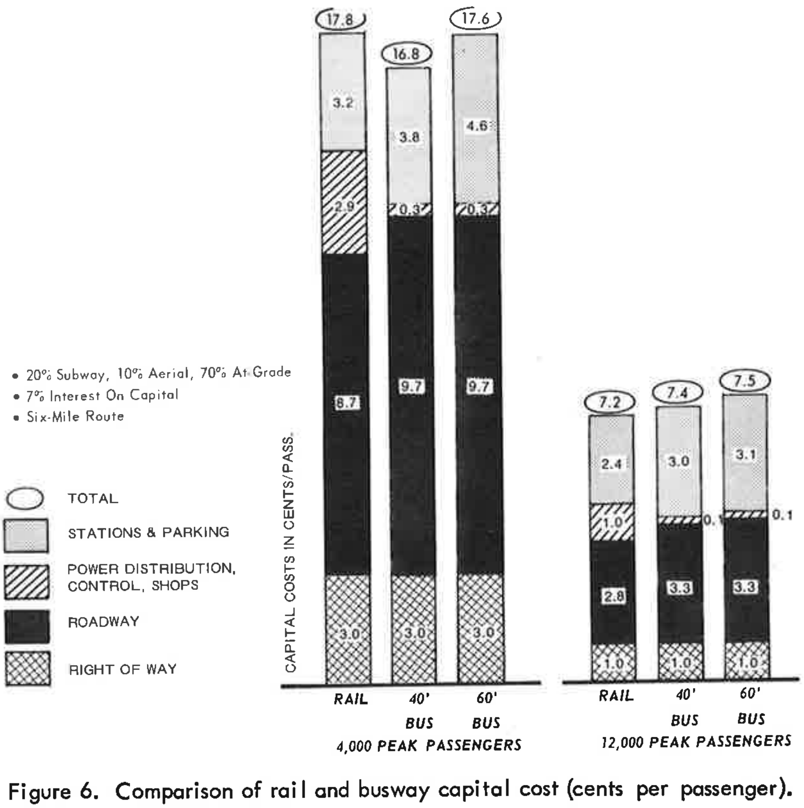



Figure 6. Comparison of rail and busway capital cost (cents per passenger).

cost elements, and roadway cost differentials between modes will become decisive. Even with 20 percent subway, roadway costs make up almost one-half of the total capital cost of the line.

The impact of interest rates on capital costs is significant. The increase from 4 percent to 7 percent over the 60-year period increases capital costs by 50 percent on an annual or per-passenger basis (see Tables 2 through 7).

TABLE 2

COST SUMMARY—RAIL TRANSIT
(Route Length: 6 Miles)

| Other Variables | | | Costs (cents per passenger) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4,000 Passengers (peak hour) | | | 12,000 Passengers (peak hour) | | |
| Type Construction | Wage Level (year) | Cost of Money | Operating | Capital | Total | Operating | Capital | Total |
| No subway | 1968 | Low | 12.7 | 7.5 | 20.2 | 8.0 | 3.2 | 11.2 |
| | | High | 12.7 | 11.3 | 24.0 | 8.0 | 4.8 | 12.8 |
| | 1983 | Low | 16.7 | 7.5 | 24.2 | 9.8 | 3.2 | 13.0 |
| | | High | 16.7 | 11.3 | 28.0 | 9.8 | 4.8 | 14.6 |
| 20% subway | 1968 | Low | 12.7 | 11.9 | 24.6 | 8.0 | 4.8 | 12.8 |
| | | High | 12.7 | 17.8 | 30.5 | 8.0 | 7.2 | 15.2 |
| | 1983 | Low | 16.7 | 11.9 | 28.6 | 9.8 | 4.8 | 14.6 |
| | | High | 16.7 | 17.8 | 34.5 | 9.8 | 7.2 | 17.0 |

TABLE 3

COST SUMMARY—BUSWAYS (40-FT BUS)
(Route Length: 6 Miles)

| Other Variables | | | Costs (cents per passenger) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4,000 Passengers (peak hour) | | | 12,000 Passengers (peak hour) | | |
| Type of Construction | Wage Level (year) | Cost of Money | Operating | Capital | Total | Operating | Capital | Total |
| No subway | 1968 | Low | 12.5 | 4.8 | 17.3 | 9.7 | 2.5 | 12.2 |
| | | High | 12.5 | 7.2 | 19.7 | 9.7 | 3.8 | 13.5 |
| | 1983 | Low | 17.6 | 4.8 | 22.4 | 13.0 | 2.5 | 15.5 |
| | | High | 17.6 | 7.2 | 24.8 | 13.0 | 3.8 | 16.8 |
| 20% subway | 1968 | Low | 12.5 | 11.2 | 23.7 | 9.7 | 4.9 | 14.6 |
| | | High | 12.5 | 16.8 | 29.3 | 9.7 | 7.4 | 17.1 |
| | 1983 | Low | 17.6 | 11.2 | 28.8 | 13.0 | 4.9 | 17.9 |
| | | High | 17.6 | 16.8 | 34.4 | 13.0 | 7.4 | 20.4 |

TABLE 4

COST SUMMARY—BUSWAYS (60-FT BUS)
(Route Length: 6 Miles)

| Other Variables | | | Costs (cents per passenger) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4,000 Passengers (peak hour) | | | 12,000 Passengers (peak hour) | | |
| Type of Construction | Wage Level (year) | Cost of Money | Operating | Capital | Total | Operating | Capital | Total |
| No subway | 1968 | Low | 11.4 | 4.9 | 16.3 | 8.6 | 2.6 | 11.2 |
| | | High | 11.4 | 7.4 | 18.8 | 8.6 | 3.9 | 12.5 |
| | 1983 | Low | 15.9 | 4.9 | 20.8 | 11.3 | 2.6 | 13.9 |
| | | High | 15.9 | 7.4 | 23.3 | 11.3 | 3.9 | 15.2 |
| 20% subway | 1968 | Low | 11.4 | 11.7 | 23.1 | 8.6 | 5.0 | 13.6 |
| | | High | 11.4 | 17.6 | 29.0 | 8.6 | 7.5 | 16.1 |
| | 1983 | Low | 15.9 | 11.7 | 27.6 | 11.3 | 5.0 | 16.3 |
| | | High | 15.9 | 17.6 | 33.5 | 11.3 | 7.5 | 18.8 |

TABLE 5

COST SUMMARY—RAIL TRANSIT
(Route Length: 12 Miles)

| Other Variables | | | Costs (cents per passenger) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4,000 Passengers (peak hour) | | | 12,000 Passengers (peak hour) | | |
| Type of Construction | Wage Level (year) | Cost of Money | Operating | Capital | Total | Operating | Capital | Total |
| No subway | 1968 | Low | 19.9 | 14.8 | 34.7 | 12.5 | 6.1 | 18.6 |
| | | High | 19.9 | 22.3 | 42.2 | 12.5 | 9.2 | 21.7 |
| | 1983 | Low | 26.4 | 14.8 | 41.2 | 15.8 | 6.1 | 21.9 |
| | | High | 26.4 | 22.3 | 48.7 | 15.8 | 9.2 | 25.0 |
| 20% subway | 1968 | Low | 19.9 | 23.3 | 43.2 | 12.5 | 9.3 | 21.8 |
| | | High | 19.9 | 35.0 | 54.9 | 12.5 | 14.0 | 26.5 |
| | 1983 | Low | 26.4 | 23.3 | 49.7 | 15.8 | 9.3 | 25.1 |
| | | High | 26.4 | 35.0 | 61.4 | 15.8 | 14.0 | 29.8 |

TABLE 6

COST SUMMARY—BUSWAYS (40-FT BUS)
(Route Length: 12 Miles)

| Other Variables | | | Costs (cents per passenger) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4,000 Passengers (peak hour) | | | 12,000 Passengers (peak hour) | | |
| Type of Construction | Wage Level (year) | Cost of Money | Operating | Capital | Total | Operating | Capital | Total |
| No subway | 1968 | Low | 19.5 | 9.1 | 28.6 | 14.7 | 4.7 | 19.4 |
| | | High | 19.5 | 13.7 | 33.2 | 14.7 | 7.1 | 21.8 |
| | 1983 | Low | 27.5 | 9.1 | 36.6 | 20.5 | 4.7 | 25.2 |
| | | High | 27.5 | 13.7 | 41.2 | 20.5 | 7.1 | 27.6 |
| 20% subway | 1968 | Low | 19.5 | 21.9 | 41.4 | 14.7 | 9.4 | 24.1 |
| | | High | 19.5 | 32.9 | 52.4 | 14.7 | 14.1 | 28.8 |
| | 1983 | Low | 27.5 | 21.9 | 49.4 | 20.5 | 9.4 | 29.9 |
| | | High | 27.5 | 32.9 | 60.4 | 20.5 | 14.1 | 34.6 |

TABLE 7

COST SUMMARY—BUSWAYS (60-FT BUS)
(Route Length: 12 Miles)

| Other Variables | | | Costs (cents per passenger) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4,000 Passengers (peak hour) | | | 12,000 Passengers (peak hour) | | |
| Type of Construction | Wage Level (year) | Cost of Money | Operating | Capital | Total | Operating | Capital | Total |
| No subway | 1968 | Low | 17.8 | 9.3 | 27.1 | 13.0 | 4.8 | 17.8 |
| | | High | 17.8 | 14.0 | 31.8 | 13.0 | 7.2 | 20.2 |
| | 1983 | Low | 25.0 | 9.3 | 34.3 | 17.5 | 4.8 | 22.3 |
| | | High | 25.0 | 14.0 | 39.0 | 17.5 | 7.2 | 24.7 |
| 20% subway | 1968 | Low | 17.8 | 22.7 | 40.5 | 13.0 | 9.7 | 22.7 |
| | | High | 17.8 | 34.1 | 51.9 | 13.0 | 14.6 | 27.6 |
| | 1983 | Low | 25.0 | 22.7 | 47.7 | 17.5 | 9.7 | 27.2 |
| | | High | 25.0 | 34.1 | 59.1 | 17.5 | 14.6 | 32.1 |

## OPERATING COSTS

Figure 7 shows comparisons of operating costs for rail and busways based on the assumption that future wage rates are 67 percent above today's. For low-volume routes operating costs vary only 5 percent between modes. Surprisingly, costs for the 60-ft bus are even lower than those for rail.

On low-volume routes and off-peak periods as many operators are required for rail as for bus if the minimum of 12 trains per hour (5-minute headway) is to be maintained. For higher volumes, however, extra passengers are carried simply by adding cars to trains, and then more operators are required for buses. Costs for bus, excluding those for operators' wages, are equal to or higher than those for rail except that costs for vehicle maintenance are higher for 40-ft buses because more of them are needed.

Operating costs for the 12-mile line (see Tables 2 through 7) are about 55 percent higher than those for the 6-mile line for all modes. This is because of economies inherent in operating larger systems (spreading overhead), the lower number of

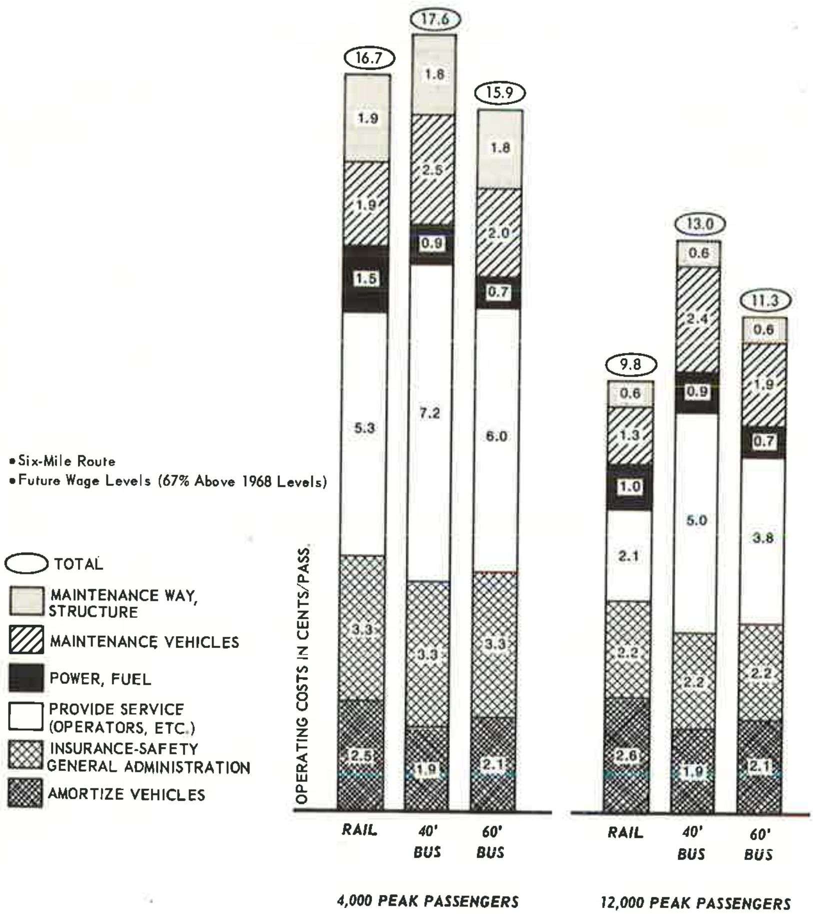



Figure 7. Comparison of rail and busway operating cost (cents per passenger).

passengers handled per station, and reduced layover as a percentage of running time in longer systems. The impact of increased wages on bus operating costs can be seen in Figure 8. Biggest increases are found in operators' wages and in maintenance. Vehicle amortization and fuel costs are unaffected.

## TOTAL COST

When the cost per passenger associated with capital cost is added to the operating cost, the total cost per passenger can be estimated for each of the various cases studied. Figure 9 shows total cost comparisons of the 6-mile line where 7 percent interest on capital and future wage rates are assumed.

For cases involving 20 percent subway, both construction and operating costs are approximately equal, although rail is slightly less costly for the higher volume cases. Doubtless the relative cost efficiencies with rail would grow at volumes above 12,000 per hour because additional volumes can be accommodated by simply lengthening trains and stations. Bus stations, on the other hand, get increasingly complex, and operating costs would soar with added drivers. Lines with greater than 20 percent subway would also tend to show cost efficiencies in favor of rail because larger tunnels and expensive ventilation systems are required for buses.

The relative costs of bus or rail with 20 percent subway are not particularly sensitive to assumptions about interest rates or wage levels. All systems have about the same capital intensiveness and/or labor intensiveness, i. e., higher interest or wage rates would increase costs of rail or bus almost equally.

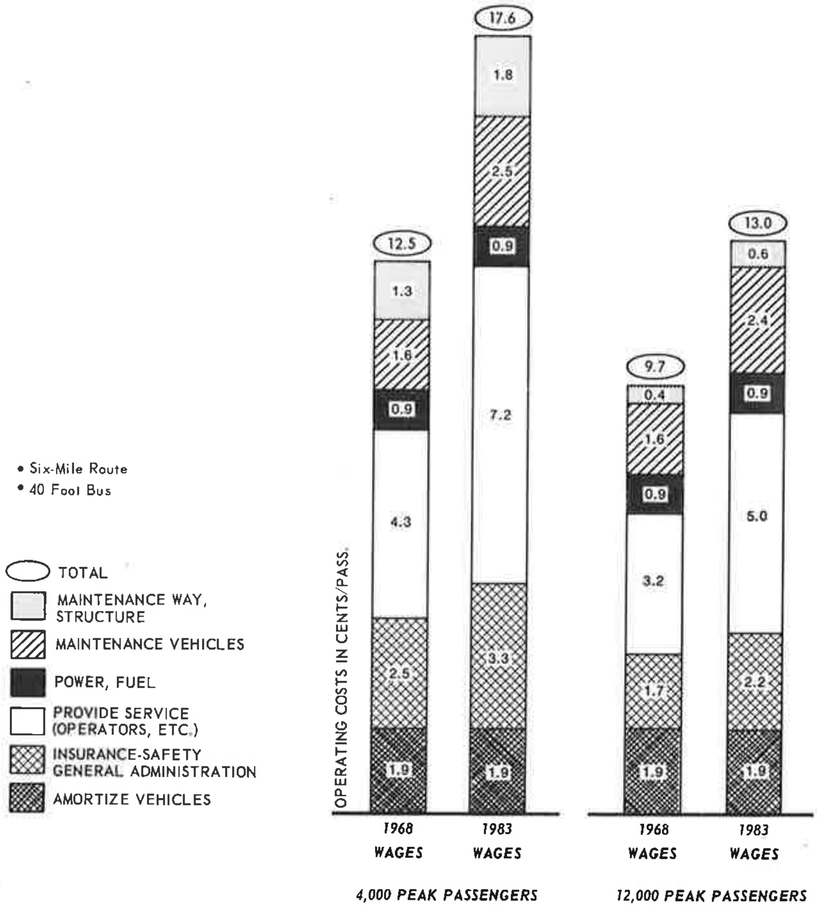



Figure 8. Busway operating cost variations (cents per passenger).

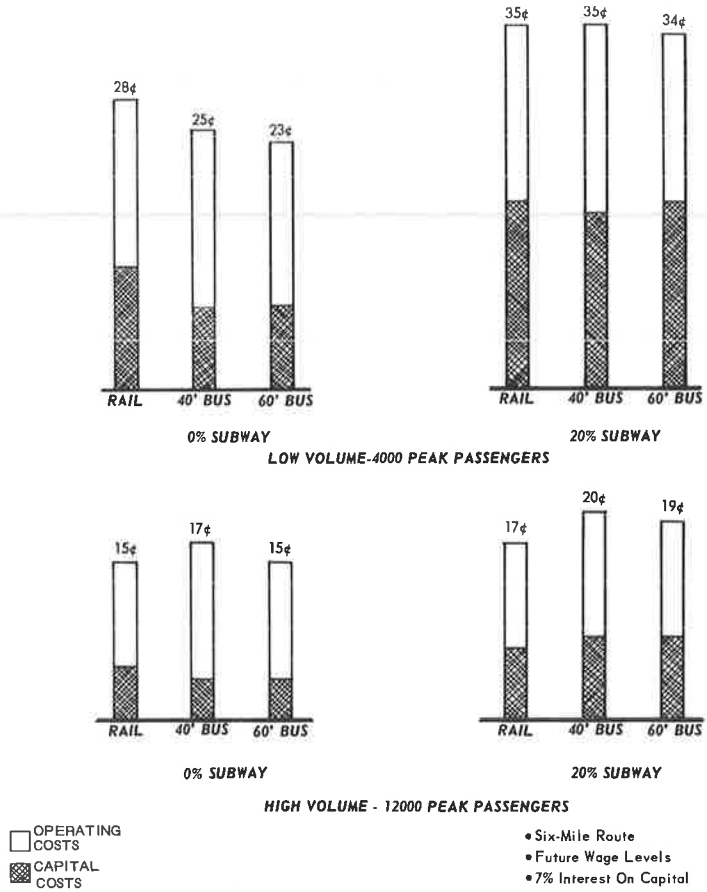



Figure 9. Total rail and busway costs (cents per passenger).

Without subways and with volumes of 12,000 passengers per hour, there appears to be no great cost advantage of one mode over another. Tables 2 and 4 show costs of rail to be slightly less (14.6 cents per passenger) than those for the articulated bus (15.2 cents) in this case. Doubtless the relative efficiencies of rail at higher volumes would show more clearly at volumes above 12,000. Relative costs for the 12-mile line and for various assumptions concerning wage escalation and interest rates can be found in Tables 5 through 7. Comments noted above are generally applicable to cases with high volumes and no subways.

Cost differentials, which appear to be significant, can be observed for nonsubway, low-volume cases (Fig. 9). Operating costs for buses and rail are about equal for this case, but the higher capital cost of rail tips the balance in favor of buses. Savings of 17 to 22 percent can be observed for buses (see Tables 2 through 7), depending on the line length, wage, and interest assumptions.

Caution must be exercised in applying these findings to other situations. Buses have never been used in a system such as that assumed in these studies. There are risks in developing an untried operation. Further, there is the belief in some quarters that busways cannot gain public support or acceptance as easily as rail. This may have some validity because the public knows buses only as presently operated, i.e., slow, subject to traffic delays, and used primarily by lower income people. Sensitivity to weather might be a factor weighing against buses in some environments.

Finally, estimates of cost were made in this study on a prelimianry basis for only one community. Before results are applied to other communities, operating and capital cost estimates should be developed for the particular operational types and project envisioned. Yet the same assumptions and unit costs were applied to all modes; one must change the assumptions radically to get different relative cost values. Most of the cases tested showed insignificant cost differentials. However, for low volumes in situations where surface or elevated rights-of-way can be obtained, the lack of need for power distribution and train control systems for buses provides a cost advantage that appears to be significant. (Why a rail system requires a control system to avoid vehicle collisions and a bus does not is a matter that seems debatable. Control systems make up about 30 percent of the costs of "power distribution control and shops" shown in Figures 5 and 6. Yet the assumptions used in this study in this regard correspond to what would happen if busways or rail were built today.)

## CONCLUSIONS

1. There are various ways to operate a busway. Any one method may be superior to another in cost or service, depending on the operating environment. Systems using the same vehicle for feeder service and trunk-line operation tend to serve the CBD well at the expense of non-CBD destinations.

2. Where 20 percent or more of a line must be in subway, the costs of subway construction become so large that they dominate other cost elements. Subway construction costs less for rail systems than for buses. Thus, if subway construction is required, rail can be considered superior to the bus.

3. On systems where 20 percent subway is required, capital costs (construction and right-of-way) represent less than one-half the total annual costs. Where no subway is involved, capital costs represent less than one-third under some conditions. This suggests that planning studies and feasibility analyses for systems of this type should put at least as much emphasis on estimates of operating costs and revenues as on capital costs.

4. Articulated buses have 40 percent more seats than the conventional bus and can provide for cost reductions of 10 to 12 percent in total capital and operating costs in some instances. Research and development programs should be inaugurated to refine such a vehicle for busway use. Three-element articulated units could conceivably be used to even greater advantage.

5. Where no subway is required on a line, volumes as high as 12,000 passengers (peak hour) can be accommodated as cheaply on buses as on rail, if large articulated buses are used. As volumes above 12,000 are reached, rail systems show increasing cost savings over bus.

6. Where no subway is required and where volumes are in the vicinity of 4,000 passengers equivalent (peak hour), either 40-ft or 60-ft buses can provide service equivalent to that of rail at cost savings of 15 to 20 percent.

7. Increasing wage rates in future years will tend to increase costs of bus operation more than those of rail, but not significantly. A 67 percent assumed future wage increment increased bus operating costs only about 1.5 cents per passenger more than it increased rail costs (for low volumes). This 1.5 cents is about 4 percent of total costs. assumptions that increase the cost of borrowing capital tend to increase rail costs more than bus. An assumed change of borrowing rates from 4 percent to 7 percent applied to both modes increases rail costs such that increased future wage rate assumptions are approximately offset; i.e., estimates made assuming today's wage rates and low interest costs will not be significantly changed by assuming future wage rates and high interest rates.

8. Many considerations other than cost are involved in the bus-rail decision in any particular community. Bus systems may be considered more risky in view of the fact that they have not been used on exclusive right-of-way systems to date; that they might provide more troublesome problems in bad weather; that they may not be as safe because collision avoidance control systems are required for rail but not for bus systems; and that public acceptance may be more difficult to obtain for bus than for rail. On the other hand, buses provide more opportunities for tailoring service to demands and for developing a system gradually rather than all at once.

# Search and Choice in Transport Systems Analysis

MARVIN L. MANHEIM, Department of Civil Engineering, Massachusetts Institute of Technology

•THE SUBJECT of this paper is the problem of searching out and choosing among transportation system alternatives. We first describe the essential features of transportation systems analysis, as a substantive problem. In the next section, we review a very popular model of the design process, the "rational model," and discuss its limitations; we then present a more complete model of the design process, the PSP model, and explore its implications. Next, we demonstrate the applicability of the PSP model to transportation planning and illustrate systematic analyses of transportation alternatives. In the last section is described specific operational techniques that have been developed to implement or extend the PSP model in the light of issues raised by the prototype analysis.

This paper builds upon initial concepts proposed earlier (15, 19). A fuller exposition of the ideas and techniques summarized here can be found in several current research reports (1, 18, and also 2 through 17).

## THE PROBLEM: TRANSPORTATION SYSTEMS ANALYSIS

### Options

A wide spectrum of aspects of a transportation system can be varied. Not all of these aspects are open to a single decision-maker, nor are all open at the same time. This spectrum of options, or "decision variables," may be summarized as follows:

1. Technology—Development and/or implementation of new combinations of transportation components enable transportation services to be offered in ways that were not previously available. Examples include containers, containerships, and piggyback trucks and rail cars; supersonic transport; new urban mass transportation concepts such as the Westinghouse Skybus and the various types of dial-a-bus systems featuring dynamically routed and scheduled vehicles (20, 21). In general, we are dealing with systems containing several different transportation technologies, i.e., with multimodal systems.
2. Networks—Options about networks include the general configuration pattern of the network as well as the approximate geographic location of the links of the network.
3. Link Characteristics—Networks consist of links as well as nodes; links correspond to routes, such as highways, airways, rail lines, and urban streets, and to terminals, interchanges, and other physical facilities.
4. Vehicles—There are options regarding the number and characteristics of vehicles in the system.
5. Operating Policies—Operating policies include the full spectrum of decisions about how the transportation system is operated: routes and schedules of the vehicles; types of service to be offered, including various services auxiliary to transportation (passenger meal services; diversion and reconsignment privileges for freight); prices to be charged (both general pricing policy and specific pricing decisions); subsidies; and taxes.

This set of transportation options fully defines the space of possible transportation plans and policies. However, these options are exercised not in a vacuum but in the context of a system of social and economic activities. We define such an activity system as all the social, economic, political, and other transactions that take place over

Paper sponsored by Committee on Transportation System Evaluation and presented at the 48th Annual Meeting.

space and time in a particular region. These transactions, both actual and potential, determine the demand for transportation, and, in turn, the levels and spatial patterns of these interactions are affected in part by the transportation services provided.

Therefore, in addition to the options about the transportation system itself, we must clearly identify those options in the activity system that will be expressed as the demands on the system:

1. Travel—These are the options open to every potential user of the transportation system: whether to take a trip at all, where to make it, when, and how, i.e., by what mode and route. The aggregate result of all the individual decisions about travel is expressed as the demand for transportation.
2. Activity System—Each of the social, economic, and political actors in the activity system has a wide range of options about how, when, and where it will conduct its activities. Over the long term, these options profoundly influence the demand for transportation. For example, as major changes in a transportation system are made over time, the spatial pattern of population and economic activity will change, as actors exercise their options for changing the location and/or scale of their activities. Forces within the economy external to the transportation system, such as housing subsidies or mortage policy, may impact on the spatial pattern of activity, and thus affect the demand for transportation.

These options are in the hands of a large variety of public and private decision-makers. In many transportation analyses, most of these activity system options must be treated as exogenous, completely uncontrollable by the transportation analyst. Still other options are controllable to some extent in explicit coordination with nontransport options such as control of land use through zoning and land development incentives.

Whether controllable or not, however, the full set of transportation and activity system options must be considered in any analysis.

## Impacts

When evaluating alternative transportation systems, one would like to consider all relevant impacts. Any change in the transportation system can potentially affect a large variety of groups and interests. The prospective impacts can be grouped as follows:

1. Users by location within the region, by trip purpose, and by socioeconomic group. Examples are suburban resident commuting to central city job and low-income, noncar-owning resident of center city traveling to health facilities.
2. Operators by mode, by link. Examples are air carrier, trucker, highway maintenance agency, port authority, and toll bridge operator.
3. Physical by type of impact, by link. Examples are families, jobs, and taxable values displaced by new construction; and pollution of immediate environment through noise, fumes, air pollution, and ground water changes.
4. Functional by location within region, by type. Examples are changes in retail sales areas of shopping center, changes in production costs, and changes in land values.
5. Governmental by location, by level. Examples are local, state, or national representatives and citizen groups (22, 23, 24).

An essential characteristic of transportation is the differential incidence of its impacts. Some groups will gain from any transportation system change; others may lose. Therefore, transportation choices are essentially sociopolitical choices: The interests of different groups must be balanced.

## Prediction: Basic Concepts

Any proposed change in a transportation system (or a completely new system) can be expressed in terms of the options identified above. The problem of prediction is to anticipate the impacts that a particular proposal will have: That is, we need procedures for predicting the impacts associated with any set of options (Fig. 1). In

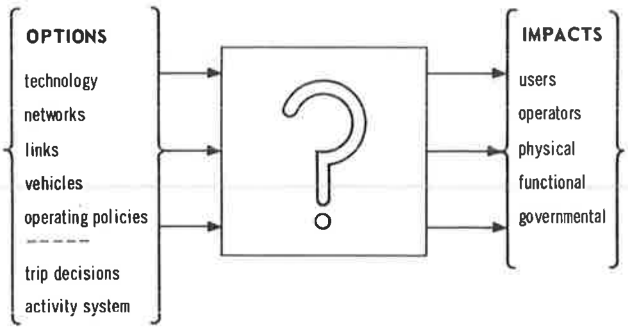



Figure 1. The prediction problem.

transportation, the impacts depend on the pattern of flows in the network that will result from the particular set of options.

The core of the transportation analysis problem is the prediction of network flows: Specification of the transportation system, T, and the activity system, A, implies the pattern of flows, F. Once the options with respect to transportation and the activity system are specified, the flows in the system are a consequence of those options. Prediction of these flows is necessary for evaluation of the impacts.

The general framework for the prediction of network flows is that of the equilibrium between supply and demand (25, 26, 27, 28, 29, 30, 31). All the transportation system options, T, can be expressed in a set of supply functions, S, and all the nontransportation options, A, in a set of demand functions, D. Then, within the constraints of the transportation channels specified as part of T, the equilibrium of S and D is a pattern of flows, F. The elements of this pattern, F, are the volumes and characteristics of the flows over each link of the transportation network: What flows over a link, from what origin zones, to what destination zones, at what speed, and cost?

The key concept allowing this formulation to be workable is that of a vector, L, of service variables. The level of service, L, that a particular set of transportation facilities provides can be expressed in terms of travel time (mean and variance), trip costs (fares and other out-of-pocket costs, as well as indirect costs such as car ownership), safety, comfort (real and psychological), and other characteristics. These service variables both characterize the transportation system and serve as the basis for the demands for transportation. The difference between perceived and actual service characteristics may be large and significant. Here, we shall make no distinction (25).

The general structure of this analysis problem can be expressed concisely (these variables are all vectors, matrices, or n-dimensional arrays).

Definition of variables:

T = specification of transportation system, in terms of full set of options;
A = specification of activity system (including exogenous characteristics);
F = pattern of flows in the system;
L = service characteristics of a particular flow or set of flows; and
V = volume of flows.

Supply Functions:

$$L = S(T, V)$$

Demand functions:

$$V = D(A, L)$$

Equilibrium:

$$\begin{Bmatrix} L = S(T, V) \\ V = D(A, L) \end{Bmatrix} \longrightarrow (V_O, L_O)$$

i.e., $(T, A) \longrightarrow F = (V, L)$

In words: The level of service, L, that the transportation system supplies is a function, S, of the transportation options, T, and the volume of flow, V. The volume of flow desiring transportation is a function, D, of the activity system options, A, and the level of service, L. The pattern of flows, F, defined as the volumes and the service levels that will actually occur for given (T, A), is the equilibrium solution to the supply and demand relations.

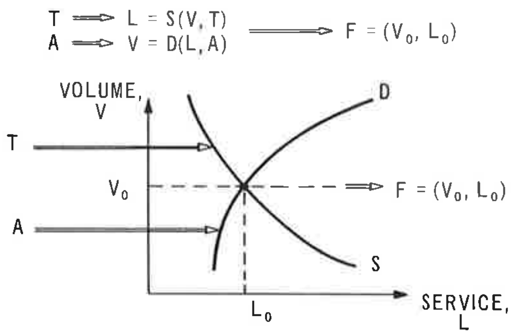



Figure 2. Simple equilibrium.

The graphical interpretation of this formulation is shown in Figure 2. In this figure V and L are assumed one-dimensional. Specification of T and A implies supply and demand functions, S and D. These in turn imply an equilibrium flow pattern, F, comprising a volume, V, and trip cost or "negative" level of service, $L_0$.

The value of this formulation is illustrated in Figure 3. $S_0$ is the supply function of the previous system, with corresponding equilibrium flow pattern $F_0 = (V_0, L_0)$. Consider a possible improved system $S_1$. If we assume the existing volume of travel, $V_0$, will occur on the new system as on the previous, we would anticipate a service level, $L_E$, i.e., a lower trip time because of the improved facility. However, assuming a constant volume level is erroneous, for the travel volume will increase because of the increased level of service (decreased trip time). The extent of this increase in travel is given by the demand function, D. Thus, the actual flow pattern resulting will be that given by the equilibrium of D and $S_1$: $F = (V_1, L_1)$. That is, the traffic volume will increase, and the level of service will be intermediate between $L_0$ and $L_E$.

Of course, because it takes time to implement transportation system improvements and because population and travel continue to increase, the demand curve, D, may meanwhile have shifted upward and to the right. Thus, the new equilibrium ($V_1$, $L_1$) may actually occur such that $L_1$ is greater than $L_0$. The level of service over the new system is actually worse than the level of service over the previous system at the initial period, but is better than the level of service that would have resulted from the old supply function and the new demand function.

The prediction of flows in a network is based, in principle, on this theory of equilibrium between supply and demand. In practice, prediction of equilibrium flows in networks is generally difficult and expensive; and, as discussed later, most present techniques for predicting network flows leave much to be desired.

The major, but by no means only, difficulty in translating this conceptual framework into practice is the role of the network in constraining the equilibrium flow pattern (10, 26, 28, 29).

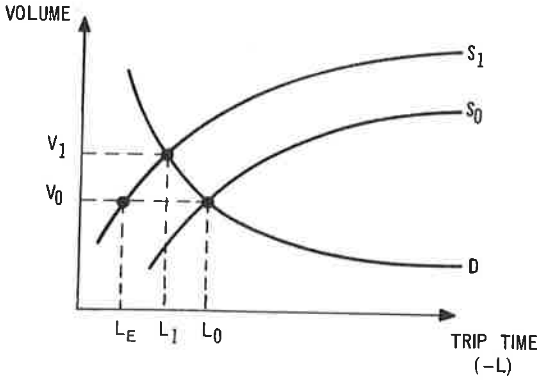



Figure 3. System improvement.

1. Multiple demand functions: The area to be studied is divided into zones; there is a different demand function for each pair of zones (origin and destination), for different groups of prospective trip-makers, and for different trip purposes (passengers) or commodity types (freight). Further, the demand for each zone pair is a function of the level of service vector, not a single "price."

2. Multiple supply functions: Each link of the network is represented by a different supply function. Note that

because L is generally a vector with time, cost, or safety as components, both the supply and demand functions are potentially very complex.

3. Finding the equilibrium pattern of flows: Instead of a simple graphical exercise, the calculation of the equilibrium flows is a difficult problem. Some of the conceptual and computational difficulties are (a) the level of service perceived by a trip between two zones depends upon which path is taken through the network; (b) the level of service over any path is a function of the levels of service over each of the links in that path (e.g., trip time equals the sum of the times over each link in the path); (c) the level of service over a link is a function of the total volume over that link (as given by that link's supply function); (d) the total volume over a link is composed, in general, of flows between many different zone pairs; (e) the actual computational procedures may be difficult and expensive; and (f) the equilibrium flow pattern is not unique. [A mechanism of trip behavior must be assumed in order to determine a unique equilibrium. One set of assumptions leads to the traffic-assignment approach of urban transportation studies; other types of assumptions lead to various mathematical programming formulations (2, 10, 26, 28, 29).]

Prediction Models

To fully implement this analysis approach, the following five major types of models are required:

1. Supply models to determine for any specified setting of the options what the level of service will be for various flow volumes. Examples are travel time over a rail link as a function of train length, schedule frequency, roadway, and volume of passengers; volume-travel time curves as used in traffic assignment procedures.

2. Resource models to determine the resources consumed—land, labor, capital—in providing a particular level of service with specified options.

3. Demand models to determine the volume of travel demanded, and its composition, at various levels of service.

4. Network equilibrium analysis to predict the volumes that will actually flow in a network for a particular set of supply and demand functions; short-term equilibrium.

5. Demand shift models to predict the long-term changes in the spatial distribution and structure of the activity system as a consequence of the short-run equilibrium pattern of flows, the feedback effect of transportation on land use.

These five are the basic prediction models in transportation. The interrelationships among them are illustrated in Figure 4.

This structuring of the transportation systems analysis problem incorporates several hypotheses. The first hypothesis is that this is a complete and useful summary of the types of options and impacts. The second hypothesis is that it is meaningful to model transportation technology from two perspectives: in terms of the service perceived by prospective users, the supply functions; and in terms of the resources consumed in providing that transportation service, the resource functions. The third hypothesis is that it is useful to separate short-term and long-term equilibrium: the short-term responses of transportation users in a transportation market with the activity system fixed, as represented by the demand functions; and the long-term responses of users and others in a larger, more general market (the total

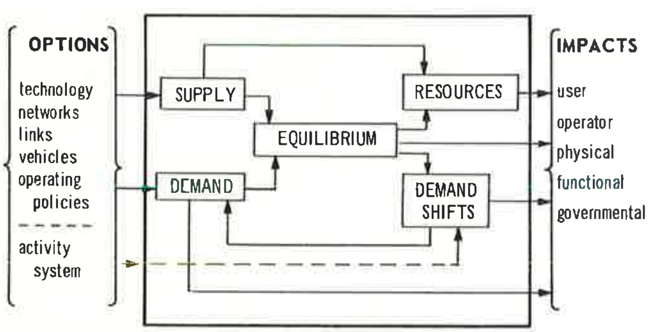



Figure 4. Basic prediction models.

economy), as represented by the demand shifts. The fourth hypothesis, which in a sense is the operational test of the second and third hypotheses, is that valid predictive models can indeed be constructed.

At present, in addition to the modest set of models developed at M. I. T. for the prototype analysis (1, 2), there are several transportation systems analysis activities in which this framework is being applied, implicitly if not explicitly. Three major areas are (a) urban transportation planning, (b) Northeast Corridor Project, and (c) Harvard Transport Research Program.

The prediction portion of the urban transportation planning process, as it has been established in almost all the metropolitan areas of the United States, consists of variants of the following sequence (32): (a) project land use, population, and employment changes; (b) predict trip ends generated in each zone; (c) predict interzonal distribution of trip ends (e.g., using gravity or opportunity models); (d) predict modal split; and (e) predict distribution of flows over the proposed network. As pointed out by Wohl and Martin, Deen, and others (25, 33, 34), there are serious internal inconsistencies in this sequence of steps, from the point of view of an equilibrium analysis. For example, the estimation of trip ends assumes implicitly a general level of service in the system, and a level of service is assumed explicitly for input to the interzonal distribution calculations (e.g., using a gravity model). The last step of the process, traffic assignment, predicts an "actual" level of service, or set of travel times, for flows in the network. However, the initial estimates of level of service used for trip generation and distribution are rarely revised to be consistent with the travel times that are predicted by the traffic assignment.

In spite of these inconsistencies, and other limitations, the structure implicit in urban transportation planning is fundamentally that of the supply-demand equilibrium framework. The supply functions are represented as volume-travel time functions or simply link capacities and travel times. The demand functions are represented by the sequence of predicting trip ends, interzonal distribution of trips, and modal split. The network equilibrium model is the traffic assignment process, with the various capacity-restraint formulations representing explicit attempts to find equilibrium in the network, given fixed demands. (All-or-nothing assignments are obviously very difficult to justify as a meaningful prediction of equilibrium flows.) The resource requirements models are represented in a variety of ad hoc calculations: right-of-way, construction, and operating costs. Demand shifts models are sometimes explicit, as when land use models are used to predict the effects of differential changes in accessibilities on the location of population and economic activities.

Perhaps one of the most significant needs in urban transportation planning is to revise the models and procedures to incorporate the equilibrium approach more explicitly. The present procedures represent a series of ad hoc approximations, as a pragmatic approach to the problem of predicting flows in networks. A new generation of urban transportation models should be developed on the sounder theoretical basis of explicit supply-demand equilibrium analysis.

In the Northeast Corridor Transportation Project's system of models, this structure is more explicit. Although not yet fully operational, this system of models is described in several documents (35, 36). There are explicit interregional and intraregional demand models for passengers and freight (37); technology models to produce supply and resource functions; a network simulator to predict network equilibrium (although it is not yet clear whether this will be a consistent equilibrium, i.e., something other than all-or-nothing); and demand shift models for forecasting changes in interregional and intraregional location and intensities of economic activities as a function of changes in transportation and other factors (38).

The Harvard Transport Research Program models are designed for use in planning investment in transportation in developing countries (39). Several explicit technology models are used for predicting cost-service characteristics of highways, rail, and intermodal transfer points. Demand is derived from a macroeconomic model, containing an interregional input-out model; these are also used to predict demand shifts. Network equilibrium is found with a modified traffic assignment approach.

These are very cursory descriptions of highly sophisticated systems of models; they simply point out how the basic framework outlined above underlies several major transportation analysis efforts. Yet, in spite of the magnitude of effort that has gone into development of the model systems described, there are still major difficulties with implementing this framework in a thoroughly satisfactory way. Constructing demand models is always difficult, particularly because of the lack of good data for model calibration, as was demonstrated in Plourde's efforts to test the Baumol-Quandt model for metropolitan travel (12, 40, 37, 41, 32). The understanding of the structure of transportation supply functions and resource requirements models, particularly with respect to basically new technologies, is very elementary, and many difficult problems remain (6, 31, 42, 43, 44, 45, 46, 47, 48). The problem of computing supply-demand equilibrium in networks was described above. There is a significant amount of effort under way in exploring various computational approaches to this problem, ranging from simulation approaches to mathematical optimization applications for certain special cases (10, 26, 28, 29). The problems become even more complex when equilibrium is to be determined for multiple time periods, where vehicle schedules may be adjusted to meet the transient fluctuations in demand (49).

Predicting demand shifts, that is, changes in land use, population, and the structure of social and economic activity generally, is perhaps the most difficult problem of all. This requires not only understanding the basic structure of the social and economic system and the influence of a variety of policy levers on that system, but also identifying the specific role of transportation in modifying growth patterns. Even in urban land use model development, an area of significant activity in the past, there is still a great gap between what would be desirable in terms of land use prediction models and what has actually been implemented (50, 39, 38).

## The Problem of Search and Choice

The system of basic models discussed in the preceding section is the core of the analysis problem, but not the whole of it. The framework of equilibrium analysis provides a basis for prediction of the impacts associated with a particular set of options. However, there still remain several major issues.

1. The problem of search: how to generate a set of options in the first place, that is, how to formulate a complete, meaningful, well-specified transportation system strategy, which is worth testing in the complex system of prediction models.
2. The problem of choice: given the predicted impacts of several alternative specifications of options, how to evaluate the impacts and choose among the alternatives. The difficulties in choice arise because an essential characteristic of transportation is the differential incidence of its impacts. Some groups will gain from any transportation system change; others may lose. In a realistic network context, there are many user, operator, and other groups, and the interactions among them are complex.

Therefore, transportation choices are essentially sociopolitical choices: the interests of different groups must be balanced. This does not mean that negative impacts of a system on some group are inevitable. It may well be possible, particularly with complementary programs, such as relocation subsidies, industrial development, and job training, to develop a concerted strategy such that no single group is hurt unduly (51). The development of a real-world, implementable transportation plan or policy requires more than just the prediction of impacts of one or two alternatives. Feasible, desirable solutions can be developed only through a careful and sensitive analysis, in a systematic way, of a variety of alternatives and their impacts.

Systematic analysis in transportation requires that a wide variety of alternative options be explored and their differential impacts traced out explicitly. In Figure 5 we illustrate the systematic exploration of the range of options for a single link. We assume that this link is of sufficient length and importance that we in fact have the full range of decision options open to us. We can choose alternative technologies for this link, for example, mass transit, highway, automatic bus on separate right-of-way, or some new technology. We can change the network configuration in this area by introducing

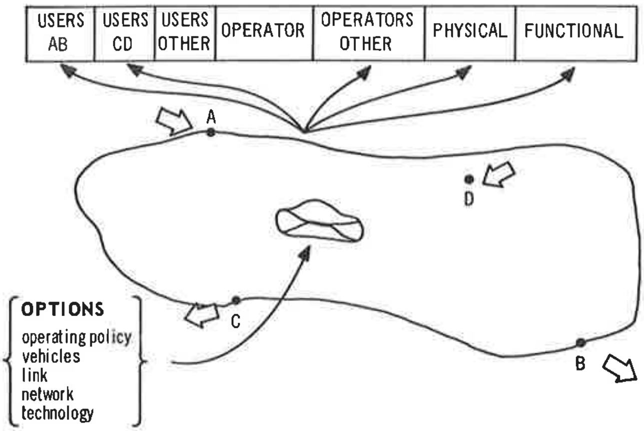



Figure 5. Changes in a network.

several links to supplement this one or change the relationship of this link with other links by adding or eliminating intersections between links. We can change the characteristics of the link itself by changing its physical location or widening it. We also have options regarding the number and type of vehicles that will run over this link, as well as speeds, schedules, stopping times, and fares.

Changes in this particular link will impact differently on each of the groups shown in the figure. In particular, we have identified several user groups, AB and CD, as well as the group of all other users, the operator of this particular facility, the operators of all other facilities, and the physical and functional impacts. In Figure 6, we show how changes in the options for this particular link, as reflected solely in speed over the link, might impact differentially on the various actors. (This is hypothetical; the actual variations would be a property of the network at hand.)

We may have already examined a particular set of options, and thus know its impact on each of the actors. However, in general, we are very uncertain about the changes in these impacts that will occur if we make relatively small changes in these options. This is precisely because of the complexity of the supply-demand interactions in the network.

This is the real difficulty in analysis of a large, complex, multimodal transportation system, such as that of the Northeast Corridor region (Boston to Washington, D. C.). Instead of changes to just one link, we have a potential of changes to, or introduction of, a very large number of links: interstate highways, other highways, conventional jet aircraft routes, and new high-speed ground transportation systems, such as tube trains

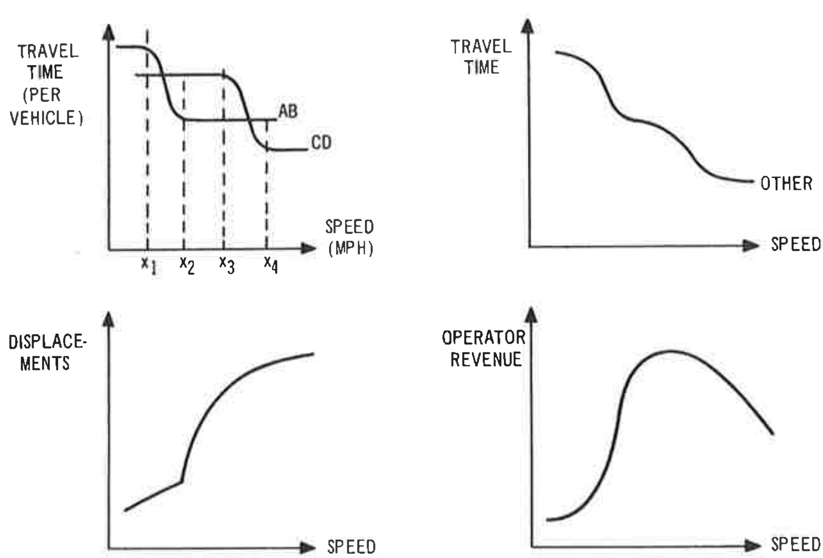



Figure 6. Differential impacts.

or tracked air-cushioned vehicle systems. Furthermore, we are concerned with the impacts on a correspondingly very large number of groups. Clearly, it is not a trivial problem to identify the set of options that has the most desired impacts.

What is required is a systematic exploration of the options that carefully traces out the differential impacts on each group. The set of prediction models serves as a vehicle for assisting in this; but additional techniques are required. Procedures are required for systematically searching out and choosing among transportation alternatives in order to develop equitable transportation system changes that explicitly recognize the differential incidence of impacts.

### Summary

Many options are open for manipulating a transportation system, and many impacts must be considered. To predict the impacts associated with a particular set of options requires prediction of the corresponding pattern of flows that will occur in the multimodal transportation network. The basic logic of this prediction is that of supply-demand equilibrium. Making this logic operational in the network context requires a complex system of models. Thus, to predict the impacts associated with only one alternative transportation system requires significant effort.

The problem of search is to generate alternative transportation systems sufficiently attractive to be worth testing by predicting the impacts. The problem of choice is to rank the alternatives based upon evaluation of their predicted impacts. Systematic analysis in transportation planning requires that a wide variety of alternative options be explored and their differential impacts traced out explicitly.

Thus, the task is to systematically search out and choose among transportation systems alternatives, with careful consideration of the differential impacts on different groups, in a network context. Therefore, inevitably, the hard realities of alternative transportation technologies and complex computer models lead into the "soft," very subtle differences of impacts: Who gets better service with a prospective change? Who gets worse? Who pays? And how do they all change the way they work, live, and relax as a result?

## THE ANALYSIS PROCESS

The topic of the preceding section was primarily the scope of the substantive problem of transportation systems analysis. Underlying this discussion was a second theme: a concern with how this substantive problem should be analyzed. Clearly, it is not sufficient to have a set of computer models for predicting impacts of a given alternative; there are a wide variety of issues involved just in the problem of generating good alternatives to test. Furthermore, if we examine the role of transportation systems analysis in the context of the political process within which transportation planning inevitably takes place, then we see that there is an even greater order of complexity involved. Thus, in order to understand how to conduct systematic analyses of transportation policy, we must step back from the substantive problem of transportation and focus on the process of analysis.

We do this, first, by pointing out some of the major issues by reviewing the rational model and its limitations. Then, we outline a conceptual model stimulated by this review of transportation systems analysis in the context of a political world. In the next section we then show the application of this PSP model to transportation systems analysis. This sets the stage for a discussion of specific operational techniques derived from the PSP model.

### The Rational Model and Its Limitations

The essence of the rational model of decision-making is expressed in the very common prescription of systems analysis (52, 53, 54, 55, 56) that includes the following steps: (a) define objectives and formulate a utility function; (b) enumerate all the possible alternative actions; (c) identify the consequences of each action; (d) evaluate the consequences in terms of the objectives via the utility function; and (e) choose that action that best achieves the objectives.

This "synoptic" model (57, 58, 59) is very limited in its application to complex public policy questions, such as transportation, for many reasons:

1. We can never know completely all the alternatives.
2. We can never define all the relevant objectives consistently and completely. First, there are too many points of view, "actors," to get agreement on objectives (though as Lindblom points out, we may get agreement on actions without agreement on objectives). Second, objectives are difficult to formulate in the abstract, and they will be substantially revised and clarified through examination of the consequences of specific alternatives (58).
3. We can never completely identify the relative values of all possible combinations of the various objectives; that is, we can never get a fully defined utility function. Therefore, evaluation of the consequences of an action is not so simple. Most naturally, we prefer to examine alternatives explicitly and to evaluate the incremental differences between them (58).
4. There will always be uncertainty in the prediction of consequences. Many relevant consequences will be left out; because of the open nature of the socioeconomic system (i.e., not a closed system), consequences of actions such as transportation will "spill over" into other areas (e.g., functional and governmental impacts). Further, we can expect to find that actions that we have implemented do lead to unanticipated consequences (60).
5. The costlines of analysis are a severe constraint. Generation of alternative actions, formulation of objectives, anticipation of the consequences of actions (prediction), and determination of the most desired action all take resources. The resources of analysis are dollars, time, manpower, and computing capability (computer time). Analysis of policy alternatives generally suffers under very severe constraints on these resources, and so analysis is inevitably incomplete.
6. Analysis is dynamic. Problems are never solved completely; massive changes in the system, such as transportation, generally take time for implementation. As specific changes are implemented, the context of the problem will change. Within the analysis process itself, the conceptions of the problem held by analysts and decision-makers will evolve. Initial statements of objectives will be revised as successive alternative actions are generated and examined. Examination of previous alternatives will suggest new ones for analysis.

Perhaps the best way of summarizing the limitations of the rational model is that "... in the face of man's limited capacities, it offers simply a prescription: 'Be comprehensive!' " (58). The comprehensive ideal fails to accept the realities of policy analysis: the costs of analysis, and the inability for cost and cognitive reasons of ever being comprehensive.

This does not mean, as some would argue, that systematic analysis must be discarded altogether, rather, that the simple five-step model of the analysis process must be replaced by a more subtle structure.

## PSP: A Dynamic Model of Decision-Making

As a partial answer to the limitations of the rational model described above, a more complete model of decision-making is necessary. Such a model has been formulated, the problem-solving process (PSP) model (15, 4). Only a few of its major characteristics will be summarized here.

The first important characteristic of the PSP model is the role of time. The analysis process itself takes place over real time; to develop, evaluate, and choose among alternatives takes time. Further, the analysis process is itself embedded in the larger evolutionary process of the real-world system of interest; the actions selected by analysis are implemented, their results observed, and new analyses lead to new, revised actions.

The second important characteristic of the PSP model is the distinction between generating and choosing among actions. We emphasize this distinction by defining search and selection as the procedures that perform these functions. Search designates any procedure used to produce one or more alternative actions. Search may be intuitive, as in the sense of design, or may be formalized, as in a linear programming model. Selection designates the process of choosing among several alternative actions. The input to selection is a set of alternative actions. The output of selection is a preference ordering, or ranking, of the actions by desirability.

To actually accomplish selection requires three basically different kinds of procedures. Prediction models are used to anticipate the consequences that an action would have if implemented in the real world, for example, to predict volume of travel on a particular transportation link. Evaluation procedures operate upon the predicted consequences to yield statements of the valuations, or relative desirabilities of those consequences of a particular action, for example, the values of user costs and benefits associated with a particular flow volume on a link, or the relative desirability of a particular regional growth pattern. Because all predicted consequences cannot be represented adequately by a single measure of value, or valuation, we do not assume that evaluation summarizes all the valuations into such a single measure. For example, we do not assume that construction dollars, loss of recreation land, and regional development patterns can all be lumped into a single measure of value, such as dollars or some overall utility measure. Therefore, after evaluation there must be choice. In choice, two or more actions are compared on the basis of the set of valuations for each—dollar costs, recreation land acreages, quality of regional pattern—and a decision made about the rankings of the actions. Choice is difficult, but necessary. (At this point, we will not distinguish between choice executed by the analytical staff and choice executed by a small group of decision-makers or the larger political process.)

The third important characteristic is a distinction between the state of the analysis process at any particular time, and the procedures that may be used to change that state. The state of the process expresses the analyst's current view of the problem. The problem-solving system contains a variety of procedures to be used in the problem-solving process when and as appropriate. Each time a procedure is used it changes the state of the process in a way appropriate to the procedure: the use of a procedure changes the analyst's view of the problem.

From the point of view of our present discussion, the major variables describing the state of the process are the actions, A, the goals, G, and the current ranking of the actions, R. The major procedures for changing that state are search, selection, and goal formulation and revision. Additional state variables include consequences, valuations, raw information, and probability distributions over uncertain variables. Additional procedures include information analysis, model construction and revision, decomposition and restructuring procedures, and metaprocedures (15). All these procedures involve some reference to the current set of goals, G.

The basic view of the problem-solving process that this implies is shown in Figure 7. Alternative actions are generated, and then a preference ordering is established over those alternatives. If the most desirable alternative is sufficiently attractive, then the problem-solving process ceases and that most desirable action is implemented in the real world; if not, then search is repeated and new actions are generated. These new actions may or may not be related to the previous actions. The sequence is repeated again and again, until finally there is one action sufficiently attractive for implementation in the real world.

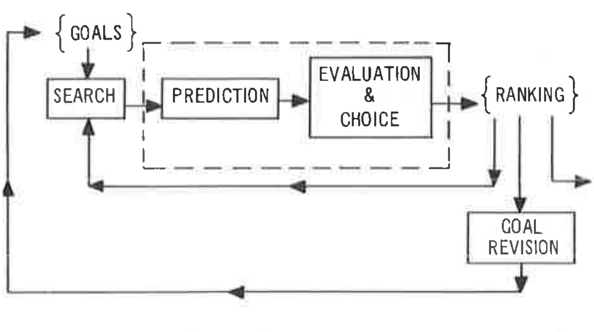



Figure 7. Basic cycle.

This image of a trial-and-error process, basic to the PSP concept, is completely contrary to the image of a problem for which the optimal solution is obtained directly by solving a mathematical (or other) model. Such optimizing methods do have an important role in the broader PSP, but such real problems as transport systems planning are too complex for such techniques to carry the whole burden. A particular optimizing method corresponds to one search-and-selection sequence; but many kinds of search and selection procedures are required in addressing the total problem.

## Evolution of Actions and Goals

The focus of PSP is on actions. Because search and selection procedures concern the basic processes of generation and selection of actions, these procedures are at the heart of the PSP. However, there are a variety of other activities that must occur in PSP to allow search and selection to operate, and to revise the context in which they operate (15). Particularly important are goal formulation and revision procedures.

The purpose of these procedures is to formulate and revise the statement of goals, G, as new actions are generated and their consequences are examined. That is, each cycle of search and selection potentially may trigger goal revision; for example,

1. An initial statement of goals is formulated.
2. The search procedures are executed, one or several times, producing one or more alternative actions.
3. The selection procedures are executed, identifying and evaluating the consequences of the actions and their performance with respect to the goals; the result is a ranking of the alternatives.
4. In the basic cycle, steps 2 and 3 may be repeated a number of times, until an action sufficiently desirable for implementation has been found, or analysis resources have been exhausted.
5. The mutability of goals must be recognized, however, and so step 3 or even step 2 may be followed by goal revision.

Goal revision may follow search immediately, particularly when the results of search are either very disappointing (it proves very difficult to generate actions that achieve the goals) or very successful (the goals are set so low it is not at all difficult to achieve them).

By this simple model, the analysis process is typified by continued and parallel evolution of the set of actions, A, and the set of goals, G. The ranking of the alternative actions will change, not only by addition of new actions to the set, but also by revision of the goals.

Search and choice interact heavily in the sociopolitical arena. To account for this interaction, a model of the analysis process more subtle than the static, rational model is required. The evolution of goals as well as actions is an important facet of the more complete model.

## Implications of the PSP Model

In this brief exposition, we have touched only quickly upon the characteristics of a more general model of the problem-solving process. Let us now explore some of its implications. Once we shift from a static conception of the analysis process to an evolutionary one in which actions and goals change over time, our perception of the interaction of search and choice changes in fundamental ways.

For one thing, we need no longer search desperately for a utility, or social welfare, function. Such functions are used to reduce a multi-dimensional set of goals to a single dimension (for example, benefit-cost ratio). In an evolutionary process, we can accept a less well-structured set of goals, because all that we require at any time is sufficient information about goals to reach a choice over alternatives, not a function over all conceivable combinations. (In fact, we may not need to have a completely consistent goal structure to produce agreement over alternatives; as Lindblom points out, actors may reach agreement on actions even though they are striving for different objectives.) Thus, a much looser, more flexible structure of goals is appropriate and useful. The concept of a "goal fabric" has been proposed to serve this purpose; it is discussed later in this paper.

A second important implication is that alternatives need not be single, massive-system proposals, but can and should be formulated as staged strategies over time. The typical urban transportation study chooses among a small number of alternative transportation system plans for a target year (1985, say). Instead, there should be a much richer number of smaller actions, each one being the building of a specific part of the network in a particular year (or other options, such as buying additional transit cars). Then the major alternative 1985 plans could be composites of a number of the specific time-staged facility actions. However, the whole approach to analysis would be different. Instead of choosing among single packages, the emphasis would be upon choosing among sequences of actions staged over time. In this way, there would be a great deal of flexibility for revision of both actions and goals as each stage of the system selected is implemented. Furthermore, the decision about appropriate facilities to construct at each successive time period could be revised as changes in the real world are observed, or changes in the goals or the available actions as new technologies are developed. Models for searching out and choosing among sequential decisions in transportation planning are under development (61).

A third implication is more emphasis on the value of information. That is, instead of collecting all the information necessary for constructing demand models and the other analysis models in one single survey, there can be a much more efficient use of resources through continuous sampling over time, with a flexible readjustment of data acquisition as new issues are identified for study. This is an important consequence of the time-staged strategy approach. Further, this implies that there should be an economic analysis of the value of information in its relevance to the search and choice issues at hand; as the actions and goals change over time, the value of different types of information will also change. Models for optimal information collection strategies in networks are under development (16).

Finally, and perhaps most important, the evolutionary image of the analysis process leads to a major new perspective on the relationship between the analysis team and the political environment (51).

The evolution of actions and goals takes place at several different levels. First of all, consider the technical analysis team actually doing transportation systems and related studies for a particular area. Within this team, the sets of actions and goals will evolve fairly rapidly: the team is engaged in day-to-day development and testing of alternatives, and as it learns more about the problem, it will be almost continuously revising its assumed goals. At a second level, this analysis team will interact periodically with the political decision-makers or other responsible public or private officials for whom the analysis team is acting as staff. As a result of these (more or less frequent) interactions the actions and particularly the goals will be further refined and revised. These decision-makers, in turn, will interact with the body politic: the variety of actors, individuals and interest groups, who comprise the full set of interests impacted by transportation systems alternatives. As a result of their interactions with the body politic, the decison-makers will revise their conceptions of actions and, more particularly, goals, and will pass these revised conceptions on to the analysis team. But also, the results of the analysis team, as communicated through the decision-makers to the public at large, will help to change and broaden the perceptions of the decision-makers and of the body politic.

The interactions in search and choice among analysis team, decision-makers, and polity should be exploited explicitly. Perhaps the most important role of a technical analysis effort is to clarify public objectives, even more than the development and implementation of specific actions. For example, one of the major contributions of the highway transportation program may have been to create a public awareness of the choice issues that need to be addressed in the core of the metropolitan area. The threat of highways through the centers of cities has set in motion political forces that have helped to raise serious discussion about the competing objectives of groups in the metropolitan area and have stimulated the search for new transportation technologies as well as new methods of highway planning.

## EXAMPLE: THE "PROTOTYPE" ANALYSIS

The preceding sections have laid out the scope of a comprehensive systematic analysis of transportation, and a theoretical model of the analysis process. The feasibility and utility of these ideas have been demonstrated by conducting a prototype analysis of passenger transportation in the Northeast Corridor region. (Clearly, the models could be used for many other contexts as well.) To the maximum extent feasible, realistic data were used. The result demonstrates how the analytical approaches and techniques can be applied to improve policy decisions. However, the result is not of sufficient detail or comprehensiveness to be used for policy decisions without further calibration and modifications of the models and substantial additional data.

### TRANSET II, a Laboratory

To do this prototype analysis, it was necessary to develop a set of models for transportation systems analysis. These models, in the form of computer programs, provided a "laboratory" for experiments in systematic analysis in transportation.

This laboratory was implemented as TRANSET II, a new subsystem of ICES, the Integrated Civil Engineering System (62). This subsystem is a problem-oriented, command-structured language for transportation system analyses, and thus is designed for ease of use by analysts without computer training. The development of TRANSET II is based upon additions and changes to an earlier subsystem for urban transportation analysis, TRANSET I (63). For example, to create a new regional transportation network by adding a link to a network previously stored in the computer, the analyst might give the computer this problem-oriented language command:

```
MODIFY NETWORK 'BASE' FORMING 'NEWRAIL' ADD LINK FROM
56 TO 97, DISTANCE 37.2, LANES 6, VOLUME/DELAY 4.
```

In this example, BASE is the previously stored network, NEWRAIL the name to be given to the new network. The modification consists of adding a link from node 56 to 97, with the indicated length, number of lanes, and supply function (volume/delay curve number 4). Such problem-oriented language capabilities enable the analyst to use the computer models in a much more flexible and efficient manner than do the more traditional forms of programs.

As a system of computer models, TRANSET II provides the capability to analyze transportation problems by predicting supply and demand equilibrium in a multimodal transportation network (2). Some of the particular features of TRANSET II are (a) the capability to express transportation policy options through technology choices, network configuration, link characteristics, fares, frequency of service, subsidy, and tax policy; (b) the use of the Baumol-Quandt abstract mode demand model; (c) incremental assignment techniques as an approach to calculating equilibrium; and (d) explicit evaluation routines for tracing out impacts on different groups.

With the cooperation of the Northeast Corridor Project of the U. S. Department of Transportation, data were obtained through which the Northeast Corridor network was modeled in two forms: 5-district and 29-district versions with the networks modeled at corresponding levels of abstraction. The resulting models then served as the basis for a number of analyses. The TRANSET II model system was used as a laboratory to conduct numerous experiments with the 5-district data. These experiments demonstrate (a) the feasibility of developing a supply-demand equilibrium model for transportation analysis; (b) the difference between equilibrium and nonequilibrium approaches to the problem; (c) how different options and impacts can be included in a single model so that their interactions can be explored systematically; (d) how trade-offs between options can be explored; (e) the technique of differential impact analysis through tracing out impacts among different actors as the options are varied; (f) sensitivity analyses; and (g) effects of alternate time-staging of actions.

Space does not permit summary of all of these experiments here (1, 2). The major points we will discuss in the following sections will be the applicability of the PSP model, as illustrated by TRANSET II, and the systematic analysis of options and impacts.

## Applicability of the PSP Model

To see the applicability of the PSP model to transportation systems analysis, we can refer to the specific problem-oriented language capabilities of TRANSET II, as now operational.

Search—No explicit search procedure is provided in TRANSET II at this time, although several are under development. At present, the analyst must generate a policy alternative, execute search, through his own judgment. He may generate a completely new action, or use parts of one or more previously generated actions as stored in the computer files (i.e., on disc). If he uses stored components of an action, he may modify them if he wishes. One particularly powerful capability of TRANSET II is the ability to name data files. Thus, the analyst may store several transportation networks under arbitrary names such as '1956-1,' 'RAIL-2,' and 'HWAY.'

To generate a completely new transportation system alternative, or new components of an alternative, to save portions or all of an alternative in computer storage, and/or to create a new alternative through modification of a previously stored component, he uses the following commands:

1. Transportation Options
   a. READ NETWORK, for general network characteristics
   b. LINKS, for network connectivity and link characteristics
   c. READ VOLUME DELAY SET, for generalized supply functions
   d. INPUT MODAL SERVICE DATA, for interzonal fares and frequencies for each mode
   e. INPUT MODAL COST DATA, for cost parameters for each mode
2. Activity System Options
   a. INPUT DISTRICT DATA, for population, incomes, and holding capacities, for each zone
   b. INPUT MODAL SPLIT PARAMETERS, for demand model parameters

In addition to specifying a completely new alternative, it is also possible to generate an action by using portions of another action previously stored in the computer, as follows:

3. To modify an action, previously stored on secondary storage as a permanent file, to create a new one
   a. MODIFY NETWORK + name of network to be changed

| | |
|---|---|
| ADD LINK<br>DELETE LINK<br>CHANGE LINK } | specification of changes |

   b. REVISE MODAL DATA + name of data to be changed

| | |
|---|---|
| MODE COST<br>MODE FREQUENCY } | specification of changes |

   c. REVISE DISTRICT DATA + name of data to be changed

| | |
|---|---|
| DISTRICT | specification of changes |

Selection-Prediction—The prediction procedures of TRANSET II are based upon the supply-demand equilibrium concept. The commands to accomplish prediction of the consequences for a specific alternative are as follows:

1. PREDICT POTENTIAL TRIPS, generate estimated trip demands
2. PREDICT ACTUAL TRIPS, predict actual network equilibrium flows
3. PREDICT DISTRICT DATA, predict future population and income for each zone, based upon predicted network flows

Selection-Evaluation—The evaluation components of TRANSET II are relatively simple. For any particular alternative action, its predicted consequences can be displayed in a variety of ways for intuitive evaluation by the analyst. User, operator, and government costs can be computed and aggregated in a variety of ways through the EVALUATE COSTS commands. Accessibilities can also be evaluated as measures of functional impacts (i.e., potential changes in the activity system). There are no capabilities at this stage for predicting physical or governmental impacts explicitly.

1. Display flow pattern consequences
   a. REQUEST FINAL LINK DATA, MINIMUM PATHS, TRAVEL TIMES, SYSTEM TRAVEL DISTRIBUTION, INTERZONAL TRIPS
   b. PRINT TRIP MATRIX
   c. In graphical form, by plotter or other display device, PLOT NETWORK; DISPLAY LINK VOLUMES, TRAVEL TIMES, SPEEDS; DISPLAY INTERZONAL VOLUMES, TRAVEL TIMES, SPEEDS
2. EVALUATE COSTS, for user, operator, and governmental impacts
3. EVALUATE ACCESSIBILITY, for functional impacts

Selection-Choice—Choice involves the comparison of alternatives to determine a preference ordering. In TRANSET II, no automatic choice capability is provided. However, a very simple but powerful set of commands provides the analyst information that is extremely useful in his judgmental decision about preferences between alternatives. These commands compare two alternatives, displaying the differences between them. Then the analyst can examine the incremental differences between the two alternatives, as well as the absolute levels of the impacts.

1. COMPARE TRIPS, for summary of the differences in flow volumes between two alternatives
2. COMPARE SURPLUSES, for differences in user benefits as provided by consumer surplus measures
3. COMPARE ACCESSIBILITIES, for differences in functional impacts

The problem-oriented language capability is particularly useful in commands such as COMPARE TRIPS, ALTERNATIVES 'AIR' AND 'RAIL'.

Utility Commands—In addition to the above, there are available in TRANSET II a variety of utility commands for editing data, obtaining intermediate results during the course of the computations, or filing data on computer disc storage (2).

The use of TRANSET II commands is illustrated in Figure 8, which shows typical computer input and output for a single selection cycle. Substantial improvements and additions to these capabilities have been made since the completion of TRANSET II (64).

## Systematic Exploration of Options and Impacts

Often, when computers are used for problem-solving, far too much emphasis is placed upon getting the computer model running; not enough attention is given to how the model is to be used once it is running. A major objective of the prototype analysis was to demonstrate how prediction models should be used in transportation systems analysis.

The basic issues are these: What different combinations of options can achieve the same impacts? What different combinations of impacts (on different groups) can be achieved (by any combination of options)? TRANSET II was used as a laboratory to conduct a number of experiments to trace out such trade-offs. These are illustrated in Figures 9 through 12. The relationships in these figures were derived from data produced by a series of computer runs. In these runs, three levels of fare and three of frequency were explored, resulting in nine combinations. This sample provided the basis for inferring the relationships shown in the figures (except Fig. 12, as noted).

Figure 9 shows how, as frequency of service between two points is increased, fare must also be increased to maintain the same volume of trips in the system (e.g., 1.10 = 1,100,000 trips per day). Thus, this figure shows how two options—fare and frequency over one single route—must be manipulated together to achieve a constant level of one impact—total trips. In Figure 10, this same approach is extended to consideration of

```
$     EXAMPLE RUN 5 -- EQUILIBRIUM AND IMPACTS
TRANSET
$     RETRIEVE CATA PREVIOULSY FILED
LOAD NETWORK '5X3,1965'
LOAD TRIPS 'BASE'
LOAD VOL/DELAY SET '3MODES,2'
LOAD GEN/RATE SET 'BASE'
GET DISTRICT DATA '5CENTERS'
GET MODAL DATA '5X3 NO.1'
$     DEFINE VARIABLE INCREMENT
USE INCREMENT 50 PERCENT
$     SPECIFY NAME OF NETWORK WHICH WILL INCLUDE THE EQUILIBRIUM CONDITIONS
SAVE ASSIGNMENT RESULTS IN NETWORK 'BASE'
$     START THE INCREMENTAL APPROACH TO EQUILIBRIUM COMPUTATION
PREDICT ACTUAL TRIPS

                    THE NETWORK IS COMPLETELY ASSIGNED AFTER   420 ITERATIONS
                          TIME USED SINCE START OF RUN IS    1.65 MINUTES

                    ASSIGNMENT RESULTS HAVE BEEN STORED ON DISK IN NETWORK BASE

$     OUTPUT ACTUAL AND ESTIMATED TRIP DEMANDS
$     CONTRARY TO HEADING, TRIPS ARE IN PASSENGERS PER DAY
PRINT TRIP MATRIX 'BASE'
                                  INTERZONAL TRIP STATUS

                                TRIP MATRIX NAME IS BASE

                 THIS IS THE TABLE TO CONVERT MACHINE ZONE NUMBERS TO  USER ZONE NUMBERS
```

| MACHINE NUMBERS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 USER NUMBERS | 1001 | 1002 | 1003 | 1004 | 1005 | 2001 | 2002 | 2003 | 2004 | 2005 |
| 10 USER NUMBERS | 3001 | 3002 | 3003 | 3004 | 3005 | | | | | |

```
                        THE FOLLOWING MATRIX USES MACHINE ZONE NUMBERS.

$     OBTAIN COST DATA BASED ON ACTUAL FLOW PATTERN
EVALUATE COSTS TIME VALUE 2.00 WAIT FACTOR .50
                        COST SUMMARIES FOR NETWORK BASE
                           AND TRIP MATRIX BASE
```

DAILY USER DATA BY MODE AND ORIGIN

| MODE | ORIGIN DISTRICT | TOTAL TRIPS (PASS/DAY) | USER FARES TOTAL ($) | USER FARES AVERAGE ($/TRIP) | USER TRAVEL TIME TOTAL (HOURS) | USER TRAVEL TIME AVERAGE (MIN/TRIP) | USER WAIT TIME TOTAL (HOURS) | USER WAIT TIME AVERAGE (MIN/TRIP) | WEIGHTED COSTS TOTAL ($) | WEIGHTED COSTS AVERAGE ($/TRIP) |
|---|---|---|---|---|---|---|---|---|---|---|
| AIR | 1 | 1552 | 0.28861E 05 | 18.60 | 0.39947E 04 | 154. | 0.99343E 03 | 38.41 | 0.38837E 05 | 25.02 |
| AIR | 2 | 1618 | 0.24215E 05 | 14.97 | 0.36320E 04 | 135. | 0.10535E 04 | 39.07 | 0.33586E 05 | 20.76 |
| AIR | 3 | 3054 | 0.46437E 05 | 15.21 | 0.69187E 04 | 136. | 0.17284E 04 | 33.96 | 0.63731E 05 | 20.87 |
| AIR | 4 | 974 | 0.14885E 05 | 15.28 | 0.22902E 04 | 141. | 0.11386E 04 | 70.14 | 0.21742E 05 | 22.32 |
| AIR | 5 | 1844 | 0.31603E 05 | 17.14 | 0.44475E 04 | 145. | 0.10327E 04 | 33.60 | 0.42564E 05 | 23.09 |
| RAIL | 1 | 1426 | 0.13982E 05 | 9.81 | 0.61999E 04 | 261. | 0.85517E 03 | 35.98 | 0.28092E 05 | 19.70 |
| RAIL | 2 | 4676 | 0.31738E 05 | 6.79 | 0.13401E 05 | 172. | 0.21759E 04 | 27.92 | 0.62891E 05 | 13.45 |
| RAIL | 3 | 5922 | 0.41680E 05 | 7.04 | 0.18018E 05 | 183. | 0.31044E 04 | 31.45 | 0.83925E 05 | 14.17 |
| RAIL | 4 | 2020 | 0.15598E 05 | 7.72 | 0.65330E 04 | 194. | 0.16744E 04 | 49.73 | 0.32013E 05 | 15.85 |
| RAIL | 5 | 672 | 0.71588E 04 | 10.65 | 0.34545E 04 | 308. | 0.83411E 03 | 74.47 | 0.15736E 05 | 23.42 |
| ROAD | 1 | 7968 | 0.36145E 05 | 4.54 | 0.42515E 05 | 320. | 0.99600E 03 | 7.50 | 0.12317E 06 | 15.46 |
| ROAD | 2 | 18470 | 0.58890E 05 | 3.19 | 0.90952E 05 | 295. | 0.23087E 04 | 7.50 | 0.24541E 06 | 13.29 |
| ROAD | 3 | 29900 | 0.95604E 05 | 3.20 | 0.13415E 06 | 269. | 0.37375E 04 | 7.50 | 0.37138E 06 | 12.42 |
| ROAD | 4 | 22580 | 0.63044E 05 | 2.79 | 0.84587E 05 | 225. | 0.28225E 04 | 7.50 | 0.23806E 06 | 10.54 |
| ROAD | 5 | 13004 | 0.45721E 05 | 3.52 | 0.59450E 05 | 274. | 0.16255E 04 | 7.50 | 0.16789E 06 | 12.91 |

DAILY USER DATA BY MODE

| MODE | TOTAL TRIPS (PASS/DAY) | USER FARES TOTAL ($) | USER FARES AVERAGE ($/TRIP) | USER TRAVEL TIME TOTAL (HOURS) | USER TRAVEL TIME AVERAGE (MIN/TRIP) | USER WAIT TIME TOTAL (HOURS) | USER WAIT TIME AVERAGE (MIN/TRIP) | WEIGHTED COSTS TOTAL ($) | WEIGHTED COSTS AVERAGE ($/TRIP) |
|---|---|---|---|---|---|---|---|---|---|
| AIR | 9042 | 0.14600E 06 | 16.15 | 0.21283E 05 | 141. | 0.59466E 04 | 39.46 | 0.20046E 06 | 22.17 |
| RAIL | 14716 | 0.11016E 06 | 7.49 | 0.47606E 05 | 194. | 0.86439E 04 | 35.24 | 0.22256E 06 | 15.13 |
| ROAD | 91922 | 0.29940E 06 | 3.26 | 0.41177E 06 | 269. | 0.11490E 05 | 7.50 | 0.11459E 07 | 12.47 |
| TOTALS | 115680 | 0.55556E 06 | 4.80 | 0.48066E 06 | 249. | 0.26081E 05 | 13.53 | 0.15690E 07 | 13.56 |

YEARLY COSTS AND REVENUES BY MODE

| MODE | TOTAL TRIPS | USER FARES | TOTAL USER COSTS | OPERATOR'S PROFIT | GOVERNMENT REVENUE | OPERATOR'S PROFIT PER PASSENGER | GOVT. REVENUE PER PASSENGER |
|---|---|---|---|---|---|---|---|
| AIR | 3300330 | 0.53290E 08 | 0.60138E 08 | -0.26871E 08 | -0.72335E 08 | -8.14 | -21.92 |
| RAIL | 5371340 | 0.40207E 08 | 0.66797E 08 | -0.55785E 08 | 0.0 | -10.39 | 0.0 |
| ROAD | 33551520 | 0.10928E 09 | 0.34378E 09 | 0.15690E 08 | 0.0 | 0.47 | 0.0 |
| TOTALS | 42223200 | 0.20278E 09 | 0.57270E 09 | -0.66966E 08 | -0.72335E 08 | -1.59 | -1.71 |

Figure 8. TRANSET II commands illustrated by typical computer input-output.

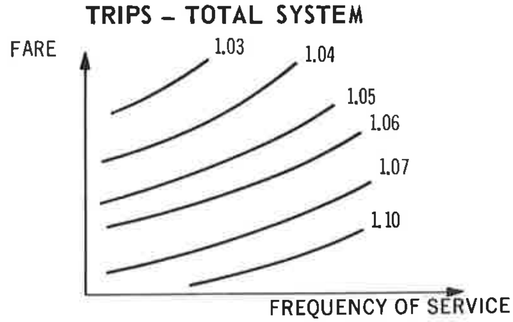



Figure 9. One impact.

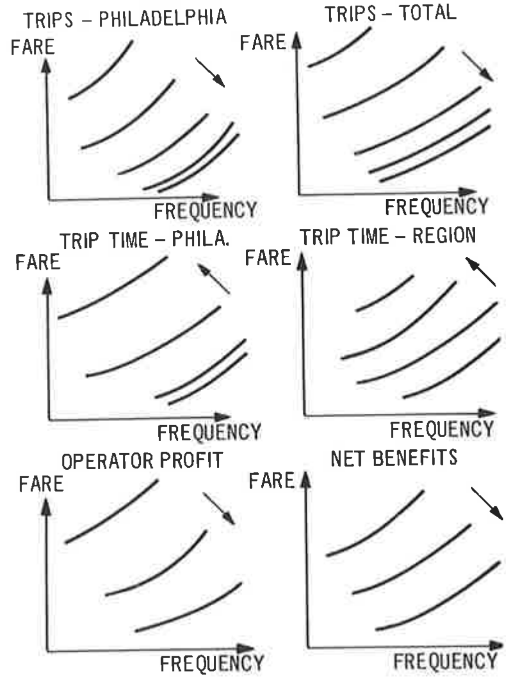



Figure 10. Differential impacts.

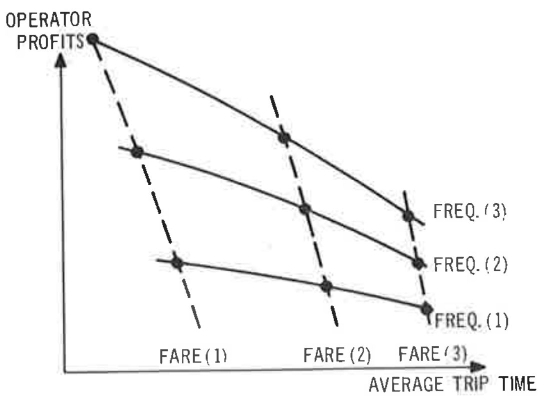



Figure 11. Trade-offs among impacts.

several different impacts simultaneously: total trips in the system; total trips to or from Philadelphia; average trip time, all trips; average trip time, Philadelphia trips; operator profit; and net benefit to the region as a whole.

In Figure 11 the impacts are shown on the axes. Thus, the same data are now shown as trade-offs among impacts, in this case, operator profit and average trip time (a user impact). All other things being equal (they are not), the point most to the upper left would be most desirable. Evaluation and choice deal with such trade-offs among impacts; whereas for search, trade-offs among options are needed.

In Figure 12, a third option, in addition to fare and frequency, is added: level of investment in the network. There are now 27 data points: 3 different levels of network and 9 combinations of fare and frequency for each. From this sample, we can now infer the locus of the most desirable alternatives, as indicated by the dotted line (again, everything else assumed equal!).

These examples illustrate how a number of runs of the computer models can be used to generate information. In this way, trade-offs among options and among impacts can be systematically analyzed. As these trade-offs are being developed in the analysis process, the information that is obtained can also be useful for search and choice. As systematic relations among the options are perceived, search procedures can concentrate on generating alternatives in the most interesting areas of the space of possible options. As achievable trade-offs among impacts are identified, the key issues of choice become clearer. It may be relatively easy to find options that produce desirable impacts for each of one group of actors; but there may be unavoidable conflict in the impacts that are achievable between two other groups of actors, for example, decreased travel time for suburban residents can only be achieved by displacing families from central city homes by freeway construction. It is precisely these differential impacts that must be traced out.

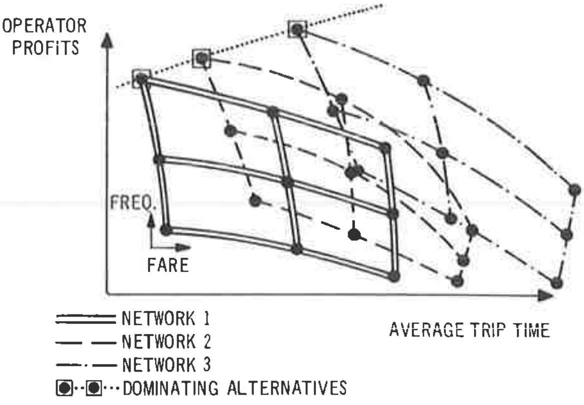



Figure 12. Dominating trade-offs.

## Conclusions From the Prototype Analysis

The prototype analysis demonstrated the feasibility of network equilibrium analysis and the applicability of the PSP model. It illustrated the systematic exploration of options and impacts. However, it also emphasized the following problems, which are explored briefly in a later section:

1. Systematic analysis generates, and requires, a large volume of information from a number of computer runs. How can the analyst deal with, and understand, this model-generated information (particularly, to be able to infer such relationships as illustrated in the figures)?
2. If differential impacts among a variety of groups are to be considered explicitly, how can evaluation and choice be reasonably done without aggregating all the impacts indiscriminately?
3. What kinds of procedures can be developed to assist in search?
4. Each run of the computer models costs time, money, and other resources. Must all search and selection cycles be at the same level of detail? If not, what errors are introduced?

## EXTENSIONS OF THE PSP MODEL

In the prototype analysis, the set of models and data was not highly refined. However, it illustrated some of the major issues in a systematic exploration of alternative transport systems. While showing the feasibility of systematic analysis, the prototype analysis also raised a number of issues about the difficulties in doing such an analysis. In this section, we will briefly summarize some of the techniques and approaches now under development to address these issues.

### DODO: A PSP-Oriented System (4)

The prototype analysis uses a simplified set of models and a very simplified representation of the Northeast Corridor transportation system. Yet, even with this simplicity, the analyst finds it difficult to deal with the large volume of information produced in doing a systematic analysis of the alternatives. If only twenty different actions have been generated and compared, the analyst has real difficulty understanding the difference and similarities among the actions: Which actions are basically different in their impacts; which are very similar? What are the feasible trade-offs in impacts, and which actions produce the most desirable combinations? Even a single run of the model system results in large masses of data that are difficult for the analyst to comprehend. Given the results of analyses of a series of complex transportation systems alternatives, how can the analyst understand the differences between these alternatives, in order to establish a preference ordering among them and in order to identify fruitful areas to search for alternatives even more promising than those he has already examined?

What is needed is a way of storing all the relevant information generated in a series of model runs, such that questions meaningful to the decision problem can be asked of the data. Some of these questions can be identified a priori and built into the system; but many significant questions will occur to the analyst only as he is examining the specific data of a series of runs. Therefore, the information system must be designed for interactive use with flexible query capabilities.

The concept of DODO was developed in response to this need. DODO is an information system intended to provide the decision-maker and analyst with the capability to analyze and structure the large amount of data that may be generated in the analysis of a complex problem. The name DODO reflects this objective: Decision-Oriented Data Organizer. An initial operational version of DODO has been developed in the context of the TRANSET II subsystem of ICES, as developed and modified for the prototype analysis. Later versions of DODO will be more general, applicable to many other design problems as well as transportation.

The design of DODO is based upon the module suggested by the PSP model. This basic decision module (BDM) consists of the quintuple (A/P/C/U/V): action, A, consequences, C, valuations, V; with consequences conditional on a (data) parameter set, P, and valuations conditional on a set of (partial) utilities, U.

In an actual problem, each of these files may comprise large volumes of data. For example, in the prototype analysis, each action, A, corresponds to specification of the options of technology mix, network, fares, frequency of service, and costs and subsidies. The consequences, C, and valuations, V, may also be large files of data, e.g., the file of travel time for system users by trip purpose by zone pair by mode, for 5 purposes, 4 modes, and 50 zones contains 50,000 items, which can be aggregated a number of ways.

In outline, the basic capabilities of DODO are as follows:

1. It is designed as a command-structured, problem-oriented language, a subsystem of ICES. Thus, the commands are easy for the nontechnical analyst to use.
2. The basic files of the system are organized in terms of the basic decision module (BDM); i.e., the basic record is the quintuple (A/P/C/U/V).
3. Each step in the analysis process either initiates a new cycle, through initiation of a new BDM, or adds information to a previously initiated BDM. That is, a log of the analysis process is built up in the form of a sequence of BDM's.
4. Actions can be grouped into arbitrary subsets at will through a capability for defining sets of BDM's according to very general criteria. This set capability allows a wide variety of relationships among actions to be established. In particular, actions can be grouped so as to isolate and display trade-off relations as demonstrated in the prototype analysis.
5. By explicitly separating the parameters, P, sensitivity analysis is easy: simply designate a new P and repeat the prediction of consequences. Similarly, the sensitivity of the preference ranking to the statement of goals can be explored: designate a new goal statement (represented in DODO by a set of utilities, U) and repeat the evaluation and choice procedures.
6. The system is designed for browsing through the results of the analysis process. For example, the analyst may suddenly percieve a new issue and wish to define a new goal variable. He can do this, through the DEFINE GOAL VARIABLE command, at which time he also specifies how it is to be computed. The analyst may then "browse" the predicted impacts of actions previously examined with this new variable; if he decides it is a meaningful variable to use, he can add it to the system on a permanent basis, thus enlarging his set of goals.

## The Goal Fabric Concept

The impacts of transportation alternatives are many. We can distinguish these impacts by their nature, the groups that are affected, and the time at which they occur. Some of these impacts are relatively easy to evaluate in quantitative terms, such as travel time and out-of-pocket costs. Others are difficult or impossible to quantify, such as quality of life, and change in regional growth pattern. Some impacts can only be ranked, i.e., placed in an ordinal scale, or perhaps given only nominal values, such as the numbers on the shirts of football players. Some impacts occur quickly and cause only short-term effect; others will not occur until a long time into the future. The groups that are affected must potentially include a number of examples of the basic types outlined earlier: users; operators; nonusers including physically impacted, functionally impacted, and governmental actors.

To define and operate upon goals, we must first formulate a list of goal variables; we must have a variable for each facet of the problem that will be relevant to the decision among alternative actions. In light of the earlier discussion, we will have goal variables for different groups that may be affected, for different kinds of impacts, and for different points in time. Once we have defined a list of goal variables, we may then attempt to use this list as a basis for choosing among alternative actions. The simplest way is to use the list as a checklist; if the level of every goal variable on the list is satisfactory, then that action is acceptable; it is unacceptable otherwise (65).

A more general approach is to establish some type of scoring scheme. Mathematically this can be expressed as

$$U_i = \sum_j a_{ij} w_j$$

where

$w_j$ = relative weight of the j th goal variable,
$a_{ij}$ = level on scale of j th goal variable achieved by action i, and
$u_i$ = total weighted score for action i.

Standard economic criteria, such as total annual cost, net present worth, or benefit-cost ratio, are variants on this scheme, as is also utility theory.

The difficulties with such a scheme are as follows:

1. We never have a full, complete list of goal variables.
2. It is very difficult to get all the goal variables defined so as to be independent, mutually exclusive, and all at the same scale of relevance.
3. We can never completely identify the relative values, $w_j$, of all possible combinations of the various objectives.
4. We prefer to make decisions based upon the differences between alternatives, not the absolute levels. It is particularly important to know how much of goal j we must give up to achieve goal k.
5. Particularly in a sociopolitical context such as transportation, it is essential to examine the differential incidence. One alternative may score high on the goal variables important to group A but low on those important to group B, and therefore, the total score, $U_i$, hides the essential issues of choice.
6. Objectives change over time. It is sometimes as important to learn about the objectives as it is to develop new actions.

Recognizing the difficulties of the "scoring scheme" approach, we attempted to develop a looser, more subtle, more flexible approach to evaluating and choosing among alternatives. The concept developed is termed the goal fabric (66). It does not solve the problem completely, but seems a fruitful direction for development.

The basic ideas of the goal fabric are as follows:

1. A list of goal variables can be generated, but this list is never complete nor fully consistent and independent.
2. A number of relationships among goal variables can be identified and used to structure the list (means, ends, specification; value dependence or independence).
3. It is not necessary to get complete information on all possible combinations of values of the goal variables (e.g., by getting dollar equivalents, or by defining a utility function); the decision-maker need supply only sufficient information to indicate his preferences between the alternatives with which he is confronted, not all possible alternatives.

The basic impacts and consequences that are treated in the prototype analysis were briefly discussed earlier. Various aggregate measures of those impacts were constructed (see Fig. 5). For example, travel times are summed over origins, destinations, and/or modes to various levels of aggregation. In such a summation, all the goal variables are weighted equally.

The general structure of the goal variables in the prototype analysis is shown in Figure 13. This figure indicates how many different goal variables the analyst may wish to examine and the complex structure of their interrelationships.

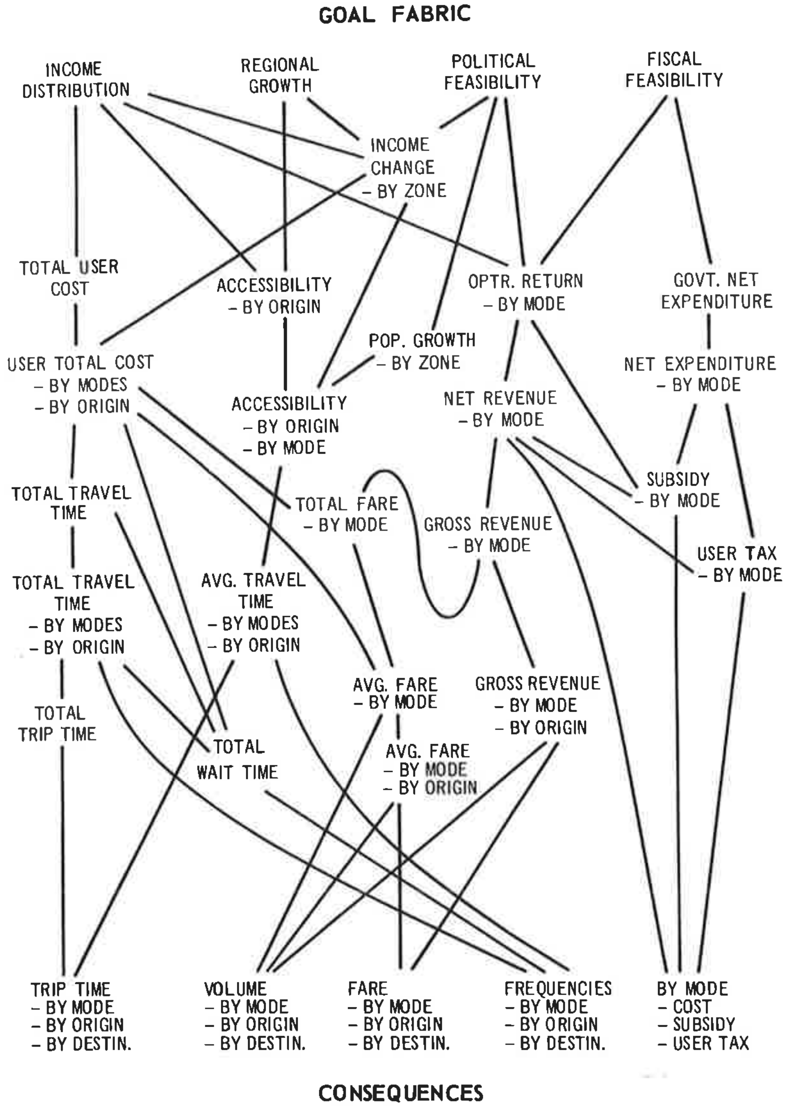



Figure 13. General structure of goal variance in prototype analysis.

The goal fabric concept is not fully tested. Capabilities for displaying and manipulating impact information in goal fabric form are being developed in DODO. Experiments with goal fabric analysis will be conducted in the context of systematic explorations of alternatives such as those outlined in the prototype analysis.

## Search

The search problem is the following: Given a number of previously generated actions and their consequences and a statement of goals or desirable directions to go in to improve the impacts of these actions, what values of the decision options will achieve the desired levels or directions of change? For example, if we want to decrease travel time without decreasing operator revenue, how do we change the system?

In general, as indicated by the prototype analysis, the range of possibilities for any particular problem is immense. In practice, of course, we sample only a very small percentage of the large number of possible alternatives. In doing this, we use a variety of short-cuts: (a) We consider the existing system, and explore possible small changes to that system. (b) We focus on a component of the problem, we decompose the problem into subproblems and work only on a piece of it. For example, we design a better vehicle and ignore the terminal, the access links, or land use impacts. (c) We abstract and simplify from the detail of the real problem and construct a model that we can manipulate to get an approximate idea of the characteristics of the desirable alternative. Each of these approaches has its value.

In actuality, we want to build up our search strategy out of all of these. Because of the massive fixed existing investment in transportation systems, we will generally implement strategies that involve a series of small changes from the existing system. However, we want to make this series of steps part of an incremental path toward some target end state, which may be radically different from the present system. We will often take a piece of the problem and focus on some component if only to understand its properties better in relationship to the overall system. And, we will definitely attempt to abstract and simplify from the problem.

This suggests the general flavor that underlies our approach to search. There is no single all-powerful procedure to use for searching out transportation systems alternatives. Rather, what is needed is a variety of tools that can be used flexibly by the transportation systems analyst as he explores the shape of the problem.

In this exploration, we assume that the analyst uses a more complex system of models for predicting the impacts associated with each alternative action. This model system, of course, involves network equilibrium prediction. Because of the variety of options and impacts, and the complexity of their interactions as represented by the system of models, the analyst will attempt to systematically trace out variations of the options and impacts.

Every model abstracts from reality and imposes its own biases on the problem. We need to be careful about the limitations and biases imposed by the system of models being used. Thus, we can conclude that systematic analysis using the predictive model system is only a guide to the search process of the analyst.

The analyst uses predictive models and search techniques to stimulate his perception and understanding of the problem. Neither the predictive model system nor the "kit" of search techniques will specify the solution to a transportation systems problem. The systematic analysis of alternatives and the results of search procedures serve to build up the analyst's image of the issues in a problem. This understanding, conscious and unconscious, provides an experience base from which he will create intuitively (67) that synthesis of technical and political elements with which he will try to solve a transportation problem. The solution comes from the analyst's understanding and imagination, not the models; but the models are an important aid.

There are a variety of search techniques available in transportation systems analysis. If we make some fairly drastic simplifications in the problem, we can apply such powerful techniques as mathematical optimization, including linear programming, integer programming, dynamic programming, and other techniques based on calculus. Alternatively, direct search or other hill-climbing approaches may be used, as well as heuristic search techniques such as pattern recognition and network aggregation. Another important family of search techniques is to provide an effective on-line computer environment with graphic display such that the analyst is able to operate efficiently in rapidly searching out and evaluating a large number of alternatives.

None of the analytical search techniques, such as mathematical optimization, is yet computationally feasible for large real-world transportation systems. However, it should be fruitful to use these techniques as approximations for search, in the following way, via the concept of network aggregation (see further the following section). In this approach, an aggregate representation of the particular detailed transportation systems problem would be constructed using a particular aggregation rule. The result of applying this aggregation rule would be a formulation of the problem appropriate for some mathematical optimization technique. This mathematical optimization technique would then be used within the context of a gross representation of the problem to find an optimum. The results of this process would then be translated back to the detailed level. Thus, mathematical optimization formulations and other analytical search techniques may be useful in the context of a broader search strategy via the concept of network aggregation.

A rich variety of search techniques of different types will probably prove more efficient as a system than any single technique used alone. Furthermore, the judgment of the analyst can and should play a strong role throughout the search process. Therefore, it is appropriate to develop a variety of different search techniques, as well as a flexible environment in which they can be used. Work is proceeding in this direction.

## Multilevel Problem-Solving: Network Aggregation

In human problem-solving, we rarely analyze real problems at only one level. It is a natural approach to problem-solving to operate at several levels of analysis. In some contexts, this corresponds to first doing preliminary design then detailed design.

When an analysis process deals with several levels, we say it is hierarchically structured. A model of hierarchically structured problem-solving processes has been proposed (68) that revolves around the concept of inclusion among actions.

By inclusion, we mean that one action may be a representation of a set of other actions. For example, a schematic diagram of a network and its associated regional development pattern (linear system, a polynucleated region) may represent a number of different detailed network and land use pattern alternatives. Conceptually, we can visualize the gross or higher level action as a set of more detailed, or lower level actions; all the lower level actions in the set differ in details but have the same basic characteristics, and so can be represented by a single higher level action.

The concept of hierarchical structure is defined by this inclusion relationship. Consider now the "basic operator," consisting of the sequence search-prediction-evaluation-choice (i.e., search followed by selection). Such a basic operator produces a characteristic kind of action. For example, in highway location, we might have three operators: one operator to produce bands of interest, a second to produce location bands, and a third to produce locations.

The purpose of multilevel structure is to enable us to search more efficiently by generating and evaluating gross alternatives as well as detailed ones. The inclusion relationship implies that only for the most detailed alternatives is it possible to predict impacts precisely. Of course, because at higher levels we are working with approximate characteristics rather than precise detailed characteristics, there is higher uncertainty about the performance of alternatives. In other words, if we are dealing with performance attributes of a particular system, then we can get a single-valued vector only at the most detailed level. At other levels, we can only deal with the distribution of these attributes (68). So what we gain in computational effort by dealing at the gross level, we lose in accuracy and certainty of results.

The complexity of the transportation systems problem suggests we may find it efficient to structure it as a multilevel process. A possible multilevel structure was initially proposed by Bruck, Manheim, and Shuldiner for the Northeast Corridor study (13).

To successfully use a multilevel framework, an understanding of the fundamental relationships among different levels is necessary. To develop this understanding in the particular context of network equilibrium in transportation, experiments are being conducted in network aggregation (8, 69).

A basic problem in the analysis of transportation systems is that of the detail with which networks are modeled. Very rarely can the analyst expend sufficient resources to model a transport network in full detail, with a link in the model for every link in the real-world network. Rather, the analyst must be satisfied with an approximate representation of the real network. For example in urban transportation systems analysis, usually the rail transit and expressway systems are represented in complete detail, but the arterial and local streets and bus lines are more usually approximated in some way. In megalopolitan or national transportation systems analysis, usually only the major intercity links can be modeled directly; the intraurban networks and secondary road systems can be represented only approximately at this scale of interest.

When modeling in full detail is not possible, the analyst must explicitly account for the uncertainties introduced by the approximation by a less-than-fully detailed (aggregate) network. The basic idea is that each link of an aggregate network corresponds to a set of links in the true network. More precisely, a subset of links at the aggregate level corresponds functionally to a subset at the "true" level (68). This simple notion implies that the aggregate link does not have a single value of travel time (length, capacity, or other parameters), but a probability distribution. The analyst should formulate and use this distribution and thus avoid possible serious errors that may result from using only the point estimate of the aggregate link's travel time (or length or capacity).

In general, many different levels of detail of network representations will be desirable. Each will have a corresponding uncertainty in analysis, consequent upon the degree of aggregation; but as aggregation and uncertainty increase, ease of computation for analysis should increase and computing cost (dollars, time) should decrease.

It is the task of the analyst to determine the desired level of detail for an aggregate network. To do so, he must estimate the relative benefits of increased accuracy associated with a detailed network versus the ease of computation and analysis effort permitted with an aggregate form. If a small number of alternative transport systems are being evaluated for a final decision regarding which alternative to implement, uncertainty in the network parameters should be minimized and the required aggregate network will retain a high level of detail. If preliminary studies are being performed on a large set of widely different transportation alternatives, then a higher level of uncertainty can be tolerated to permit many analyses, so that possible alternatives can be reduced to a small number for detailed analysis. The usefulness of aggregation in a transportation planning environment is reflected in the ability to analyze a larger number of alternative transportation systems for a given set of analysis budget and time constraints than would be possible if aggregation techniques were not employed.

## SUMMARY

We began with a brief description of the problem of transportation systems analysis and emphasized the number and variety of options and impacts, the complexity of the system of predictive models, and the need to trace out differential impacts systematically. After review of the rational model, we then described a more general model of the design process, PSP, and its implications. To show the juxtaposition of these two themes, we next discussed a prototype analysis of the Northeast Corridor passenger transportation system. We showed the PSP structure in the computer language developed to do transportation analyses and illustrated the systematic analyses that could be done. There were many issues raised, however; and in the last section, we described extensions of the PSP model in the form of specific operational techniques to address these issues: DODO, goal fabric, search, and network aggregation.

Work is continuing to test, refine, and further develop the techniques and concepts described here. To summarize our basic conclusions: (a) The core of the transportation systems problem is the prediction of equilibrium flows in networks. (b) The systematic analysis of transportation problems requires careful, sensitive exploration of trade-offs among options and impacts. (c) The prototype analysis demonstrates the feasibility of this approach, but also raises issues. (d) It is useful to view transportation systems analysis as a problem-solving process. (e) A variety of specific operational techniques can be developed to make transportation analysis a more efficient process.

## ACKNOWLEDGMENTS

The basic support for this research was provided by the U. S. Department of Transportation, Office of Systems Analysis. Additional support for portions of this research was provided by the General Motors Grant for Highway Transportation Research; by the Special Assistant to the Joint Chiefs of Staff for Strategic Mobility, Department of Defense; and by the Ford Foundation through a grant to the Urban Systems Laboratory of M. I. T. The support of these agencies is gratefully acknowledged. Clearly, however, the opinions expressed herein are those of the author and do not necessarily represent the views of any research sponsor.

This paper summarizes work to which many individuals have contributed, as indicated by the references and footnotes. The author expresses his appreciation for the stimulation and assistance of his colleagues and students, and particularly to Earl R. Ruiter.

## REFERENCES

1. Manheim, Marvin L., Ruiter, Earl R., and Bhatt, Kiran U. Search and Choice in Transport Systems Planning: Summary Report. Dept. of Civil Eng., M. I. T., Cambridge, Vol. I, Research Rept. R68-40, 1968.
2. Ruiter, Earl R. A Prototype Analysis. Dept. of Civil Eng., M. I. T., Cambridge, Vol. II, Research Rept. R68-41, 1968.
3. Ochoa-Rosso, Felipe. Applications of Discrete Optimization Techniques to Capital Investment and Network Synthesis Problems. Dept. of Civil Eng., M. I. T., Cambridge, Vol. III, Research Rept. R68-42, 1968.
4. Manheim, Marvin L., Follansbee, Kenneth G., and Walter, Ronald A. Modelling the Evolutionary Nature of Problem-Solving. Dept. of Civil Eng., M. I. T., Cambridge, Vol. IV, Research Rept. R68-43, 1968.
5. Ochoa-Rosso, Felipe, and Silva, Angel (17). Optimum Project Addition in Urban Transportation Networks via Descriptive Traffic Assignment Models. Dept. of Civil Eng., M. I. T., Cambridge, Vol. V, Research Rept. R68-44, 1968.
6. Stafford, Joseph H., Prokopy, J. C., and Aivazian, V. Production Functions: Models of Transport Technology. Dept. of Civil Engineering, M. I. T., Cambridge, Vol. VI, Research Rept. R68-45, 1968.
7. Bhatt, Kiran U. Search in Transportation Planning: A Critical Bibliography. Dept. of Civil Eng., M. I. T., Cambridge, Vol. VII, Research Rept. R68-46, 1968.
8. Chan, Yupo, Follansbee, Kenneth G., Manheim, Marvin L., and Mumford, John. Aggregation in Transport Networks: An Application of Hierarchical Structure. Dept. of Civil Eng., M. I. T., Cambridge, Vol. VIII, Research Rept. R68-47, 1968.
9. Jackson, L. F., Schwartz, Jesse G., and Bivins, David. Equilibrium Models. Department of Civil Eng., M. I. T., Cambridge, Vol. IX, Research Rept. R68-48, 1968.
10. Schwartz, Jesse G. Network Flow and Synthesis Models for Transportation Planning: A Survey. Dept. of Civil Eng., M. I. T., Cambridge, Vol. X, Research Rept. R68-49, 1968.
11. Silverstone, Stuart M., and Mumford, John. Development of Graphic Communication in a Computer-Based Transportation Planning Process. Dept. of Civil Eng., M. I. T., Cambridge, Vol. XI, Research Rept. R68-50, 1968.
12. Plourde, Rodney. Consumer Preferences and the Abstract Mode Model. Dept. of Civil Eng., M. I. T., Cambridge, Vol. XII, Research Rept. R68-51, 1968.
13. Bruck, Henry W., Manheim, Marvin L., and Shuldiner, Paul. Transport Systems Planning as a Process: The Northeast Corridor Example. Dept. of Civil Eng., M. I. T., Cambridge, Vol. XIII, Professional Paper P67-23, 1967.
14. Manheim, Marvin L., and Hall, Frederick. Abstract Representation of Goals: A Method for Making Decisions in Complex Problems. Dept. of Civil Eng., M. I. T., Cambridge, Vol. XIV, Professional Paper P67-24, 1967.
15. Manheim, Marvin L. Problem-Solving Processes in Planning and Design. Dept. of Civil Eng., M. I. T., Cambridge, Vol. XVI, Professional Paper P67-3, 1967.
16. Johnson, William F. Decision Theoretic Techniques for Information Acquisition in Transportation Network Planning. Dept. of Civil Eng., M. I. T., Cambridge Vol. XVII, Research Rept. R69-1, Jan. 1969.
17. Guenther, Karl W. Predictive Models for Vehicle Operating Consequences. Dept. of Civil Eng., M. I. T., Cambridge, Vol. VIII, Research Rept. R69-2, 1969.
18. Manheim, Marvin L., Ruiter, Earl R., et al. Techniques for Searching Out Transportation Systems Alternatives: Summary Report. Dept. of Civil Eng., M. I. T., Cambridge, Research Rept. R68-50, 1969.
19. Manheim, Marvin L. Transport, Problem-Solving, and the Effective Use of Computers. Highway Research Record 148, 1966, pp. 49-58.
20. Project Metran. M. I. T. Press, Cambridge, 1966.
21. Tomorrow's Transportation. U. S. Department of Housing and Urban Development, 1968.
22. Kanwit, Edmond L. Some Aspects of the Social Impact of Urban Transportation. In Transportation: A Service, Proc. New York Academy of Sciences, 1968.

23. Thomas, Edwin N., and Schefer, Joseph L. Information Requirements for Evaluating the Social Impacts of Transportation Investments. In Transportation: A Service, Proc. New York Academy of Sciences, 1968.
24. Nash, William W., Jr. Disaggregative Impact Studies: Their Potential Use in Transport Decisions. In Transportation: A Service, Proc. New York Academy of Sciences, 1968.
25. Wohl, Martin, and Martin, Brian V. Traffic System Analysis for Engineers and Planners. McGraw-Hill, New York, 1967.
26. Hershdorfer, Alan. Predicting the Equilibrium of Supply and Demand: Location Theory and Transportation Flow Models. Transportation Research Forum, 1966.
27. Manheim, Marvin L. Principles of Transport Systems Analysis. Transportation Research Forum, 1966.
28. Beckman, Martin. On the Theory of Traffic Flow in Networks. Traffic Quarterly, Jan. 1967.
29. Jewell, W. S. Models for Traffic Assignment. Transportation Research, Vol. 1, No. 1, 1967.
30. Nelson, James R. Pricing Transport Services. In Transport Investment and Economic Development (Fromm, Gary, editor), Brookings Institution, Wash., D. C., 1965.
31. Heflebower, Richard B. Characteristics of Transport Modes. In Transport Investment and Economic Development (Fromm, Gary, editor), Brookings Institution, Wash., D. C., 1965.
32. Martin, Brian V., Memmott, F. W., and Bone, A. J. Principles and Techniques of Predicting Future Demand for Urban Area Transportation. M. I. T. Press, Cambridge, Rept. 3, 1961.
33. Deen, Thomas B. Discussion of "An Evaluation of Alternative Land Use and Transportation Systems in the Chicago Area" by Campbell, E. Wilson, and "Transportation System Development and Evaluation as Practiced in Seattle" by George, Stephen, Jr. Highway Research Record 238, 1968, pp. 120-122.
34. Tanner, J. C. Traffic Generation and Forecasting. Traffic Eng. and Control, Vol. 8, No. 3, July 1966, pp. 164-165.
35. Bruck, H. W., Manheim, Marvin L., and Shuldiner, Paul W. Transport Systems Planning as a Process: The Northeast Corridor Example. Proc. Transportation Research Forum, 1967.
36. Study Design. Northeast Corridor Project, Tech. Paper 5, June 1966.
37. Quandt, R. E., and Baumol, W. J. Abstract Mode Model: Theory and Measurement. Jour. Regional Science, Vol. 6, No. 2, 1966.
38. Bruck, H. W., Putman, Stephen H., and Steger, Wilbur A. Evaluation of Alternative Transportation Proposals: The Northeast Corridor. Jour. Am. Inst. of Planners, No. 32, Nov. 1966, pp. 322-333.
39. Roberts, Paul O., and Kresge, David T. Models for Transport System Simulation. Harvard Univ., Cambridge, Mass., and Brookings Institution, Wash., D. C., Discussion Paper 65, Nov. 1967.
40. Kraft, Gerald, and Wohl, Martin. New Directions for Passenger Demand Analysis. Transportation Research, Vol. 1, No. 3, 1967.
41. Soberman, Richard M. Approaches to Regional Transportation Models. Proc. Transportation Research Forum, 1966, pp. 357-366.
42. Hay, William W. An Introduction to Transportation Engineering. John Wiley and Sons, New York, 1961.
43. Meyer, John R., Peck, M. J., Stenason, J., and Zwick, C. The Economics of Competition in the Transportation Industries. Harvard Univ. Press, Cambridge, Mass., 1960.
44. Soberman, Richard M. Transportation Technology for Developing Regions. M. I. T. Press, Cambridge, 1967.
45. Spiegelberg, C. H., et al. Evaluation Study of Transport Vehicles. Boeing Co., Renton, Wash., Report D6-2067, 1964.

46. Morlok, Edward K. An Analysis of Transport Technology and Network Structure. Northwestern Univ., Evanston, Ill., unpublished PhD dissertation, June 1967.
47. Lang, A. Scheffer, and Soberman, Richard M. Urban Rail Transit. M. I. T. Press, Cambridge, 1964.
48. Meyer, John R., Kain, J. F., and Wohl, Martin. The Urban Transportation Problem. Harvard Univ. Press, Cambridge, Mass., 1965.
49. Devanney, J. W., III. Transportation Scheduling Under Multi-Dimensional Demand Levels. M. I. T., Cambridge, unpublished PhD dissertation, 1967.
50. Urban Development Models. HRB Spec. Rept. 97, 1968, 266 pp.
51. Manheim, Marvin L., et. al. The Impacts of Highways Upon Environmental Values. Urban Systems Laboratory, M. I. T., Cambridge, unpublished report.
52. Schoeffler, Sidney. Toward a General Theory of Rational Action. Kyklos, Vol. VII, No. 3, 1954, pp. 245-271.
53. Hall, Arthur D. A Methodology for Systems Engineering. D. Van Nostrand Co., Princeton, N. J., 1962.
54. Hitch, Charles J., and McKean, Roland N. The Economics of Defense in the Nuclear Age. Atheneum, New York, 1965.
55. Baumol, William J. Economic Theory and Operations Analysis, 2nd edition. Prentice-Hall, Englewood Cliffs, N. J., 1965.
56. Ackoff, Russell L. Scientific Method. John Wiley and Sons, New York, 1962.
57. Braybrooke, David, and Lindblom, Charles E. A Strategy of Decision. Free Press, New York, 1963.
58. Lindblom, Charles E. The Intelligence of Democracy. Free Press, New York, 1965.
59. Steger, Wilbur A., and Lakshmanan, T. R. Plan Evaluation Methodologies: Some Aspects of Decision Requirements and Analytical Response. HRB Spec. Rept. 97, 1968, pp. 33-76.
60. Meyerson, Martin, and Banfield, Edward C. Politics, Planning, and the Public Interest. Free Press, New York, 1955, p. 319.
61. Pecknold, Wayne M. The Evolution of Transport Systems: An Analysis of Time-Staged Investment Strategies Under Conditions of Uncertainty. PhD dissertation in progress.
62. Roos, Daniel. ICES System Design. M. I. T. Press, Cambridge, 1967.
63. Ruiter, Earl R. ICES TRANSET I: Transportation Network Analysis—Engineering Users' Manual. Dept. of Civil Eng., M. I. T., Cambridge, Research Rept. R68-10, Mar. 1968.
64. Ruiter, Earl R. ICES DODOTRANS I: Decision-Oriented Data Organizer for Transportation Analysis—Users' Manual. Dept. of Civil Eng., M. I. T., Cambridge, Research Rept. R69-42, 1969.
65. Alexander, Christopher. Notes on the Synthesis of Form. Harvard Univ. Press, Cambridge, Mass., 1960.
66. Manheim, Marvin L., and Hall, Fred L. Abstract Representation of Goals. Dept. of Civil Eng., M. I. T., Cambridge, Professional Paper P67-24, 1968.
67. Hadamard, Jacque. The Psychology of Invention in the Mathematical Field. Dover Publications, New York, 1954.
68. Manheim, Marvin L. Hierarchical Structure: A Model of Planning and Design Processes. M. I. T. Press, Cambridge, 1966.
69. Chan, Yupo. Optimal Travel Time Reduction in a Transport Network: An Application of Network Aggregation and Branch and Bound Techniques. Dept. of Civil Eng., M. I. T., Cambridge, 1969.

## *Discussion*

STANLEY L. GORDON, System Analyst, Texas Instruments Incorporated—Professor Manheim has aptly presented two major themes in his paper. First, he has carefully described the multidimensional nature of transportation decision-making, i.e., options on the physical system, impacts on transportation subscribers, and prediction of impacts. Second, he has presented a well-thought-out description of the problem-solving process (PSP) that involves search (the generation of alternative transportation systems) and choice (evaluation and selection of alternatives).

Both topics, transportation and the PSP, are complex activities on which much attention has been focused. The unification of the many viewpoints that already exist about these two areas, in addition to fresh thinking on these subjects, present an exciting challenge to the analyst.

The major benefit of a fuller and more complete understanding of both transportation and the PSP as it applies to transportation is the greatly increased likelihood that a viable, adaptive transportation system will be selected to meet transportation requirements for the future. The need for a vast improvement in transportation decision-making exists because of the highly probable occurrence of gross suboptimization and the relative inability of anyone, individually or collectively, to significantly change a system once it has been implemented. It is because of the fixed nature of transportation systems and their enormous expense that calls for an advance in the state of the art in transportation system decision-making.

In this light, the major thrust of Professor Manheim's paper is the problem of search and choice—how alternatives are generated and how a best one (in some sense) is chosen. Professor Manheim has done well in "bounding" the problem of search and choice and in providing a framework for others to follow. Search is described as the generation of alternative transportation systems that appear to be sufficiently attractive for evaluation, given predicted impacts upon various transportation subscribers. Choice is the preference ranking of these alternatives based on an evaluation of the predicted impacts.

A model of the search and choice process consists of the definition of objectives or goals to be met and a utility function sufficiently well defined to render a preference ordering among alternatives. Next, enumeration of alternative transportation systems is performed. Third, the consequences of implementation of each transportation system are predicted. Fourth, the consequences are evaluated in terms of objectives through the utility function. Finally, one system that best achieves the objectives is chosen.

Additional flexibility in this model is achieved through appropriate iterations in which objectives and/or goals undergo revision. A very important concept is that of selective acquisition of input data for the model in which continuous sampling over time is preferred to an enormous single data-gathering effort. This ensures a flexible re-adjustment of data acquisition. Next, Professor Manheim acknowledges the fact that the model must account for multilevel problem-solving by generating and evaluating gross level alternatives as well as detailed ones. Finally, but most important, the model and the analysis team must properly interface with the social and political dimensions of transportation. The interface, in turn, serves to clarify public objectives.

I believe Professor Manheim has contributed something of much value to analysts dealing with decision-making problems in general and transportation problems in particular. His paper certainly establishes a point of departure for further work in transportation systems decision-making.

# Transportation Systems in the Future Development of Metropolitan Areas

## The Permanent Corridor Concept

JACK E. LEISCH, De Leuw, Cather and Company of Canada Ltd.

It is expected that 85 percent or more of the population will reside in urban areas by the turn of the century, and it is mandatory that transportation networks in metropolitan regions be established and the land for them be reserved now in order to serve the travel requirements in the future. This cannot be effectively accomplished on the basis of current practice. This paper develops a concept of permanent transportation corridors and network configurations that allow for long-range planning. It indicates that an optimum network of major transportation facilities, initially spaced at 4 to 6 miles and eventually spaced at 2 to 3 miles in built-up areas, can function into the future at reasonable levels of service for any predicted rates of population growth, urban area expansion, and increased travel demand. The concept provides for major transportation corridors that form the framework for all urban development. The corridors are permanent, but the facilities within the corridors are not. The corridors are established at sufficient width and appropriate spacing so that they are (a) capable of accommodating travel generated by likely future population densities and of allowing balanced transportation service, (b) adaptable to design or redesign to carry a freeway, a rapid transit facility, or a combination of the two, and (c) capable of conversion in the future to meet the design requirements of new technological developments. The application of this concept could resolve the urban planning dilemma that seems insurmountable.

•LONG-RANGE planning for expansion and redevelopment of cities has been carried out to various degrees, but there is little evidence that such plans are being successfully implemented. In the North American society, positive or restrictive controls needed to implement long-range planning are not apt to be imposed. The most widespread controls—zoning laws—are frequently changed in response to new requirements and unforeseen pressures for development. Opportunities for attractive short-range land development projects that can be readily programmed and financed frequently set the planning pattern. Therefore, the long-range aspects of master plans are seldom retained.

As a result, long-range city planning is often not even attempted. In practice, relatively short planning periods are used, generally 20 years. This appears to be logical in terms of programming and financing, both for public and private development projects, and is widely accepted as the period within which future growth can be predicted with some degree of confidence.

For individual transportation improvements, such as a freeway, a rapid transit line, or a terminal facility, the 20-year planning period is generally considered to be appropriate. The implementation process, which includes planning studies, design and plan preparation, procurement of land, relocation of residents, clearing of the site, and construction, normally takes 5 to 10 years before the facility is placed in operation. Staging procedures and programming or financing difficulties may cause large-scale facilities and area-wide systems to take more than 20 years to complete.

---

Paper sponsored by Department of Urban Transportation Planning and presented at the 48th Annual Meeting.

For the broader aspects of transportation planning—the planning of total systems, including various modes of transportation—planning periods considerably longer than 20 years are required. Unlike detailed land use designations and individual land development projects, which are confined to relatively compact areas, major transportation facilities are parts of a highly interrelated system that encompasses the whole of the metropolitan area. In the past, portions of transportation systems have been destroyed or badly distorted by land development that blocks one or more corridors in the system.

On the basis of current practice, planning beyond 20 years is difficult and cannot be accomplished effectively. Moreover, few authorities are able to reserve or protect land in advance for transportation facilities within a 20-year period. Reservation and procurement of rights-of-way normally are not instituted until the project has been programmed and funds earmarked for construction. Sometimes exact locations are not determined until public hearings have been held. Under this procedure, sections of major transportation facilities designated for completion at the end of a 20-year planning period or those considered essential beyond 20 years may fall by the wayside.

Fully effective legislation for the protection of future rights-of-way is nonexistent. Another difficulty is that funds normally are not available for the purchase of land that may not be required for construction until 10, 15, or 20 years later. Municipalities in particular lack the necessary funds, and even the current programs must be carefully budgeted.

The problem of reserving rights-of-way for transportation facilities needed within a 20-year planning period is difficult enough, but what is even more discouraging is that there are practically no means of protecting locations of routes that are based on needs beyond 20 years. In the interim, large development projects and multistory buildings can destroy the continuity of routes and cause an imbalance in a carefully worked out network. One philosophy holds that if certain facilities are negated by development, other locations will be found and, besides, some of the planned facilities may not be required after all, or may be required somewhere else. This is a negative attitude but, unfortunately, it is widely held in the absence of a positive solution.

Another discouraging note is the disagreement among certain planners, and between highway and public transit interests, as to what modes of transportation should be provided in the future. Some are looking to a significant shift from travel by private automobile to travel by public transit with reduction or elimination of freeway development. Some stress that freeway development must be accelerated if cities and their core areas are to thrive. Others claim it is useless to construct additional freeways because they will continue to be congested and, if freeway expansion is not stopped, cities will be stifled and come to ruin because too much space will be taken over by freeway facilities. Anticipated technological improvements and possible future changes in transportation modes are used as additional excuses for not currently planning for the long-range future.

These conflicting philosophies and points of view are detrimental to sound transportation planning and to redevelopment and expansion of urban areas, and frequently they are used as the rationale for doing nothing or doing little in terms of long-range planning. This apparent dilemma must be resolved so that future cities and their transportation systems can be properly planned and development programs implemented. A completely new and bold approach is necessary.

One way of resolving the problem is first to break down the following notions and contentions (referred to here as deterrents) that are holding back progress and, in the process, to develop a positive approach.

## PLANNING DETERRENT 1

> Deterring Attitude: Development of freeways will have to be minimized, and future travel must be taken over by public transit, if cities are to survive.

Travel in every city is accommodated by some combination of private automobiles and public transit. The proportion handled by each mode varies from one urban area to another. It is suggested here that the prevailing modal split in a particular city is an inherent characteristic of that city and is a by-product of its historical development, land use arrangement, and population density.

Cities that matured before the advent of the automobile tend to be oriented more heavily toward public transit, whereas cities that grew up afterward are more likely to depend on private motor vehicles. Thus, the ratio of public transit to private vehicle usage is not necessarily a function of city size alone.

The fact that it would seem logical and economical to have travel in urban areas accommodated predominantly by public transportation is no indication that this actually could take place in the future. The personal mobility provided by the motor vehicle has become a way of life, and the public expects reasonable freedom in the use of the automobile. Denial of this mobility by withholding transportation improvements can stunt the growth of a community, change the pattern of its growth, or cause that growth to take place somewhere else. The dramatic example of the relative decline of the CBD reflects in part the lack of personal mobility in the transportation service to the city core as compared to the outer areas. Development of facilities to serve the natural desire of the motorist and transit user must keep pace with expanding population and commercial and industrial activity.

Urban transportation is not an either-or situation—either rapid transit or freeways. (Rapid transit is defined as any form of efficient, high-speed public transportation.) Freeways do not obviate the need for transit, nor do rapid transit facilities serve as a substitute for freeways. Highways and public transit are complementary and not competing facilities. Both must be developed to a proper balance of service according to the characteristics of the particular city.

In large metropolitan areas, certain predominant transportation corridors are now served, and others will eventually need to be served, by both rapid transit and freeways. This trend is clearly evident in many North American cities. Thus, to build a rapid transit facility rather than a freeway is normally valid only as a stage in the long-range development of transportation facilities. Although this may defer freeway construction, which may be desirable in some cases, eventually both will be needed.

The ratio of public transit to private vehicle usage is generally decreasing. There is evidence, however, that in certain corridors that serve high-density areas, a significant increase in the number of transit riders can be effected where modern and highly attractive transit service is provided for travel to and from the downtown area. In terms of the total metropolitan area, several such corridors would not have a pronounced effect on the modal split for the metropolitan area as a whole. One reason is that the ratio decreases in parts of the metropolitan area where development is too dispersed to be effectively served by transit. Another reason is in the normal makeup of trip purposes and their origins and destinations. Even if all logical trips were assigned to public transit, it would not cause the modal split to be greatly increased.

In view of these conditions and trends, it would be considered highly successful if, in the particular city, the current overall modal split could be maintained in the future, and this should be the objective. To accomplish this would require significant improvement and expansion of public transit facilities as the city grows. In many cities, provision of a balanced transportation plan would call for comparable expenditures for public transit and highway facilities. However, the proportion of the system to be served by transit in the future need not affect the long-range planning of transportation networks if a concept of permanent major transportation corridors is adopted. (Major transportation corridor is defined here, in its narrow sense, as the actual strip of land or right-of-way for the purpose of accommodating a major ground transportation facility.)

The accommodation of both a highway and a public transit facility in one right-of-way is not new. If this treatment is extended to a total network of major routes, each link or segment of the system could provide, in the long-range plan, for any one of several types of development—a highway facility, a public transit facility, or a combination of the two (Fig. 1). Certain functional relationships would favor one or the other mode in the early planning stages. In later stages the other mode could be added to the same corridor.

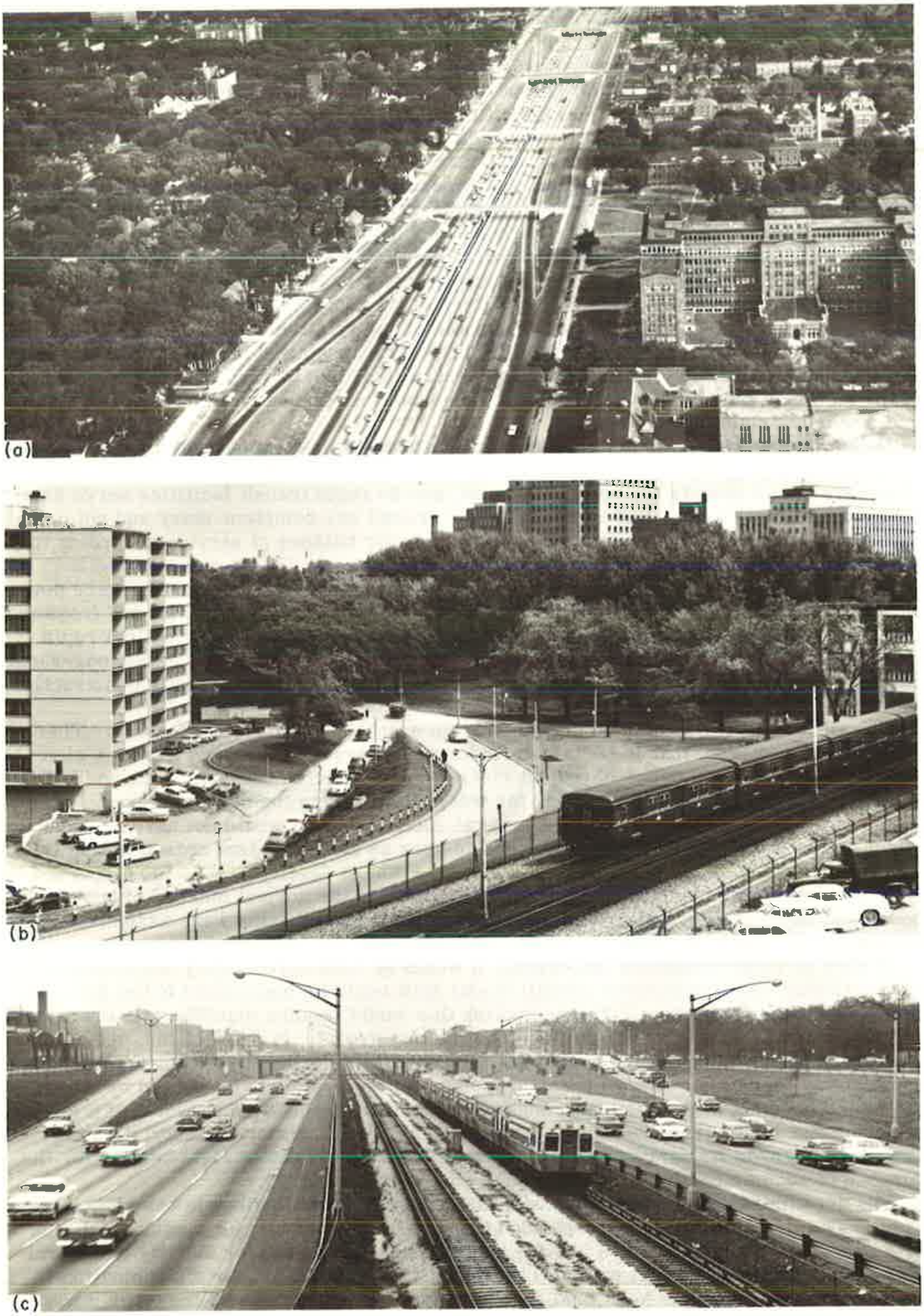


Figure 1. A major transportation corridor containing (a) a freeway, (b) a rapid transit facility, and (c) a combination of the two.

Where the combined types of facilities are called for, they would generally be located in radial corridors outside the city core. In approaching the urban center, the freeway and the rapid transit line would separate a mile or so in advance of the CBD, with the transit going underground on a different alignment to provide direct service within the district. Transit service may be required to other points of concentration between the corridors. Here, again, such service would be a distribution function and not a direct function of the major transportation corridors and would be provided by a spur or distributor facility deviating from and returning to the major corridor.

By the time it is decided whether a given element of the system should be a freeway or a rapid transit line, or both, the right-of-way could have been reserved and one of the forms of transportation could have been in service for a decade or more and accommodating an important travel demand. Conversion or modification within the right-of-way could take place later. The question, therefore, of whether a freeway or a rapid transit facility is to be provided for should not deter long-range planning of transportation networks.

## PLANNING DETERRENT 2

> Deterring Attitude: Future transportation could change radically as a result of technological advances in the design of vehicles, means of propulsion, and electronic control.

Technological advances of the future are often used as reasons for not extending transportation plans to include "horizon-year" or long-range requirements. It should be recognized, however, that even if some of the revolutionary technology were available today, it would take many years, probably a generation or more, to replace present facilities. The gradual implementation of new developments in transportation, communication, and other public services is evident through history. Obviously, no government would scrap its highway or transit system overnight to provide radically different roadbeds, vehicles, and operating controls. Existing facilities would be gradually retired near the end of their useful lives.

For this reason, planning for the next 20 years should not be deterred, provided that the facilities can be transformed to new or different modes at the end of this period. Again, this may be accomplished through the application of the permanent corridor con-

Figure 2. A transformed facility within a major transportation corridor where a 12-lane facility replaced a 4-lane, 10-year-old freeway, generally within the original right-of-way.

cept. The permanence of major transportation corridors is one of the striking features of the history of cities. Many routes established more than a century ago have been retained and transformed, some many times over, to modern carriers of traffic. Today they accommodate freeways or rail facilities.

Several examples of railroad rights-of-way in their second generation of use are the Pennsylvania Turnpike, parts of the BARTD transit system in San Francisco, the GO commuter system in Toronto, and the high-speed rail system along the Eastern Seaboard between Washington and Boston.

Likewise, numerous early expressway rights-of-way are now being rehabilitated into modern freeway or freeway and transit facilities to incorporate new technology and accommodate changes in travel characteristics (Fig. 2). Outstanding examples of extensive transformation are the MacDonald-Cartier Freeway (Highway 401) in Toronto and the New Jersey Turnpike in the vicinity of Newark. There is a great need now to establish new major corridors to satisfy the ever-changing travel demands of the future.

The mode of ground transportation, the vehicles, and their characteristics of operation may be different in detail two decades from now, but the corridors would be essentially the same, particularly in relation to the other permanent features of urban development. If the corridors were established initially at sufficient width to accommodate future changes in vehicle characteristics and modes of operation, then long-range planning for transportation networks and corridor location would be realized automatically for periods well beyond 20 years.

## PLANNING DETERRENT 3

> Deterring Attitude: Because freeways are filled with traffic and become congested as soon as they are built, it would be impractical to provide all the freeways needed in the future to serve an expanding metropolitan area.

This notion, expressed by many planners, is a further deterrent to the development of long-range transportation plans. Frequently new freeways have been fully loaded within a few years after they have been opened to traffic. It is expected that travel in urban areas will continue to grow in the foreseeable future, as the population and economic activity increase. Therefore, the crowding of new freeways can be expected for many years, until the travel demand is more nearly in balance with freeway construction and until the individual freeways are tied together to function as systems.

It is well known that heavy volumes on many urban freeways are the result of distorted traffic patterns caused by partial freeway systems constructed as parts of staged development programs. Eventually, when the freeway network is complete and the system is in balance, traffic volumes will stabilize or decrease on individual routes, even though the level of traffic over the whole city will have grown substantially.

The number and spacing of freeways required to give satisfactory service are dependent on (a) population densities in combination with type and intensity of urban development, (b) topography, land use structure, and makeup of supporting street systems, (c) continuity of routes within the overall system and extent of through travel, and (d) size of freeways (numbers of lanes) and their interrelation with public transit facilities. Studies have indicated that traffic in metropolitan areas requires an ultimate network of freeways spaced generally at 2- to 3-mile intervals in the central area, and 4 to 5 miles in the outer areas (1, 2, 3, 4). It can be shown that an optimum network of freeways properly spaced, and supported by appropriate street systems and public transit facilities, can be expected to function into the future at a reasonable level of service regardless of population growth, urban area expansion, and increasing travel demand. This phenomenon is evident from an analysis of various metropolitan area studies in which freeway spacing was found to be independent of city size.

The relationships in Figure 3 derived from proposed freeway systems (see Appendix) planned for various cities to satisfy travel requirements about 20 years in the future indicate that spacings of freeways are related primarily to population densities. For example, where the population density of the central city is an average of approximately

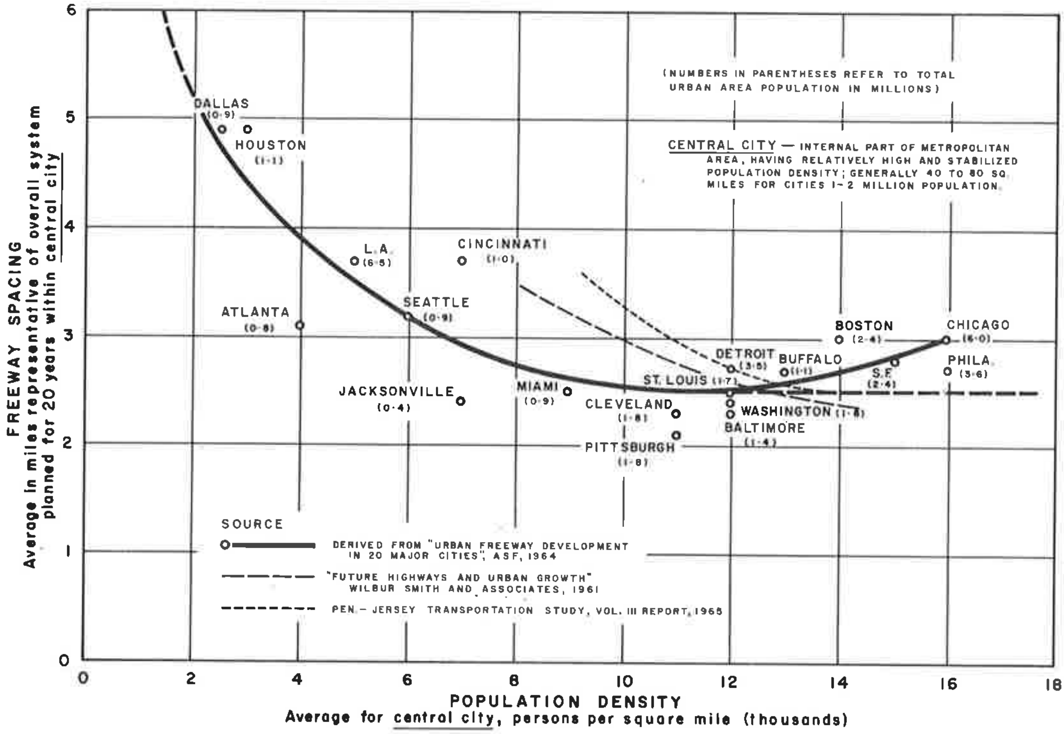



Figure 3. Freeway spacing requirements indicated by planning studies in 20 cities.

12,000 persons per square mile, seven cities show a need for freeways spaced at 2- to 3-mile intervals. Included among these are the cities of Buffalo, metropolitan area population of 1.1 million; Baltimore, 1.4 million; Pittsburgh, 1.8 million; and Detroit, 3.5 million. It can be seen that the spacing of freeways required in a stabilized built-up area tends to remain the same even though the population in the total metropolitan area may double or triple.

The analysis further reveals that cities with central area densities ranging from 7,000 to 16,000 persons per square mile also require an average spacing of internal freeways of 2 to 3 miles. These cases include metropolitan areas with populations of 0.5 to 6.0 million (Jacksonville, Florida, and Chicago, Illinois). Considering the wide range of metropolitan area sizes and population densities, such freeway spacing is extremely versatile and can be assumed to be optimum. This is in line with operational experience that generally points up that the minimum practical spacing of freeways is 2 to 3 miles. This average minimum spacing is capable of functioning well for a range of development and population densities because of the flexibility afforded through (a) variations in freeway size (capacity), and (b) distributional opportunities that are self-compensating in closely knit systems.

The 2- to 3-mile pattern of freeways in the central city can continue to function adequately in the future, provided that the freeway system is gradually expanded outward to serve the outer sectors of the metropolitan area as it continues to expand. This is illustrated in Figure 4. Once the population density is stabilized and the optimum freeway system is achieved in the central city, the volume of traffic on individual freeways would tend to be stabilized also.

In Figure 4, the designation and development of major transportation corridors to accommodate progressively the growing requirements for travel in an expanding metropolitan area are shown diagrammatically. Major transportation facilities in operation are indicated by heavy solid lines, and corridors designated for the future are shown by

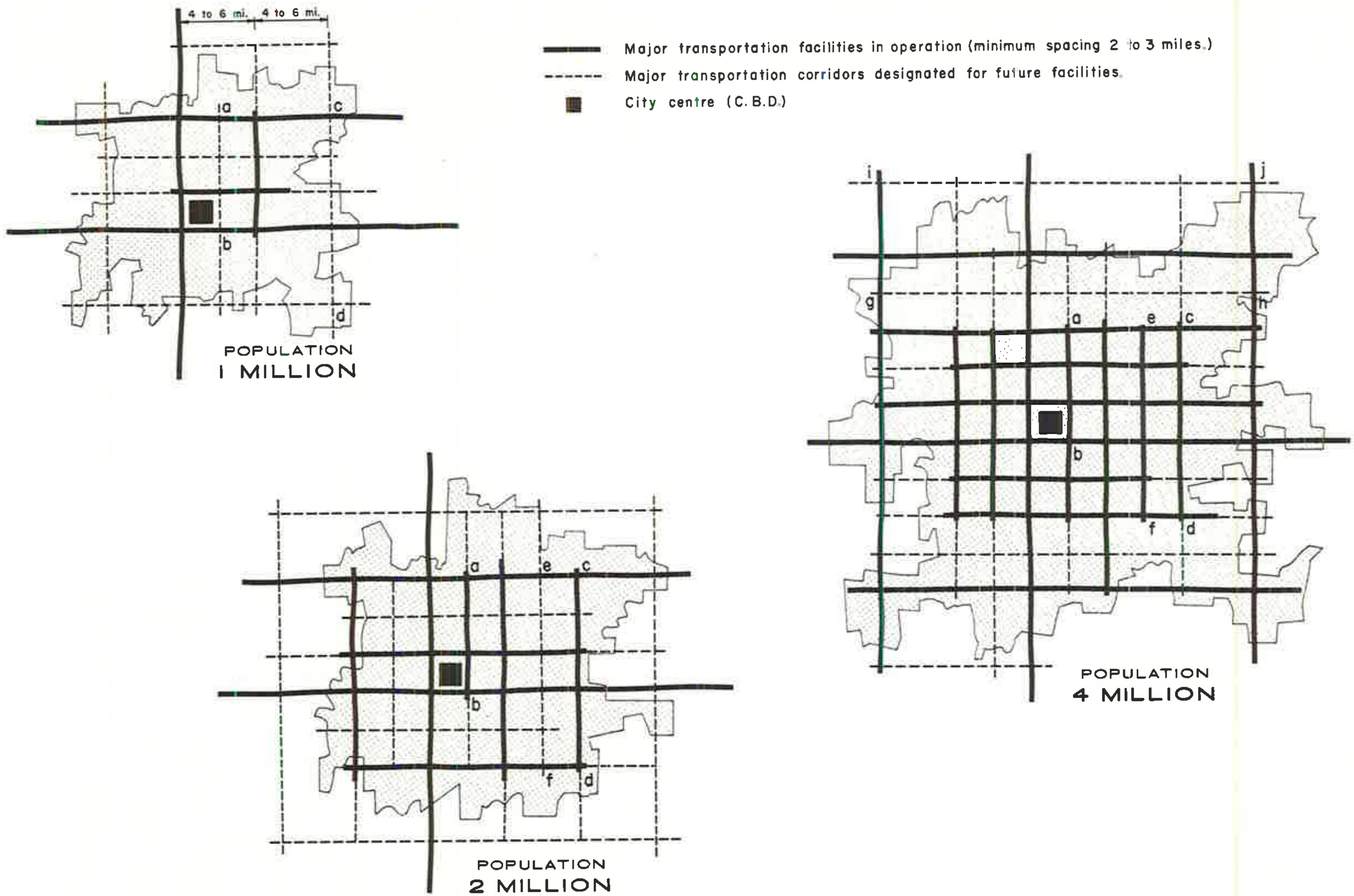



Figure 4. Formation and expansion of major transportation corridors as the metropolitan area grows.

dotted lines. The illustration represents a metropolitan area that has a present population of 1 million people and that doubles and redoubles its population in the future. In its growth from 1 to 2 million people, the reserved corridors are transformed to operating facilities as additional corridors are designated for further expansion. Such progressive designation of corridors and their development is further shown for the stage when the urban area reaches a population of 4 million.

At the stage when the population is 1 million, freeways are developed largely at 4- to 6-mile spacings. Additional facilities in the outer areas are designated for the future at 4- to 6-mile spacings, and several internal corridors are reserved at additional 2- to 3-mile intervals. Progressive development with city growth is indicated by the intermediate facility a-b, designated in the 1-million population stage. It is developed as an operating facility when the 2-million population is reached. Similarly, an outer facility, c-d, is constructed in the expanded area of the city. At this stage a new intermediate transportation corridor, e-f, is reserved for the future and is later shown as an operating facility when the urban area has reached a population of 4 million. Even at this stage, additional corridors would continue to be designated, such as g-h and i-j.

In this plan the ultimate freeway system is achieved first within the central portion of the urban area, prior to its completion in the partially developed outer districts. The built-up central area will continue to redevelop from time to time, but its level of activity, even with some increase in density, will tend to become stabilized and the outlying areas will tend to attract the additional activities of the region.

Further freeway segments are then introduced in the intermediate areas, completing that portion of the system. At the same time, freeways are extended and added in the outer areas (Fig. 5). This process continues: the transportation network expands outward as the metropolitan area grows, while the internal facilities, without basic change, continue to function. The established freeways within the city need not carry any more traffic because the internal system is progressively supplemented by additional freeways, which function as parallel routes providing bypass opportunities and further dispersion of traffic.

Figure 5. Freeways in the process of being extended and added within designated major transportation corridors, as urban development pushes outward from the city.

This analysis gives the answer to Planning Deterrent 3 and, recognizing what the optimum freeway system can accomplish in its stage development, suggests the basis for long-range planning of freeway networks. The application of the permanent corridor concept would provide ultimately for the full development of freeways at intervals of 2 to 3 miles in built-up areas and 4 to 6 miles in outlying areas. The 4- to 6-mile spacing would eventually reduce to 2 to 3 miles as the areas grow and additional outer facilities are introduced at the intermediate intervals of 4 to 6 miles. The plan makes it possible to serve the increasing traffic demands of the metropolitan area without limitations on its size, rate of growth, or period of expansion.

In this discussion a grid pattern was chosen to simplify the presentation. In actual practice there would be some degree of irregularity in the network, dictated by land use and development features and by topographic and other physical controls. The main point is that the indicated average corridor spacing would be generally achieved in the ultimate plan, even though there may be a lack of parallelism in the system, and the size of any one cell may be somewhat larger or smaller than the average indicated.

## PLANNING DETERRENT 4

> Deterring Attitude: Transportation planning must be closely integrated with landuse planning and, because specific land uses and detailed urban development cannot be predicted much beyond 20 years, transportation planning should not be carried out for periods greater than 20 years.

The interdependence of urban development and transportation demands joint consideration of the two in any long-range planning procedure. However, there is no assurance that any specific urban development for a period beyond 20 years will be implemented. Therefore, any long-range transportation plan will be, for the most part, ineffective if it is based on an urban plan developed by the usual short-range city planning procedures.

Another problem in current practice is the complexity associated with the determination of future transportation networks. When the basic system is established, an attempt is made to coordinate land uses and urban development with transportation, together with considerations of the modal split. This becomes an overwhelming ordeal. The procedure is further burdened by the present emphasis on data collection, research, mathematical models, statistical procedures, systems analysis techniques, and adaptation to computer processing. Such sophistication appears to be appropriate in resolving the more detailed features of the transportation plan after the basic system has been determined.

The crux of the problem is in the difficulty of handling the multitude of interacting variables and in predicting the future. To overcome this difficulty, it is necessary to break the process down into its component parts, the first of which is the network determination. Every part, then, would be contingent on a relatively small number of variables, allowing for simplification in each of the various study stages. Urban transportation planning concepts and techniques have been under intensive study and development, but they are relatively new and in a state of flux. There is a tendency here, as in other fields of endeavor, for the procedure to become unduly complex. As with any new development—whether it be a machine, a manufacturing process, or an analysis procedure—there comes a time when the number of moving parts must be reduced to achieve better results and greater dependability. The time for simplification in the transportation planning process has arrived, and more appropriate methods must now be established.

The procedure proposed here reduces the problem to three distinct steps or study stages: (a) long-range network determination of permanent major transportation corridors, (b) system determination including modes and combinations of modes for a 20-year planning period, and (c) design of a specific part of the system or an individual facility for a 20-year period or less. Each successive planning step is dependent upon the preceding step and is a progression of continuous planning and design. The degree of refinement and detail is increased with each succeeding study stage. At each step, there must be appropriate coordination with other relevant disciplines including city planning, economics, sociology, and architecture.

Planning stage 1, the determination of a long-range network of major transportation corridors, is comparable to basic structural design of a large office or apartment building. In the planning of the structure, the size and shape of the lot, number of occupants, parking spaces per occupant, size requirements for rooms, and perhaps several other general features provide sufficient information to design the framework of the building. The exterior finish and architectural touches and the type of window frames, wall coverings, elevators, plumbing fixtures, and managerial and operational considerations are all important but need not have any bearing on the location, size, and interrelationship of the columns, girders, and beams that form the framework. If designed to carry the likely ranges of loads, each floor or space in each bent or cell can be utilized in any of many different ways and developed or redeveloped as desired in the future within the basic structural framework of the building. Likewise, the framework of a transportation system, comprising the permanent major transportation corridors, can be established with respect to broad controls, and individual urban development plans can then be detailed, implemented, adjusted, or replaced in the future within this framework.

In actual application the major transportation framework is one of several frameworks that form the total urban structure. There are other urban features that have a fundamental and permanent significance. Some of these features are in the category of large capital improvements under government control, such as streets, schools, parks, sewers, and water supply. These features form additional frameworks and, together with transportation, may be planned jointly to shape the basic structure of tomorrow's metropolitan areas. Although this interrelationship is not specifically a part of this discussion, nevertheless its significance is stressed here to point out the necessity for coordinated effort in effective, long-range planning.

The establishment of the major transportation corridors would not limit the changes that may take place inside each cell or unit of land area between the corridors of the

Figure 6. Streets occupy 25 to 30 percent of usable land within cities; freeways at optimum spacing of 2 to 3 miles occupy only 7 to 8 percent of this land.

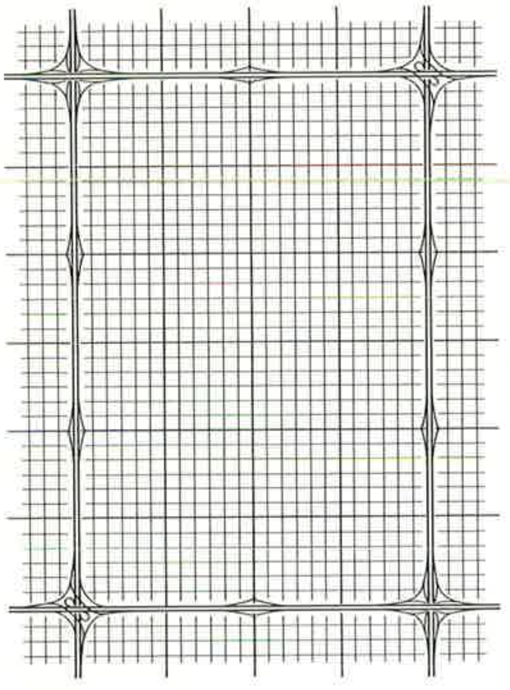


Figure 7. Spatial relations of freeways, streets, and urban area development for 2- x 3-mile major corridor module (diagrammatic configuration, but to proper scale).

system. The various segments of the corridor system would be sufficiently flexible and each facility sufficiently adjustable in mode and capacity within its corridor that all possible future changes in the magnitude and pattern of travel could be accommodated.

The establishment, therefore, of the major transportation corridors in coordination with other urban structure frameworks as everlasting features in the ultimate development of the city is the key to the solution. The system in this form is not limited to specific types of transportation facilities nor to specific land development plans. These other, more detailed considerations may be decided later in the process as part of planning stage 2. The corridors, supported by appropriate distribution facilities including an arterial street system, would assume an optimum average spacing of 2 to 3 miles in built-up areas and 4 to 6 miles in less dense or partially built-up areas.

The network would by no means take on a regular rectangular pattern, although arranged at optimum spacing as a whole. Land use forms, environmental areas and their interrelationship, major travel desire lines, topography, historic, and sociological aspects, and other basic features peculiar to the area would play a major role in shaping the final network.

The corridors and the cells within the corridor network produce the permanent configuration of the city, just as the major streets platted hundreds of years ago in the larger cities have remained as fixed elements of the city structure. The widths of the new major corridors would be of 1- or 1½-block units, or in the general range of 400 to 600 ft wide. Such corridors, assumed in their ultimate position to have average spacings of 2½ miles and to have an average width of 500 ft plus space for interchanges, would occupy 7 to 8 percent of the usable land area within the city. This is a small portion relative to the street system itself, which generally occupies 25 to 30 percent of the area within cities (Fig. 6). Figure 7 shows diagrammetically but to proper scale the proportionality among freeways, streets, and occupied space. It does not appear that major transportation corridors at an ultimate spacing of 2 to 3 miles would take over the city.

Another way of expressing the space eventually occupied by the major corridors would correspond to a plan without freeways but with all street reservations widened from an average of 70 ft to 85 ft. It is likely that the space given up to major transportation corridors could be compensated for by increasing the height of some buildings in the blocks adjacent to the corridors or through development of air rights or other joint land-use arrangements. As an example, a 15-story building placed over the corridor within every fourth block or three such buildings along each mile of the facility could compensate for the entire corridor right-of-way. There is a tendency in densely built-up areas for redevelopment to take place contiguous to freeways and rapid transit lines, not only in blocks adjacent to them but in those several blocks away. This, combined with increased property values and air-rights development, could compensate many times over for the tax rateables initially set aside by the reservation of the corridor (Fig. 8).

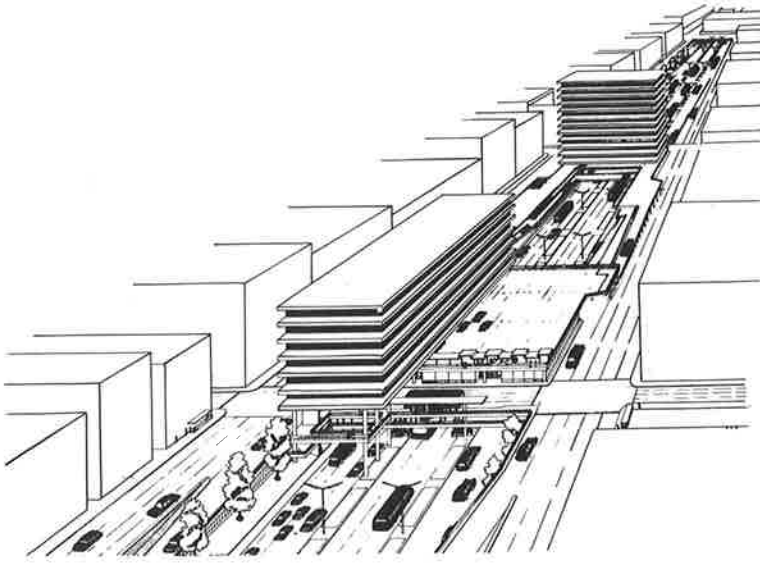


Figure 8. Increased property values of redevelopment in the vicinity of freeways and rapid transit facilities combined with development on air-rights tend to compensate for tax rateables removed by the transportation improvements.

The cellular urban form, framed by major transportation facilities, is a fundamental concept of city planning. For example, this is recognized as the basic planning principle in the Buchanan report (5) for the planning of new towns or the redevelopment of existing towns. The following quotation from this report emphasizes this principle:

> There must be areas of good environment—urban rooms—where people can live, work, shop, look about, and move around on foot in reasonable freedom from the hazards of motor traffic, and there must be a complementary network of roads—urban corridors—for effecting the primary distribution of traffic to the environmental areas. These areas are not free of traffic—they cannot be if they are to function—but the design would ensure that their traffic is related in character and volume to the environmental conditions being sought. If this concept is pursued it can easily be seen that it results in the whole of the town taking on a cellular structure consisting of environmental areas set within an interlacing network of distributory highways. It is a simple concept, but without it the whole subject of urban traffic remains confused, vague, and without comprehensive objectives. Once it is adopted, then everything begins to clarify.

It is precisely this simplicity and clarity that make the major transportation corridor concept a logical approach to future planning. The key, however, to practical and successful application of this principle is the establishment of permanent major transportation corridors to form an optimum network that could be achieved eventually.

From the point of view of community planning, the setting out of the major transportation corridors would evolve from the design of small environmental areas, the grouping of these areas into a unit cell or district, and then the establishing of major highways on the boundaries of the cell. Arranging the cells contiguous to each other in a logical and desirable pattern would form a network of major transportation facilities. In the process, the interconnection of the various segments of the network must provide continuous and direct routes for intracity travel.

This approach points out that the function of the network would be to serve the environmental areas and not the reverse. However, development within these areas would be affected by the service provided by the network. An environmental area consisting of given land uses calls for a certain system of transportation facilities but, when this system is provided, it in turn affects land development and causes changes within it.

To account for the interaction between urban development and transportation and to resolve the total planning problem, it is essential to have a number of disciplines working together: city planners; highway, traffic, and transit engineers; architects; economists; sociologists; and political scientists. Together they would plan for an appropriate balance of transportation and land development influences.

Such cooperative effort is being piloted in Baltimore and other cities in the detailed planning and design of individual corridors. But, the need to utilize and coordinate these abilities and skills during the initial, conceptual stages of overall metropolitan area planning (planning stage 1) is even more important. This is essential in order to set out the basic or rudimentary transportation framework to which specific forms of transportation facilities can be ultimately patterned and to which eventual urban development and all its ramifications can be related.

The application of multiple disciplines at the conceptual stage should be broad and quite different from the detailed planning and design of individual corridors. Obviously, in order for the interdisciplinary efforts to be effective at the later detailed stages of planning (as in Baltimore), it must first be applied in its broader aspects to the overall makeup of urban transportation systems. Furthermore, stages 2 and 3 are continuous planning processes that would update and refine the plan and implementation programs at intervals of 5 to 7 years, subject to continual planning reviews.

In summary, the shape of the network and the spacing and width of its segments are predicated on the broader aspects of the following controls: land use, population densities, topographic features, historical considerations, aesthetic opportunities, future growth and development aspects, and resulting travel demands. In the analysis and formulation of these controls, all of the appropriate disciplines are utilized. These controls, most of which may be expressed as a series of urban structure frameworks, are broadly applied and coordinated with transportation to physically establish the corridors without regard for the particular mode or modes of transportation that ultimately may be required. At this stage, the specific arrangements of future urban development, which may or may not be predictable, play no part in the planning analysis.

In the succeeding planning stages 2 and 3, which consider periods of 20 years or less, the details of urban development are considered. In these stages the mode, type, and size of the transportation facility are determined for the planning or design year. However, because the corridors have been properly established with respect to location and width in study stage 1, the design within the corridors allows for possible adjustment, expansion, or conversion of the facility to its future modal and operational requirements. The major transportation corridors and the network they form are permanent, but the facilities within them are not. The urban development within the cells of the transportation network and within the other frameworks of the urban structure is not considered to be permanent and is subject to change with time.

Thus, in answer to planning deterrent 4, long-range transportation planning can be carried out effectively by establishing a permanent network of major transportation corridors. This can be accomplished with full recognition of the fact that transportation planning is an integral part of total urban development planning, which cannot be specific as to form and detail on a long-range basis.

## PLANNING DETERRENT 5

> Deterring Attitude: Long-range transportation plans have little value because there is neither appropriate legislation nor funding capability for the reservation and advance procurement of future right-of-way.

Planning deterrent 5 is the only real deterrent to long-range transportation planning. Because it is the only real deterrent, all possible effort should be made to resolve the problem it presents.

Government at all levels has been grappling with the problem of reserving rights-of-way for future transportation facilities. Even though, in principle, the idea of establishing major transportation corridors does not differ materially from the traditional concept of platting and dedicating street reservations, the idea has not yet been applied to major transportation networks for expanding urban areas. This practice of street platting was established many centuries before the automobile. To account for current trends of growth in urbanization, similar dedication of larger thoroughfares or major transportation corridors, spaced at 2- to 6-mile intervals, appears logical as it was for the streets of yesteryear that were spaced at 400- to 600-ft intervals.

It is logical that a network of major transportation corridors can and should be established as a foundation and a framework for future travel and to give direction to further expansion and redevelopment of the metropolitan regions. To make such long-range planning possible, needed legislation must be vigorously pursued. Furthermore, it is suggested that the federal government may be the appropriate agency to provide the drive, incentive, and guidelines for establishing major transportation corridors as a matter of national welfare and security.

The Interstate System in the United States and freeway development in the provinces of Canada demonstrate the capability to carry out large freeway programs. Now the emphasis must be placed on transportation within the metropolitan areas. It is expected that 85 percent or more of North America's population will live in urban areas by the turn of the century, and it will be a race with time to provide the transportation facilities that will be needed.

As part of the formidable job of directing and shaping urban growth, it is essential that the dilemma of transportation planning be resolved. This can be accomplished only by a fresh and bold approach—designating, protecting, and procuring major transportation corridors for the cities of tomorrow. Old precedents, concepts, and legislative measures, many of which were developed during horse-and-buggy days, must be changed and modernized.

A positive and systematic means of procuring land for the future transportation corridors must be formulated and adopted. An important feature would involve legislation providing, possibly, for several steps in procuring rights-of-way and for interim uses of designated lands. Incentives to permit the retention of existing development and to foster certain new development in the interim period may be essential. Arrangements for palatable and equitable means of relocating people must be devised. Many disciplines have to be brought together on the legal, economic, and social problems. Once effective legislation is established, it will be necessary to develop and institute a long-range financing program for securing the corridor lands.

Implementation, the reservation of land for long-range plans, is the only real deterrent, and it must be overcome to secure the future of our cities!

## THE SOLUTION—SUMMARIZED

The solution to the dilemma of how to accomplish long-range transportation planning for the expansion and redevelopment of cities has been developed in the discussion of the various deterrents and, with some amplification, is summarized here.

Effective long-range planning is essential if we are to properly shape and build the cities of tomorrow. Yet, current planning methods and controls have not been and apparently are not likely to be successful in carrying out long-range plans. Furthermore, planning for transportation for periods beyond 20 years has been confronted with a series of road blocks that are producing confusion, indecision, lack of confidence, and deferment of plans. The public will pay dearly for this in the future, even before 20 years. The deterrents to long-range planning are widely held attitudes, summarized briefly as follows: (a) public transit should replace freeways; (b) transportation will change radically because of new technology; (c) if freeway development were continued to its ultimate conclusion, an impractical number of freeways would be required; and (d) long-range transportation planning cannot be carried out without detailed land use planning.

Each of these deterrents can be overcome by one bold stroke in metropolitan area planning—the application of the concept of permanent major transportation corridors. Just as cities in their early growth were developed on the basis of dedicated streets, so may the metropolitan regions of today expand and redevelop within a framework of designated major transportation corridors. Each corridor would be preserved and eventually utilized for a major transportation facility—a freeway, a rapid transit line, or a combination of the two—and so dimensioned that it could be converted to any form or mode in accordance with future technological developments. The spacing of corridors and their reservation are fundamental and should be adhered to if fully effective transportation is to be achieved in the future.

With the setting of the corridors, the die is cast. The basic structure is established, and land development can precede or follow the construction of transportation facilities. Detailed land use changes, new development, and redevelopment take place within the cells, the areas between the major transportation corridors. The cells are comprised of one or more environmental areas that can be adjusted or revised within the cell and within the frameworks of the urban structure.

If the permanent major transportation corridor concept is accepted, then only one real deterrent remains: Land cannot be reserved for transportation facilities planned for periods beyond 20 years because of the lack of appropriate legislation and the lack of funds for advance procurement of rights-of-way.

This problem is the responsibility of the various levels of government as a matter of national welfare and security, and the solution must be found if transportation planning and long-range urban planning are to succeed. The solution to the urban transportation problem, in conjunction with the expansion, development, and redevelopment of urban centers, clearly evolves in recognizing the following principles and features:

1. Complete systems of major transportation corridors for long-range requirements should be designated and established now as permanent features of the future metropolitan regions.
2. The major transportation corridors form the basic structural framework for both short- and long-range future urban development and redevelopment.
3. The major transportation corridors are permanent, but the modes and types of facilities they accommodate are not permanent.
4. The cells of land between the major transportation corridors are permanently situated within the corridor network, but the detailed land uses and development within the cells need not be permanent.
5. The corridors are spaced at intervals of approximately 2 to 6 miles, providing for an ultimate average spacing of 2 to 3 miles in built-up areas.
6. As urban areas expand and joint megalopolitan regions are formed, additional corridors are established in the outer reaches of the regions (Fig. 7).
7. The corridor networks, although arranged on an optimum average spacing as a whole, by no means assume a regular rectangular pattern. Land use forms, environmental areas and their interrelation, major travel desire lines, topography, historic and sociological considerations, and various other basic features expressed as urban structure frameworks contribute to the shaping of the corridor networks.
8. The corridors are established at sufficient width (400 ft or more) and appropriate spacing so that they are (a) capable of accommodating travel generated by likely future population densities and of allowing for balanced transportation service, (b) adaptable to design or redesign to carry a freeway, a rapid transit facility, or a combination of the two, and (c) capable of conversion in the future to meet the design requirements of new technological developments.
9. A 20-year plan is based on the established network of permanent transportation corridors and is used for the programming and development of (a) individual transportation facilities within the corridors and (b) detailed plans for land use along the corridors and within the cells formed by the corridors. The detailed planning is updated on the basis of planning reviews at 5- to 7-year intervals.
10. In establishing the long-range plans of major transportation corridors and the more detailed 20-year plans, city planners, highway engineers, traffic engineers, public transit engineers, architects, sociologists, economists, and political scientists all participate in the coordinated effort of solving the joint problems of urban development and transportation.
11. New special legislation and implementation methods are needed to reserve the rights-of-way for the major transportation corridors, to foster certain development within the corridors in the interim periods, and to procure the land when required. A

completely new and direct approach is mandatory. Once it is accomplished, the problem will become clear and its solution direct.

With appropriate changes of attitude and suitable legislation, the permanent corridor concept could form the basis for effective long-range planning. Ultimately, the overall cost would be decidedly lower and the utility, efficiency, and economy of the system much superior to those allowed by the planning processes in use today. In any case, there is no way to accommodate the expanding urbanization cheaply. The problem is great, so its solution must also be great.

## ACKNOWLEDGMENTS

Acknowledgment is gratefully expressed to Robert Conradt, De Leuw, Cather and Company of Canada Ltd., and to Robert L. Hart, Hart, Krivatsy, and Stubee, for assistance in the formulation of concepts and in helpful review and consultation in the preparation of this paper. Photographs were provided by Associated Electrical Industries Limited, California Division of Highways, Ontario Department of Highways, Eric Trussler, U.S. Bureau of Public Roads, and Michigan State Highway Department.

## REFERENCES

1. Creighton, Roger L., Hock, Irving, Schneider, Morton, and Joseph, Hyman. Estimating Efficient Spacing for Arterials and Expressways. HRB Bull. 253, 1960, pp. 1-43.
2. Future Highways and Urban Growth. Wilbur Smith and Associates, Feb. 1961.
3. Peterson, James M. Freeway Spacing in an Urban Freeway System. Jour. Highway Div., Proc. ASCE, Vol. 86, No. HW3, Sept. 1960.
4. Penn-Jersey Transportation Study, Vol. III, May 1965.
5. Steering Group and Working Group of the Minister of Transport. Traffic in Towns (the Buchanan Report). Her Majesty's Stationery Office, London, 1963.

# *Appendix*

TABLE 1

SPACING OF FREEWAYS BASED ON PLANNED SYSTEMS[a]

| City | Population of Total Urbanized Area (millions) | Central City | | Size of Representative Square-Shaped Area of Central City Used in Analysis | | Mileage of Freeways Within Representative Square | | Average Spacing of Planned Freeways[b] (mi) |
|---|---|---|---|---|---|---|---|---|
| | | Area (sq mi) | Population Density (1,000 persons per sq mi) | Side (mi) | Area (sq mi) | Planned (mi) | In Operation (mi) | |
| New York | 14.0 | 315 | 25 | 15 | 225 | 150 | 125 | 3.0 |
| Los Angeles | 6.5 | 455 | 5 | 24 | 576 | 312 | 125 | 3.7 |
| Chicago | 6.0 | 224 | 16 | 16 | 256 | 172 | 76 | 3.0 |
| Philadelphia | 3.6 | 127 | 16 | 10 | 100 | 75 | 20 | 2.7 |
| Detroit | 3.5 | 140 | 12 | 12 | 144 | 105 | 37 | 2.7 |
| San Francisco | 2.4 | 48 | 15 | 7 | 49 | 35 | 12 | 2.8 |
| Boston | 2.4 | 48 | 14 | 8 | 64 | 42 | 11 | 3.0 |
| Toronto[c] | 1.8 | 55 | 16 | 7.5 | 56 | 28 | 13 | 4.0 |
| Washington | 1.8 | 61 | 12 | 8 | 64 | 54 | 24 | 2.4 |
| Pittsburgh | 1.8 | 54 | 11 | 7 | 49 | 46 | 14 | 2.1 |
| Cleveland | 1.8 | 81 | 11 | 9 | 81 | 72 | 15 | 2.3 |
| St. Louis | 1.7 | 61 | 12 | 7 | 49 | 40 | 14 | 2.5 |
| Baltimore | 1.4 | 79 | 12 | 8 | 64 | 56 | 19 | 2.3 |
| Houston | 1.1 | 328 | 3 | 15 | 225 | 92 | 42 | 4.9 |
| Buffalo | 1.1 | 39 | 13 | 6.5 | 42 | 31 | 15 | 2.7 |
| Cincinnati | 1.0 | 77 | 7 | 8 | 64 | 35 | 13 | 3.7 |
| Dallas | 0.9 | 280 | 2.5 | 16 | 256 | 105 | 38 | 4.9 |
| Seattle | 0.9 | 88 | 6 | 9 | 81 | 50 | 20 | 3.2 |
| Miami | 0.9 | 34 | 9 | 6 | 36 | 29 | 8 | 2.5 |
| Atlanta | 0.8 | 128 | 4 | 11 | 121 | 79 | 18 | 3.1 |
| Jacksonville | 0.4 | 30 | 7 | 6 | 36 | 30 | 12 | 2.4 |

[a]Derived from Urban Freeway Development in Twenty Major Cities, Automotive Safety Foundation, Aug. 1964. Indicative of freeway systems proposed for the central city (as of 1964); representative of recommendations to meet requirements in about 20 years.
[b]Approximate spacing predicated on a selected square-shaped area representative of the central city; based on the formula $S = 2A/M$, where S is average (equivalent uniform) spacing of freeways in miles, A is square-shaped area containing the freeways in sq mi, and M is the length of freeways within the square-shaped area in miles.
[c]Toronto included, although not part of the ASF report on 20 cities.

## *Discussion*

J. J. BAKKER, Associate Professor, University of Alberta—Jack Leisch makes an excellent case for wide rights-of-way for freeways so that future expansion will be possible. However, he confuses this need for adequate rights-of-way with a cure-all solution for all transportation ills. The cure-all solution proposed is that rapid transit and freeways should be located within the same right-of-way. The permanent corridor concept, however, will work only if the entire system uses individual units such as cars, buses, or multimode vehicles. It is not so suitable for people who may want to transfer from the pedestrian mode to the transit mode, or from an individual unit mode to a rapid transit mode.

The walking distances involved to and from stations in the median of an expressway detract from the transit facility. Moreover, the transit facility often detracts from the smooth operation of the freeway or expressway ramps because of loading and unloading operations at the stations. The transit facility should be surrounded by high-density uses (residential or employment) that cannot easily be provided in the right-of-way of a freeway, air (pollution) rights notwithstanding, or the transit facility should serve parking lots that should be close to a freeway but not in the same right-of-way.

To use the MacDonald-Cartier Freeway (Highway 401) in Toronto as an example of the cure-all solution shows up the fallacy. This highway is a circumferential-distributor type of road. In Toronto there is another east-west expressway (Queen Elizabeth-Gardiner) along the lake front. In between is the Bloor Street subway. A rapid transit facility along the alignment of the MacDonald-Cartier would be unwarranted. A freeway along Bloor Street is not feasible from the point of view of land use and economics. In Toronto both transit and freeways have a role, but not in the same right-of-way.

If the spacing of freeways is going to be 3 to 4 miles, then the transit facility should be located between two freeways. Transit, if it is going to be used by people, will have to be convenient to people and not just to highway planners.

JACK E. LEISCH, Closure—Mr. Bakker seems to have missed the whole point of the permanent corridor concept. It is not merely a case for wide rights-of-way providing a "cure-all solution," as he puts it, and having the criterion that both transit and freeways are located within the same reserve. The permanent corridor concept clearly designates the establishment of corridors "at sufficient width and appropriate spacing so that they are adaptable with full flexibility for design or redesign to carry a freeway, or a rapid transit facility, or a combination of the two." Only some corridors may have both.

The concept embodies a long-range planning procedure that serves as a framework for all facets of urban development. It is a device by which the rebuilding and expanding of old cities and the development of new cities can be accomplished in an orderly fashion, with the assurance that the city can continue to function effectively for all time in the movement of people and goods.

Mr. Bakker purports that where a transit and a highway facility are situated within a common right-of-way, it is not suitable for people wishing to transfer from the pedestrian or private automobile mode to the transit mode. Although some of the earlier combined facilities revealed certain difficulties in operation, it has been shown that these now can be overcome by new design. Whereas transfer facilities used to be combined with interchanges, they are now designed to be altogether divorced from each other with the result that the same degree of freedom is made available to transfer operations as if the freeway and the transit facility were miles apart. Parking facilities, where required, would be located adjacent to or in combination with freeways, and would also be removed from interchanges so as not to interfere with normal traffic movements. Furthermore, accessibility to the rapid transit facility combined with a freeway can be just as effectively accomplished as it can to a separate transit facility.

It is fundamental that a rapid transit facility, if it is to be effective, must serve a general corridor of high-density population. The concentrated development considered necessary to support a rapid transit line takes place within a band at least a half-mile wide and not necessarily surrounded by or within the right-of-way of the transportation facility. The same degree of development concentration, including air rights, can be provided generally when the rapid transit is combined with a freeway or when it is a separate facility, and the corridor concept makes provision for both.

Mr. Bakker's reference to the Toronto transportation situation is again an indication of his apparent misunderstanding of the corridor concept as presented. He states that a rapid transit line in combination with the MacDonald-Cartier Freeway (Highway 401) east-west corridor is unwarranted. It was never intended here that a rapid transit line would be combined with the freeway. This is one of the three possible forms of corridor concept development—a freeway without a rapid transit facility. The east-west Bloor Street subway is a clear case of the second form of corridor development—a rapid transit facility without a freeway. However, another east-west corridor, between Bloor Street and Highway 401, is called for in the future development of Toronto. If designated and preserved now as a major transportation corridor (400 ft or more in width), it would probably be developed as a freeway to start with but could be expanded ultimately to include a transit line; or it could be completely converted in time to whatever form or mode of transportation is required in the future. It is also of interest that plans are currently under preparation in Toronto for construction of two combined facilities (freeway with a rapid transit line in the median), the Spadina Expressway and the Scarboro Expressway.

If the spacing of freeways is to be at 3 to 4 miles, Mr. Bakker indicates that the transit facilities should be located between the freeways. In the corridor concept, the major transportation corridors are spaced at 2 to 3 miles in highly built-up areas and 4 to 6 miles in outlying areas. As and when required each may be utilized to carry a freeway, a transit line, or a combination of the two. This spacing is already optimized for a network of major transportation facilities to carry large numbers of people at high speeds. The modes can be determined or changed in the future. In special cases a subsidiary facility may be required in between the 2 to 3 miles indicated. If so, it would be for a short distance as a distribution facility. This may be in conjunction with a high-traffic generator, such as a downtown area, a university, a recreational complex, or some other form of concentrated development. In such case the intermediate location of the distribution facility, either transit or highway, would deviate from the corridor to give this type of service and then, within a mile or two, would rejoin the corridor.

In the modern approach, transportation planners are working together with other disciplines to resolve the problems of future urbanization and to improve the quality of urban living. Transit, highways, and other modes are being considered in appropriate balance to serve the convenience of the people and to meet the natural desires and normal demands of the public.

# A Test of the Concept of a Household Shopping-Travel-Behavior Corridor

C. T. MOORE and J. B. MASON, University of Alabama

ABRIDGMENT

•THE QUESTIONS of whether household movement patterns depict a corridorized character is still largely unknown. In addition, the fact that a corridor may exist based on movement patterns does not answer the question of whether the construction of a new highway results in positive economic or social benefits within the existing corridor frame or within a new corridor frame.

There is no question that the construction of a new highway creates the potential for new land uses in an area near a new highway and hence presents the opportunity for new functional activities. On the other hand, are the economic developments, which take place near the new highway, transfers out of existing functional capacity in the corridor or primarily transfers of economic development from other proximate corridors? The fact that the new highway will generate as well as divert traffic from existing highways and arterials suggests that at least some increases in demand will take place within the area (for example, the demand for gasoline). The key factor in determining whether a new highway generates additional demand and not mere transfers of functional activity and consumer movements is to establish the investment and consumption responses that the new highway stimulates.

The research reported here is the result of a study conducted in Jefferson County, Alabama (conterminous with the Birmingham SMSA), during 1965 to ascertain whether there is justification for the concept of a household travel-behavior corridor. The findings are based on the results of 700 household interviews. The households were identified by using probability sampling techniques. Specifically, the research was designed to yield at least partial answers to the following questions. (a) Is a corridorized concept of travel behavior justified? (b) What is the general shape of such a corridor (eliptical, elongated)? (c) Do gravity models aid in explaining travel-behavior patterns within such an area? (d) Does travel behavior within the corridorized area depict a rational cost-conscious process? (e) Do demographic and economic variables aid in explaining the process of travel behavior? (f) What are the frequency, number, and duration of trips generated from such an area?

Conclusions were as follows. (a) The shopping-travel behavior of households depicted a rational, cost-conscious approach to the process of obtaining goods. (b) Relationships to functional retail centers were isotropic in character, and retail functional hierarchies apparently were easily perceived by households. (c) Corridor formulations were realistic and will be useful for additional analysis. (d) Gravity formulations were of only limited usefulness in explaining the origins and destinations of trips originating from the total area of analysis. (e) Demographic and economic variables did not adequately explain travel behavior within the corridorized areas.

Findings of this research are available in more detail in the following three volumes published by the Alabama State Highway Department, Montgomery: Land Use Analysis and Consumer Household Shopping Travel Behavior in a Highway Corridor Area: Theoretical Framework, Summary, Findings, and Conclusions, HPR Report 31-A, 1968; Land Use Analysis in a Highway Corridor Area, HPR Report 31-B, 1968; and Consumer Household Shopping Travel Behavior in a Highway Corridor Area, HPR Report 31-C, 1968.

---

Paper sponsored by Committee on Origin and Destination.

# Freeway Corridor Planning

DARWIN G. STUART, Barton-Aschman Associates

ABRIDGMENT

•FREEWAY corridor planning is identified as a means for achieving integrated transportation and land-use development. It can strengthen the process of specific route location for both urban and interurban freeways. Six different kinds of corridor planning are discussed, and each may be applied as a basic technique for coordinating freeways with adjacent land use. A general planning procedure for joint development within freeway corridors is outlined.

Recreation corridors are concerned mainly with the coordinated development of regional parks and interurban freeways. They offer major potential for meeting growing urban recreation demands, especially through the provision of high-level freeway access. They might involve both outdoor recreational facilities, such as picnic grounds, hiking trails, swimming beaches, or amusement parks, and cultural-recreational facilities, such as museums, stadiums, exhibition halls, or performing arts centers. Freeway service corridors are concerned with the provision of properly located service facilities for highway users, including safety rest areas, gas-food-lodging services, specialized truck parking facilities, overnight parking for camping trailers, and scenic overlooks. Industrial corridors represent a means for promoting coordinated metropolitan development, primarily in suburban and urban fringe areas. They can provide guidance for the substantial trend toward outlying industrial location and relocation at sites provided with convenient freeway access.

Integrated urban corridors focus on the need to design urban freeways in full coordination with multiple-purpose adjacent land use, often involving the multiple use of rights-of-way. Perhaps the most important needs and opportunities for freeway corridor planning lie in this area. Appropriate joint development projects might involve air rights development, neighborhood parks, major public buildings (medical facilities, educational buildings, government facilities, and cultural and public assembly facilities), various utilities systems, housing, and various private developments (office buildings, hotel, shopping centers, industrial parks, and distribution centers). Freeway-renewal corridors represent a special opportunity for achieving integrated urban corridor development. Within near-in central city areas, important potentials for freeway corridor planning can often be identified in association with urban renewal needs. Finally, coordinated transportation corridors are concerned with the accompanying provision of needed transit and parking facilities, including exclusive transit lanes for rail or bus, special turn-outs or passenger stops for express bus operations, and interchange parking facilities for car pooling and transit park-and-ride.

Perhaps the most critical problems in achieving a wider use of this corridor planning concept arise in connection with the practical aspects of interagency cooperation. Some kind of structured intergovernmental and public-private coordination is needed for identifying and organizing the major potential roles and decision points in the processes of freeway corridor planning and joint development. One such planning procedure might have two major administrative thrusts. First, it would generally place the initiative for inviting specific joint project proposals with the transportation agency involved, and, second, it places the initiative for independently studying and recommending joint project opportunities with an area-wide (state, county, metropolitan, or city) planning agency. In this way, much of the leadership and commitment necessary to draw other public agencies and private developers into a corridor planning joint development process could be provided.

---

Paper sponsored by Committee on Origin and Destination.

# Transportation Planning for New Towns

ROBERT L. MORRIS, Alan M. Voorhees and Associates, Inc.

The development of new towns is a growing phenomenon. These self-contained communities offer the transportation planners an opportunity to build on the basic research that has been carried out in urban areas in recent years. Employment opportunities juxtaposed with residential development, in conformance with accepted site-planning concepts, can reduce trip lengths. This is particularly true in the larger new towns. Street standards related to function can optimize construction costs and minimize the negative effects of noise and induced commercial development on residential areas. In spite of generally moderate intensity of land use, the cluster development concept can make good-quality public transit operation feasible. Exclusive rights-of-way, with linkages more direct than those that can be provided by automobile, ensure good transit service at relatively low cost. Parking can be designed according to realistic demands for space as functions of land use and characteristics of the residents. Good site planning can permit joint use of parking facilities.

•THE NEW TOWN concept that burst into full bloom in Europe shortly after World War II has become of increasing importance in American urban development. The role of new towns as a partial answer to growing urban problems received congressional recognition in the Housing Act of 1968. Title IV of this landmark legislation, known as the New Communities Act of 1968, authorizes federal guarantees to lenders who supply private capital for appropriate new town developments. It is clear that the handful of starts in this direction in the United States will soon be accompanied by a great many comparable developments. As is the case with the older established cities, the viability of the new communities will depend to a large degree on the effectiveness of their transportation systems.

Planning for the new towns involves two transportation considerations: an internal system to serve the new town and an external system to link the new town with the "mother" city. That is, new towns are clearly related to large metropolitan areas from which are drawn the population and business activities. Although they are designed as complete, self-contained communities, the new towns must have a basic link with the large established urban areas. (No inference should be drawn from the expression "self-contained" that these communities exist in isolation, independent of other urban areas nearby. Rather, they should be regarded as communities offering all the opportunities and advantages that can be found in any fully developed city.) For example, Reston is a part of the metropolitan Washington area, Columbia is related to Baltimore and Washington, Irvine Ranch is linked to Los Angeles, and Litchfield Park is connected to Phoenix.

## LAND USE

The basic element of the new town, of course, is the residential area. Generally speaking, these communities have been planned for middle-income families. The

Paper sponsored by Committee on Community Values and presented at the 48th Annual Meeting.

TABLE 1

POPULATION, HOUSING, AND EMPLOYMENT RELATIONSHIPS IN NEW TOWNS

| Town | Population | Dwelling Units | Employment |
|---|---|---|---|
| Basildon | 75,000 | 16,199 | 20,136 |
| Crawley | 63,700 | 13,753 | 17,595 |
| Harlow | 75,800 | 19,203 | 15,524 |
| Stevenage | 61,700 | 15,415 | 18,453 |
| Tapiola | 16,000 | 4,575 | 6,000 |
| Fort Lincoln (planned) | 16,000 | 4,500 | 5,000 |
| Columbia (planned) | 110,000 | 30,107 | 40,000 |
| Reston (planned) | 75,000 | 21,000 | 23,000 |

poverty element is often excluded by the planning and financing process rather than by policy. At the opposite end of the scale, one rarely finds real luxury in a new town. The reasons for the concentration on the middle-income levels of society are complex, primarily financial, and beyond the scope of this paper. Nevertheless, housing types are varied, ranging from single-family, detached dwelling units to high-rise apartments.

The concept of the self-contained community is carried out by providing job opportunities for those residents who wish to work near their homes. Such opportunities are of great importance to many people. For instance, the Fort Worth Community Planning Survey (1) in 1963 discovered that 57 percent of the residents of that city rated inadequate job opportunities as the most serious regional problem. A similar study in Albany, New York, revealed comparable findings. Many new towns plan employment approximately equal to the number of workers who will be living in the community.

Because it is not possible to match housing with jobs and because of the penchant of many people for living in places somewhat remote from their employment, many home-to-work trips will have one trip end outside the new town. As the size of the new town increases to a population of 100,000 or more with comparable job opportunities, the percentage of home-to-work trips remaining within the new town will likely increase.

Other uses of the urban land are related primarily to the residential areas and to lesser degrees to each other. In planning for Belconnen, Australia, a matrix of functional interrelationships for a typical town center was developed (Table 2). Most intimately related to the residential areas is the educational system, whose elementary schools serve their traditional role as a neighborhood focal point. Linking and serving the various land uses is the transportation system. Because mobility is important to the American public, and perhaps to others as well, the quality and adequacy of this system is of great importance to prospective residents. A comprehensive attitude survey taken in Greensboro, North Carolina (2), showed that, among 13 values rated by the residents, good roads and sidewalks ranked first. Social values came close behind, with convenient public transportation in fifth place. Accessibility seemed less important. Schools close enough so the children can walk to them ranked eighth, while closeby shopping facilities

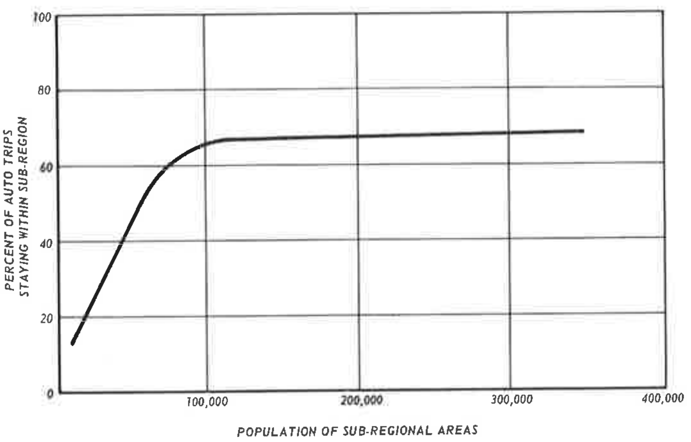



Figure 1. Effect of size of subregional areas on internal trips.

TABLE 2

FUNCTIONAL INTERRELATIONSHIPS OF TOWN CENTER COMPONENTS

| Plan Elements | Retail | Commercial Services, Banks, and Offices | Public Services and Civic Buildings | Entertainment, Hotels, and Commercial Recreation | Social Institutions, Churches, and Clubs | Hospital | Recreation, Stadium-Extension | College | Government Offices | Service and Trade | Spine Transit | Collector-Distribution Transit | Peripheral Freeway | Town Arterial System |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Retail | Concentration required; activity should not be dispersed | Provides market support; pedestrian proximity required; heavy interaction; moderate competition for parking | Provides market support; pedestrian proximity desired; medium interaction | Provides good market support; pedestrian proximity required | Little or no support | None | Little or none | Some market support; pedestrian interaction likely without close proximity | Provides market support; heavy competition for available parking | Storage and warehouse support; incompatible, good service vehicle access desired | Relatively unimportant | Enforce market support | Relatively unimportant | Prerequisite for market support |
| Commercial Services | Provides market support; close proximity required | Concentration desirable | Some support; pedestrian proximity desired | Provides service function; close proximity desired | Little or none | Little or none (doctors' offices) | Incompatible | Little or none | Some market support | Incompatible | Important for regional business contacts; access to labor force | Some support | Important for regional business contacts | Prerequisite for market support |
| Public Service Buildings | Little or none | Little or none | Concentration desired | Service function; close proximity desired | Little or none | None | None | Little or none | Little or none | None | Little or none | Important | Little or none | Prerequisite for market support |
| Entertainment | Some support; pedestrian proximity desired | Good market support; pedestrian proximity required | Minor market support | Concentration desired | Some support for clubs | None | Some support, however, many incompatible factors | Good support; pedestrian interaction desired | Some support | None | Little or none | Little or none (some for young age groups) | Fairly important | Prerequisite for market support |
| Social Institutions | Little or none | Little or none | Little or none | Little or none | Can be dispersed | Little or none | Little or none | Little or none | Little or none | Little or none | Little or none | Little or none except for youth clubs | Little | Low traffic generator but good access desired |
| Hospital | None; incompatible | None; incompatible | None; incompatible | None; incompatible | None | Single unit | None | None | None | None | Important | Important | Important | Important |
| Recreation | Some joint use of parking facilities possible | Some joint use of parking facilities possible | Some joint use of parking facilities possible | Service support | None | None; incompatible | Single unit | Could be important; possible joint use of facilities | Some joint use of parking facilities possible | None | Important | Important | Important | Important |
| College | Service function; relatively close proximity desired | Service function; relatively close proximity desired | Little or none | Service function; relatively close proximity desired | Little or none | None | Possible joint use of facilities | Single unit | Could be important; research support | None | Important | Important | Minor | Important |
| Government Offices | Employee service; pedestrian proximity desired | Employee service; pedestrian proximity desired | Employee service; pedestrian proximity desired | Employee service; pedestrian proximity desired | None | None | Possible joint use of parking facilities | Possibly important | Could be dispersed | None | Very important | Very important | Very important | Very important |
| Service and Trade | None; incompatible | None; incompatible | None; incompatible | None; incompatible | None; incompatible | None; incompatible | None; incompatible | None; incompatible | None; incompatible | Could be dispersed | Little importance | Little importance | Very important | Very important |

TABLE 3
AVERAGE WALKING DISTANCES

| Trip Purpose | Walking Distance (ft) | |
|---|---|---|
| | Children | Adults |
| Play | 400 | |
| Recreation | 800 | |
| School | 1,600 | |
| Services | | 200 |
| Business and Shopping | | 300 |
| Work | | 400 |

ranked eleventh. The street system and the transit system planners must take these values into account.

## THE PEDESTRIAN

Despite the importance of mobility, walking is a prominent means of transportation in new towns as well as in older cities. The United Nations' Symposium on the Planning and Development of New Towns found that ". . . those institutions, services, and amenities that are visited by the public frequently, but not necessarily daily, such as cinemas, clubs, health clinics, shopping centers, . . . should be sited within easy walking distance of all dwellings they serve (3)." The question of walking distance is, of course, a relative one. The studies for Fort Lincoln, in the District of Columbia, indicated that walking distances are a function of age and trip purpose. The approximate ranges for these walking trips, shown in Figures 2 and 3, are based on very small samples and should not be interpreted as the result of definitive studies.

Planning for a pedestrian network deserves the same consideration as that given to the street plan. Where feasible and where particularly heavy conflicts exist, grade separations should be provided between vehicular and pedestrian ways. This is particularly important where school children are involved. Good site planning can minimize the number of such grade separations and thereby reduce the total cost of public improvements. Cost reductions are of major consideration to developers who must provide huge sums of money well in advance of their having sufficient cash flow to pay debt service on the loans.

## THE ROAD SYSTEM

New town development creates an unusual opportunity for building a road system that is tailor-made for the specific functions that it will serve. The very low cost of land in new towns, compared with land value in established urban areas, makes freeways feasible for substantially lower volumes than those considered in existing cities.

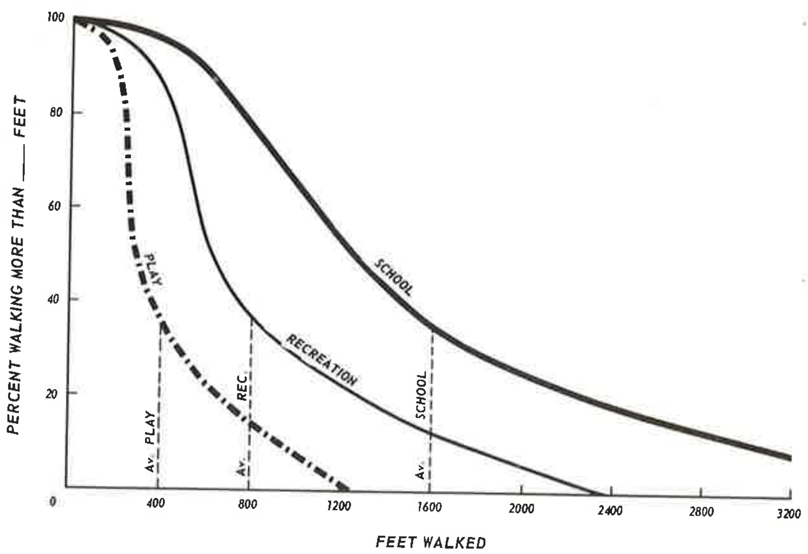



Figure 2. Children's walking trips.

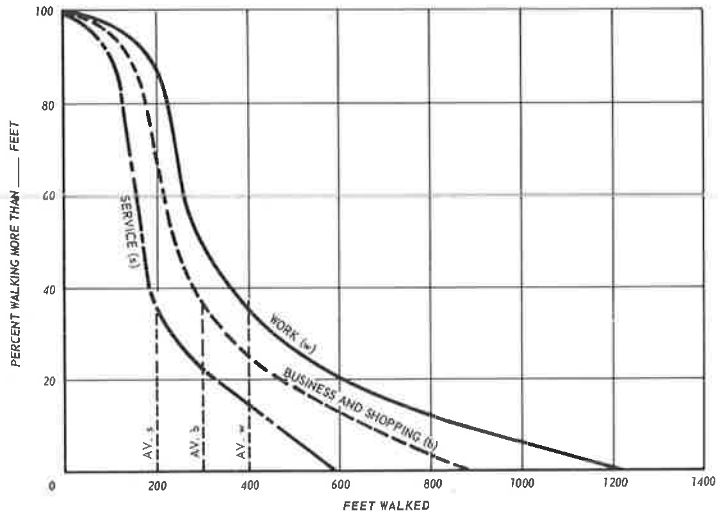



Figure 3. Adults' walking trips.

The general plan for Canberra, Australia (4), demonstrates the practicality of freeway development for daily traffic volumes in excess of 20,000 vehicles (Figs. 4, 5 and 6). The data for these figures are based on experience in Australia and Chicago. Capital costs do not include land. Construction costs are capitalized at 5 percent compound interest over 30 years. Average operating costs include the running cost of vehicles, cost of passengers' time, accidents, and road maintenance.

Richard Llewelyn-Davies has noted that ". . . the basis generally used for computation of network capacity is that there should be no delays to traffic by congestion at peak hour at the turn of the century when motor car ownership will have reached, or nearly reached, saturation point (5)." Admittedly, this is a very high standard. It may be achieved in Great Britain, where new town developers have the advantage of being able to borrow directly from the treasury of England. Where new towns are constructed by private enterprise, such high standards that provide capacity well in advance of need are not generally practicable. Nevertheless, a limited number of freeway miles is feasible not only to serve the highest internal volumes but also to link the new town with the mother city.

In addition to freeways, other streets should be developed to serve the specific needs of the new community. These streets can be classified according to the amount of traffic that is generated on them. The lowest class of streets will serve 25 dwelling units or less. Observed parking practices and probability estimates indicate that on streets in this class the chance that one driver will meet an oncoming car where two cars are parked opposite each other occurs about once every two months for an average driver.

$$\text{Probability of meeting} = 1 - e^{-(V_a/t)(2L/V)} + e^{-[(V_a+V_b)/t](2L/V)} - e^{-(V_b/t)(2L/V)}$$

where

$V_a$ = peak 15-min volume in "a" direction or 70 percent of the peak 15-min volume in both directions,
$V_b$ = peak 15-min volume in "b" direction or 30 percent of the peak 15-min volume in both directions,
t = time in sec (900 for 15 min)
L = length of street (500 ft)
v = velocity of vehicles in ft per sec (36.7 ft per sec or 25 mph)

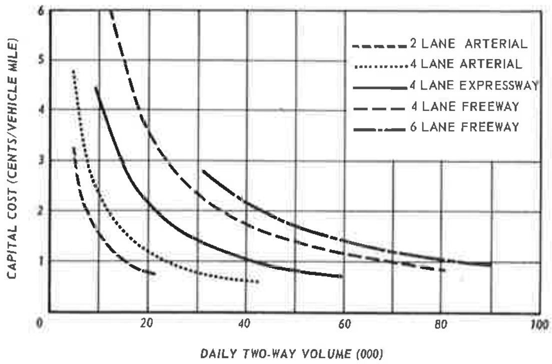



Figure 4. Capital cost versus traffic volume.

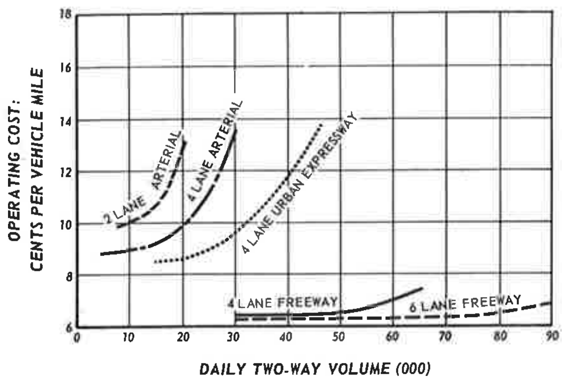



Figure 5. Unit operating cost versus traffic volume.

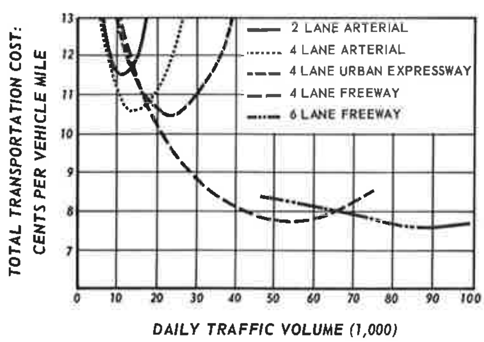



Figure 6. Unit transportation cost versus traffic volume.

This would indicate that on such streets two parking lanes and one moving lane are required. When a street serves more than 25 dwelling units or more than 250 average daily traffic (ADT), two moving lanes are required. When volumes exceed 1,500 ADT, vehicles will meet more frequently so that wider moving lanes are desirable.

The cluster-development concept, which is the basis for new town site planning, requires street design directly related to function. Streets can be used as parking lots or as parking connectors. With appropriate standards, the street will serve the desired function. These standards must realistically relate to criteria developed from field studies; this was done in Broward County, Florida (6). (See Table 4.) There it was found that as traffic increases on urban streets, two things occur: the noise level increases concomitantly, and land along the street tends to be used for commercial purposes. Figure 7 indicates the relationship between noise and daily traffic, and Figure 8, the relationship between traffic volume and commercial development in Jacksonville. These factors must be considered in street standards. A solution to the noise problem is to provide proper setbacks for buildings (Fig. 9). For residential areas, 70 decibels is considered the maximum acceptable noise level. Therefore, residential development should not be permitted to face streets that reach a volume of 10,000 vehicles per day.

Increasing volumes of traffic will likely result in pressures for commercial development. It is, therefore, essential to plan proper cross-sectional designs and controls of access as well as building setbacks for these heavily traveled streets. Table 4 indicates the street standards that have been designed to meet the criteria described above.

## THE TRANSIT SYSTEM

The principal function of a transit system is to improve circulation within the new town and to provide opportunities for those who do not have access to automobiles; it does not supplant automobile service. A good road system must be provided throughout the community because it provides people with the freedom to travel anywhere they want. Only in areas of great intensity of pedestrian activity, such as the town center, should the primary emphasis be on transit.

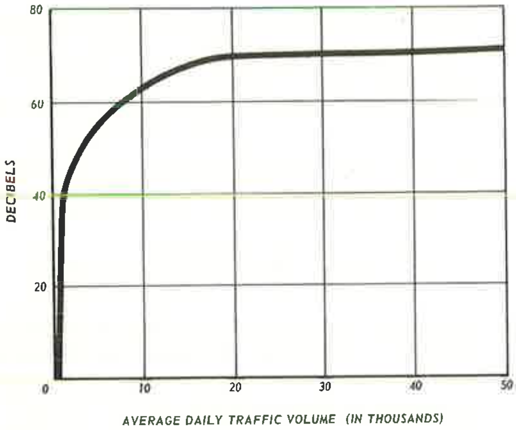



Figure 7. Relationship of traffic noise to traffic volume.

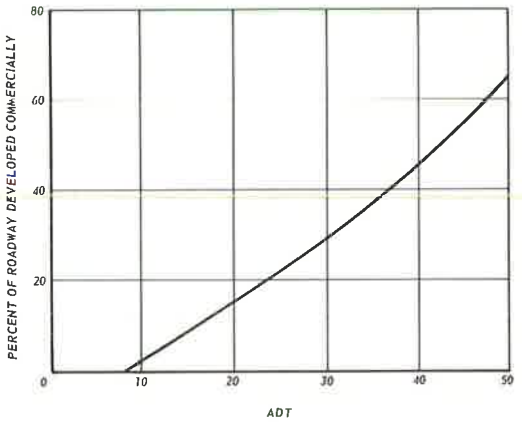



Figure 8. Relationship of commercial development and traffic volume.

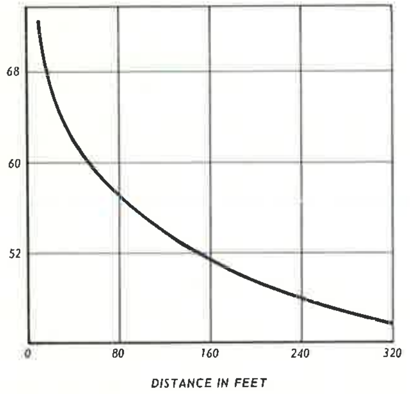



Figure 9. Relationship of noise level and distance from noise (10).

It is well recognized that successful transit systems require appropriate densities of development. The sparser the development, the less frequently can transit service be provided and, therefore, the less desirable will it be. For this reason, it may seem at first incongruous to discuss public transit with relation to new towns. However, many new towns have moderate densities, as shown in Table 5.

Notwithstanding their lack of intensive development, the new towns generally follow the cluster-development pattern, with large areas of open space and residential development concentrated in proportionately small areas of the total community. Such cluster development not only lends itself readily to mass transit operations, but also permits more economical road systems. Columbia, with cluster development has only 8 percent of land consumed by streets. In a typical subdivision, where travel is dominated by automobile movements, 20 percent of the land might be in streets. In many suburban areas, about 10 trips per day are produced by an average dwelling unit; 6 of these are automobile-driver trips, 3 are automobile-passenger trips, and depending on the degree of transit service, 1 is a transit trip.

TABLE 4

STREET STANDARDS

| Traffic Volume (ADT) | Moving Lanes | | Emergency Storage Lanes[a] | Design Speed | Building Setback |
|---|---|---|---|---|---|
| | No. Required | Width | | | |
| Under 250 | 1[b] | 10 | 8 | 25 | 20 |
| 250- 1,500 | 2 | 10 | 8 | 30 | 25 |
| 1,500- 5,000 | 2 | 12 | 8 | 35 | 30 |
| 5,000-10,000 | 4 | 12 | 10 | 40 | 40 |
| 10,000-20,000 | 4 | 12 | 12[c] | 50 | 60 |
| Over 20,000 | 4 | 12 | 12[d] | 60 | 100 |

[a]Required only if adjoining land use dictates.
[b]Two lanes if no emergency or storage lane.
[c]Plus 20-ft divider.
[d]Plus 40-ft divider.

TABLE 5

NEW TOWN DENSITIES

| Town | Planned Population | Area (acres) | Persons per Gross Acre |
|---|---|---|---|
| Tapiola | 16,000 | 6,150 | 26 |
| Hemel Hempstead | 80,000 | 4,980 | 16 |
| Harlow | 75,800 | 3,000 | 25 |
| Columbia | 110,000 | 13,690 | 8 |
| Reston | 75,000 | 6,800 | 11 |
| Fort Lincoln | 16,000 | 340 | 47 |
| Litchfield Park | 100,000 | 13,000 | 8 |
| Cumbernauld | 70,000 | 4,150 | 17 |
| Typical subdivision | — | — | 10 |
| District of Columbia | — | — | 20 |

The more variety that is built into the transportation system and the land it serves, the greater will be the ability of people to reach their own goals. In the typical suburb, the predominant trip purpose—about 35 percent—is for work or related activities. About 20 percent of the trips is for shopping, 25 percent for social-recreation purposes, and 20 percent for personal business and other reasons. The last category includes a substantial number of trips by housewives transporting children to various activities. A transportation system that relieves the housewife of these chores would give her more freedom, and it would also give the children greater opportunity to do things on their own. Such a system would undoubtedly generate trips beyond those that might be calculated empirically. If a child could go to the library any time he desired, he would certainly make use of the transportation system that affords this accessibility. The system would thus provide for many more social opportunities. Therefore, transportation in new towns should offer a variety of services—transit as well as automobile.

With good site planning, a transit system can provide internal transportation for trips to school, convenience shopping, work, social-recreation activities, and cultural and health facilities. In fact, a well-planned community with satisfactory transit service can reduce the need for automobile ownership and thereby save some development costs in terms of roads and parking spaces. Admittedly, this is a theory that has yet to be fully substantiated. As Columbia, Maryland, and Runcorn, England, among others, develop, it will be interesting to observe if practice follows theory.

The configuration of the new town transit system will be related to land development and vice versa. Each, in turn, will be governed to a large degree by topography. However, it can be demonstrated that the concept of a double loop, in the form of a figure eight, is more effective than the more common concept of radials. These configurations, depicted in Figure 10, each involve about the same length of routing, and each require the same number of vehicles for comparable service. However, the double loop permits passengers to go anywhere on the system without a transfer. Models developed in Washington, Caracas, and Toronto have shown that the inconvenience of making transfers is a strong deterrent to mass transit usage. (Each minute involved in transferring equals 2 to $2\frac{1}{2}$ min travel time.) Because it offers convenience and minimal travel time, a figure eight transit system was called for in the original plan for Columbia (7).

It must be emphasized that, for the transit system to be effective, land use development must be intimately related to the routing and particularly to the location of the stations. Generally speaking, a range of 25 to 30 percent of the residents can be expected to use public transit where there is satisfactory service. Good land use planning will provide a high degree of convenience for 25 percent of the residents in relationship to the transit stations. With a really good system, additional trips will be attracted beyond these 25 percent close-in residents so that a total of 40 percent might be attracted. Concentrations of high-rise apartments in the immediate proximity of the stations, with townhouses and single-family developments at more remote distances, provide an ideal solution. For example, in Columbia, 40 percent of the residents will be within a 2-min walk of a minibus stop.

The new town offers a unique opportunity to carry out basic planning concepts. A genuine choice between automobile travel and truly comparable public transit travel can be offered to the residents. It is possible to build communities that are free of automobile congestion and that, at the same time, offer fast, convenient, and comfortable transit service. The better the transit system is—that is, the more attractive it is to the residents of the community—the more likely it is to be a financially viable system.

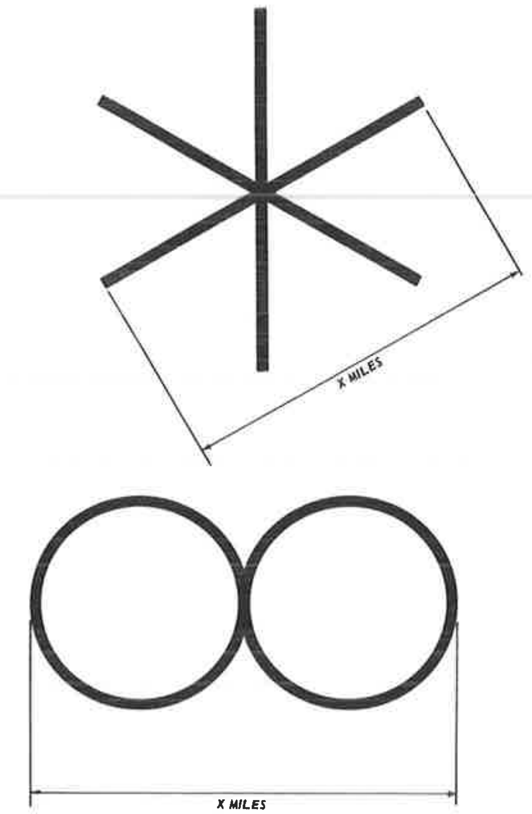



Figure 10. Basic configurations of transit systems.

It may not be essential, however, that the system be supported out of the fare box. This argument has been presented many times in regard to existing cities. The value of a transit system to a resident of a community who never uses the system is easily established. Increased property values, accessibility to the resident's home for those who do not have automobiles, and decreased traffic interference for the resident's automobile trip are all tangible, quantifiable benefits that can be taken into account to justify community support for the transit system.

Development of the new town provides a unique opportunity to relate land use to transportation as it can never be done in an established community. The provision of intense development around transit stations justifies adding an increment of these properties to the tax burden to assist the transit operation. Furthermore, the concomitant development of a road system and a public transit system permits alignments and configurations that can make transit more direct and therefore, to some degree, more appealing for some trips. A good example is the plan for Belconnen, the new town just outside Canberra, Australia (Fig. 11). Here the public transit line that links the new town with the mother city provides station stops directly in the intensively developed core of the town center. The convenience of this direct access, as compared with the walking trip that would be required from the automobile parking lot, adds to the attractiveness of the transit system. Public transit in Belconnen serves primarily as a link between towns. Other examples of internal transit systems in Figure 11 show alignments for Columbia, Runcorn, and Ft. Lincoln. In each of these three examples it can be seen that the public transit system offers more direct service for many trips than the road system can provide.

The equipment that can be used for public transit depends on a number of factors. Runcorn proposes using standard buses. The original plan for Columbia contemplated the use of minibuses, although this equipment for the long-range development of the community is being reconsidered. In each of these instances, exclusive rights-of-way were proposed for public transportation to ensure reliability and high-quality service. Buses that operate in mixed traffic are always subject to delays because of conflicts with other vehicles and pedestrians. Even on streets with adequate capacity, delays can be expected during periods of rain or snow or when accidents occur. (A study made in St. Louis (8) showed that 28 percent of all buses ran late during the afternoon rush hour and over 15 percent ran early. Even during the off-peak period, 12 percent of the buses ran late.) Clearly, the schedule-makers face a dilemma. If schedules are slowed down to permit more on-time operation, then the buses would not be providing service as fast as they are capable. Furthermore, costs would be increased because more driver-hours would be required to provide the same number of bus-miles. Providing service regularity is particularly troublesome because a bus running only slightly behind schedule sets in motion forces that tend to make it even slower. The time gap between an on-time bus and a following late bus is greater than normal. During this longer gap, more people arrive at the bus stop, board the late bus, and cause it to fall

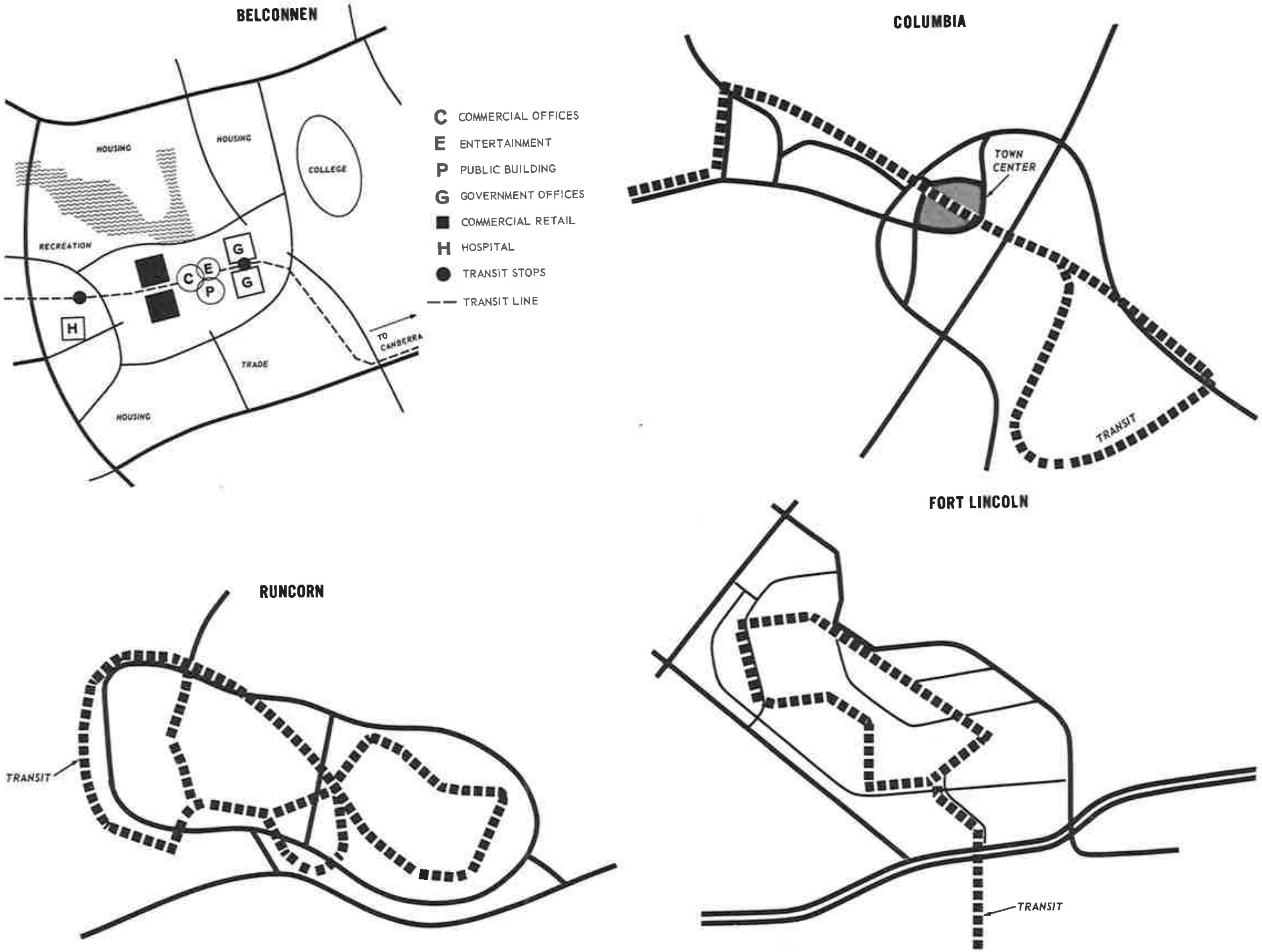



Figure 11. Transit plans for Belconnen, Columbia, Runcorn, and Fort Lincoln.

further behind schedule. The result of this is a tendency for buses to become bunched or to move in platoons instead of in evenly spaced units.

Buses on their own rights-of-way would largely avoid this problem. The use of exclusive busways would decrease travel time for bus passengers in two ways: by increasing scheduled speeds and by decreasing distances traveled. The usual transit delays, which account for about 15 percent of total bus travel time, caused by congestion, pedestrians, signals, and stop signs could be largely eliminated by the use of exclusive busways. For example, scheduled speeds of 15 mph have been anticipated for Columbia transit where, without busways, 12½ mph would be more likely.

Busways can also provide more direct routing patterns through some residential areas than would otherwise be possible. Main arterial roadways would purposely not be routed through residential concentrations or the city center. If busways were eliminated, then either additional main roads would have to be cut through these areas, or buses would have to be circuitously routed. Additional local streets of a continuous nature could not be added, because automobile traffic would tend to use these streets and to make them, in effect, through streets but without adequate capacity.

During the testing process that preceded the actual development of the proposed Columbia system, it was assumed that the busways were not developed. Therefore, buses were routed on the street system through Village I and II to the central business district to test the effect of the circuitous routing required. The route was 9,450 ft in length. The route over the more direct busways that serve the same area was only 7,000 ft. Thus, the street route was 35 percent longer than the busway route. For this particular service area, total travel time by busways was 50 percent less than that by street routing because of the reduction in traffic complex and more direct routing. At the same time, a 50 percent increase in operators' wages was required to provide the service on the streets. While it is conceivable that these differences might be narrowed by the judicious redesigning of the street system, it is apparent that the busways significantly improve the performance of the transit service to the consumer and reduce the cost of providing the service by the operator. The quantified effects of exclusive busways in Columbia, given in Tables 6 and 7, show that busways can make the difference between a profitable, or at least breakeven, operation and a deficit operation.

The Fort Lincoln plan contemplates an attractive, small-scale automated train that virtually dictates an elevated system throughout the community. Although automated systems promise savings in labor costs over the long-term operation, the more flexible bus system more readily lends itself to staged construction.

## PARKING

Parking requirements for residential land use are related to a variety of factors such as family size, family age, availability of transit, family income, level of automobile service, character and density of area, general climatic conditions, and kind of housing. Many of these factors are difficult to quantify. However, there is strong evidence that some basic relationships hold. Figure 12, derived by implication from studies by the Tri-State Transportation Commission, indicates a relationship among income, transit service, and automobile ownership. In planning for Fort Lincoln, measurements were made in the Washington, D. C., area of automobile ownership acccording to characteristics of housing and their residents. The findings are given in Table 8.

Demands for parking space at other land uses can be similarly measured. A study by the Urban Land Institute (9)

TABLE 6

DAILY BUS OPERATING COSTS

| Cost | For 20 Buses With Busways | For 28 Buses Without Busways | Percentage Increase, No Busways[a] |
|---|---|---|---|
| Maintenance | $ 240 | $ 693 | 25 |
| Operators | 640 | 896 | 40 |
| Supervisors | 20 | 28 | 40 |
| Administration | 100 | 140 | 40 |
| Payroll taxes | 48 | 67 | 40 |
| Vehicular taxes | 4 | 6 | 40 |
| Depreciation | 140 | 196 | 40 |
| TOTAL | $1,240 | $1,693 | |

[a]Bus units increase from 20 to 28; bus miles increase from 4,800 to 6,000 per day; bus hours increase from 320 to 448 per day.

TABLE 7

ANNUAL TRANSIT COSTS AND REVENUES[a]

| Costs and Revenues | With Busways | Without Busways |
|---|---|---|
| Revenues | $495,000 | $372,000 |
| Operating costs[b] | 372,000 | 508,000 |
| Cost of busways | 100,000[c] | — |
| Maintenance of busways | 5,000[d] | — |
| Total costs | 477,000 | 508,000 |
| Revenues less costs | 18,000 | -136,000 |

[a]Daily costs and revenues are annualized by a factor of 300.
[b]Includes equipment depreciation.
[c]Calculation assumes 10 miles of exclusive busway estimated to cost $125,000 per mile or $1,250,000.
[d]Maintenance costs for busway are estimated at $500 per mile per year.

TABLE 8

AUTOMOBILE OWNERSHIP RELATED TO RESIDENT CHARACTERISTICS

| Kind of Housing | Income Level | Age Group | Automobile Ownership |
|---|---|---|---|
| High-rise | High | Elderly | 0.33:1 |
| High-rise | High | Other | 1.30:1 |
| High-rise | Middle | Elderly | 0.20:1 |
| High-rise | Middle | Other | 1.10:1 |
| High-rise | Low | Elderly | 0.10:1 |
| High-rise | Low | Other | 0.20:1 |
| Low-rise | High | | 1.50:1 |
| Low-rise | Middle | | 1.30:1 |
| Low-rise | Low | | 0.40:1 |

TABLE 9

AVAILABILITY OF RESIDENTIAL OFF-STREET PARKING SPACE IN DAYTIME

| Kind of Housing | Percentage in Good Transit | Percentage in Fair Transit |
|---|---|---|
| Public | 25 | 30 |
| Middle-Income | 45 | 55 |
| High-Income | 65 | 70 |

indicated that 5.5 spaces per 1,000 sq ft of retail space would be appropriate for regional shopping centers. This might be a proper value in the town center of a new town where there is not good transit service. The same Urban Land Institute report showed that up to 20 percent of office space can be superimposed on the retail space without increasing the demand for parking. Other opportunities exist for multiple use of parking space. Where daytime and nighttime activities are in close proximity, there is normally no need to duplicate parking space for the nighttime uses. Table 9 gives findings, again from the Washington, D.C., area, as to the availability of parking space of residential areas during the daytime. This would indicate that joint use can be made of parking spaces for residential and commercial or other uses.

Walking distances are an important factor in the location of parking. If a walk in excess of 250 ft is required, there is a strong likelihood that streets more convenient

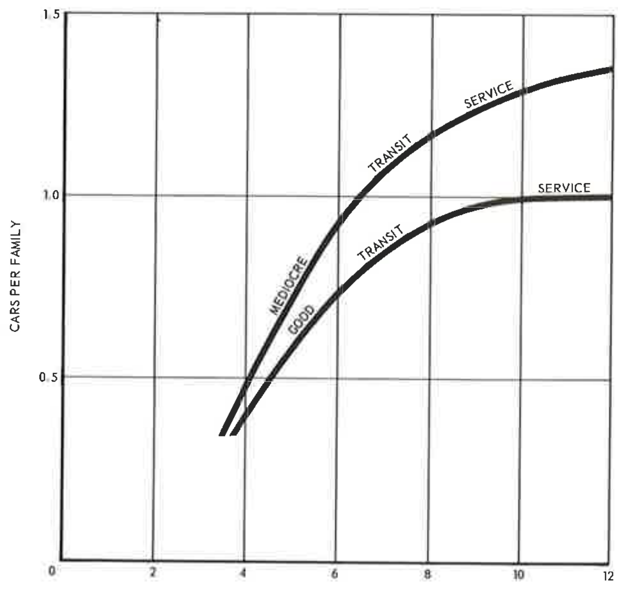



Figure 12. Automobile ownership versus income as a function of transit service.

than the planned parking lots may be used for parking. Many streets that do not have active uses fronting on them would not normally require parking space and the design cross section would not provide for this use. The excessively long walking trip, then, from the parking lot or garage to the building, which it is intended to serve, tends to be self-defeating.

Appearance is always an important consideration in site planning. Large areas of parked automobiles exposed to public view are not generally considered desirable. At the same time, low land values dictate against the construction of expensive garages. Careful site planning with good screening of both walls and natural growth can do much to improve the appearance of parking lots.

## SUMMARY

The planning of new towns is a combination of art and science. To a large degree, the art aspect has dominated because of lack of information on basic travel habits and living patterns in these new communities. Much has been learned, of course, in the mistakes as well as with the successes of older established cities. Perhaps the most important lesson that can be learned from the new towns is the advantage of flexibility in planning. Standards for development, including streets, parking, and densities, that vary according to actual conditions can provide economic benefits while enhancing the functional aspects of the community.

## REFERENCES

1. Community Planning Survey, Fort Worth Metropolitan Area Transportation Study. Alan M. Voorhees and Associates, Inc., and Burgwin and Martin, Consulting Engineers, April 1963.
2. Godshalk, David R. Comparative New Community Design. AIP Jour., Nov. 1967.
3. Report of the United Nations, Symposium on the Planning and Development of New Towns. Moscow, Aug. 24 to Sept. 7, 1964, p. 11.
4. Canberra General Plan Concept. Alan M. Voorhees and Associates, Inc., Jan. 1967.
5. Llewelyn-Davies, Richard. Town Design. The Town Planning Review, Oct. 1966.
6. Principles and Standards for Street Development in Broward County. Alan M. Voorhees and Associates, Inc., Oct. 1966.
7. The Proposed Transit System for Columbia. Alan M. Voorhees and Associates, Inc., Oct. 1964.
8. St. Louis Metropolitan Area Transportation Study. W. C. Gilman and Company, 1959.
9. Parking Requirements for Shopping Centers. Urban Land Institute, Bull. 53, Nov. 1965.
10. Winterbottom, D. M. The Nuisances of Traffic in Residential Areas. Traffic Quarterly, July 1965.

# Sensitivity to Various Parameters of a Demand-Scheduled Bus System Computer Simulation Model

JEFFREY M. BRUGGEMAN, Peat, Marwick, Livingston and Company; and
KENNETH W. HEATHINGTON, Department of Civil Engineering, Purdue University

A computer simulation model has been developed that simulates a demand-scheduled bus system offering door-to-door service. Sensitivity analyses were performed on various parameters in the model. The parameters studied were link travel time, maximum pickup time, shape of serviced area, frequency of calls, bus capacity, and length of trips. The outputs considered sensitive to a change in these parameters were cost of operation per passenger-mile, waiting time to be picked up, passenger travel time on the system, and the total time required to make a trip.

In general it was found that, as the link travel time increased, the cost per passenger-mile rose sharply, the waiting time was relatively constant, the travel time moderately increased, and the total time required to complete a trip also increased moderately. As the maximum pickup time increased, the cost per passenger-mile decreased; and the waiting, travel, and total time increased linearly. The shape of the geographical area served did not influence the waiting, travel, and total time as might be expected.

As the demand for service increased, the waiting time remained relatively constant; the travel time and total time had only a slight increase and then leveled off. The cost per passenger-mile decreased significantly up to a demand of approximately 175 calls per hour and then began to level off. The total operating cost of the system increased with an increase in demand for service. An increase in bus capacity had little effect on the waiting, travel, and total time. However, there was a slight decrease in cost per passenger-mile with an increase in bus capacity. When short trips (1 to 4 blocks) were excluded from being served, the cost per passenger-mile decreased somewhat. The waiting time remained relatively constant, but the travel and total time had a slight increase.

•THE RESTORATION of public transportation to a place of prominence in urban areas has become a pressing and perplexing problem facing transportation planners. Mounting evidence of steadily declining ridership is an indication that conventional transit modes do not meet the needs of current situations. Thus, a vast number of proposals for new transit modes has appeared in recent years. These proposals vary widely in scope, complexity, feasibility, and cost; they range from those that would tend to alter urban development to those that make more modest marginal improvements in existing techniques. One such proposal (BUSTOP) is that of a demand-scheduled bus system; the operational characteristics are discussed by Heathington, Miller, Knox, Hoff, and Bruggeman (1).

The concept of a demand-scheduled bus (DSB) system falls more nearly into the latter group of proposals that make marginal improvements. Requiring few major technological developments and operating on existing city streets, a DSB system will neither create havoc in the existing urban structure nor require exorbitant funding to

Paper sponsored by the Committee on Future Concepts and presented at the 48th Annual Meeting.

implement. On the other hand, a DSB system will not be able to supplant the automobile, although it can help to relieve congestion at crucial times and locations.

Although the actual operation and control of a DSB system differs among various proposals, the underlying concept is that a transit system can be developed that utilizes small bases that operate on local streets and respond to a specific call for service (1, 2, 3, 4). Such service would thus be similar to that provided by taxicabs, except that savings in cost can accrue by assigning more than one passenger to a vehicle.

Several different control schemes have been proposed, but it is now generally accepted that some type of computer control is needed in order to assign a passenger more efficiently to one of several available vehicles. Several different types of service have also been proposed, such as the so-called many-to-one system of diffused origins and a single destination (3). The most general scheme of diffused origins and destinations, the many-to-many system, has been proposed and is developed in more detail in this report (1).

Although the notion of a demand-scheduled bus system had been conceived some years before, the idea did not become popular until the publication of the details of the Genie system proposed at MIT in 1966 as part of Project Metran (2, 3). In the short time since then, much interest and attention has been given to further research on DSB systems at MIT and elsewhere (4). General Motors included an evaluation of a Genie-type system as part of its research on furture urban transportation technology (5). General Research Corporation has worked on developing an analytical model of the Genie performance (6). All of these studies have concentrated on the many-to-one system mentioned above.

## DEVELOPMENT OF BUSTOP

In April 1967, a group of transportation researchers at Northwestern University conceived of the idea of developing a generalized computer simulation model of a DSB system offering service to diffused origins and destinations (the many-to-many problem). It was felt that the potential of a DSB system was much broader than that of one serving only as a feeder to a line-haul transit station or some other high-demand location. A DSB system can, it is hoped, not only provide service to those people without access to an automobile but also divert some marginal automobile users. In addition, an investigation of the many-to-one problem could easily be carried out with a more general model that simulates the many-to-many problem.

The results of this effort, known as Project BUSTOP, go into greater detail about the need for and use of a demand-scheduled bus system. The present paper concerns itself with some sensitivity analyses of the specific simulation model previously developed.

BUSTOP employs a different control philosophy than that of the Genie model that imposes a rather rigid set of vehicle requirements and operating rules and obtains passenger service as an output. This has the effect of giving fair control over the operating costs of the system, but little guarantee regarding service to the customer. BUSTOP, on the other hand, specifies a rather strict set of passenger service criteria, which can be altered parametrically, and determines the system requirements as output. This approach seems more appealing if one is to develop a public transportation system in competition with other modes.

Specifically, the control logic specifies a minimum and maximum time (MINP, MAXP) for a passenger to be picked up after his call is received. Then, a latest possible delivery time is calculated from the maximum pickup time plus a travel time depending on the length of the trip. This travel time is equal to twice the link travel time (LTT) between origin and destination for trips under a mile (10 links), and 5 min plus the link travel time for trips over a mile. This decision rule can be varied parametrically as desired.

The service area selected is that of a rectangular grid network with constant travel times on each link. Vehicle control is maintained by a simple "first north, then east" rule that is acceptable because of the assumption that travel times are equal in all directions. For initial calibration, a square mile area was chosen with 9 blocks to the mile and a link travel time of 30 sec, corresponding to an overall speed of 13 mph. All these quantities can be varied parametrically within the model.

Any demand distribution can be accepted as an input to the model. For simplicity, a uniform distribution of calls over a given time period was chosen, as well as a uniform distribution of origins and destinations. As mentioned above, the many-to-one problem can be simulated by substituting a single location for all destinations. Similarly, a high-demand strip development can be simulated by adjusting the demand distribution to reflect such activity.

It is assumed that all vehicles are located initially at a central terminal, from which they are "generated" as needed. Passengers are assigned to the closest vehicle that can service them without violating the time constraints of any other committed passengers. Pickup and delivery may be made in any order, and all possible combinations are tested. Only if none of the vehicles on the system can service the passenger is a new vehicle generated. When a vehicle has delivered its last assigned passenger, it returns to the terminal and becomes the first to be "regenerated" when a call situation requires an additional vehicle.

No attempt at "optimal" assignment has yet been made. Because of the stochastic processes within the model, an optimal assignment at any given time is likely to be less than optimal after the next call is received. Likewise, no consideration to reassigning passengers has been made; once a passenger has been assigned to a vehicle, that vehicle is obligated to pick up and deliver him. Assignments on an optimal basis might prove to be worthwhile; however, the effect of reassignment on passenger level of service is a very difficult question to answer and probably should await the development of a model to predict demand.

## SENSITIVITY ANALYSIS

It is believed that the development of a workable DSB system should proceed in three phases. The first phase is the development and application of a large-scale computer simulation model of a DSB system. Through sensitivity analyses on the model, the influence of many parameters can be evaluated, which would not be feasible in a field-test situation.

Parallel to the development of the simulation model should come the development of a demand model for passenger usage of the system. This model must be sensitive to changes in the various control parameters of the simulation model, as they affect passenger demand. These parameters must include not only absolute quantities such as waiting and travel time but also such things as reliability of service.

The third phase must be the development of adequate cost data based on various levels of hardware requirements, control procedures, and levels of service. Only after these three steps have been completed should a full-scale demonstration or operational project be undertaken.

The response or sensitivity of the simulation model to changes in various parameters is one of the most basic studies needed to be undertaken. Some preliminary sensitivity investigations of the primitive model discussed above were undertaken in an attempt to get at these responses. Although the simulation model under consideration is much simpler than that of an actual proposed operating system, it is felt that some valuable insight into the operation of such a system was gained.

Several of the parameters mentioned above were varied, one at a time, and the output from the model recorded. The simulation period chosen was one hour of real time. Operation of an "up-and-down" situation was studied; that is, the simulation was started "cold" with all vehicles in the terminal and was continued past the one-hour cutoff point for calls until all passengers were delivered and the vehicles had returned to the terminal. Other simulation techniques could have been chosen, such as a "steady-state" period out of the middle of the simulation run or perhaps a cutoff of the simulation at the end of the hour. The up-and-down technique was selected, however, because it was desired to study the behavior of the model under both of the end situations.

Output from the model is in two forms: passenger data and vehicle operation data. Passenger data consist of average waiting, travel, and total time per passenger on the system and are presented here directly in that form. These results could be used as feedback to a demand model in order to determine the effect of different passenger statistics on generated demand. The distribution of these service times was also available but was not used in this analysis, although it would form a measure of system dependability as an input to a demand model.

Vehicle performance data were mainly of two types: total number of vehicles used and total number of vehicle-minutes on the system. Vehicle occupancy, because it is a transitory phenomenon in DSB operation, was not judged to be a meaningful output. For purposes of presentation, these vehicle output results were converted to crude operating costs. Vehicle costs were assumed to be $5,000, amortized over a 10-year period at 6 percent interest, and based on usage of only 4 peak hours per day (1,000 hours per year). Driver-labor cost was set at $3.00 per hour per vehicle generated and came to $3.35 per simulation period that allowed for an average of 7 extra minutes required by the up-and-down operation. Finally, a cost of 2.2 cents per vehicle-minute, corresponding roughly to 12 cents per vehicle-mile, was selected to cover the cost of vehicle operation. Additional costs, such as garage facilities at the terminal, control facilities, taxes, licenses, and administration, were felt to be relatively constant over the range of situations considered and were not included in the evaluation.

These costs are meant to be only illustrative. For this reason, the actual output from the various simulation runs is included in the Appendix so that the reader may insert more precise estimates of costs and achieve a more meaningful cost evaluation. Such evaluations however, will only alter the magnitude of the relationships here presented; the fundamental nature should remain relatively unchanged.

Each data point included in the Appendix and used in the analysis is the average of 5 runs of the simulation model with different "seeds" for the random generation of calls. However, the set of calls is the same between points that are based on the same number of calls and size of area. A total of 4 data points was calculated from each run of the simulation model, which took, at a call frequency of 100 per hour, roughly 105 sec of central processor time on a CDC 6400 computer. This means that data for over 20 hours of real time were obtained in just over a minute and a half of computer time, with no attempt made at optimizing the existing FORTRAN IV code. The time increased roughly proportionally with an increase in call frequencies.

Several input parameters to the model remained invariant throughout all the simulation runs. The minimum waiting time was set at 1 min and the loading and unloading time at 15 sec. It was not felt that realistic selection of other values for these parameters would be of sufficient magnitude to noticeably affect the result. The simple travel-time rule discussed above was used throughout, although other rules could have certainly been incorporated. No runs were made with other than uniform call distributions.

## SIMULATION RESULTS

The first parameter investigated was link travel time. Values from 18 sec (22 mph) to 36 sec (11 mph) were chosen. (These different travel times could be interpreted equally well as resulting from closer street spacing and hence a smaller area.) Two slightly different service policies were selected: the first was based on a division between long and short trips by distance, and the second was based on a time value of 5 min. Both are equivalent with a link travel time of 30 sec. The results for the two cases differed only slightly and are shown in Figure 1. Cost per passenger-mile rose steeply with increasing travel time, as expected. Average waiting time increased very little, which was somewhat surprising, and remained approximately half of the maximum value of 6 min. Travel time increased linearly, and total time, of course, followed the same general shape.

The effect of changing the maximum allowable waiting time was not as well defined. The cost of operation showed a tendency to decrease, though not nearly as much as might be anticipated, and the variation about the line shown in Figure 2 was quite high. Of equal interest was the behavior of the passenger time distributions. All three closely followed a linear pattern. The average waiting time increased, although much

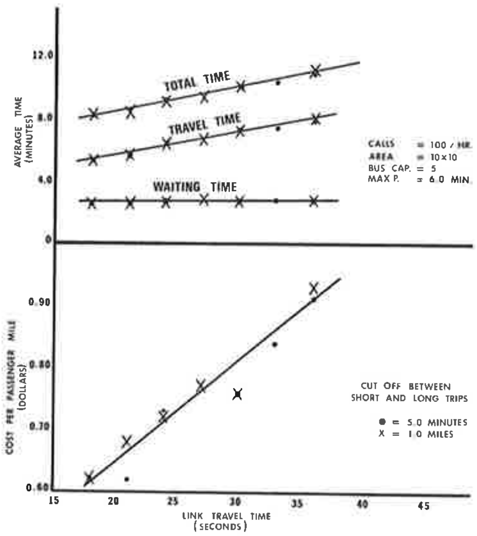



Figure 1. Sensitivity to changes in link travel time.

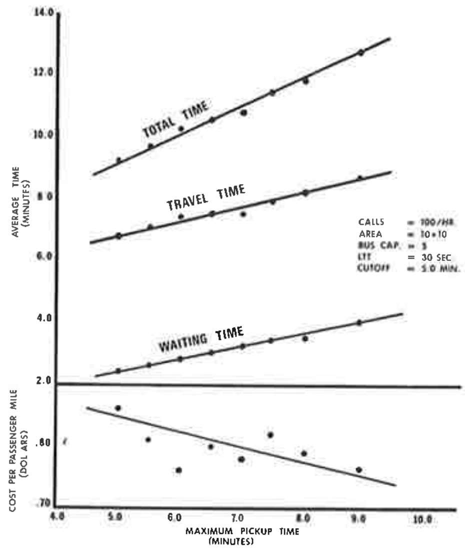



Figure 2. Sensitivity to changes in maximum pickup time.

less rapidly than might have been anticipated. More remarkable, however, is the fact that the average travel time increased slightly more sharply than the waiting time, even though no changes were made in travel parameters. This indicates a higher flexibility in vehicle routing, although at a significant inconvenience to passengers already assigned. Thus, only a very minor cost savings developed, possibly at the expense of a considerable reduction in patronage caused by the increased travel time, plus the added uncertainty as to actual pickup time. The results seem to indicate, at least for the demand used here (100 calls per hour), that increasing the maximum pickup time would not be a profitable change in system operating characteristics.

A third analysis was performed using differently shaped areas and the results are shown in Figure 3. The areas selected were just slightly smaller than the basic 10 by 10 grid used in the other analyses. Although a slight increase in cost per passenger-mile over the square area was noted, no further increase was detected even for quite elongated shapes. Almost no change in passenger output was detected among any of the areas. These results are rather significant in that they show that DSB operation in districts of different shape, but similar land area, will be almost identical. One difficulty, of course, is that in extremely elongated areas, either (a) the service to the extreme points must be lowered, (b) a rather high uniform maximum pickup time must be set, or (c) more than one vehicle "generator" must be provided. The latter is probably the preferred solution.

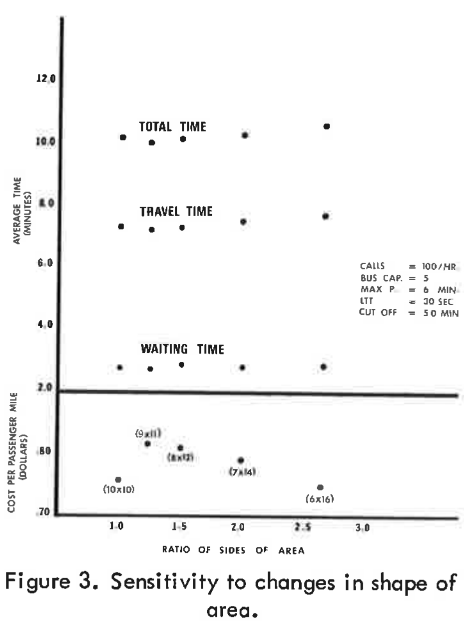



Figure 3. Sensitivity to changes in shape of area.

Demands of from 50 to 250 passengers per hour were examined. The results shown in Figure 4 indicate that the total operating cost increases linearly beyond 75 passengers per hour. However, the

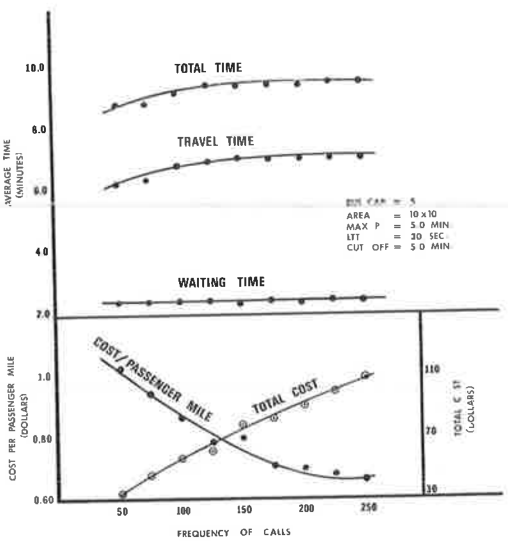



Figure 4. Sensitivity to changes in demand.

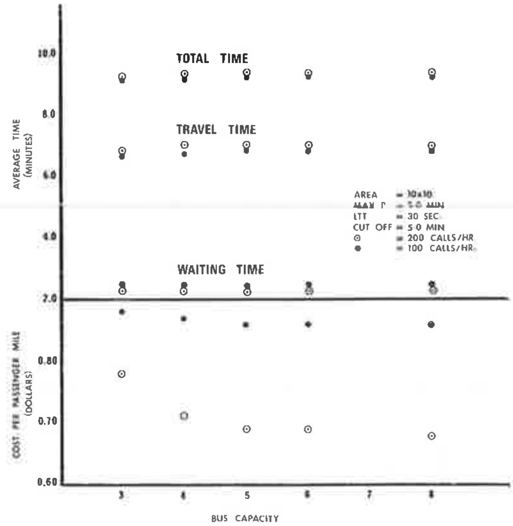



Figure 5. Sensitivity to bus capacity.

average cost per passenger-mile decreases significantly up to about 200 passengers per hour, beyond which little change occurs. The effect on passenger service is rather interesting. No effect on average waiting time was observed. Travel time was found to increase slightly up to about 150 calls per hour, after which no effect was observed. These results seem to indicate that the system will provide virtually identical service over a fairly wide range of demands.

The effect of vehicle size was examined at two different demand levels: 100 and 200 calls per hour. Figure 5 illustrates the results. The cost figures here are somewhat spurious, because it was assumed that the capital and operating costs for the different size vehicles were the same. Again, the reader is referred to the data in the Appendix for use in determining the cost functions for different vehicle sizes based on different unit operating costs. At 100 calls per hour, so few buses are ever filled that only a very slight increase in equipment is required when three- and four-passenger vehicles are used. At 200 calls per hour, a more noticeable variation in necessary equipment of different sizes was noted, although the advantages of six- and eight-passenger vehicles over five-passenger ones were very small. However, the expected cost per passenger-mile was much lower in all cases at the higher demand level and indicated higher utilization of available vehicle capacity. Passenger output showed almost no appreciable change among vehicle sizes or among different call frequencies. Although insufficient data are available for any really meaningful conclusions, one might infer that, for any given demand level, an increase in vehicle capacity beyond a certain point will not improve system performance. However, any cost savings involved in using a minimum size vehicle must be offset against the loss in flexibility potential.

Finally, it was hypothesized that model performance would be improved if unrealistically short trips were excluded from the analysis. This was examined by successively eliminating all trips of under 1, 2, 3, and 4 blocks in length. Call frequencies of 100 and 200 calls per hour were generated for each case, and the results are shown in Figure 6. Although total operational cost changed very little, the cost per passenger-mile increased when only the shortest trips were eliminated and then decreased steadily as longer trips were eliminated. Passenger waiting time remained virtually constant, but travel time increased slightly, reflecting the increased desired travel distance.

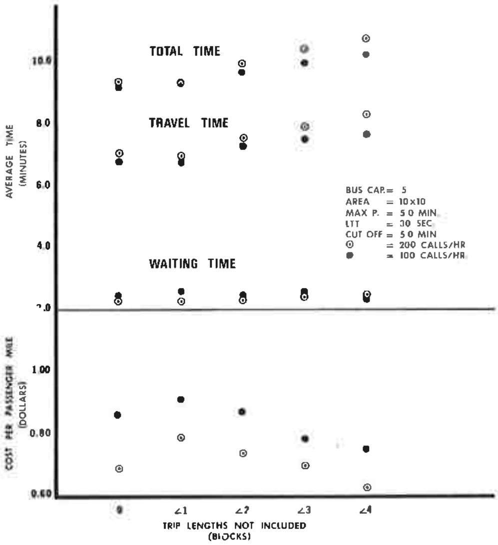



Figure 6. Sensitivity to reduction in short trips.

## MODEL PERFORMANCE

The output of the model in terms of cost is most closely tied to driver salary. This is an unfortunate inevitability in DSB operation, because such a system cannot be readily automated as can a more conventional rail system. Also, a critical element here is the generation of an additional vehicle to meet a specific demand. Although possibly needed for only one passenger, the vehicle remains on the system and lowers the overall passenger-to-vehicle ratio for all other vehicles. Perhaps some form of a buffer should be established in the model to raise the allowable pickup time for a call that would otherwise cause a new vehicle to be generated. Although this would reduce overall system reliability, the cost savings might be worthwhile. An alternative is to have some vehicle, designated as an "emergency" vehicle, that would respond to calls only when no other vehicle can handle them; this vehicle would not be checked for availability under normal circumstances. Any of these variations could be readily incorporated into the model, although a meaningful evaluation of their effect could be difficult without a highly sensitive demand model.

Other system alterations could be incorporated. The insertion of irregular, though reasonably well-behaved, boundaries or barriers could be made. The coding of an actual street system with varying link travel times is not possible at this stage in the model development, but the use of average times with a reasonably small list of special situations could be incorporated. Thus, operation on a real street system might be roughly approximated, without the development of a general model, although the size of the service area might have to be rather small.

Sophisticated alterations on the demand side, including varying demand temporally or spatially, the inclusion of priorities or varying service levels, or the use of special purpose vehicles for different trip types could be incorporated with varying levels of difficulty. Once again the evaluation of such situations could require a highly sophisticated demand model and a much more adequate set of system operating costs.

Although the output seems to indicate that such a system will be quite costly, even with the assumption that there is a low unit cost, it should be noted that an extremely high level of service has been maintained. Also, a quite small area and very low demands have been used, which do not allow for any economies of scale. Likewise, the assumption that demand distribution is uniform, though the most general, is probably also the most inefficient, because it does not provide for economies in concentrating origins and destinations. Further investigation of these and other extensions must await the development of a more sophisticated model.

## CONCLUSIONS

In general, the results of this study show that a very flexible, highly efficient simulation model of a demand-scheduled bus system can be constructed and operated for the many-to-many problem. Parametric examination of the model yields the significance of various control parameters in determining system efficiency. The model is highly sensitive to link travel time, but is relatively insensitive to the maximum allowable pickup time or the shape of the area. The effect of vehicle size does not seem to become apparent until a substantial number of vehicles are operating at capacity. Finally, the cost per passenger-mile decreases with increasing demand, but appears to approach some minimum value under a given set of system characteristics.

Authors' Note—Since this paper was written, a U.S. Department of Housing and Urban Development publication on DSB operation has appeared in the series "Study in New Systems of Urban Transportation." This publication, "Study of Evolutionary Urban Transportation," was prepared by Westinghouse Air Brake Company (WABCO) and others. Although a considerably different algorithm was used, the findings quite noticeably support those of this paper. In addition, the WABCO report contains considerable cost data that might be applied to the simulation results contained in the Appendix of this paper.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance and encouragement of Donald S. Berry of the Department of Civil Engineering, Northwestern University, and the contribution of the Department of Civil Engineering in providing funds for the computer time necessary to complete this project. Also acknowledged is the assistance of Mrs. Jean Haefner, Mrs. Lee Egger, and Northwestern University in producing and duplicating the manuscript.

## REFERENCES

1. Heathington, K.W., Miller, J., Knox, R.R., Hoff, G.C., and Bruggeman, J.M. Computer Simulation of a Demand-Scheduled Bus System Offering Door-to-Door Service. Highway Research Record 251, 1968, pp. 26-40.
2. Jones, P.S. Miniature Bus Systems. Proc. SRI Urban Transportation Alternatives Symposium, Menlo Park, 1964.
3. Hanson, M. E., Ed. Project Metran, M.I.T., Cambridge, 1966.
4. Wilson, N. H. M. Computer Aided Routing System. M.I.T., Cambridge, Jan. 1967.
5. Canty, E.T., Sobey, A.J., and Wallace, J.P. New Systems Implementation Study. Transportation Research Dept., General Motors Corp., Vols. 1, 2, and 3, Feb. 1968.
6. Brennand, J.R., and Sirbu, M.A. An Analytical Model of Genie Performance. General Research Corp., Rept. IMR-689, Aug. 1967.

# *Appendix*

The summary of simulation runs are given in the following table. Maximum pickup time (Maxp), link travel time (LTT), and cutoff point (COP) are shown in minutes. Waiting time, travel time, total time, and bus-minutes are totals for all passengers and vehicles over the simulation period. The total amount of desired travel (Lnks) is given in blocks. Gross occupancy (Occ.) is given as the ratio of passenger-minutes to bus-minutes. (Passenger-minutes is the same as total travel time.)

SUMMARY OF SIMULATION RUNS

| Run | Calls | B-Cap | Area | Maxp | LTT | COP | Lnks | Wait | Trvl | Totl | Bus | B-Min | Occ. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 100 | 5 | 10x10 | 6.0 | 0.30 | 5.0 | 658 | 263 | 557 | 819 | 8.2 | 360 | 1.57 |
| 2 | 100 | 5 | 10x10 | 6.0 | 0.40 | 5.0 | 658 | 270 | 651 | 921 | 9.6 | 445 | 1.48 |
| 3 | 100 | 5 | 10x10 | 6.0 | 0.50 | 5.0 | 658 | 280 | 736 | 1017 | 9.8 | 504 | 1.47 |
| 4 | 100 | 5 | 10x10 | 6.0 | 0.60 | 5.0 | 658 | 297 | 799 | 1096 | 11.6 | 612 | 1.31 |
| 5 | 100 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 658 | 241 | 678 | 919 | 11.2 | 533 | 1.28 |
| 6 | 100 | 5 | 10x10 | 6.0 | 0.50 | 5.0 | 658 | 280 | 736 | 1017 | 9.8 | 504 | 1.47 |
| 7 | 100 | 5 | 10x10 | 7.0 | 0.50 | 5.0 | 658 | 323 | 755 | 1077 | 10.0 | 500 | 1.52 |
| 8 | 100 | 5 | 10x10 | 8.0 | 0.50 | 5.0 | 658 | 359 | 824 | 1183 | 10.2 | 516 | 1.61 |
| 9 | 100 | 5 | 10x10 | 5.5 | 0.50 | 5.0 | 658 | 259 | 702 | 961 | 10.4 | 519 | 1.36 |
| 10 | 100 | 5 | 10x10 | 6.5 | 0.50 | 5.0 | 658 | 300 | 752 | 1052 | 10.4 | 511 | 1.48 |
| 11 | 100 | 5 | 10x10 | 7.5 | 0.50 | 5.0 | 658 | 345 | 796 | 1141 | 10.6 | 525 | 1.53 |
| 12 | 100 | 5 | 10x10 | 9.0 | 0.50 | 5.0 | 658 | 404 | 878 | 1281 | 9.6 | 529 | 1.67 |
| 13 | 100 | 5 | 9x11 | 6.0 | 0.50 | 5.0 | 657 | 278 | 723 | 1002 | 10.6 | 516 | 1.41 |
| 14 | 100 | 5 | 8x12 | 6.0 | 0.50 | 5.0 | 662 | 288 | 730 | 1018 | 10.6 | 504 | 1.46 |
| 15 | 100 | 5 | 7x14 | 6.0 | 0.50 | 5.0 | 686 | 280 | 747 | 1027 | 10.6 | 518 | 1.45 |
| 16 | 100 | 5 | 6x16 | 6.0 | 0.50 | 5.0 | 722 | 287 | 773 | 1060 | 10.6 | 519 | 1.50 |
| 17 | 100 | 5 | 10x10 | 6.0 | 0.30 | 3.0 | 658 | 271 | 553 | 825 | 8.2 | 354 | 1.58 |
| 18 | 100 | 5 | 10x10 | 6.0 | 0.40 | 4.0 | 658 | 271 | 653 | 924 | 9.4 | 447 | 1.47 |
| 19 | 100 | 5 | 10x10 | 6.0 | 0.45 | 4.5 | 658 | 284 | 668 | 952 | 10.0 | 489 | 1.38 |
| 20 | 100 | 5 | 10x10 | 6.0 | 0.60 | 6.0 | 658 | 297 | 813 | 1111 | 12.0 | 605 | 1.35 |
| 21 | 75 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 500 | 188 | 477 | 665 | 9.4 | 422 | 1.14 |
| 22 | 150 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 979 | 355 | 1053 | 1407 | 15.4 | 731 | 1.59 |
| 23 | 200 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1301 | 473 | 1396 | 1869 | 17.4 | 912 | 1.55 |
| 24 | 250 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1630 | 600 | 1773 | 2372 | 20.6 | 1117 | 1.61 |
| 25 | 50 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 332 | 121 | 318 | 440 | 6.8 | 296 | 1.08 |
| 26 | 125 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 820 | 315 | 859 | 1175 | 12.4 | 637 | 1.37 |
| 27 | 175 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1143 | 429 | 1222 | 1651 | 15.6 | 822 | 1.50 |
| 28 | 225 | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1457 | 545 | 1587 | 2133 | 18.6 | 1045 | 1.54 |
| 29 | 100 | 3 | 10x10 | 5.0 | 0.50 | 5.0 | 658 | 247 | 660 | 907 | 11.4 | 542 | 1.23 |
| 30 | 100 | 4 | 10x10 | 5.0 | 0.50 | 5.0 | 658 | 241 | 675 | 916 | 11.4 | 535 | 1.27 |
| 31 | 100 | 6 | 10x10 | 5.0 | 0.50 | 5.0 | 658 | 241 | 678 | 919 | 11.2 | 533 | 1.28 |
| 32 | 100 | 8 | 10x10 | 5.0 | 0.50 | 5.0 | 658 | 241 | 678 | 919 | 11.2 | 533 | 1.28 |
| 33 | 200 | 3 | 10x10 | 5.0 | 0.50 | 5.0 | 1301 | 470 | 1360 | 1830 | 20.0 | 982 | 1.40 |
| 34 | 200 | 4 | 10x10 | 5.0 | 0.50 | 5.0 | 1301 | 475 | 1394 | 1860 | 17.8 | 935 | 1.51 |
| 35 | 200 | 6 | 10x10 | 5.0 | 0.50 | 5.0 | 1301 | 477 | 1394 | 1871 | 17.2 | 923 | 1.53 |
| 36 | 200 | 8 | 10x10 | 5.0 | 0.50 | 5.0 | 1301 | 473 | 1388 | 1861 | 17.0 | 916 | 1.54 |
| 37 | 100 | 5 | 10x10 | 6.0 | 0.35 | 5.0 | 658 | 267 | 605 | 872 | 8.0 | 389 | 1.57 |
| 38 | 100 | 5 | 10x10 | 6.0 | 0.55 | 5.0 | 658 | 298 | 753 | 1051 | 10.8 | 547 | 1.39 |
| 39 | 100 | 5 | 10x10 | 6.0 | 0.35 | 3.5 | 658 | 264 | 596 | 860 | 8.8 | 402 | 1.50 |
| 40 | 100[a] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 682 | 254 | 676 | 931 | 12.2 | 593 | 1.18 |
| 41 | 100[b] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 713 | 242 | 728 | 971 | 12.4 | 582 | 1.26 |
| 42 | 100[c] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 754 | 251 | 746 | 997 | 11.4 | 587 | 1.28 |
| 43 | 100[d] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 793 | 239 | 781 | 1020 | 11.6 | 582 | 1.34 |
| 44 | 200[a] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1337 | 460 | 1390 | 1849 | 20.4 | 1029 | 1.37 |
| 45 | 200[b] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1421 | 477 | 1503 | 1980 | 20.6 | 1059 | 1.43 |
| 46 | 200[c] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1525 | 499 | 1577 | 2076 | 21.4 | 1060 | 1.50 |
| 47 | 200[d] | 5 | 10x10 | 5.0 | 0.50 | 5.0 | 1620 | 484 | 1670 | 2154 | 19.6 | 1057 | 1.59 |

[a]No calls less than 1 block included.
[b]No calls less than 2 blocks included.
[c]No calls less than 3 blocks included.
[d]No calls less than 4 blocks included.

# An Analytical Technique for Identifying Freeway and Expressway Systems as Part of the Rural State Highway System

FREDERICK J. WEGMANN, Department of Civil Engineering, West Virginia University; and
DONALD S. BERRY, Department of Civil Engineering, Northwestern University

A link-analysis technique, which relies on multivariate statistical procedures, is presented as a possible guide for designating major rural highways into freeway, expressway, and other major rural road systems. The link-analysis technique groups road links into systems on the basis of the functional characteristics of travel on each link. After the goals and objectives of the major highway systems are formulated, a criterion function is identified to select highway links to be included in each of the three systems. Measures such as average traffic volume, average trip length, and percentage of commercial vehicles are used to place road links into systems based upon a modified regionalization grouping procedure and multivariate discriminant analysis. Application of the model requires the availability of a statewide traffic model to define the existing or projected traffic demands for each road section. Also presented are the results of testing the technique on the major rural road system in southern Illinois.

CURRENTLY, questions are being posed as to the extent and nature of the intercity highway network after 1973, the anticipated completion date of the Interstate System. At present, many states have initiated comprehensive need studies in an attempt to define the structure of the after-1973 highway network, and some states have proposed construction of a supplemental freeway system (1). For example, Wilbur Smith and Associates recently recommended construction of a 2,176-mile (at an estimated cost of $4.2 billion) supplemental freeway system for Illinois (2).

The highway planning engineer must then be concerned with determining the level of highway service to be provided in major rural traffic corridors in order to best accommodate future travel demands and also to achieve the goals established for the highway system. (Level of service can be defined in terms of operating speed, travel time, safety, volume-to-capacity ratio, and vehicle operating cost. Level of service can be implemented by degree of access control, geometric design standards, and operational measures.) Through a routine process of designating the level of highway service to be provided on each road link, the highway planner is involved in the first step of identifying statewide freeway and expressway systems.

Proper selection of a supplemental freeway or an expressway system requires that the system be identified by a uniform measure that can be applied on a statewide basis. Individual preferences and decisions should be reserved for evaluation of alternative plans that have been generated. The link-analysis technique (multivariate statistical procedure) presented in this paper identifies road links exhibiting similar travel functions by placing all road links into groups that statistically maximize within group homogeneity. The uniform identification of road links displaying similar functions

Paper sponsored by Committee on Highway Costs and Programming and presented at the 48th Annual Meeting.

guides the highway planning engineer in the identification of statewide freeway and expressway systems. For final designation of the systems, utilization is made of the logic that road links displaying similar travel functions should receive identical levels of service and that road links located in the groups reflecting the greatest functional importance should be provided with the highest level of service.

Rightfully, the highway engineer still retains his responsibility of evaluating the statewide highway plan on the basis of economic consequences, noneconomic consequences, and the goals established for the transportation system. Thus, the proposed link-analysis technique serves only as a guide by statistically identifying road links that accommodate similar travel functions. It is intended that the link-analysis model be sufficiently flexible to permit designation of road systems for any time period, as long as the corresponding input data (statewide traffic model) are available. From these conditions it should be possible to adopt the link-analysis technique for guiding decisions concerning the nature and extent of the intercity road building program after 1973.

## OBJECTIVES

The purpose of this paper is to formulate and test link-analysis techniques that may be useful as guides for the functional classification of major rural highways into the following systems: (a) freeways (no intersections at grade, full control of access); (b) expressways (generally a divided highway, some intersections at grade, partial control of access); and (c) other major rural roads (generally undivided, most intersections at grade, partial or no control of access).

## LINK-ANALYSIS TECHNIQUE

### The Systems Approach

The link-analysis technique which is developed in this paper is based on a systems approach to transportation planning. This systems approach involves identifying goals and objectives of functional highway systems and establishing criteria for use in maximizing attainment of the objectives. The flow of the link-analysis technique is shown in Figure 1.

### Goals and Objectives

As a first step, goals of the highway system should be formulated, as should also be the specific objectives of each of the three functional highway systems. Knowing goals and objectives, it is then possible to identify criteria that will be useful in selecting highway links to be included in each of the three systems.

Areas that have been explored in the past to formulate explicit goals for a statewide highway system include: promote safety, minimize disruption to economy and dislocation of people, promote faster travel, lower vehicle operating costs, maximize return on investment in transportation facilities, promote development, improve the environment and conserve resources, provide transportation service to all persons and areas, promote national defense.

Objectives of providing different major functional rural highway systems might include the following:

1. Provide a relatively uniform level of service for each system, with due regard to the functions of each system: (a) higher levels of service for systems planned predominantly for high-volume travel and/or for long trips (interstate and between major metropolitan areas); and (b) lower levels of service for systems expected to provide both land-access functions and to accommodate intermediate-length trips.
2. Provide for spacings and network configurations for each functional major rural highway system so as to maximize volumes of long-distance travel (and commercial travel) on those systems with the highest levels of service, with due consideration to improving the environment and minimizing disruption to the environment and minimizing disruption to the economy and dislocation of people.

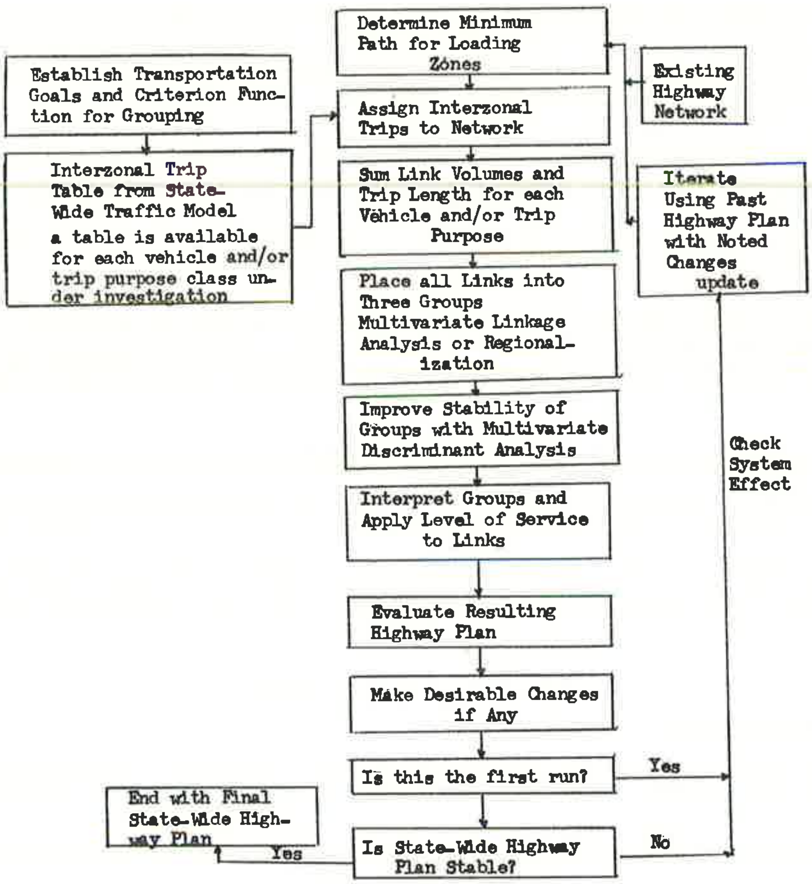



Figure 1. Link-analysis technique for identifying major rural highway systems.

3. Allocate resources to each functional system so as to maximize the return on the investment in transportation facilities.

4. Plan systems for both present and future demand, considering present development and travel desires, potential future growth and development, needs for recreation, and needs for preserving capacity and safety through access control.

Criteria for Evaluation

These objectives must be translated into quantitative measures or criteria for use in evaluating alternative functional highway classification plans, as shown in Figure 1. The criteria provide standards that can be used to measure the functional importance of various segments of the different alternative road systems.

For example, it may be assumed that certain road links displaying a high total volume in which long-haul, interstate trips, commercial vehicles, and work trips predominate will justify the highest level of service. Other measures might include the amount of recreational travel, governmental and business travel, social travel, industrial travel by segments such as mining, lumbering, or agriculture, and military travel on each road link (3, 4). The selection among the many measures available depends primarily on the capabilities of the corresponding statewide traffic model and the public policy decisions underlying the proposed function of a highway network. (The Automotive Safety Foundation identifies roads of statewide interest by their location in connecting the seats of state and county governments and other principal communities of the state and adjacent states, by the greater volume of traffic served, by their superior service to national defense, and by their usage by long-distance commercial and passenger motor vehicle travel. The past criterion formulated by the Automotive Safety Foundation serves only as a guide. The link-analysis approach assumes that with the availability of statewide transportation goals the highway planner can specify an appropriate criterion function.)

Identifying expressway and freeway systems is then the process of grouping road links into one of three classes (freeway, expressway, state arterial) depending on the measures listed above. Any measures felt to be important could easily be included in the analysis if the pertinent data for each road link are available. The primary emphasis is placed on measures that reflect a road link's function in terms of traffic movement and not of land access.

## Major Steps in Applying the Link-Analysis Technique

Methods of Grouping—Given the measures of trip length, volume, percentage of commercial vehicles, percentage of work trips, and percentage of interstate trips, the road links then must be placed into three distinct groups. It is necessary that each road link be placed into only one group and that all road links be placed into one of the groups.

Many methods of grouping items into classes exist such as elementary-linkage analysis (5), regionalization (6, 7), cluster analysis (8, 9) and integer programming (10). Regionalization and elementary-linkage analysis are selected as being most applicable to highway classification because they are both able to handle multivariate data and can be easily manipulated with the electronic digital computer. Only a modified form of regionalization will be discussed in this paper.

Modified Regionalization—Regionalization as applied to functional classification will differ slightly from its traditional approach as first published by Berry (6, 7, 12). In the ensuing analysis, the measures identified for grouping have been selected a priori. Each measure is given equal weight and if two measures are redundant or measure the same fundamental concept then that concept has been assigned a weighting double that of the other variables. Also, it is assumed that the number of groups to be identified is known prior to the grouping analysis. (This study is concerned with the development of a methodology or an approach toward grouping road links. Thus, the objective of the study is not to define how road links can be differentiated, but rather to present an approach toward placing road links into preselected groups. Through factor analysis, redundancy in measures selected in the criterion function can be identified, but without a dependent variable or correct answer. Selecting appropriate measures and weighting these measures are basically policy decisions. However, in further research it would be interesting to determine how a resulting classification plan is affected by the selection and weighting of the policy variables.)

Regionalization is a procedure used to combine n items, each associated with m measures or variables, into a smaller number of items by a stepwise-grouping process. The stepwise grouping combines n items into n-1 items, n-2 items, . . . 1 item by combining those two items that demonstrate the least multivariate distance. Two grouped items then have their multivariate distances replaced by the distance to their centroid. Distance is measured as

$$D^2_{jk} = \sum_{r=1}^{m} (X_{rj} - X_{rk})^2$$

where

$X_{rj}$ = variable r for link j, r = 1. . . m,

$X_{rk}$ = variable r for link k, r = 1. . . m, and

$D^2_{jk}$ = squared distance between link j and k.

Distance measures the amount of similarity between links based upon m observations. Each link can be located as a point in m dimensional space. A pair of links with identical

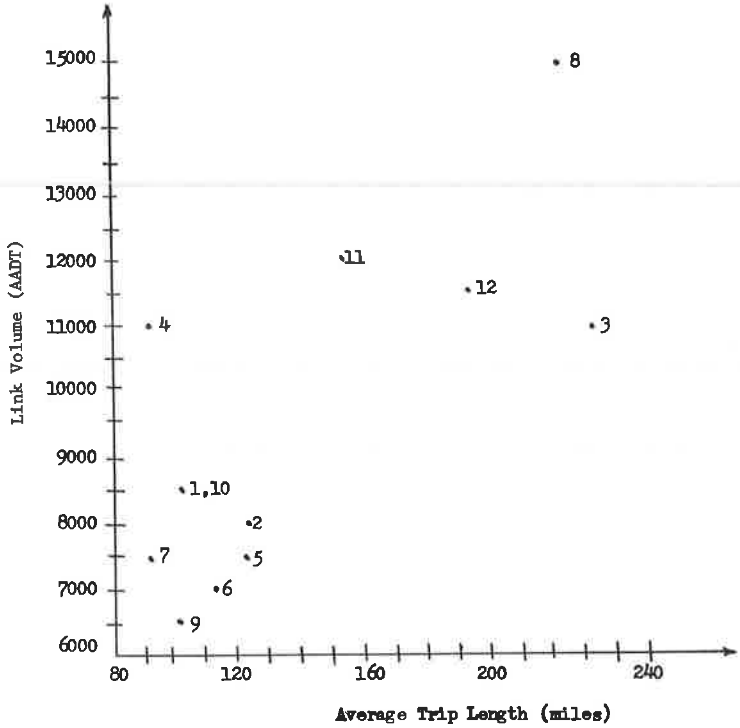



Figure 2. Road links in two-dimensional space for example road network.

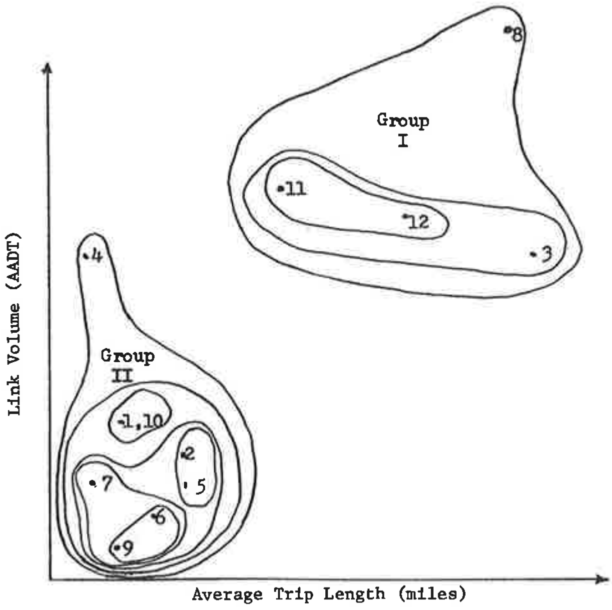
scores on all dimensions has zero distance. Points close in m dimensional space have small distances, and those links can be considered similar. Dissimilar points have high distances in m dimensional space. Grouping is conducted to minimize within group distance. Ahmad (11) presented a detailed outline of the procedure involved in a stepwise regionalization grouping process.

A simplified example will be worked for a twelve-link road network with associated measures of average trip lengths and volumes. Although grouping on the basis of two measurements provides little challenge, the approach can easily be extended up to m measures or dimensions, which is beyond visual capabilities. In two-dimensional space, these road links will be plotted as shown in Figure 2.

These observations are standardized to zero mean and unit variance in order to remove the effect of measurement units upon the grouping. Standardization permits volume measured in vehicles per day to be associated with trip length measured in miles. Otherwise, volume with a higher absolute unit would dominate the grouping procedure, and road links would become grouped solely on the basis of this one measurement.

Output from a stepwise-grouping routine yields the hierarchical ordering of these group combinations shown in Figure 3. Two distinct clusters are identified, and members in Group I (3, 8, 11, and 12) would be assigned a higher level of service by virtue of a higher loading on the link volume and average trip length variables. Road link 4 is a choice candidate for entrance into Group I and should receive further attention through a discriminant analysis.

Discriminant Analysis—Discriminant analysis can be used to differentiate among groups of road links based upon a series of associated measures available for each link. A linear combination of measures is selected that best discriminates among the groups. Discriminant analysis requires that the group membership, group means, and variance must all be specified. Thus, discriminant analysis can be used to check and improve existing groups, but not to create groups. Also, discriminant analysis can be




| Step | Groups Combined | Step | Groups Combined |
|---|---|---|---|
| 1 | 1 & 10 | 7 | 11 & 12 |
| 2 | 2 & 5 | 8 | 3 & 11 |
| 3 | 6 & 9 | 9 | 1 & 4 |
| 4 | 6 & 7 | 10 | 3 & 8 |
| 5 | 2 & 6 | 11 | 1 & 3 |
| 6 | 1 & 2 | | |

Note: Groups are redefined by the lowest numbered member in the group.

Figure 3. Hierarchical grouping of road links to two distinct groups.

used to place an unclassified item from the same population into one of the existing groups. One interesting aspect of discriminant analysis is the determination of each road link's probability of group membership in each of the designated groups. Thus, road links are regrouped on the principal that each item be assigned to that group for which it demonstrates the highest probability (13). Like grouping, discriminant analysis is adaptable to computer processing. The theory and application of discriminant analysis is covered by Hoel (14) and Cooley and Lohnes (13).

Previously, either of two grouping techniques, regionalization or elementary-linkage analysis, was utilized to form initial groups. Discriminant analysis can then be used to determine if road links 1, 2, 4, 5, 6, 7, 9, and 10 are best assigned to Group I and if road links 3, 8, 11, and 12 are best assigned to Group II. As previously mentioned, the location of link 4 is of particular interest in the testing procedure. Also, an unclassified road link can be classified through discriminant functions.

An initial discriminant function was developed as both volume and average trip length were found to be significant variables for discriminating between the two groups. As shown in Figure 4, the two dimensions are collapsed into a one-dimensional index through the discriminant function. Visually, it appears that road link 4 is best associated with Group II and not Group I. The discriminant-analysis output indicated that the probabilities of group membership are high for all road links in their initial group except for road link 4. The assignment of link 4 to Group II is then recommended, and the discriminant analysis is recomputed to check on the desirability of such a change in

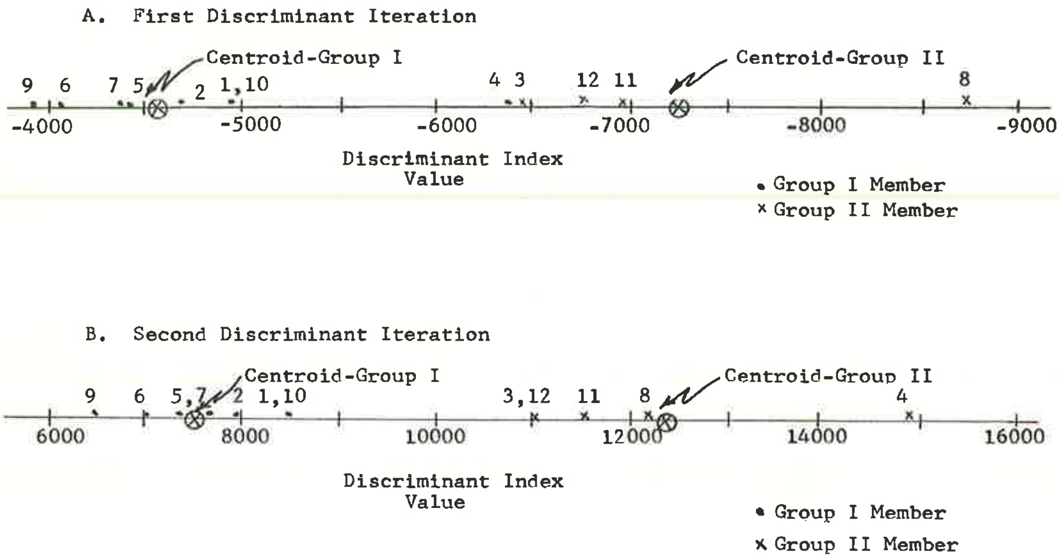



Figure 4. Collapsed two-dimensional discriminant.

group membership. Group membership probabilities indicate that link 4 is indeed best assigned to Group II.

A highway plan has then been developed based upon the assumption that Group I links form the freeway trunkline highway network and Group II links form the arterial system.

In addition to its use in improving existing groupings, discriminant analysis can be used to assign an unclassified road link to a system. For example, suppose some road link, 13, in the study area has a given volume and trip length. Should it be provided with an arterial or freeway level of service? The second discriminant function could be used to indicate to which group the new road link belongs.

Evaluation of Grouping Analysis—Problems arise with this grouping technique and, in fact, with all presently developed grouping methods because a truly optimal stage cannot be determined nor defined. Individual groupings cannot be evaluated against some desired optimum goal, and the most effective improvements cannot be identified until acceptable replication of this goal is achieved. It is never known if some better answer, measured in terms of minimizing within group variance, might exist nor how this better answer can be achieved within a feasible number of examinations. However, it is felt that for the purposes of designating road links to systems, an optimal grouping is not generally required. Any conscientious pursuit of the optimum represents an appreciable improvement over just relying upon subjective decisions in formulating a major statewide highway plan. The analytical-grouping technique can be used to create uniformity of plan formulation over both time and space. It represents an improvement over what the human mind could conceivably comprehend on a statewide basis. As improved algorithms are developed for grouping, they can easily be substituted into the model.

Extension of the link-grouping method would also allow minor reappraisals of the highway network. Based upon unexpected changes in traffic flow patterns, a link could be checked against existing group membership requirements in order to review if it might warrant reclassification. Naturally, these changes must be kept minor; otherwise flow on the entire network will change as will the definition of group membership. In the short run, such an approach might be employed as a hedge against pressure from special interest groups and as a periodic reevaluation of the plan.

Interpretation of Groups—The two distinct groups of road links formulated above must still be assigned a level of service in accordance with the principle that those links demonstrating the greatest functional use, measured in terms of the defined criterion function, be assigned the highest level of service. To these two groups, a level of service package defined either as a freeway, expressway, or trunkline arterial road will be applied. Identification of a desirable state highway system depends on a limited sensitivity analysis. Answers should be gained to the question: What is the consequence of either limiting or eliminating investment in a state freeway or expressway system?

Evaluation—Finally, the link-analysis plan must be evaluated and tested for system effects and network consistency. The system effects are used in this discussion to represent the diversion of traffic among road facilities as the level of service provided on these roads is changed. Thus, upgrading a two-lane rural highway to freeway standards will divert some traffic off parallel routes. Sudden changes in level of service must be avoided so that a motorist is not subjected to situations of high accident potential. Approximate rules of network consistency can be developed from existing design practices, accident records, and logic. A sensitivity analysis can be applied to test the impact of these consistency rules on the final classification plan.

Output from the link-analysis plan results in a proposed plan that must still be evaluated against stated objectives, local needs, and aspirations as expressed through the transportation goals.

Before a plan can be accepted, it must be determined whether upgrading the level of service on one road link will not significantly disrupt travel patterns. Thus, these travel patterns must be recomputed utilizing the proposed new freeway and expressway plan and thus new travel times in place of the original highway network. If no substantial changes in link assignment are encountered, then the plan can be assumed acceptable. Otherwise, the link-analysis and evaluation process must once more be repeated until two successive iterations yield the same result. Alternatively, the iterative technique can be used to test the implication of various alternative road plans that have been proposed. In this sense the model becomes a testing device, but originally it served as a guide to the development of these alternative plans. In addition, transportation planners are given the opportunity to work with future travel projections and not just present travel trends.

Ultimately, it is not only desirable to reassign travel to the proposed highway network but also to identify how highway investment has altered trip generation rates and trip distributions. In rural areas, highway accessibility can have a definite influence on both the desirability of making a trip and the desirability of making a trip to a given trip end. Designation of freeway and expressway corridors will alter link travel times and will alter the time separation of urban places; this can influence the relative attractiveness of making a trip and selecting a trip end. Tracing these system effects back to the trip generation and trip distribution models is feasible within the context of the proposed methodology.

## APPLICATION OF LINK-ANALYSIS TECHNIQUE TO STUDY AREA

### Study Area

In order to illustrate the application of the link-analysis technique to designating major rural highway systems, the model was applied to a study area located in southern Illinois. All major rural roads in the study area, as defined by the intercity travel network (2), were identified either as a freeway, expressway, or arterial trunkline highway based upon 1985 projected travel demands (15). Because travel data were assigned only for a portion of routes in Illinois, and routes in adjacent states were not considered, differences from the Smith plan (2) can be expected.

The study area's road network includes all the Interstate and U.S. numbered highways, and most of the Illinois numbered routes. In all, 131 road links are included in the link analysis. These major highway links were identified from the traffic model as the road section between intersections with other roads. For purposes of traffic assignment, 511 road links and 347 nodes were considered.

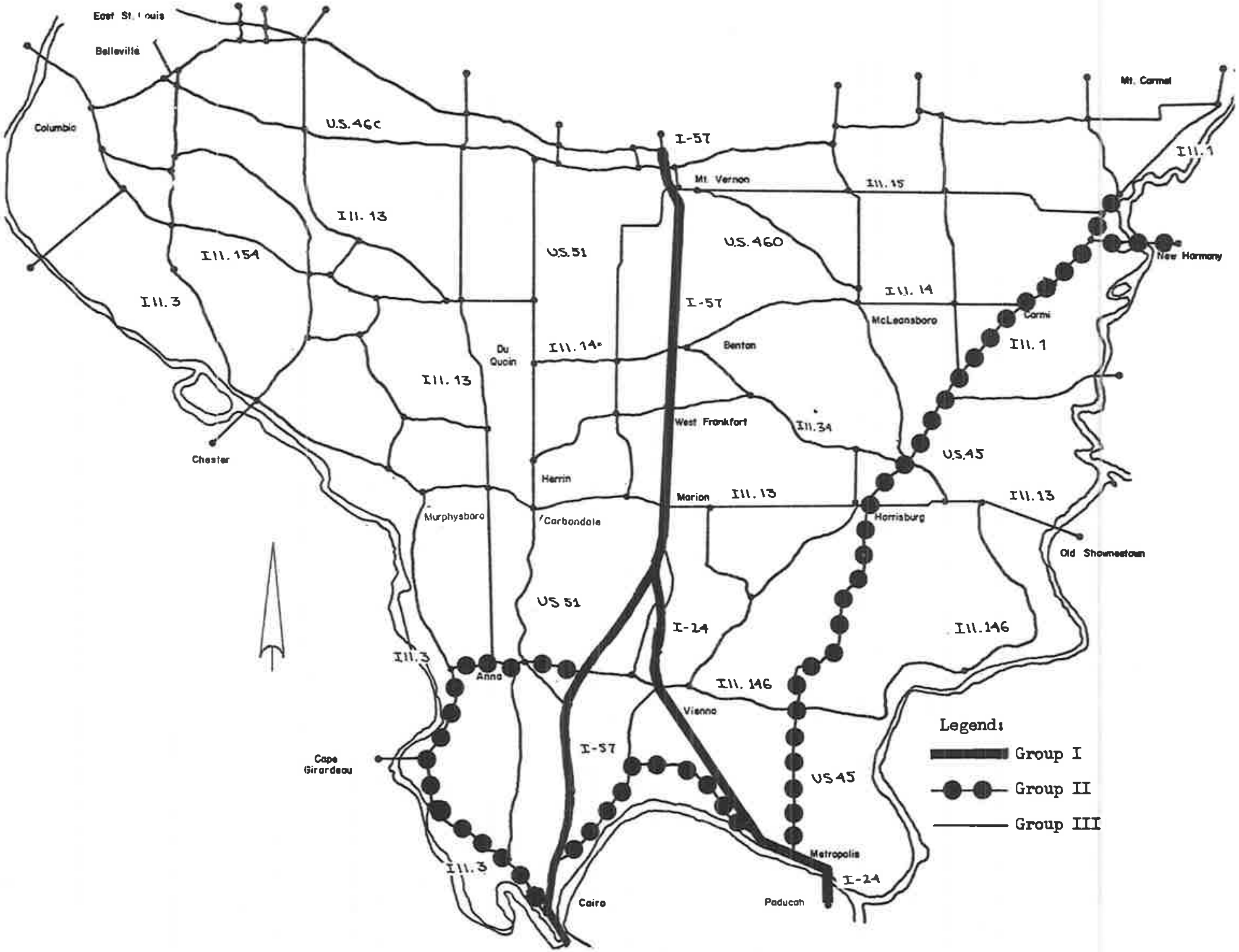



Figure 5. Final grouping on all road links from link-analysis model.

TABLE 1

MEAN VALUE OF EACH MEASURE FOR THREE GROUPS OBTAINED FROM LINKAGE ANALYSIS

| Group Number | Number of Members | Volume (AADT) | Percent Trucks | Average Trip Length (miles) | Percent External Trips |
|---|---|---|---|---|---|
| I | 18 | 25,190 | 29.9 | 238.4 | 85.3 |
| II | 19 | 4,863 | 29.7 | 90.3 | 75.6 |
| III | 94 | 2,300 | 29.2 | 57.9 | 18.8 |
| Average for all groups | | 5,817 | 29.4 | 87.5 | 36.2 |

## Method of Applying Link Analysis to Study Area

The 131 network links were grouped on the basis of four measures available as output from the traffic model. These four measures included:

1. Link Volume—The average annual daily traffic (AADT) on the maximum utilized portion of each road link was presented as a measure. The AADT was obtained directly from the traffic model.
2. Trip Length—The average trip length of vehicles using the maximum utilized portion of each road link was presented as a measure. However, the traffic model provides estimates of trip lengths for travel only within the State of Illinois. Thus, trip length cannot reflect the true importance attributed to roads serving interstate travel.
3. Percent Trips External to Illinois—The percentage of trips having at least one trip end external to Illinois was presented as a measure.
4. Vehicle Type—The percentage of trucks on the maximum utilized portion of each road link was presented as a measure.

Due to the availability of only four measures, the modified-regionalization concept of grouping was adopted for the link-analysis mode. Discriminant-analyses iterations were then conducted to determine the final grouping of road links. The availability of a traffic model having a greater number of vehicle class and trip purpose stratifications would have made it possible to substantially increase the number of measures included in the analysis.

## Application of Link-Analysis Approach to Study Area

A major rural highway plan was developed for the study area based on the characteristics of each road link's volume, percentage of trips external to Illinois, and average trip length. The plan was prepared from a grouping analysis conducted using the criteria that those road links are combined that are most similar (measured as minimum distance between group centroids). Six distinct road systems were noted as output from the first grouping analysis. Included in the six groups were three groups that consist of two members or less. These minor groups were placed within the three primary groups, and their group placement was tested through six discriminant iterations that were conducted until each road link was placed in the group for which it has the highest probability of group membership. Figure 5 represents the output from the sixth discriminant analysis; three road systems are identified. Interpretation of the average volume, trip length, and other characteristics for the links in each group defines the functional hierarchy of the groups. For example, Table 1 identifies the characteristics associated with each of the three road systems identified from the link-analysis plan shown in Figure 5.

Group I displays the greatest functional importance based on the variables selected in the criterion function, and should then be provided with the highest level of service. Thus, Group I links would represent the freeway system, Group II the expressway system, and Group III the arterial system. Knowing the availability of funds and transportation goals, the engineer could also evaluate the following alternative systems through a sensitivity analysis:

Alternative 1 (3 systems)
Group I, freeway; II, expressway; III, arterial.
Alternative 2 (2 systems, funds readily available)
Group I, freeway; II, freeway; III, arterial.
Alternative 3 (2 systems, funds restricted)
Group I, freeway; II, arterial; III, arterial.
Alternative 4 (2 systems, funds severely restricted)
Group I, expressway; II, arterial; III, arterial.

Through further analysis it was possible to subdivide the previously defined second group into two separate groups. Thus if priorities had to be assigned to a supplemental expressway or freeway system, the link analysis approach could be extended to provide stratification within a composite group. The final step would include an evaluation of the resulting plan vs other alternative plans and the transportation goals.

### Limitations

Limitations of the link-analysis approach include the following points:

1. The link analysis technique was never tested for system effects and considered only four travel parameters, some of which are highly dependent.
2. The link analysis model was applied totally on a regional basis without consideration for the rest of the state. Thus, it is quite possible a Group II road link in southern Illinois would be identified as a Group III road on a statewide basis.
3. The link analysis technique was concerned exclusively with the travel function of highway classification. No attempt was made to reflect the land-access nature of these road sections or to identify how these roads might contribute to the region's economy.
4. No attention was given to urban situations except to provide connecting links to the rural system.
5. No attempt was made to reflect how trip generation and distribution might be affected by the relative highway network selected for testing. It is anticipated these two considerations would tend in the direction of justifying the network being tested.

## CONCLUSIONS

The following conclusions may be drawn:

1. Intercity traffic models that estimate projected travel demands are applicable to identifying freeway and expressway systems as part of the rural highway system.
2. The link-analysis technique can be used to identify freeway and expressway systems consistently on a statewide or regional basis and can include future traffic projections as well as tests of system effects. The link-analysis model is sensitive to transportation goals and local needs and aspirations; it should prove useful in developing policy decisions regarding the status of the after-1973 highway program.

## ACKNOWLEDGMENTS

The opinions, findings, and conclusions expressed in this report are those of the authors and not necessarily those of Northwestern University, the Illinois Division of Highways, or the U. S. Bureau of Public Roads.

This report was prepared as part of the Illinois Cooperative Highway Research Project 55, Warrants for Highway-Highway Grade Separations, a cooperative investigation sponsored by the Illinois Division of Highways with the cooperation of the U.S. Bureau of Public Roads and conducted by the Department of Civil Engineering, Northwestern University. The help of G. T. Satterly, Jr., J. H. Shortreed, C. W. Ockert, and C. R. Walker is gratefully acknowledged. Also the assistance of R. Whiteside, Wilbur Smith and Associates, and R. D. Brown, Jr., Illinois Division of Highways, is acknowledged for providing the Illinois intercity travel data. Discussions with the following graduate students assisted in development of the statistical analysis: J. Drake, J. Dickey, J. Stowers, J. Schofer, and M. Wachs.

## REFERENCES


1. Brown, R. D., Jr. After 1972—What? Illinois Highway Engineer, Vol. 16, No. 4, Oct. 1964, pp. 5-8.
2. Illinois Highway Needs and Fiscal Study. Wilbur Smith and Associates, New Haven, Conn., Oct. 1967.
3. State Interest in Highways—A Report on Highway Classification. Washington State Council for Highway Research, 1952.
4. Ellis, F. F. A Proposed System of Highway Classification for the State of New Mexico. New Mexico State Highway Department, Santa Fe, 1965.
5. McQuitty, L. L. Elementary Linkage Analyses for Isolating Orthogonal and Oblique Types and Typal Relevancies. Psychological Measurement, Vol. 7, No. 2, Summer 1957, pp. 207-229.
6. Berry, B. J. L. The Mathematics of Regionalization. Paper prepared for the Brno Meeting of the Commission on Methods of Economic Regionalization, International Geographical Union, Sept. 1965.
7. Berry, B. J. L. A Method for Deriving Multi-Factor Uniform Regions. Przeglad Geograficzny, Vol. 33, No. 2, 1961, pp. 263-279.
8. Fruchter, B. Introduction to Factor Analysis. D. Van Nostrand Company, Inc., Princeton, N.J., 1954, pp. 12-17.
9. Rubin, J. Optimal Classification Into Groups: An Approach for Solving the Taxonomy Problem. International Business Machines Corporation, New York, March 1966.
10. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. American Statistical Association Journal, Vol. 60, March 1963, pp. 236-244.
11. Ahmad, Q. Indian Cities Characteristics and Correlates. University of Chicago Press, 1963, pp. 228-235.
12. Berry, B. J. L. Commercial Structure and Commercial Blight. University of Chicago Press, 1963, pp. 228-235.
13. Cooley, W. W., and Lohnes, P. R. Multivariate Procedures for the Behavioral Sciences. John Wiley and Sons, Inc., New York, 1962, pp. 116-150.
14. Hoel, P. G. Introduction to Mathematical Statistics. John Wiley and Sons, Inc., New York, 1962, pp. 179-184.
15. 1985 Assignment Network—State of Illinois. Wilbur Smith and Associates, New Haven, Conn., 1966.

# Analytical Estimation of Highway Impedances Within Urban Areas

PAUL R. RASSAM and RAYMOND H. ELLIS,
Peat, Marwick, Livingston and Company

This paper proposes an Urban Network Impedance Model (UNIM) to estimate highway impedances within urban areas without employing current network simulation procedures. Point-to-point travel time, distance, and cost are the principal outputs of UNIM. Inputs to the model are the geographic locations of the two points, the service characteristics of the highway network as specified by a function that characterizes the distribution of isotropic speed throughout the metropolitan area, and an assumption concerning an idealized geometric configuration of the highway network. A method for calibrating the speed is presented.

UNIM was tested, using Baltimore, Maryland, as a case study, by comparing the impedance estimates of the model to those provided by conventional network analysis procedures. Test results suggest that the model provides a reliable and efficient procedure to estimate impedances within urban areas. UNIM could be used in the initial calibration of a distribution model, to provide the accessibility inputs to a land use model, or for estimating site accessibility in location studies. Further refinements of the model are possible and would lead to more accurate results and to an even broader range of applications.

•A SIMULATION framework and an analytical formulation are two possible approaches to the estimation of highway impedances. The concept of impedance is used as a substitute for travel time, cost, or distance. The objective of this paper is to present the Urban Network Impedance Model, developed to evaluate analytically highway impedances within urban areas.

Traditional simulation methods for estimating point-to-point impedances involve the representation of the transportation system as a set of links and nodes and the association of some impedances with each element. Spider networks have also been constructed in which all of the transportation facilities between adjacent zones are represented by one set of impedances. Impedances between any two nodes may then be estimated by building minimum impedance trees through the network. System usage and service are derived in the simulation analysis by assigning interzonal traffic movements to the coded network. Thus, such a coded network is used to estimate interzonal impedances and to simulate the operation of the system.

Acquisition and processing of the large amount of data required to adequately represent a network are sources of many problems, e.g., exceeding the available resources or the capacity of the computer. Thus, an alternative technique for estimating usage of a facility has been proposed (1, 2). Furthermore, a simulation strategy is inherently cumbersome, and analytical strategies have been suggested for the analysis of flow and service in idealized situations (3).

The approach of this study follows the efforts of various researchers in this direction. Most of Smeed's models assume that drivers choose a path that minimizes the

Paper sponsored by the Committee on Future Concepts and presented at the 48th Annual Meeting.

distance between origin and destination (4, 5). Other authors have used a similar approach (6, 7, 8). But, such models ignore the relationship between the service provided by a highway and the volume of traffic using it. Lam has proposed a method that incorporates the volume-capacity relationship into an analytical assignment model (9, 10).

The problem addressed in this discussion is somewhat different from that structured above. Previous work has concentrated on the assignment of the total amount of traffic in an area to an idealized network and to the simultaneous estimation of link impedances. The model presented hereafter is concerned with point-to-point impedances. Such data are required for the accessibility inputs to a number of urban transportation models, for example, those used for both land use forecasting and trip distribution.

UNIM was initially developed to estimate the time, distance, and cost of traveling to a given terminal site from various points within an urban area in the Northeast Corridor. The location of the terminal is an exogenous input to the analysis, because it is unknown prior to the specification of a given design. It is conceivable that access link characteristics can be obtained by coding the internal transportation network of each analysis district and building trees. However, in view of the area covered by the Northeast Corridor, this method would involve coding an extremely large network and would be prohibitively expensive. This suggests that the problem of estimating access impedances should be treated analytically. Such an approach is particularly applicable in estimating access link characteristics, because the problem does not explicitly involve traffic assignment.

## PRELIMINARY OBSERVATIONS

In this research, the model to synthesize highway impedances requires three initial information inputs: (a) the geographic locations of the two points between which impedances should be estimated; (b) the service characteristics of the highway network, as specified by a function that characterizes the distribution of speed throughout the urban area; and (c) a definition of the paths between the two points.

Our initial effort to structure a highway impedance model used only the first two inputs, i.e., location and speed distribution. The travel path selected was the one corresponding to a minimum time, given a specified speed distribution over a continuous transport plane. This approach led to a calculus-of-variations problem that can be solved explicitly only by using rather limiting assumptions regarding the speed distributions.

In particular, the problem has been solved in a manner feasible for use in an operational impedance model for the following speed distributions:

$$v = ar$$

and

$$v = ar^2$$

where

r = the radial distance from a given point to the center point of the speed distribution (usually this center point would be the central business district of the metropolitan area);

a = a parameter to be calibrated for the metropolitan area; and

v = an estimate of the isotropic speed (i.e., the speed is independent of the direction of travel) at radial distance r from the center.

The geometry of the path connecting the two points was found to be a spiral for the linear speed distribution and a circle for the parabolic speed distribution. This result provides an interesting basis for evaluating Miller's discussion regarding spiral road networks (11). However, the assumptions regarding the speed distribution to be used in the calculus-of-variations model are unduly restrictive. In particular, the speed at the center of the distribution (i.e., the central business district) is set at zero. As a result, none of the paths may pass through the center.

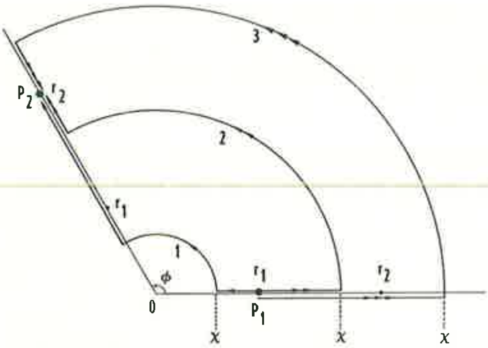



Figure 1. Alternative paths between two points.

In order to avoid the numerical solution of differential equations and to use a more realistic speed distribution, it is necessary to specify explicitly the paths between the two points. In UNIM, it is assumed that movement is restricted to either a radial or a circular path. Further, movement along these directions may occur at any point within the metropolitan area. This formulation allows realistic speed distributions as inputs; it was solved using a logistic curve to define speed.

## FORMULATION OF UNIM

Consider a city whose CBD center is designated by O and two points $P_1$ and $P_2$ in this city (Fig. 1). Let

$$r_1 = OP_1 \qquad (r_1 \geq 0)$$

$$r_2 = OP_2 \qquad (r_2 \geq 0)$$

$$\phi = (\vec{OP_1}, \vec{OP_2}) \qquad (0 \leq \phi \leq \pi)$$

At a given point, P, the speed, v(r), is assumed to depend only on the radial distance, r = OP. The function v(r) is supposed to be defined, continuous, and derivable twice over the range of r. It is hypothesized that the speed increased from a nonzero value, $v_0$, at the center to a constant value, $v_\ell$, at the edge of the city. In mathematical terms, this can be expressed as follows:

$$v(r) \geq v_0 > 0$$

$$\lim_{r\to\infty} v(r) = v_\ell$$

$$\frac{dv(r)}{dr} \geq 0$$

$$\lim_{r\to\infty} \frac{dv(r)}{dr} = 0$$

As mentioned previously, travel can only take place along a radial and/or a circumferential path. Because of the symmetry around O, it can be assumed that $r_2 \geq r_1$. Let x be the radius of the circumferential portion of the path ($x \geq 0$). Between any pair of points, $P_1$ and $P_2$, there are three distinct types of travel paths that may be used:

Case 1—(a) inward radial travel from $P_1$; (b) circumferential travel at a radius x, where $O \leq x \leq r_1 \leq r_2$; and (c) outward radial travel to $P_2$;

Case 2—(a) outward radial travel from $P_1$; (b) circumferential travel at a radius x, where $r_1 < x \leq r_2$; and (c) outward radial travel to $P_2$;

Case 3—(a) outward radial travel from $P_1$; (b) circumferential travel at a radius x, where $r_1 < r_2 < x$; and (c) inward radial travel to $P_2$.

Along a circumference of radius x, the speed remains constant and the corresponding travel time is $x\phi/v(r)$. Along a radial path, the speed varies and the travel time is given by the integral $\int dr/v(r)$. The travel times for the three basic cases are respectively

Case 1 $\tau_1(x) = \int_x^{r_2} \frac{dr}{v(r)} + \frac{x\phi}{v(x)} + \int_x^{r_1} \frac{dr}{v(r)}$

Case 2 $\tau_2(x) = \int_x^{r_2} \frac{dr}{v(r)} + \frac{x\phi}{v(x)} + \int_{r_1}^{x} \frac{dr}{v(r)}$

Case 3 $\tau_3(x) = \int_{r_2}^{x} \frac{dr}{v(r)} + \frac{x\phi}{v(x)} + \int_{r_1}^{x} \frac{dr}{v(r)}$

Three possible forms of the speed function have been considered:

Hyperbola $v(r) = \frac{r + a}{br + c}$

Exponential $v(r) = a(b + e^{-pr})$

Logistic curve $v(r) = a(b + e^{-pr})^{-1}$

where a, b, c, and p are parameters. The first two functions keep a constant concavity, whereas the last one offers the advantage of a varying concavity. Because of the analytical flexibility to fit empirical data, which is provided by the provision for an inflection point, the logistic curve was selected for the speed function.

The parameters of the logistic curve have the following significance:

$$v_0 = \frac{a}{b + 1}$$

$$v_\ell = \frac{a}{b}$$

i.e.,

$$b = \frac{v_0}{v_\ell - v_0}$$

$$a = \frac{v_0 v_\ell}{v_\ell - v_0}$$

UNIM is based on the assumption that the travel path selected is that which corresponds to the minimum travel time. Three values of x, which result in the minimum travel times via Path 1, Path 2, and Path 3 respectively, are calculated. These minimum travel times, $\tau_1(x)$, $\tau_2(x)$, and $\tau_3(x)$ are evaluated, and the minimum minimorum is selected as the travel time between the two points. This minimum minimorum travel time completely defines the geometry of the travel path, thus allowing the calculation of the travel distance and travel cost along this path between $P_1$ and $P_2$.

Substituting for v(r) in the expressions of $\tau_1$, $\tau_2$, and $\tau_3$ and integrating, we obtain

$$\tau_1(x) = \frac{1}{a}\left\{ b\,[(\phi - 2)x + r_2 + r_1] - \frac{1}{p}[e^{-pr_2} + e^{-pr_1} - (px\phi + 2)e^{-px}] \right\}$$

$$\tau_2(x) = \frac{1}{a}\left[ b\,(\phi x + r_2 - r_1) - \frac{1}{p}(e^{-pr_2} - e^{-pr_1} - px\phi e^{-px}) \right]$$

$$\tau_3(x) = \frac{1}{a}\left\{ b\left[(\phi + 2)x - r_2 - r_1\right] - \frac{1}{p}\left[e^{-pr_2} + e^{-pr_1} - (px\phi - 2)e^{-px}\right] \right\}$$

The values of x, which correspond to the minimums of $\tau_1(x)$, $\tau_2(x)$, and $\tau_3(x)$, are found by solving the equations

$$\frac{\partial \tau_i(x)}{\partial x} = 0 \qquad i = 1, 2, 3$$

subject to:

$$\frac{\partial^2 \tau(x)}{\partial^2 x^2} > 0 \qquad i = 1, 2, 3$$

A complete analytical solution of the minimum minimorum problem has been developed. It shows that the value of b is a critical element in the selection of the minimum time path [the ratio $v_0/v_\ell$ is equal to $b/(b + 1)$]. For certain values of b, the absolute minimum time can be found without comparing the minimum of $\tau_1$, $\tau_2$, $\tau_3$. Figure 2 illustrates the limiting values of b that define the possible paths among which the solution is found. When the pair ($v_0$, $v_\ell$) falls in Regions I or II, the minimum corresponds to that of $\tau_1$; when the pair ($v_0$, $v_\ell$) is in III, the minimum corresponds to the smallest of the minimums of $\tau_1$ and $\tau_2$; finally when ($v_0$, $v_\ell$) is in IV, $\tau_1$, $\tau_2$, and $\tau_3$ have to be considered.

The travel distance corresponding to the minimum minimorum travel time is:

$$\ell(x^*) = r_1 + r_2 + (\phi - 2)x^*$$

where $x^*$ = the radius of the circumferential element of the minimum minimorum travel time path.

The travel cost corresponding to the minimum minimorum travel time is established as follows. It is assumed that the unit cost of travel is a function of the instantaneous travel speed according to the following function:

$$c(v) = a_0 + a_1 v + a_2 v^2 + a_3 v^3$$

where

$c(v)$ = the unit costs of travel at speed v; and

$a_0$, $a_1$, $a_2$, $a_3$ = parameters to be calibrated.

Because speed, v, is a function of the distance from the center, r, the unit cost of travel can be expressed as a function of r. The travel cost between $P_1$ and $P_2$ is obtained by taking the line integral of the unit cost function over the minimum minimorum travel time path:

$$C = \left| \int_{x^*}^{r_1} c(r)dr \right| + \phi x^* c(x^*) + \left| \int_{x^*}^{r_2} c(r)dr \right|$$

where C = the travel cost between $P_1$ and $P_2$.

## TESTING OF UNIM

In order to test the methodology, it was decided to compare impedance values obtained from the model to impedances estimated from coded networks. Because of the availability of data, Baltimore was chosen as a test city. Before exercising UNIM, it is necessary to calibrate the speed function for the metropolitan area.

A ring-sector zonal structure was superimposed on a map of Baltimore. This investigation utilized the 1962 network coded during the course of the Baltimore Transportation Study. Off-peak speeds for the 1962 network were established through field

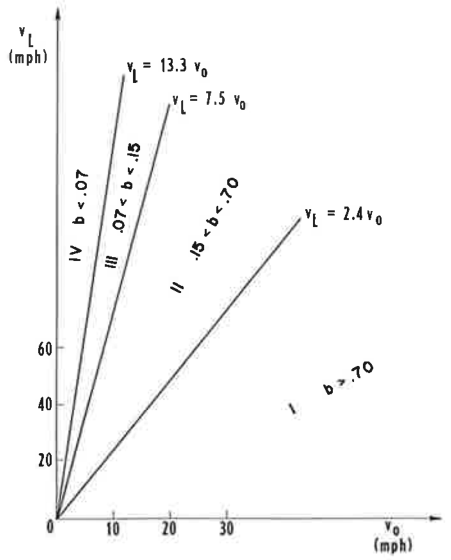



Figure 2. Limiting values of b.

observations made during the summer and fall of 1961. The average speed in each of the 15 rings was estimated as shown in Figure 3. It is evident that the speed in a given sector generally increases with the distance from the center. The average speed in the innermost rings is relatively high, e.g., about 20 mph, but this may have been caused by inaccuracies in the network coding.

The formula $v = a(b + e^{-pr})^{-1}$ can be linearized as follows:

$$\frac{1}{v} = \frac{b}{a} + \frac{1}{a} e^{-pr}$$

where $\frac{a}{b} = v_\ell$, the speed reached in the fringes. Hence:

$$\frac{1}{v} - \frac{1}{v_\ell} = \frac{1}{a} e^{-pr}$$

$$\ln\left(\frac{1}{v} - \frac{1}{v_\ell}\right) = -\ln a - pr$$

This relationship is linear with respect to $\ln\left(\frac{1}{v} - \frac{1}{v_\ell}\right)$ and r. In order to perform a regression analysis to determine a and p, $v_\ell$ has to be assigned. It should be noted that other techniques are available for calibration.

Three values were assigned to $v_\ell$: 40, 45, and 50 mph. The parameters of the speed function were estimated for each of these limiting values using the complete sample of 15 rings, using a sample of 11 for which the first four rings were left out because of the seemingly high values of v at the CBD center, and using a sample of 13

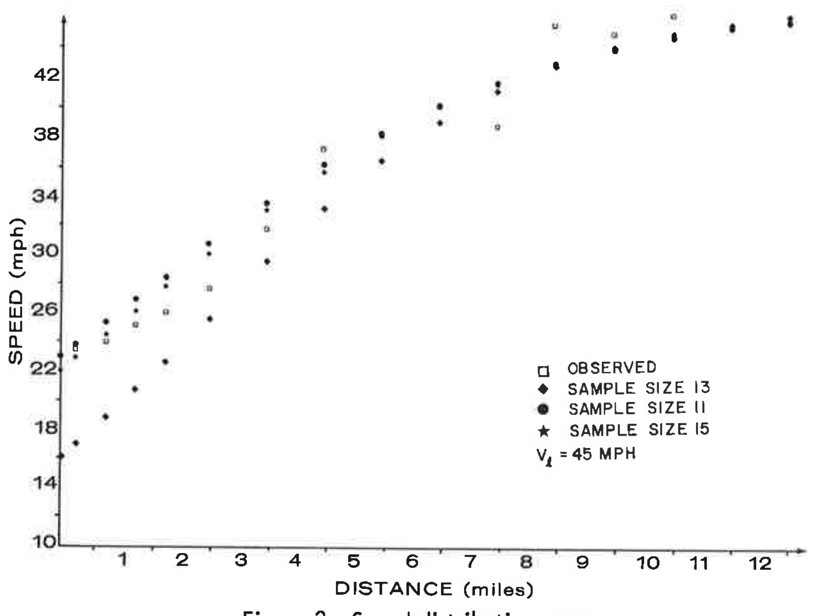



Figure 3. Speed distribution curve.

TABLE 1

RESULTS OF REGRESSIONS

| Speed Equation | Sample Size | $v_\ell$ (mph)[a] | R | $\ell n$ a | p | Equation | $v_0$ (mph)[b] |
|---|---|---|---|---|---|---|---|
| 1 | 11 | 40 | 0.94 | 3.00022 | 0.3289 | $v = 20.09\ (0.522 + e^{-0.3289r})^{-1}$ | 13.2 |
| 2 | 15 | 40 | 0.96 | 3.30896 | 0.2944 | $v = 27.36\ (0.684 + e^{-0.2944r})^{-1}$ | 16.3 |
| 3 | 11 | 45 | 0.86 | 3.67674 | 0.2799 | $v = 39.52\ (0.878 + e^{-0.2799r})^{-1}$ | 21.1 |
| 4 | 15 | 45 | 0.93 | 3.59518 | 0.2887 | $v = 36.42\ (0.809 + e^{-0.2887r})^{-1}$ | 20.1 |
| 5 | 11 | 50 | 0.49 | 4.71429 | 0.1246 | $v = 111.50\ (2.230 + e^{-0.1246r})^{-1}$ | 34.5 |
| 6 | 15 | 50 | 0.76 | 4.08172 | 0.1946 | $v = 59.26\ (1.185 + e^{-0.1946r})^{-1}$ | 27.1 |
| 7 | 13 | 40 | 0.96 | 2.73577 | 0.3582 | $v = 15.43\ (0.386 + e^{-0.3582r})^{-1}$ | 11.1 |
| 8 | 13 | 45 | 0.92 | 3.03539 | 0.3514 | $v = 20.80\ (0.462 + e^{-0.3514r})^{-1}$ | 14.2 |

[a]Assumed speed at the edge of the city. [b]Computed speed at the center of the city.

formulated by adding the following data points to the sample of 11: 10 mph for the first ring and 12.5 mph for the third ring.

The results of the regressions are presented in Table 1. The equations corresponding to a limiting speed of 50 mph were rejected for low correlation coefficients and high speeds at the center. It was also found that arbitrarily adding data points for the first and third rings did not significantly change the coefficients of the equation. The equations corresponding to $v_\ell$ = 45 mph are graphed in Figure 3.

The model was first applied to estimate impedances from a set of districts within the Baltimore metropolitan area to Friendship Airport. Each of the six equations obtained for the speed distribution was used in this test of UNIM. To this effect, three programs were written in FORTRAN IV for the IBM 360/40 computer. The first program transformed the grid coordinates of 586 Baltimore zone centroids into a system of polar coordinates, the second program computed the time and distance from each centroid to Friendship Airport, and the third program aggregated zonal-level times and distances to the corresponding values for the study districts defined in the Baltimore-Washington Airport Survey. UNIM's estimates of travel time and distance, using Eq. 1 (Table 1) are compared in Figures 4 and 5 to corresponding network values. The calculated travel times are generally within 10 percent of the corresponding network values, whereas travel distances are systematically underestimated.

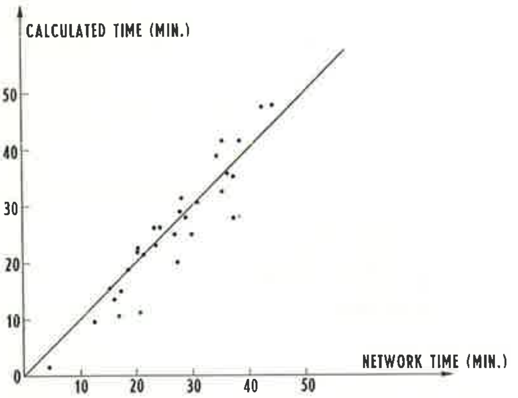



Figure 4. UNIM estimate of travel times vs network.

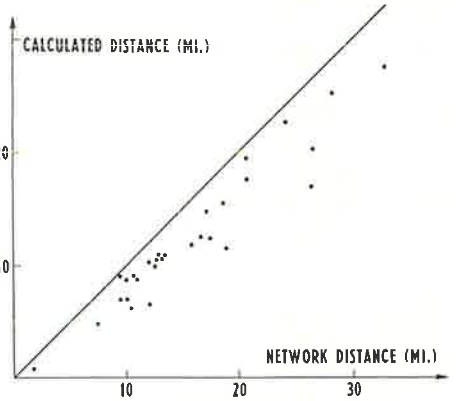



Figure 5. UNIM estimate of travel distances vs network.

TABLE 2

EVALUATION OF UNIM OVER 231 INTERCHANGES

| Speed Equation | Time | | | | | Distance | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean (min) | Standard Deviation (min) | Correction Factor | R | Standard Error (min) | Mean (mi) | Standard Deviation (mi) | Correction Factor | R | Standard Error (mi) |
| 1 | 25.1 | 11.1 | 0.98 | 0.99 | 4.5 | 11.8 | 6.5 | 1.33 | 0.96 | 4.0 |
| 2 | 23.2 | 10.5 | 1.06 | 0.99 | 4.4 | 11.7 | 6.3 | 1.35 | 0.96 | 3.9 |
| 3 | 19.8 | 9.1 | 1.24 | 0.98 | 4.0 | 11.5 | 6.0 | 1.38 | 0.96 | 3.8 |
| 4 | 20.0 | 9.2 | 1.23 | 0.98 | 3.9 | 11.6 | 6.1 | 1.36 | 0.96 | 3.9 |
| 7 | 25.6 | 10.9 | 0.97 | 0.99 | 4.5 | 12.1 | 6.9 | 1.28 | 0.96 | 4.0 |
| 8 | 21.8 | 9.5 | 1.12 | 0.99 | 3.9 | 11.9 | 6.7 | 1.30 | 0.96 | 3.9 |
| 231 network observations | 25.9 | 11.8 | | | | 14.7 | 8.7 | | | |

In view of the promising results of this initial test, it was decided to carry out a more comprehensive evaluation. Twenty-two points uniformly distributed over the Baltimore urban area were selected. UNIM was used to obtain the time and distance corresponding to the 231 different possible interchanges. These results were compared to the impedances obtained from the coded 1962 network.

Results of the evaluation are presented in Table 2. The correction factor is obtained by means of regressing the observed network values against the values estimated by UNIM. Because the regression line was forced through the origin, only one correction factor was obtained. Finally, the mean and standard error of estimate of the set of estimated impedances for each speed model were obtained.

Several inferences may be drawn from these results. It would appear that Eq. 1, (Table 1), which uses a limiting speed of 40 mph, could be usefully applied to reproduce off-peak highway impedances in the Baltimore area. The proportion of variance explained and the standard error of estimate are not appreciatively sensitive to variations in the assumed limiting speed, although the same is not true of the correction factor.

The correlation coefficients are all over 0.95, the standard errors of estimate are relatively low, and deviations that occur between an estimated and a network value can be satisfactorily corrected by means of a linear factor. This suggests that UNIM does explain a substantial portion of the variance in interzonal impedances within the test area. As noted in the first test, the model's estimate of travel times is better than its estimate of travel distances. In particular, in the case of Eqs. 1, 2, and 7 (Table 2), the travel time correction factors are very close to one. This is not true, on the other hand, for the travel distance correction factors, which are about 1.3.

## CONCLUSIONS

On the basis of these tests, it would appear that UNIM provides an acceptable method for calculating point-to-point highway impedance in urban areas without requiring the preparation and processing of coded networks. Further, there seems to be a close identity between the impedance values estimated by UNIM and network values. The model predicts travel time relatively accurately, although travel distance is systematically underestimated by about 30 percent.

It does not appear overly difficult to acquire the input data necessary to calibrate the speed function used in UNIM. An obvious data source is the network link data (i.e., time, distance, and speed) that are generally acquired during the inventory phase of an urban transportation study. These data could be supplemented by the results of various travel time and delay studies, and/or spot speed studies that may have been conducted within the metropolitan area. Once experience is gained with developing speed functions for a number of urban areas, it will probably be possible to define a generalized speed curve that would vary as a function of the characteristics of the metropolitan area.

UNIM would appear to have application to a number of problems encountered in transportation system analysis. For example, it could be used to provide the accessibility inputs required for land use planning models, to supply the initial impedance estimates to calibrate a trip distribution model, or to estimate the accessibility of various sites in location studies. A computer software package has been developed so that UNIM may be conveniently applied to estimate point-to-point impedances or weighted site accessibilities at a cost level that is substantially lower than current network analysis procedures. The weights used in calculating accessibilities could be any activity distribution for the metropolitan area and may be defined on either a continuous or a discrete basis.

Numerous research projects, both empirical and theoretical, could be defined to further exploit the potential implicit in UNIM. Initially, it would be useful to apply and evaluate the performance of the model in a large number of metropolitan areas and to refine current procedures for acquiring the data needed to calibrate the speed function. It is conceptually easy to extend the model by utilizing three speed functions: one for circular travel, one for inward radial travel, and one for outward radial travel. Finally, it is possible to formulate UNIM on the assumption that a city has a grid rather than a radial-circumferential highway network.

## ACKNOWLEDGMENTS

This research was performed as part of Peat, Marwick, Livingston and Co.'s work under a contract with the Transport Systems Planning Division of the Office of High Speed Ground Transportation, U. S. Department of Transportation. The authors gratefully acknowledge the assistance of Paul W. Shuldiner, Transport Systems Planning Division, and Philip Barbato, Technical Monitor for this contract. The authors also express their appreciation to Littleton C. MacDorman and Morris J. Rothenberg of Peat, Marwick, Livingston and Co. for their helpful contributions and to Susan Antoniewicz for her expert typing and her patience.

## REFERENCES

1. Schneider, Morton. Direct Estimation of Traffic Volume at a Point. Highway Research Record 165, 1967, pp. 108-116.
2. Caswell, Stearns. A Theoretical Model for the Determination of Expressway Usage in a Uniform Region. Highway Research Record 238, 1968, pp. 79-102.
3. Smeed, R. J. The Traffic Problem in Towns. Proc. Manchester Statistical Society, 1961.
4. Smeed, R. J. Road Development in Urban Areas. Jour. Institution of Highway Engineers, Vol. 10, 1963, pp. 5-30.
5. Smeed, R. J., and Holroyd, E. M. Some Factors Affecting Congestion in Towns. Engineering for Traffic Conference, London, 1963, pp. 83-91.
6. Haight, F. A. Some Probability Distributions Associated With Commuter Travel in a Homogeneous Circular City. Operations Research, Vol. 12, 1964, pp. 964-975.
7. Tan, T. Road Networks in an Expanding Circular City. Operations Research, Vol. 14, 1966, pp. 607-613.
8. Tan, T. A Mathematical Model for Commuter Traffic in Satellite Towns. Transportation Science, Vol. 1, 1967, pp. 6-23.
9. Lam, T. N. Flow Dependent Traffic Assignment on a Circular City. Inst. of Traffic and Transportation Engineering, Univ. of California, Berkeley, Graduate report, 1967.
10. Lam, T. N., and Newell, G. F. Flow Dependent Traffic Assignment on a Circular City. Transportation Science, Vol. 1, 1967, pp. 318-361.
11. Miller, Allan J. On Spiral Road Networks. Transportation Science, Vol. 1, 1967, pp. 109-125.

# The Access and Development Prototype Project

MORTON SCHNEIDER, Creighton, Hamburg Inc.

•THIS PAPER is divided into two main parts. The first is a summary sketch of a project (1) based on a paper (2) presented at the Conference on Urban Development Models held at Dartmouth College in June 1967. (The conference was sponsored by the U.S. Bureau of Public Roads, the U.S. Department of Housing and Urban Development, and the Automotive Safety Foundation and conducted by the Highway Research Board.) The second elucidates certain aspects of the original paper, adding new theoretical material. The two parts do not have a great deal to do with each other.

## THE PROTOTYPE PROJECT

The paper, Access and Land Development, attempts to show some general and computable relationships among travel, land development, and the time and cost characteristics of transportation. The purpose of this project was to develop a computer program capable of performing the calculations proposed there, partly to see if they could be performed, and partly to see if, having been performed, they appeared to mean anything. The computer program developed in the project is a prototype, which is to say that it does not try to accommodate the full range of possibilities that might be found in the real world. Its design—and prototype character—is due mostly to the decision to avoid, at this stage, the distracting, difficult, and expensive problems associated with realistically complicated transportation networks.

### Review of Theory

The general sense of the theory can be seen in what it does, in the information on which it operates, and the results it produces. For a region of interest, the basic items of given information are (a) the total amount of floor area to occur in the region; (b) the transportation network, that is, the travel times and travel costs of all means of transportation that connect places with each other; (c) a measure of the inherent attractiveness of land—essentially just proportional to area, but also distinguishing, for example, fetid swampland from firm, salubrious acreage; and (d) any constraints, planned or expected to operate on individual sites, limiting or compelling their development—park land reservations, perhaps, or the continued existence of structures already in place.

The principal quantities calculated from these are (a) the amount of floor area that can be expected naturally to arise at each site in response to the access there, consistent with development at all other sites and with any constraints; (b) the number of trips that can be expected to originate at each site; and (c) the number of trips that can be expected to use each element (of all modes) of the transportation system.

The size of the region of interest, the fineness of partitioning into "sites" or "zones," the inclusiveness and detailing of the transportation system are all simply matters of precision and the mechanics of implementation as far as the theory is concerned. There is no innate specialization either to urban areas or to great regions.

The reasoning behind the theory proceeds in three more or less distinct phases. The first works with quite general considerations to show that trip generation at any site is related to the environment in which the site is embedded. Here the concept of "attractive stuff" is introduced as a necessary attribute of sites; it is called R because it has to be called something, but the physical meaning is left undetermined.

Paper sponsored by Committee on Land Use Evaluation.

Using only the assumption that travel can be described by a symmetric (but otherwise unspecified) trip-distribution function, the result is derived that the number of trips originating at any site is proportional to the amount of R at the site and to a quantity that can very naturally be thought of as the accessibility of the site, or the access integral around it (2, Eq. 4, p. 65).

The second phase takes up the ancillary question of the exact form of the trip-distribution function and, especially, the problem of mode of travel. It proposes that all trips do not have the same sensitivity to travel time and cost and that, because of this, there can be more than one effective "minimum path" between two points. Compounded into a mathematical treatment, this gives a definite all-mode trip distribution function (2, Eq. 11, p. 70).

The third phase identifies attractiveness, R, as the sum of two components, one due to land area and the other to floor area, where floor area is used as a convenient surrogate for capital improvement of all kinds. It then argues that floor area appears at a site in order to accommodate activity there, but generates new activity by its appearance. However, if total activity is governed by the trip generation relationship already given, there will be some definite amount of floor area just right for the land area (attractiveness) and accessibility of the site (2, Eq. 20, p. 173).

## The Prototype Computer Program

The inputs and outputs of the prototype program are pretty much those listed above except that there is no link-by-link traffic volume output, but, instead, some trip-distribution tables. The transportation network is limited to a highly stylized three-moded system, designed around computational convenience. Each mode may be assigned any speed and any cost per mile, and any initial time and cost penalty. The modes compete but do not interconnect. They are assumed to operate along grid lines, but a grid cell need not be connected to all three modes.

In other respects, the prototype is a fairly usable model. It is organized on a regular grid system, in which a zone is simply a grid square. It will handle up to 2,500 such zones in any arrangement that fits within a 50 by 50 square. The scale of the grid system is arbitrary, so that a zone can be one mile on a side, or one-tenth of a mile, or ten miles—or any other size—with the dimensions of the whole region correspondingly expanded or contracted. Any zone can be constrained to have any given amount of floor area development, or less than, or more than any given amount.

The program is in FORTRAN IV. It has been operated successfully on both a Control Data 3800 computer and an IBM 360/50. The running time is roughly proportional to the square of the number of zones (on the CDC 3800, something like 40 seconds for a 230-zone case).

An output format worth mentioning is a supplemental maplike printout intended to give a quick perception of the calculated floor-area development pattern. It does.

## Plan of Tests

If the theory were perfectly credible (or the world perfectly credulous)—that is, if there were no question of its validity—there would still be several steps necessary to make a going thing of it besides writing a computer program. These steps establish a very rough framework for exercising the prototype program to achieve some feeling for the behavior and validity of the whole idea.

The theory leaves three basic quantities undetermined, subject to empirical measurement. Two of these are the constants "a" and "b" in the trip distribution function. These constants may be thought of as the average values of time and cost respectively in trip-making. As "a" grows smaller, trips become willing to take more time, and as "a" grows larger, less. In just the same way, trips tend to spend more money when "b" is small and less when "b" is large. The magnitudes of "a" and "b" together determine average trip length and the distribution of trip lengths in a given system. But the relationship of "a" and "b" to each other determines the distribution of trips among competing paths, or modes. So it should be possible to estimate "a" and "b" through these two properties of length and mode distribution by comparing calculations with actual trip lengths and modal splits.

The third unfixed quantity is the relative scale of attractiveness due to land area, $R_a$, and that due to floor area, $R_f$. The units of R can be arbitrarily chosen, just as a foot is an arbitrary unit of length, but how much R resides in a square foot of land area and how much in a square foot of floor area is up to empirical measurement. If, for convenience, one unit of R is defined to be the amount of R produced by one square mile of lând, then the empirical problem becomes that of finding the rate of R for floor area. This is a rather more subtle problem than that of estimating "a" and "b," but it can be solved by making use of the not too obvious property of the theory that the pattern of floor-area development changes with a change in total $R_F$. In general, concentration effects in a given system increase with total development (as indeed seems to be the case in the real world). The procedure for scaling $R_f$ would be to calculate the distribution of $R_f$'s for some real region in successive trials, using different values of total $R_F$, until the shape of the $R_f$ distribution agreed with the shape of the actual floor-area distribution—that is, until the ratios to each other of $R_f$'s around the region were the same as the corresponding floor-area ratios, or, the same thing, the ratio of $R_f$ to the floor area it was trying to represent was the same everywhere. The latter ratio, of course, is the sought-for conversion factor.

One further complication is that it can be difficult, though not strictly impossible, to calculate well trip distributions in order to estimate "a" and "b" without knowing how to measure R, and it is impossible to calculate floor-area distributions well, in order to measure R, without good values of "a" and "b." A certain amount of successive refinement is involved.

These are the things that would be done if things could be done that way. In practice the prototype tests, although they have this general scheme in mind, suffer from an exiguity of appropriate data and from their own ineptitude at representing complicated reality, and tend to limp and stumble from case to case. It is hard to know how to report the results of this kind of project in capsule form. The program works. Sometimes it is even fun. The mathematics make computational sense and look as though they might be reflecting the real world. Calculations exhibit agreement with data, but there are few points of contact. The approach to modal split seems to work well, but the trip-distribution function may be having some trouble in its ambitious effort to comprehend all trip behavior from the very short to the very long in one function. Interestingly enough, the scale of $R_f$ seems to turn out to be about the same as that of $R_a$; a square mile of floor area apparently contains about one unit of R, just as does a square mile of land area.

On the whole, the idea looks promising to the intuition, and not too much more than that can be expected at this point. Perhaps its most engaging feature continues to be theoretical richness, the fact that it makes predictions, as opposed to statistical recapitulations, and focuses the viewpoint (distorting it at the same time, possibly). Two very general predictions deserve pointing out: if the quality and extensiveness of transportation does not change much, centralization will occur as the world accrues more and more fixtures of development; if transportation becomes more extensive rapidly enough, decentralization will occur.

It might be mentioned in passing that the theory does not particularly require every square foot of floor area to have the same R value as every other square foot, or every square foot of land area to be the same as every other. But it would be far beyond the resolving power of this project to distinguish differences of that order.

## ELUCIDATIONS

The expression for floor area at a site is

$$R_f = \frac{R_a I R_F}{J - I R_F} \tag{20}$$

where I is the accessibility integral around the site and J is the integral of these accessibilities throughout the region.

Regional "Temperature"

An examination of Eq. 20 reveals that the ratio $J/R_F$ is a special quantity. Development distributions calculated separately for different regions (excluding constrained areas) have to give the same $J/R_F$ in order to be consistent with the distribution obtained by lumping the regions together in a single superregion. The ratio can be formed for regions of any size at all, large or small, and will be the same for every place entering into a distribution. Presumably regions could have different $J/R_F$ ratios if the interaction between them were weak, or if there were barriers or hidden constraints due to policies on travel, commerce, or immigration, but presumably also these differences would tend to diminish with time.

Indeed, this whole idea of land development can be restated in the form of a principle: the world acts to keep $J/R_F$ the same everywhere within well-communicating regions. This has something of the ring of an entropy law to it, with $J/R_F$ vaguely analogous to temperature. It says nothing about the dynamics of the process, only its goal. Carrying out the analogy, borders, oceans, and geographic and political barriers can be thought of as insulating walls slowing down or impeding the entropic process. The analogy is an attractive one, and may even be valid as far as it goes. But other important factors—things such as access to goods and resources, productivity, various kinds of costs—are not explicitly taken into account (although there seems to be theoretical elbow room for them to work their way in).

Continuing with the idea of regional "temperature," it is easy to let T stand for $J/R_F$ and to rewrite Eq. 20 as

$$R_f = \frac{R_a I}{T - I} \qquad (20a)$$

This form clearly suggests that if T is known, the distribution of development can be calculated without knowing the total development in advance: development becomes a matter of interaction between access at each point and the state parameter, T, instead of the distribution in detail of an exogenous total. Here T begins to function in a rather lofty role, trying to dictate the shape of the world. Presumably T changes with history, rising as the extensiveness and quality of transportation increases, though capable also of falling in a period of transportation stagnation combined with growth in total development. Almost certainly, T has been rising in the modern era (perhaps it has always been rising) and, from Eq. 20a, a site needs more access to get developed, or to stay developed, than it used to. Moreover, T increases especially fast when places of more than average desirability become accessible, another long-term tendency that works some hardship on fixed-access sites and on those that are less congenial.

The notion that T is some kind of fundamental system measure has considerable appeal to vagrant speculation. Perhaps deeper social, economic, and historical meanings can be found for it. Perhaps the dynamics of the development process can be understood in terms of flows of development governed by T differences from place to place. Perhaps T represents something to which a value can be attached in human events, so that a place of higher T may be said to be better off than one of lower T. But in a much more practical vein, the concept of T serves to make a connection among regions that could be treated together only in a system too big to handle comfortably: calculations done for areas all over the country can be made consistent with each other without having to stuff the whole continent into a computer at once.

The Dense Branch: Fiat Development

Given a region containing a transportation system but no development, there will be some value of J that may be called J*. If development is added uniformly everywhere, with the transportation system remaining fixed, it can be seen from definitions that J becomes, for any total development, $R_F$, and with total land, $R_A$:

$$J = (J^* - Z)(1 + R_F/R_A)^2 + Z(1 + R_F/R_A)$$

Here Z is a contribution to J* due to access from within the region to the world outside. (In a real, nonuniform case, J would increase more rapidly with $R_F$ than this because of centralization effects.)

And T becomes:

$$T = \frac{1}{R_F}(J^* - Z)(1 + R_F/R_A)^2 + \frac{Z}{R_F}(1 + R_F/R_A)$$

This is a two-valued function, approaching infinity when $R_F$ is either very small or very large, with a minimum at

$$R_F = R_A \sqrt{[J^*/(J^* - Z)]}$$

So starting with an empty system and knowledge of the world value of T, it appears that two quite different development patterns can be calculated, one at a lower total development and one at a higher. The lower density solution apparently is the proper one, representing normal incremental growth, and the higher solution can be interpreted as a kind of critical mass effect in which the greater development generates enough self-access to be stable. As Z, the contribution of the outside world, becomes large compared to J*, the minimum point separating the two branches of the T curve moves toward $R_F$, implying that the dense-branch solution for a small area well-connected to the rest of the world probably entails unreasonably high density. But for a substantially isolated region, the existence of the dense-branch solution offers the possibility that unexpectedly large self-sustaining development can occur, whether by fiat, special inducement, historical quirk, or otherwise.

Even assuming the general validity of Eq. 20a, the question of dense-branch development is an open one. It may very well be nothing but a whim of the mathematics. On the other hand, it may have real significance, especially when considering parts of the world, ancient or modern, having low values of T, which would allow smaller dense-branch development. Presumably dense-branch development could be accomplished suddenly by fiat through deliberate acts of political or economic power. It could also occur in more gradual evolution if T were to fall, go around the minimum, and rise again, dragging a city or two along with it. Conceivably the world might be made up of both low-branch and dense-branch development, with the distinction constituting a rather elegant definition of urbanization.

Everything else staying the same, the lower the low-branch solution the higher the dense-branch solution, and the greater the difference between them. The difference is mathematically capable of being small, but it looks as though it is commonly very large, possibly too large to be interesting, and growing all the time. It is intriguing to note that history records many examples of fiat development, most of which, like Karakorum, dwindled away soon after the fiat was removed. The two notable exceptions that come to mind are Constantinople and Alexandria—both enormously impressive in scale, even by modern standards. Perhaps it takes that kind of effort to make the great leap from low branch to dense branch, from normal village to hypernormal city.

Of course, both of these cities were well situated in the first place, unlike Karakorum, and their situations were improved in the course of founding by introduction of streets, roads, and shipping. This is a two-edged observation. On the one hand, it may mean that their low-branch solutions were fairly high and their dense-branch solutions fairly low, making them suitable candidates for bridging the gap. On the other hand, it opens the interpretation that no gap-jumping was involved at all, but that they are merely unusually dramatic instances of ordinary development.

Whether or not the dense branch has anything to do with fiat development, the founding of cities (or anything else) cannot really be decided without more and better numbers. In the absence of quantitative demonstration, it does not seem urgently needed to explain anything, yet there are things that it might very nicely explain.

## The Cubic-Equation Form

From the theoretical point of view, which prefers to think of a site as indefinitely small and contributing virtually nothing to its own accessibility integral, Eq. 20 is the natural and lucid way to write the expression for floor area. There is another form, though, not so natural but still lucid in its own devious way, that turns out to be in general better behaved, and in some cases quite necessary for practical computing and numerical analysis.

Equation 20 is not a pure expression for $R_f$ when dealing with sites or zones of finite size. $R_f$ does not really stand isolated on the left of the equation, but occurs on the right side as well through its contributions to I and J. From the definition of I, it is easy to see that

$$I = I_0 + f_0R_f$$

where $I_0$ is I with the contribution due to $R_f$ extracted, and $f_0$ is the intrazonal value of the trip-distribution function, its average value within the origin site itself; from the definition of J, it is possible (ignoring contraints, which complicate things somewhat) to work out that

$$J = J_0 + 2R_fI_0 + f_0R_f^2$$

where $J_0$ is J with the contribution due to $R_f$ extracted. Putting these expansions of I and J back into Eq. 20, and picking through the notational shards for $R_f$, results in the cubic equation

$$f_0R_f^3 + R^2(2I_0 - f_0R_f) + R_f(J_0 - I_0R_F - R_af_0R_F) - I_0R_aR_F = 0$$

This actually represents a large set of simultaneous equations, of course, one for every site, with every $R_f$ adding something to every $I_0$ and $J_0$. There is not much to be gained from trying to write out the solutions of such equations—each one can be solved numerically and a set of consistent solutions can be found by iteration. A cer-tain amount of analyzing can be done, however, and, without pretending to say anything rigorously general or exhaustive about any large set of simultaneous cubic equations, the following properties appear to hold.

Each equation always has at least one entertainable solution, a root that is real, positive, and consistent with real and positive solutions for every other equation; often there can be two such roots, and sometimes three.

If only the smallest entertainable solutions are accepted for all equations that have more than one, the result is a unique set of solutions for the system—development at every site is uniquely determined. This appears to correspond to all real cases and is the policy adopted in the prototype program.

If an upper solution is used, there can be more than one development pattern depending on which site is chosen for the upper solution; these upper solutions entail concentrations of development at densities well above the point at which congestion effects and construction costs would decisively supervene in the present world. The physical meaning would seem to be that various stable development patterns involving extreme concentration could be achieved, ignoring congestion, if someone had the power to deliberately rearrange things that way. These upper solutions should not be confused with the dense-branch solutions discussed earlier. The contexts are entirely different; no doubt there is some kind of formal indirect relation, but no likely one worth pursuing.

In a finite region, with the transportation system remaining fixed, as more and more total development is added the lowest solution for the most central place will rise while the next higher solution will fall toward it. At some point these two solutions will meet, and then, if a little more development is thrown into the total, they will both abruptly become nonreal, leaving only a third, uppermost solution. Evidently a highly saturated region can become suddenly supersaturated, requiring the most central place to grow cataclysmically by sucking in development from all other places. Again, these effects occur at unfeasibly high densities. The physical meaning is problematical, but it is certainly a warning of some kind to the world.

It should be pointed out that solutions other than those in the "lowest entertainable" set are always of clouded significance, though it is not impossible that a meaning can sometimes be found for them. The treatment here is essentially numerical, making the assumption that intrazonal structuring does not exist and need not be considered. Obviously, the smaller the zone size, the better this assumption. But if Eq. 20 is taken to be a point function applying to infinitely small areas of infinitely small $R_a$, the cubic equation cannot even be developed for a point of finite $R_f/R_a$ density, because such a point makes no appreciable contribution to I or J. In general, however, the three values of $R_f$ as $R_a$ approaches zero are the "proper" value taken directly from Eq. 20 (the limiting case of the lowest entertainable solution), and two other values, not necessarily real, positive, or meaningful, given by

$$\frac{1}{2f_0}\left[\left(f_0R_f - 2I\right) \pm \sqrt{4I^2 - 4f_0J + f_0^2R_F^2}\right]$$

## The Quadratic Form

If Eq. 20a is subjected to the same kind of treatment that led to the cubic-equation form of Eq. 20 in the preceding section, the result is a quadratic equation,

$$f_0R_f^2 + R_f(I_0 + f_0R_a) + I_0R_a = 0$$

with solutions

$$R_f = \frac{1}{2f_0}\left(M \pm \sqrt{M^2 - 4f_0I_0R_a}\right)$$

where

$$M = T - (I_0 + f_0R_a)$$

Both of these solutions are always real and positive (unless the given value of T is too small for the system, in which case both are meaningless). The lower solution, the one using the minus sign, is the proper, well-behaved value. For an individual zone, the upper solution here corresponds to the dense branch of the T curve, but it is not clear that a consistent set of such solutions can be formed in a many-zoned region, or that that is the way to go about finding the dense-branch development pattern. It seems more likely that the dense-branch pattern would be found by nucleating the region with development and then working with the lower quadratic solutions—the dense-branch and the upper quadratic solution can be firmly associated with each other only when the entire region is treated as a single zone. If a large region were treated as a single zone, the lower and upper quadratic solutions would give the total development for the lower and dense branches respectively, but because they ignore internal structuring they would not necessarily be very accurate. The upper solution for the dense branch especially would suffer from this. In general, the lower solution would be too low and the upper solution would be too high.

## Closing Note

With the exception of the cubic-equation form, which occurs as an algorithm in the prototype program, none of the above ideas has been computationally implemented or explored to any great extent. Most likely they will be given more attention, perhaps along with other extensions that may turn up, under a further contract with the U.S.

Bureau of Public Roads, the primary intent of which is to develop a computer model with more realistic abilities. The author is grateful to the Bureau, not only for its support, but for its patience and consideration during the course of this project.

## REFERENCES

1. Transportation and Land Development. Creighton-Hamburg, Inc., final report of project under U.S. Bureau of Public Roads contract FH-11-6792, 1969.
2. Schneider, M.S. Access and Land Development. HRB Spec. Rept. 97, pp. 164-177, 1962.

through these two properties of length and mode distribution by comparing calculations with actual trip lengths and modal splits.

The third unfixed quantity is the relative scale of attractiveness due to land area, $R_a$, and that due to floor area, $R_f$. The units of R can be arbitrarily chosen, just as a foot is an arbitrary unit of length, but how much R resides in a square foot of land area and how much in a square foot of floor area is up to empirical measurement. If, for convenience, one unit of R is defined to be the amount of R produced by one square mile of land, then the empirical problem becomes that of finding the rate of R for floor area. This is a rather more subtle problem than that of estimating "a" and "b," but it can be solved by making use of the not too obvious property of the theory that the pattern of floor-area development changes with a change in total $R_F$. In general, concentration effects in a given system increase with total development (as indeed seems to be the case in the real world). The procedure for scaling $R_f$ would be to calculate the distribution of $R_f$'s for some real region in successive trials, using different values of total $R_F$, until the shape of the $R_f$ distribution agreed with the shape of the actual floor-area distribution—that is, until the ratios to each other of $R_f$'s around the region were the same as the corresponding floor-area ratios, or, the same thing, the ratio of $R_f$ to the floor area it was trying to represent was the same everywhere. The latter ratio, of course, is the sought-for conversion factor.

One further complication is that it can be difficult, though not strictly impossible, to calculate well trip distributions in order to estimate "a" and "b" without knowing how to measure R, and it is impossible to calculate floor-area distributions well, in order to measure R, without good values of "a" and "b." A certain amount of successive refinement is involved.

These are the things that would be done if things could be done that way. In practice the prototype tests, although they have this general scheme in mind, suffer from an exiguity of appropriate data and from their own ineptitude at representing complicated reality, and tend to limp and stumble from case to case. It is hard to know how to report the results of this kind of project in capsule form. The program works. Sometimes it is even fun. The mathematics make computational sense and look as though they might be reflecting the real world. Calculations exhibit agreement with data, but there are few points of contact. The approach to modal split seems to work well, but the trip-distribution function may be having some trouble in its ambitious effort to comprehend all trip behavior from the very short to the very long in one function. Interestingly enough, the scale of $R_f$ seems to turn out to be about the same as that of $R_a$; a square mile of floor area apparently contains about one unit of R, just as does a square mile of land area.

On the whole, the idea looks promising to the intuition, and not too much more than that can be expected at this point. Perhaps its most engaging feature continues to be theoretical richness, the fact that it makes predictions, as opposed to statistical recapitulations, and focuses the viewpoint (distorting it at the same time, possibly). Two very general predictions deserve pointing out: if the quality and extensiveness of transportation does not change much, centralization will occur as the world accrues more and more fixtures of development; if transportation becomes more extensive rapidly enough, decentralization will occur.

It might be mentioned in passing that the theory does not particularly require every square foot of floor area to have the same R value as every other square foot, or every square foot of land area to be the same as every other. But it would be far beyond the resolving power of this project to distinguish differences of that order.

## ELUCIDATIONS

The expression for floor area at a site is

$$R_f = \frac{R_a I R_F}{J - I R_F} \tag{20}$$

where I is the accessibility integral around the site and J is the integral of these accessibilities throughout the region.

## Regional "Temperature"

An examination of Eq. 20 reveals that the ratio $J/R_F$ is a special quantity. Development distributions calculated separately for different regions (excluding constrained areas) have to give the same $J/R_F$ in order to be consistent with the distribution obtained by lumping the regions together in a single superregion. The ratio can be formed for regions of any size at all, large or small, and will be the same for every place entering into a distribution. Presumably regions could have different $J/R_F$ ratios if the interaction between them were weak, or if there were barriers or hidden constraints due to policies on travel, commerce, or immigration, but presumably also these differences would tend to diminish with time.

Indeed, this whole idea of land development can be restated in the form of a principle: the world acts to keep $J/R_F$ the same everywhere within well-communicating regions. This has something of the ring of an entropy law to it, with $J/R_F$ vaguely analogous to temperature. It says nothing about the dynamics of the process, only its goal. Carrying out the analogy, borders, oceans, and geographic and political barriers can be thought of as insulating walls slowing down or impeding the entropic process. The analogy is an attractive one, and may even be valid as far as it goes. But other important factors—things such as access to goods and resources, productivity, various kinds of costs—are not explicitly taken into account (although there seems to be theoretical elbow room for them to work their way in).

Continuing with the idea of regional "temperature," it is easy to let T stand for $J/R_F$ and to rewrite Eq. 20 as

$$R_f = \frac{R_a I}{T - I} \tag{20a}$$

This form clearly suggests that if T is known, the distribution of development can be calculated without knowing the total development in advance: development becomes a matter of interaction between access at each point and the state parameter, T, instead of the distribution in detail of an exogenous total. Here T begins to function in a rather lofty role, trying to dictate the shape of the world. Presumably T changes with history, rising as the extensiveness and quality of transportation increases, though capable also of falling in a period of transportation stagnation combined with growth in total development. Almost certainly, T has been rising in the modern era (perhaps it has always been rising) and, from Eq. 20a, a site needs more access to get developed, or to stay developed, than it used to. Moreover, T increases especially fast when places of more than average desirability become accessible, another long-term tendency that works some hardship on fixed-access sites and on those that are less congenial.

The notion that T is some kind of fundamental system measure has considerable appeal to vagrant speculation. Perhaps deeper social, economic, and historical meanings can be found for it. Perhaps the dynamics of the development process can be understood in terms of flows of development governed by T differences from place to place. Perhaps T represents something to which a value can be attached in human events, so that a place of higher T may be said to be better off than one of lower T. But in a much more practical vein, the concept of T serves to make a connection among regions that could be treated together only in a system too big to handle comfortably: calculations done for areas all over the country can be made consistent with each other without having to stuff the whole continent into a computer at once.

## The Dense Branch: Fiat Development

Given a region containing a transportation system but no development, there will be some value of J that may be called J*. If development is added uniformly everywhere, with the transportation system remaining fixed, it can be seen from definitions that J becomes, for any total development, $R_F$, and with total land, $R_A$:

$$J = (J^* - Z)\ (1 + R_F/R_A)^2 + Z\ (1 + R_F/R_A)$$

Here Z is a contribution to J* due to access from within the region to the world outside. (In a real, nonuniform case, J would increase more rapidly with $R_F$ than this because of centralization effects.)

And T becomes:

$$T = \frac{1}{R_F}(J^* - Z)\ (1 + R_F/R_A)^2 + \frac{Z}{R_F}\ (1 + R_F/R_A)$$

This is a two-valued function, approaching infinity when $R_F$ is either very small or very large, with a minimum at

$$R_F = R_A \sqrt{[J^*/(J^* - Z)]}$$

So starting with an empty system and knowledge of the world value of T, it appears that two quite different development patterns can be calculated, one at a lower total development and one at a higher. The lower density solution apparently is the proper one, representing normal incremental growth, and the higher solution can be interpreted as a kind of critical mass effect in which the greater development generates enough self-access to be stable. As Z, the contribution of the outside world, becomes large compared to J*, the minimum point separating the two branches of the T curve moves toward $R_F$, implying that the dense-branch solution for a small area well-connected to the rest of the world probably entails unreasonably high density. But for a substantially isolated region, the existence of the dense-branch solution offers the possibility that unexpectedly large self-sustaining development can occur, whether by fiat, special inducement, historical quirk, or otherwise.

Even assuming the general validity of Eq. 20a, the question of dense-branch development is an open one. It may very well be nothing but a whim of the mathematics. On the other hand, it may have real significance, especially when considering parts of the world, ancient or modern, having low values of T, which would allow smaller dense-branch development. Presumably dense-branch development could be accomplished suddenly by fiat through deliberate acts of political or economic power. It could also occur in more gradual evolution if T were to fall, go around the minimum, and rise again, dragging a city or two along with it. Conceivably the world might be made up of both low-branch and dense-branch development, with the distinction constituting a rather elegant definition of urbanization.

Everything else staying the same, the lower the low-branch solution the higher the dense-branch solution, and the greater the difference between them. The difference is mathematically capable of being small, but it looks as though it is commonly very large, possibly too large to be interesting, and growing all the time. It is intriguing to note that history records many examples of fiat development, most of which, like Karakorum, dwindled away soon after the fiat was removed. The two notable exceptions that come to mind are Constantinople and Alexandria—both enormously impressive in scale, even by modern standards. Perhaps it takes that kind of effort to make the great leap from low branch to dense branch, from normal village to hypernormal city.

Of course, both of these cities were well situated in the first place, unlike Karakorum, and their situations were improved in the course of founding by introduction of streets, roads, and shipping. This is a two-edged observation. On the one hand, it may mean that their low-branch solutions were fairly high and their dense-branch solutions fairly low, making them suitable candidates for bridging the gap. On the other hand, it opens the interpretation that no gap-jumping was involved at all, but that they are merely unusually dramatic instances of ordinary development.

Whether or not the dense branch has anything to do with fiat development, the founding of cities (or anything else) cannot really be decided without more and better numbers. In the absence of quantitative demonstration, it does not seem urgently needed to explain anything, yet there are things that it might very nicely explain.

## The Cubic-Equation Form

From the theoretical point of view, which prefers to think of a site as indefinitely small and contributing virtually nothing to its own accessibility integral, Eq. 20 is the natural and lucid way to write the expression for floor area. There is another form, though, not so natural but still lucid in its own devious way, that turns out to be in general better behaved, and in some cases quite necessary for practical computing and numerical analysis.

Equation 20 is not a pure expression for $R_f$ when dealing with sites or zones of finite size. $R_f$ does not really stand isolated on the left of the equation, but occurs on the right side as well through its contributions to I and J. From the definition of I, it is easy to see that

$$I = I_0 + f_0 R_f$$

where $I_0$ is I with the contribution due to $R_f$ extracted, and $f_0$ is the intrazonal value of the trip-distribution function, its average value within the origin site itself; from the definition of J, it is possible (ignoring contraints, which complicate things somewhat) to work out that

$$J = J_0 + 2R_f I_0 + f_0 R_f^2$$

where $J_0$ is J with the contribution due to $R_f$ extracted. Putting these expansions of I and J back into Eq. 20, and picking through the notational shards for $R_f$, results in the cubic equation

$$f_0 R_f^3 + R^2(2I_0 - f_0 R_f) + R_f(J_0 - I_0 R_F - R_a f_0 R_F) - I_0 R_a R_F = 0$$

This actually represents a large set of simultaneous equations, of course, one for every site, with every $R_f$ adding something to every $I_0$ and $J_0$. There is not much to be gained from trying to write out the solutions of such equations—each one can be solved numerically and a set of consistent solutions can be found by iteration. A certain amount of analyzing can be done, however, and, without pretending to say anything rigorously general or exhaustive about any large set of simultaneous cubic equations, the following properties appear to hold.

Each equation always has at least one entertainable solution, a root that is real, positive, and consistent with real and positive solutions for every other equation; often there can be two such roots, and sometimes three.

If only the smallest entertainable solutions are accepted for all equations that have more than one, the result is a unique set of solutions for the system—development at every site is uniquely determined. This appears to correspond to all real cases and is the policy adopted in the prototype program.

If an upper solution is used, there can be more than one development pattern depending on which site is chosen for the upper solution; these upper solutions entail concentrations of development at densities well above the point at which congestion effects and construction costs would decisively supervene in the present world. The physical meaning would seem to be that various stable development patterns involving extreme concentration could be achieved, ignoring congestion, if someone had the power to deliberately rearrange things that way. These upper solutions should not be confused with the dense-branch solutions discussed earlier. The contexts are entirely different; no doubt there is some kind of formal indirect relation, but no likely one worth pursuing.

In a finite region, with the transportation system remaining fixed, as more and more total development is added the lowest solution for the most central place will rise while the next higher solution will fall toward it. At some point these two solutions will meet, and then, if a little more development is thrown into the total, they will both abruptly become nonreal, leaving only a third, uppermost solution. Evidently a highly saturated region can become suddenly supersaturated, requiring the most central place to grow cataclysmically by sucking in development from all other places. Again, these effects occur at unfeasibly high densities. The physical meaning is problematical, but it is certainly a warning of some kind to the world.

It should be pointed out that solutions other than those in the "lowest entertainable" set are always of clouded significance, though it is not impossible that a meaning can sometimes be found for them. The treatment here is essentially numerical, making the assumption that intrazonal structuring does not exist and need not be considered. Obviously, the smaller the zone size, the better this assumption. But if Eq. 20 is taken to be a point function applying to infinitely small areas of infinitely small $R_a$, the cubic equation cannot even be developed for a point of finite $R_f/R_a$ density, because such a point makes no appreciable contribution to I or J. In general, however, the three values of $R_f$ as $R_a$ approaches zero are the "proper" value taken directly from Eq. 20 (the limiting case of the lowest entertainable solution), and two other values, not necessarily real, positive, or meaningful, given by

$$\frac{1}{2f_0}\left[\left(f_0R_f - 2I\right) \pm \sqrt{4I^2 - 4f_0J + f_0^2R_F^2}\right]$$

<u>The Quadratic Form</u>

If Eq. 20a is subjected to the same kind of treatment that led to the cubic-equation form of Eq. 20 in the preceding section, the result is a quadratic equation,

$$f_0R_f^2 + R_f(I_0 + f_0R_a) + I_0R_a = 0$$

with solutions

$$R_f = \frac{1}{2f_0}\left(M \pm \sqrt{M^2 - 4f_0I_0R_a}\right)$$

where

$$M = T - (I_0 + f_0R_a)$$

Both of these solutions are always real and positive (unless the given value of T is too small for the system, in which case both are meaningless). The lower solution, the one using the minus sign, is the proper, well-behaved value. For an individual zone, the upper solution here corresponds to the dense branch of the T curve, but it is not clear that a consistent set of such solutions can be formed in a many-zoned region, or that that is the way to go about finding the dense-branch development pattern. It seems more likely that the dense-branch pattern would be found by nucleating the region with development and then working with the lower quadratic solutions—the dense-branch and the upper quadratic solution can be firmly associated with each other only when the entire region is treated as a single zone. If a large region were treated as a single zone, the lower and upper quadratic solutions would give the total development for the lower and dense branches respectively, but because they ignore internal structuring they would not necessarily be very accurate. The upper solution for the dense branch especially would suffer from this. In general, the lower solution would be too low and the upper solution would be too high.

<u>Closing Note</u>

With the exception of the cubic-equation form, which occurs as an algorithm in the prototype program, none of the above ideas has been computationally implemented or explored to any great extent. Most likely they will be given more attention, perhaps along with other extensions that may turn up, under a further contract with the U. S.

Bureau of Public Roads, the primary intent of which is to develop a computer model with more realistic abilities. The author is grateful to the Bureau, not only for its support, but for its patience and consideration during the course of this project.

## REFERENCES

1. Transportation and Land Development. Creighton-Hamburg, Inc., final report of project under U.S. Bureau of Public Roads contract FH-11-6792, 1969.
2. Schneider, M.S. Access and Land Development. HRB Spec. Rept. 97, pp. 164-177, 1962.

Although the scatter of the data from this simplified relation is great, the equation agrees reasonably well with the observation that many of the test road pavements reached a PSI rating of 1.5 when the mean rut depth reached 0.6 to 0.7 in. (virtual structural failure).

## Causes of Pavement Deterioration and Failure

The subjective PSR gives no clue as to the cause of the deterioration of a pavement from the high value it presumably possessed when it was built. The empirical PSI indicates the major and minor factors in the loss of the initial serviceability, but does not define the mechanism by which they develop. In addition, the scatter of the data suggests that there are additional factors in the PSI.

Although deterioration or failure of the pavement to perform its function may be reflected in the surface condition, the seat of the trouble can be in any of the layers which make up the flexible pavement system: the surface course, the base course, the subbase (if any), and the subgrade or embankment. Furthermore, the initial failure of one may lead to a failure, often in a different form, in another. For example, cracking of the asphaltic surface may let surface water into the subgrade and cause its softening and eventual shear.

The mechanisms for pavement deterioration are suggested in Table 1. Some are primarily related to traffic load, whereas others are either independent of the traffic load or are related to the load only in that the failure is intensified or aggravated by the load rather than caused by it. Deterioration and failure in the surface were not within the scope of this investigation. The causes are listed because they had to be considered in diagnosing the mechanism of deterioration of existing pavements and deciding which failures were the result of inadequate pavement thickness. Deterioration and failure of the base and subbase were similarly beyond the scope of this study, except when the base was so similar to the subgrade in its properties that the base and subgrade had to be considered as a unit, as with topsoil and sand bases. The failure of the higher types such as sand-asphalt and soil-cement was not investigated.

A major, and possibly the most important, function of the pavement is to distribute the concentrated wheel load so that the stress does not exceed the supporting capabilities of the subgrade. Of course, the deformation and failure of the subgrade are reflected in the surface condition and thereby in the PSR or PSI. The elastic deformation, consolidation (densification) and shear (bearing capacity) failure of the subgrade are directly related to the wheel load and the resulting stress distributed through the surface and base courses.

TABLE 1

MECHANISMS FOR PAVEMENT DETERIORATION

| Surface | Base and Subbase | Subgrade |
|---|---|---|
| Elastic deformation—rebound[a] | Elastic deformation[a] | Elastic deformation[a] |
| Densification (consolidation)[a] | Densification (consolidation)[a] | Densification (consolidation)[a] |
| Thermal expansion and contraction | Shear failure (bearing capacity)[a] | Shear failure (bearing capacity)[a] |
| Longitudinal shear failure (shoving)[a] | Deterioration of aggregate | Swell-shrink |
| Curvilinear shear (bearing failure)[a] | Deterioration of cementing agent | Pumping |
| Deterioration of bitumen | Swell-shrink | Settlement of deep strata |
| Separation of courses | Pumping | Mass shear failure (landslide) |
| Bleeding | | Local mass shear (due to weak culverts, trenches) |

[a]Primarily related to traffic loads.

The remaining mechanisms are not directly caused by the wheel loads. Swelling and shrinking of the subgrade depend on the moisture changes as well as the mineralogy of the soil. The effect may be bumps and hollows at irregular intervals not related to the traffic or load pattern. Swelling may have a secondary effect in that softening or weakening of the subgrade can lead to deflection or shear failure that is load related. Similarly, shrinkage has a secondary effect in producing tension and shear cracks in the pavement courses above. Although these are not necessarily load related, they may be aggravated by the load. Therefore, it is difficult to isolate the effects of deterioration due to swelling and shrinking, although the basic mechanism is different from the others.

Pumping is a complex phenomenon indirectly related to load; it arises from the effect of free available moisture on a susceptible subgrade or base. The load of the moving wheel causes the pavement components to deflect. After the load passes, the components rebound. If the upper layers rebound faster or more than the lower layers (which is likely because in the typical flexible pavement system, the upper layers are more rigid and possibly more nearly elastic), a temporary void is formed between the layers. If free water is available, it is sucked into the void, only to be expelled at the next loading. If the base or subgrade is easily softened or eroded, the pumping of water in and out creates an erosion cavity and eventually a structural failure.

Settlement of the roadway (ordinarily an embankment), because of consolidation of deeper strata, landslides and localized shear failures caused by weak culverts or improperly compacted backfills behind bridge abutments or in trenches, can cause disruption of the pavement surface and a loss of serviceability. None of these, however, are directly related to the design or adequacy of the pavement. Furthermore, the traffic loads are often not major factors in these phenomena because they may be small compared to the weight of the soil mass that is involved. Pavement deterioration due to these phenomena, therefore, must be discounted in evaluating observed pavement conditions for the purpose of developing a pavement design.

The major subgrade mechanisms that contribute to pavement deterioration are deformation and shear failure.

Subgrade Deflection.—The deflection of the subgrade under traffic load results from stresses particularly vertical, transmitted through the pavement system. Both theory and stress measurements show that vertical stresses become smaller with increasing depth below the pavement surface and with increasing horizontal distance from the center of the line of load application, depending on the elastic characteristics of the subgrade and base course (4).

These stresses have a two-fold effect on the subgrade (and on the other pavement layers). First, they produce a downward deflection of the subgrade surface due to the deformation of the soil without appreciable volume change. This can be visualized as the shortening and lateral building of the column of soil immediately below the load similar to the shortening of any axially loaded structural member. If it is assumed that the subgrade is a semi-infinite isotropic homogeneous elastic mass with a modulus of elasticity of E and is momentarily incompressible and that a uniform pressure of q is applied to a square area of width b, the deflection $\rho$ due to deformation will be

$$\rho = \frac{0.6\,qb}{E} \tag{2}$$

that is, the deflection is the same as for a free-standing column of soil whose height is 0.6 times the width of the column. Of course, neither the distribution of the load nor the shape of the loaded area of the subgrade is as simple as the conditions assumed in this equation. More accurate, and more elaborate, mathematical representations of the deformation deflection of a subgrade are available. All are of the same general form as Eq. 2; therefore, this suffices as a model for illustrating the effects of some of the different factors involved. The deflection in any case is directly proportional to the pressure and the size of the loaded area and inversely proportional to the modulus of elasticity.

The second deflection mechanism is the consolidation or densification of the subgrade. Although the theories of soil settlement due to reduction in the volume of the voids have been primarily applied to foundations of structures, they apply also to the consolidation of the subgrade. The relation between void ratio change and stress increase is more complex than that for elastic deformation and, therefore, a simple expression for consolidation settlement is not available even for homogeneous soils. However, consolidation settlement does increase with increasing stress, not in direct proportion but more nearly in proportion to the log of the increase compared to the original stress due to the soil weight.

Under repeated loadings, progressive consolidation occurs. With each successive cycle of load and unload, the reduction in voids rapidly becomes less. Settlement appears to continue indefinitely, but at an ever decreasing rate. Subgrade deformation and consolidation cause an elongated depression in the wheelpath that is entirely below the original surface level. The deformation deflection is temporary and is recovered after the wheel passes. The major effect is an "alligator" cracking of the surface course if the deflection is sufficiently great. The estimated limiting deflection, based on the U. S. Navy airfield design, is 0.2 in., although some highway departments have suggested limiting deflections of 0.05 in. for major highways. Consolidation deflection causes a permanent rut entirely below the original surface. The rut may be accompanied by longitudinal and possibly transverse cracks. In addition, long longitudinal waves in the rut may be observed where there is severe consolidation.

Shear Failure.--Shear failure of the subgrade, similar to the bearing capacity failure of a foundation, can result if the stresses transmitted to the subgrade through the base and surface courses exceed the strength in a sufficiently large zone. If it is assumed that the subgrade is homogeneous and its properties can be described by the unit weight $\gamma$, the cohesion c, and the angle of internal friction $\varphi$, and if it is assumed that the pressure transmitted to the subgrade is vertical and uniform over an area of width b, the pressure at which the soil will shear $q_o$ is defined by

$$q_o = \frac{\gamma b}{2} N_\gamma + c N_c + q' N_q \qquad (3)$$

In this expression $N_\gamma$, $N_c$, and $N_q$ are dimensionless functions of the foundation shape and angle of internal friction and $q'$ is the weight of the pavement and base above the subgrade.

Many variations of this expression, originally proposed by Terzaghi (5), have been published. The differences are in the mode of loading and assumed character of the zone of shear failure and they are manifested in differences in the values of the N-factors. So far no analysis has been developed for a nonuniform loading of indefinite width such as that transmitted through the pavement to the subgrade. However, it is to be expected that the general form of the equation will be little changed; instead, the values of N will reflect the nonuniform loading. Therefore, bearing capacities for subgrades computed by Eq. 3 and utilizing the N-values for one of the existing methods of analysis should be approximately proportional to the true bearing capacities. Or, conversely, the safety factor with respect to shear failure computed by Eq. 3 and utilizing certain existing N-factors and the average stresses transmitted to the subgrade through the pavement system should have some reasonably constant relation to the true safety factors.

Whereas the strength parameters c and $\varphi$ reflect complete soil failure, they may not indicate the development of limited but accumulating shear under repeated loads that are not great enough to produce complete failure. Although little is known about the effects of repeated loading on progressive shear, the indications are that the magnitude of progressive failure increases with the increasing ratio of the actual stress to the failure stress. That is, progressive failure increases with a decrease in safety factor.

Shear in the subgrade is accompanied by a broad deep depression or rut in the wheelpath with the upheaval occurring beyond it. Longitudinal cracking may be severe and eventually leads to transverse cracking which forms a blocky pattern (6). Shear along the pavement edge may be accompanied by curved cracking and outward movement of the base and surface, and sometimes by severe outward tilting.

Apparent Safety Factor

The stresses computed by any of the elastic theories apply only to the state of elastic equilibrium on which that theory was based. If the elastic state is altered by nonlinear strain or by failure, the stress distribution may be altered. If failure develops suddenly from an elastic state, however, the stresses just before reaching failure are probably not greatly different from those of elastic equilibrium. If it further can be assumed that the pressure, $q_o$, required for complete failure and that required to initiate failure are approximately the same or proportional, then it is possible to compute an apparent safety factor against failure by

$$SF_a = \frac{\sigma_a}{q_o} \tag{4}$$

In this expression, $\sigma_a$ is the average vertical stress transmitted to the soil surface by the pavement system, as computed by an appropriate elastic theory, and $q_o$ is the ultimate bearing capacity computed by Eq. 3, utilizing appropriate factors. The apparent safety factor is not the true safety factor (i.e., failure does not necessarily occur at a safety factor of 1), but it is reasonable to assume that both safety factors are proportional.

Summary

Pavement deterioration and failure is the result of a series of complex processes, none of which are clearly understood and only part of which are directly related to the loads supported. Although exact methods of analyzing the mechanical processes of subgrade deflection and shear failure are not available, approximations can be made that point out the relative importance of the different factors involved and also indicate the relative magnitude of possible deformation and the safety against shear failure.

The greatest unknown factors are those which involve the environment: temperature, frost action, groundwater, surface water infiltration, and other moisture changes. These profoundly influence the deformation and shear failure characteristics of all pavement components but particularly those of the subgrade. At the present time, little is known about the direct effects of the environment on the soil and too few facts are available to permit valid empirical correlations to be made.

## SURVEY OF GEORGIA PAVEMENTS

A survey of Georgia pavements was undertaken in 1961 to locate typical areas in all four of the geologic regions (Coastal Plain, Piedmont, Blue Ridge, and Appalachian Ridge—Valley) in which comparable pavements had both exhibited good performance and deteriorated badly. An inquiry was sent to each of the Georgia State Highway Department field divisions asking for their suggested locations for study. From these a list of 84 was compiled for examination and testing.

A field examination was made of each location in the late summer, fall, and early winter of 1961. The pavement was examined visually and data on the roadway environment and pavement condition were obtained. A survey of the traffic was made during the period of pavement examination in which the total number of vehicles and the number of heavy trucks in the lane under study was counted and the percentage of heavy trucks estimated. Although such a short count is not a valid indication of the total traffic, it does give some picture of the character of the traffic on pavements for which no accurate information was available.

The typical depth of rutting was measured using a 4-ft straight-edge placed over the wheelpaths. The segment so measured was then photographed and a sketch made of the pattern of cracking (if any). The dates of construction and repair (if any) were obtained from the field division engineer. He also provided information on the design and construction of the pavement, where it was available. The pavement was marked at the location where samples were to be made, usually in the zone of the failure but not where the failure itself might have disrupted the soil. Finally, a serviceability rating was assigned utilizing the criteria described by Carey (3) and based on the visual observations of the surface condition and its riding qualities.

## Sampling

Samples were secured in most of the locations by the Georgia State Highway Department Division of Materials and Tests. The sampling program was necessarily interspersed with the routine drilling and sampling work for new construction, and thus was spread out over several months. Practically all sampling was done in the late winter and spring of 1962 when the soil moisture conditions were likely to be at the highest.

The bituminous pavement was cored where possible and its thickness measured. The thickness of each deeper pavement course was measured and the materials were described visually. Undisturbed samples were secured of each base course layer that contained no gravel and of the top 2 to 3 ft of the subgrade, utilizing 3-in. O.D. thin-wall sample tubes. The samples were sealed in the field with plastic end caps and brought to the Georgia Institute of Technology Soil Engineering Laboratory.

## Laboratory Tests

The samples were cut into 6-in. sections using either a high-speed abrasive saw or a metal-cutting band saw. Unfortunately, some of the samples were unsatisfactory because of gravel which caught under the edge of the tube and disturbed the soil or because of faulty sealing. Most, however, were suitable for testing.

Because of the limited amounts of sample available, only one form of test could be utilized. Considered the most representative of field conditions was the undrained triaxial test, utilizing the full sample diameter (approximately 2.8 in.) and no changes in moisture. Where possible, three confining pressures, 10, 20, and 40 psi, were employed; however, in some cases only 10 and 20 psi were used when the amount of sample was limited. The samples, each about 6 in. long, were loaded axially at a controlled strain rate of 0.8 to 1 percent/min. The test data were analyzed on a computer and the results plotted in the form of stress-strain curves from which the initial tangent modulus of elasticity for a confining pressure of 10 psi was found.

## AASHO SUBGRADE TESTS

The AASHO test road was constructed to provide as uniform a subgrade as possible, so that initial subgrade variability would not influence the pavement performance. Therefore, the materials utilized were as nearly uniform as possible in composition, and the construction was controlled so that the moisture contents and densities could be kept within narrow limits.

Tests of the subgrade (embankment) base and surface materials were summarized in the AASHO Road Test reports and in other published data on the road (7). These included control tests for quality and physical tests by the U. S. Bureau of Public Roads to determine the structural properties of the compacted materials. In addition, samples were furnished to many state highway departments to be tested by the procedures commonly employed for their own design work. The results of these tests have also been published (8). Limited tests were made of soils in certain of the road embankments which had been removed from test routinely at the programmed end of testing earlier because of deterioration. These included moisture, density, CBR and K-factor tests (1).

## Sampling

None of the published data included strength tests of samples of the subgrades as constructed. Four undisturbed samples were secured by the Illinois Division of Highways on about May 1, 1963, well after the spring thaw. These were of the subgrade, commencing 3 in. below the subbase, and were made with 24-in. long, 2-in. O.D. thin-walled tubes. They were sealed at the site and shipped to the Georgia Institute of Technology Soil Engineering Laboratory.

## Laboratory Tests

Triaxial tests were made of all four samples utilizing the same method and pressures as for the tests of Georgia subgrades. The results were expressed in stress-strain curves and Mohr diagrams, and are summarized in Table 2. As can be seen, the results are not uniform. There is considerable variation in the densities, moistures, and strengths. With the exception of the samples from Sta. 60 + 00, where gravel made a full program of tests impossible, the samples exhibited comparable angles of internal friction of slightly more than 20° and cohesions between 6.2 and 20 psi. A composite plot of all data shows the weaker materials exhibit an average cohesion of 9.5 psi and an angle of internal friction of 20°. These values were used in subsequent analyses.

For comparison, the BPR tests of subgrade samples, laboratory-compacted to 95 percent of AASHO T99-49 maximum (the specified value), gave c and $\varphi$ values of 11 psi and 31°, respectively, for Borrow Pit 1 and 8.9 psi and 21° for Pit 2. The corresponding CBR values were 2.7 and 2.5. The cooperative test (8) results were comparable in those cases where the soil was tested under similar conditions of compaction and moisture content.

# ANALYSIS OF GEORGIA PAVEMENT PERFORMANCE

The performance of the Georgia pavements was analyzed utilizing the pavement descriptions and serviceability rating. The theoretical bearing capacity and deflection of the subgrade were computed by the methods described. These were correlated to form a semirational basis for pavement design evaluation.

## Depth-Stress-Width

A previous paper (4) presented data on the vertical stresses at different depths beneath different pavement systems utilized in Georgia. These tests all indicated that the vertical stress was greatest immediately under the tire and became rapidly less with increasing horizontal distance and increasing vertical depth below the ground surface. For the purpose of analysis, it was assumed that the significant vertical stresses at any depth were those equal to or greater than one-half the maximum vertical stresses at that depth.

The average significant vertical pressures for a 9-kip dual-wheel load (4) were increased by ten-ninths to give the average significant vertical pressures for a 10-kip

TABLE 2

TRIAXIAL TEST DATA AASHO TEST ROAD

| Station | Position (ft) | $\gamma_d$ (pcf) | w (%) | c (psi) | $\varphi$ (deg) |
|---|---|---|---|---|---|
| 60 + 00 | 13.5 ft R of center WB | 125 | 11 | 51 | 0 |
| 149 + 00 | 13.5 ft L of center EB | 114 | 14 | 7.6 | 24 |
| 229 + 00 | 13.5 ft L of center EB | 82 | 28 | 6.2 | 22 |
| 360 + 00 | 13.5 ft R of center WB | 113 | 13 | 21.5 | 20 |
| Composite[a] | | - | - | 9.5 | 20 |

[a]Weaker materials.

dual wheel (the present Georgia load limit of 20 kips per axle). A plot of these stresses (Fig. 1) shows the significant vertical stresses beneath different bases as a function of depth beneath the pavement surface. For most Georgia pavement systems, the curve for the 3-in. asphaltic surface and 8-in. topsoil-soil-macadam, or silt base applies. This is almost identical to the stress distribution computed by the Boussinesq theory and should apply reasonably well to all but soil-cement and sand-asphalt bases. The curves for the latter are shown in Figure 1.

The width of the zone of significant vertical stresses was also found from the stress distribution curves (Fig. 2). The curve can be approximated by the straight line whose equation is

$$b = 15 + 0.72\,z \tag{4}$$

where b is the equivalent width in inches and z is the depth below the pavement surface in inches.

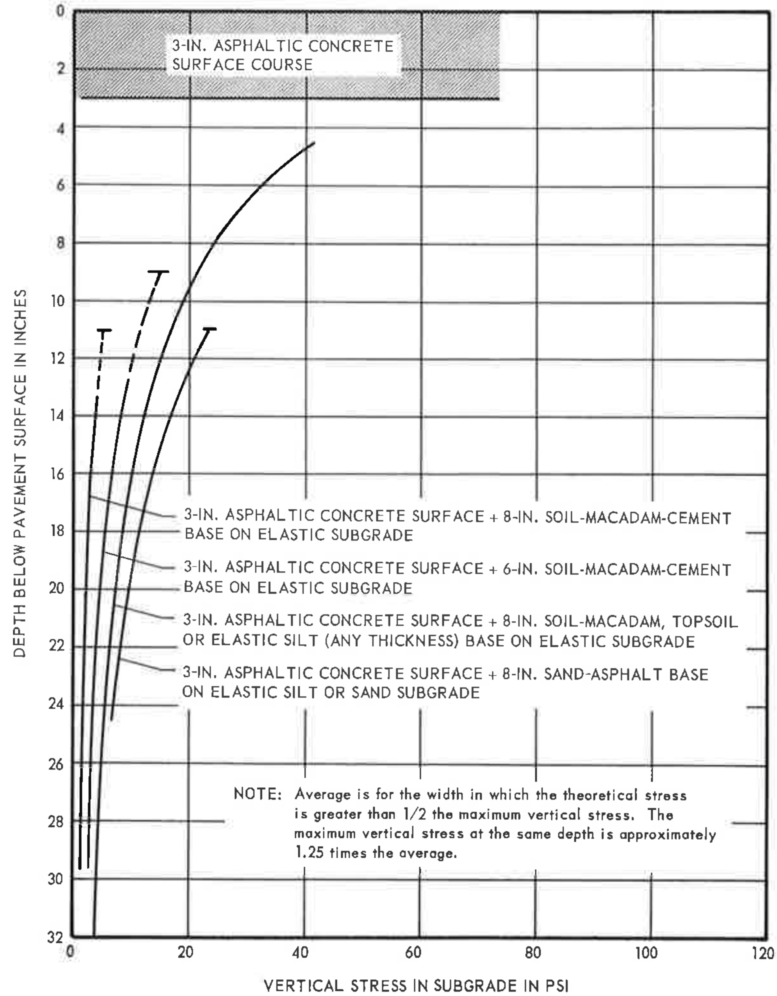



Figure 1. Average significant vertical stress in subgrade for different Georgia base courses and 20-kip axle loads on dual tires.

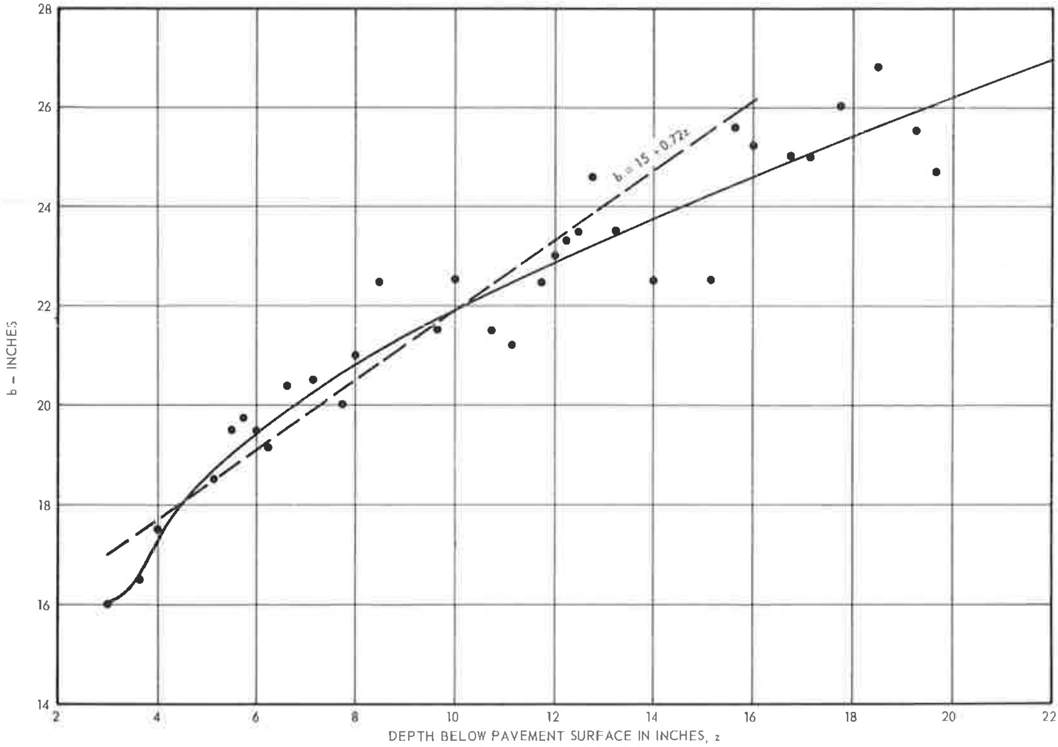



Figure 2. Equivalent width of area of significant vertical pressure in Georgia subgrade.

## Deflection—Bearing Capacity

The elastic deformation of each subgrade under a 20-kip axle load was computed from Eq. 2 and the modulus of elasticity for the subgrade at a containing pressure of 10 psi. (The width, b, utilized in the computation was shown in Figure 2. The stress was shown for the depth of the top of the subgrade by Figure 1.) Of course, it would be false to conclude that this represents the true base deflection of the subgrade. However, it should be proportional to the true deflection if the modulus of elasticity determined by the laboratory tests is correct.

The bearing capacity of the subgrade, and in some cases the bearing of each different subgrade layer where the test data differed greatly, was computed from Eq. 3. The c and $\varphi$ values were those of the soil tests and the b was found from Figure 2. The values of the bearing capacity factors were those computed from the simple Bell-Terzaghi equations as modified by the author (9). For use in these computations the relation was simplified slightly, based on the observation that the total thickness of pavement and base for Georgia is ordinarily 10 to 12 in. In such cases constants can be introduced in the terms involving b, d, and $\gamma$ with little sacrifice in accuracy (considering the greater error involved in utilizing this or any other existing bearing capacity expression in analyzing pavement capacity).

The vertical stress exerted on the subgrade by the 20-kip axle was found from Figure 1. The ratio of the computed bearing capacity to the stress is the apparent safety factor which is probably not the true safety factor, but should be proportional to it. Further, it would be reasonable to conclude that the lower the safety factor, the greater the possibility of shear failure of the subgrade and the greater the amount of progressive shear.

Traffic

A short-term traffic count was made of each sample section. Data on estimated daily total traffic were obtained from the Georgia Division of Highway Planning. Most estimates were based on actual traffic counting at the regular stations in the area. However, some estimates, particularly for the secondary roads in remote areas, were based largely on experience. In no case was the pavement failure close enough to a point of long-term traffic study that the count can be considered accurate. Both the short-term count at the sample section and the Georgia State Highway Department estimate were utilized in determining the number of trucks per day (other than pickups) on the lane under study. This was converted to an equivalent number of 20-kip axles, utilizing a relationship established by the Alabama State Highway Department in their Loadometer studies (10). The total number of trucks multiplied by the weighting factor gives the equivalent 20-kip axles. The values of the factors used were 0.43 for Interstate and primary roads and 0.32 for secondary roads. The Alabama Loadometer studies were for an 18-kip load. The distribution factors of equivalent 20-kip loads in Georgia would probably be slightly smaller, but in the absence of data, the Alabama figures were used. The daily equivalent 20-kip axle-load figure multiplied by the number of days the pavement was in service gives the total axle loads at the time of the evaluation.

Considering the amount of estimating used to establish this traffic figure, it is likely that it may differ from the true value by 50 percent. An even greater variation is likely on the secondary roads with light traffic where even a moderate use by pulpwood trucks or other local highly specialized vehicles represents the major loading of the pavement.

Serviceability-Safety-Traffic

The serviceability for each pavement area was checked by photographs, measured rut depths, and crack patterns. Greatest weight was given to those factors which reflect the subgrade behavior. For example, although the overall serviceability of a

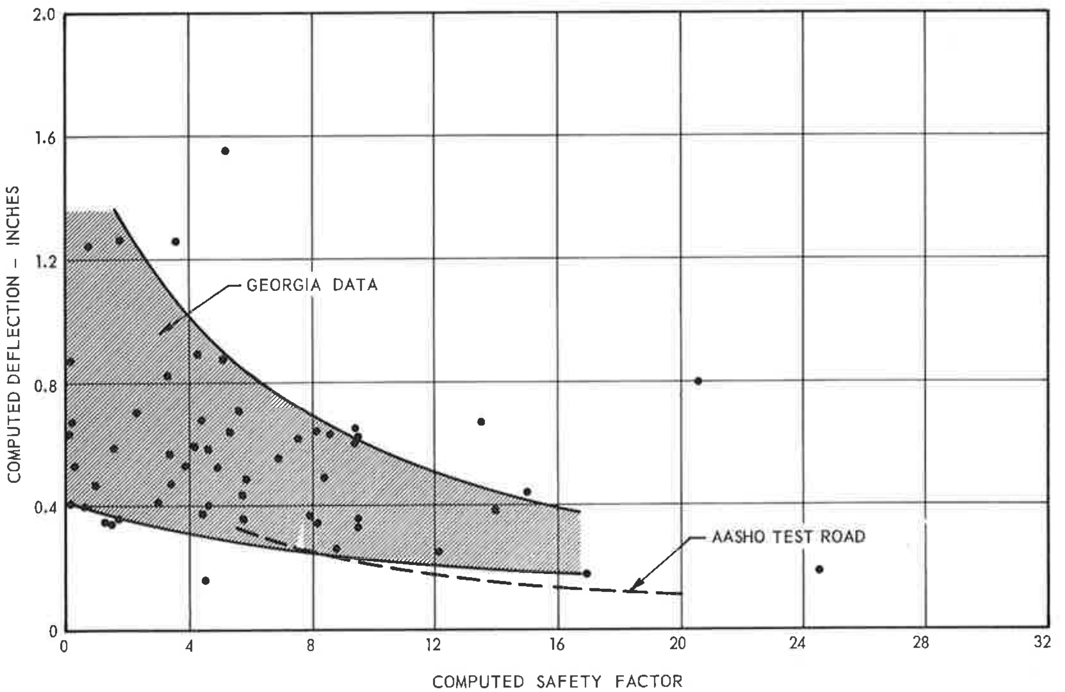



Figure 3. Computed elastic deflection vs computed safety factor of subgrades of Georgia pavements with 20-kip axle loads.

pavement suffering from the peeling of an overlay due to bad bond with the old pavement might be low, the serviceability of the pavement considering the rut depth and longitudinal profile might be high. Because this investigation is concerned with the design of a pavement to fit the subgrade, the subgrade behavior was given greatest weight.

Plots of the serviceability as a function of computed bearing capacity, apparent safety factor, and traffic were made to determine which of these factors was most significant in determining the behavior of the Georgia pavements. A plot of computed deflection vs computed safety factor is shown in Figure 3. Although there is considerable scatter, the relation shows that those pavements having the greatest safety factor against shear failure also exhibit the least elastic deflection; i. e., those soils having the greatest strength are also likely to be the most rigid. This relation also suggests that either deflection or bearing capacity alone might be a satisfactory criterion for design in that one reflects the other to some degree. Because of the limited time available for study and the many factors in both deflection and bearing capacity for which no data were available, no attempt was made to analyze the cause of the scatter.

The plot of serviceability vs apparent safety factor (Fig. 4) also exhibits considerable scatter. However, a general trend is apparent with serviceability decreasing with decreasing safety factor. If the traffic is considered, the trend becomes fairly well defined, with the lighter traffic requiring smaller safety factors than the heavier. Curves were drawn reflecting the largest safety factors required to maintain a given serviceability, for different levels of traffic. In reality, therefore, each curve represents an envelope. There are a few points that do not fit these relations. Some of these with high serviceabilities undoubtedly represent different qualities of initial construction, rather than any deterioration of the pavement. A few exhibit lower serviceabilities because of deterioration other than of the subgrade.

A major unknown factor in the scatter is the fact that the soil test data may not reflect the environmental conditions representative of the greatest degree of deterioration

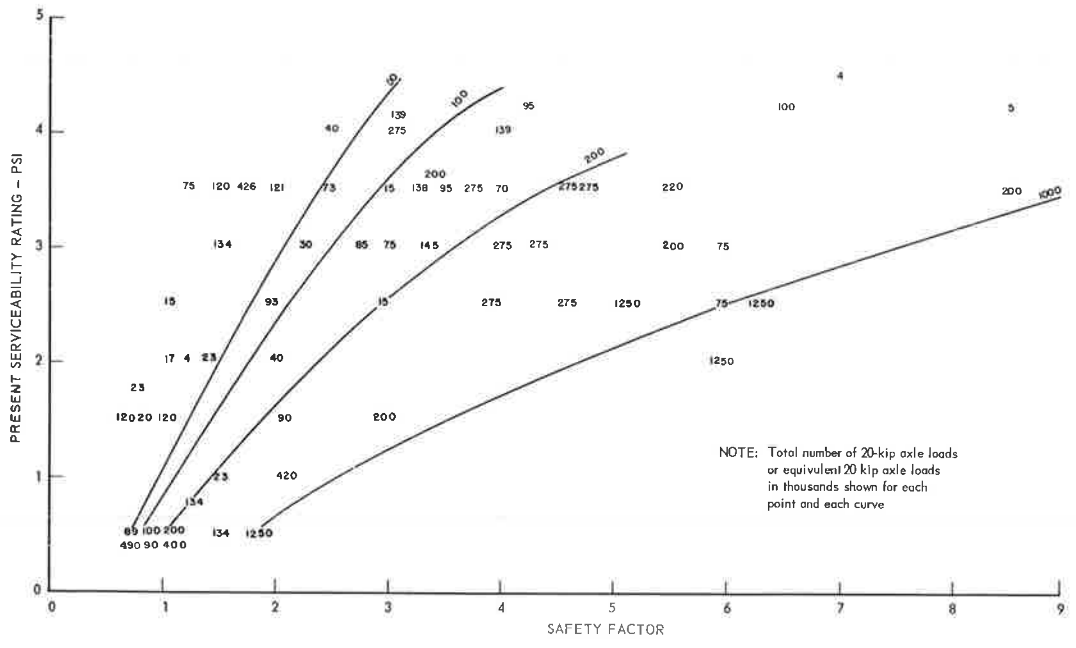



Figure 4. Required safety factors for different present serviceability indexes at end of service period for Georgia pavements and 20-kip axle loads.

and failure. For example, if the subgrade moisture increases in the winter and spring and if failure is most rapid during this period, then the tests should be made of samples obtained during this critical period. Although this was done, it is not known whether the soil at each location was sampled at its worst condition. Moreover, this might not be fair if the traffic during this period of soil weakening was materially less than average. Considering the variable factors which could not be evaluated in this investigation, the degree of correlation shown in Figure 4 is surprising.

## AASHO DATA ANALYSIS

### AASHO Flexible Pavement Evaluation

The evaluation of the AASHO flexible pavement tests is given in detail in Report 5 (1). A brief review of the program, however, is necessary to provide the background for this analysis.

The entire flexible pavement test program utilized a single subgrade soil, a silty clay classified as A-6 by the AASHO system. This was compacted under close control to densities between 95 and 100 percent of AASHO T99-49 maximum so as to provide as uniform a subgrade as possible and to eliminate the factor of variable subgrade support. The controlled variables were pavement component thickness and traffic. Although a few different base course materials were tested in limited sections, the major emphasis was on the effects of different combinations of surface, base, and subbase thickness under axle loads ranging from 2,000 to 48,000 lb, and with nearly continuous traffic. The serviceability of each pavement section was measured from time to time and a plot of serviceability as a function of the total number of axle loads was made for each pavement section. The results were analyzed statistically to develop empirical relations between axle load, number of axles, pavement design, and performance. The tests effectively demonstrated that serviceability decreased with increasing load and numbers of loads, and decreasing pavement thickness. Curves showing these relationships were developed by assuming a mathematical form and by finding the best fit for the assumed curve by statistical methods.

The method of analysis employed in the AASHO studies does not take into consideration the mechanisms contributing to deterioration or the relative contribution of each. The effect of possible variable subgrade support is not considered. The effect of environment, particularly moisture variation, is also ignored in the primary analysis. Therefore, the AASHO test results cannot be directly applied to the design of Georgia Highways (1, p. 3). Instead, the AASHO data for the 18-kip axle loads (which are nearly equal to the present Georgia legal limit of 20 kips) were analyzed in the same manner as the Georgia data in this report.

### Deflection-Bearing Capacity-Traffic

The elastic deflection and bearing capacity of the AASHO subgrade were computed in the same way as for the Georgia subgrades. A single value was utilized for c, $\varphi$, and E in all segments, corresponding to the poorer soils tested (the composite on Table 2).

A plot of computed elastic deflection vs computed safety factor (Fig. 5) is well defined, as might be expected, because the only variable involved is pavement thickness. This does, however, suggest the validity of using a single index, either bearing capacity or deflection, as a basis for evaluating subgrade support. A comparison of Figure 5 with Figure 3 is of interest: the AASHO curve in Figure 3 approximately coincides with the lower limit of the Georgia data, suggesting that many of the Georgia soils are more elastic than those of the AASHO subgrade.

A plot of the safety factor of the AASHO pavements vs number of 18-kip axle loads required to reduce the serviceability to 1.5 is shown in Figure 6. A well-defined trend is evident, showing that as the safety factor increases, so does the number of axle loads required to reduce the serviceability to 1.5. Conversely, if a serviceability of 1.5 is demanded at the end of the service life of a pavement, the required safety factor must increase with increasing traffic.

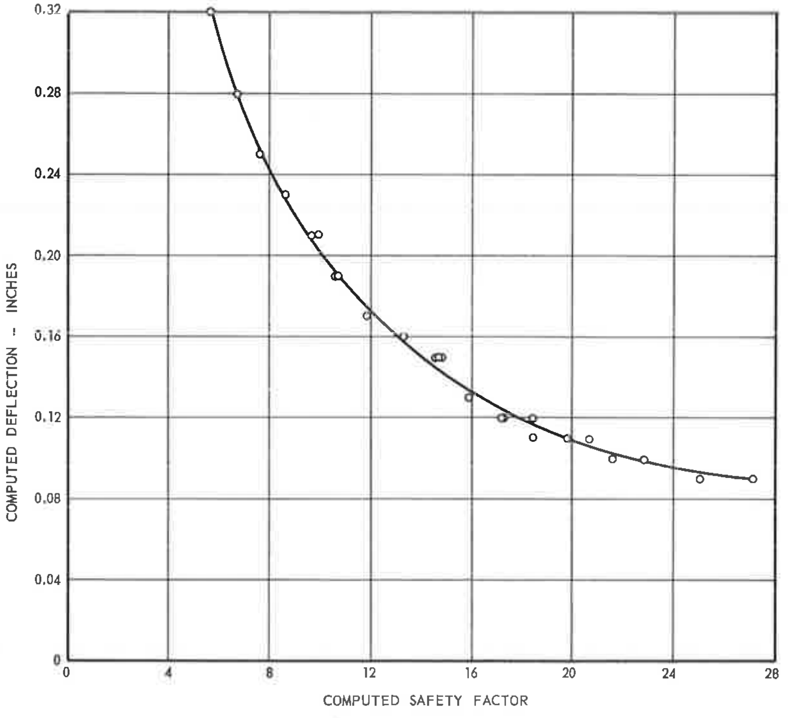



Figure 5. Computed elastic deflection vs computed safety factor of subgrades of AASHO Road Test and 18-kip axle loads.

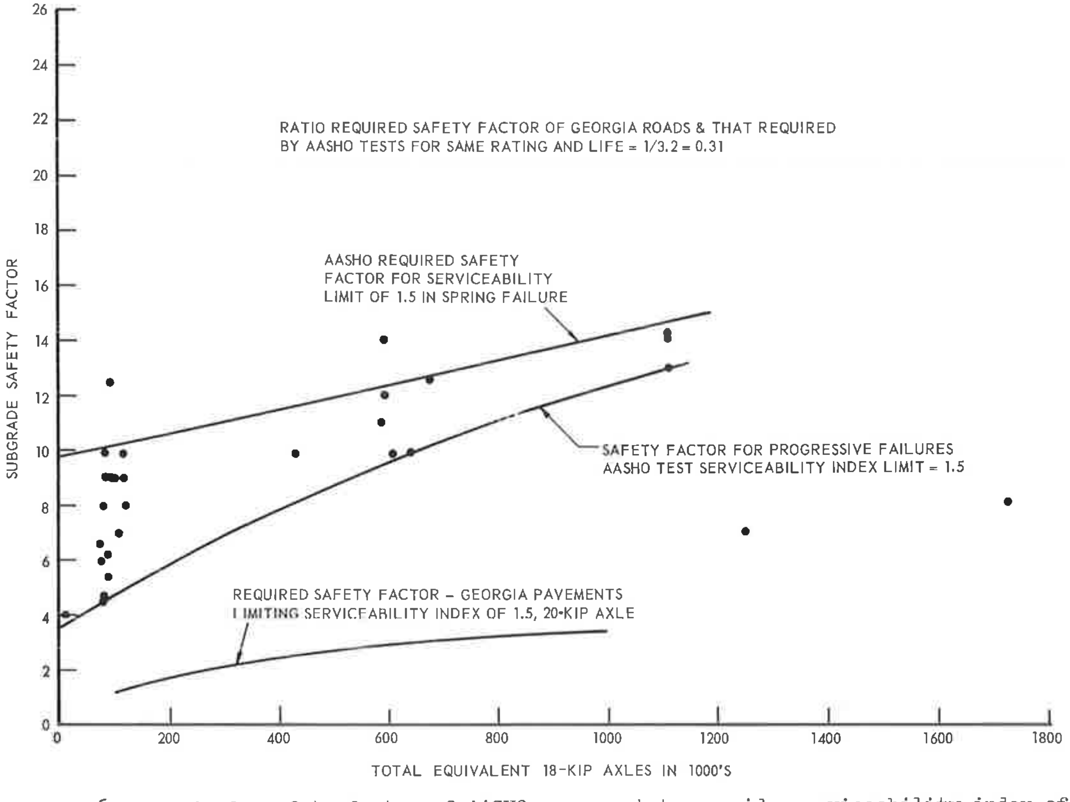



Figure 6. Required safety factor of AASHO pavement to provide serviceability index of 1.5 after different numbers of 18-kip axle loads.

The relationship exhibits considerable scatter, particularly at the lower numbers of axle loads. A study of the individual points shows that those exhibiting the higher safety factors failed suddenly in the spring of 1959 immediately after the thaw period. Two interpretations may be placed on this: (a) the failures were not the result of progressive failure or repeated load; or (b) the soil strength at this time was less than that indicated by the tests of samples made in the late spring of 1963. Those pavements which survived the spring breakup of 1959 exhibited a much better correlation between safety factor and traffic. Of these, however, those points above the lower line represent, for the most part, rather sudden failures corresponding to the spring breakup. The lower curve would seem to represent the more valid relationship between safety factor and traffic. If strength data were available for each test section for the period in which failure developed, the scatter would probably have been much less.

## PAVEMENT DESIGN

### AASHO Pavement Design

A tentative pavement design method was derived from the AASHO Road Test results by the AASHO committee on design and a draft was presented by Liddle (2). The basis for this development is outlined by Langsner, Huff, and Liddle (11).

The major Road Test correlation is pavement serviceability deterioration (from the initial constructed value) as a function of the pavement design, the axle load, and the number of axle loads. Thus, for a given initial serviceability and a desired serviceability level at the end of the pavement life, and for a required axle load and total traffic, the required pavement design can be found. The correlation is entirely empirical, based on curve fitting, and does not necessarily reflect any consideration of the mechanisms that contribute to failure. The correlation is valid only for the test road subgrade and only if the subgrade properties are uniform and unchanging. An attempt was made to include the variation of the subgrade with the season by assigning a greater weight to the number of load applications occurring during seasons of more rapid deterioration than to those during seasons of less rapid deterioration. The method of determining the factor (1) apparently was purely empirical; the weighting factors were adjusted until the serviceability and total load application data fit the assumed mathematical model with the least variation. This procedure does not indicate the mechanism by which the deterioration is accelerated. In fact, it applies the correction to the traffic rather than to the pavement support factors to which it more logically should apply. Therefore, although it may improve the fit of the AASHO data to an assumed mathematical curve, there is no reason to believe that it might be valid elsewhere.

The pavement design in the main load-performance-traffic-design relationship is expressed in terms of the structural number $\overline{SN}$ or equivalent thickness D; both definitions and symbols are used for the same thing, the first in Liddle's paper (2) and other design memoranda, and the second in the AASHO Report (1). This is related to the actual pavement components by

$$D = \overline{SN} = a_1D_1 + A_2D_2 + A_3D_3 \qquad (5)$$

where $D_1$, and $D_2$, and $D_3$ are the thicknesses in inches of the surface course, the base course and the subbase, respectively. The coefficients $a_1$, $a_2$, and $a_3$ are assumed to be indexes of the relative load-spreading or supporting qualities of each corresponding pavement course. The values found for the AASHO Road Test components varied with the traffic, load and the component thicknesses. The values for the 18-kip axle-load section are given in Table 3.

TABLE 3

VALUES OF COEFFICIENTS FOR 18-KIP AXLE-LOAD SECTION

| Course | Coeff. | Unweighted | Weighted |
|---|---|---|---|
| Asphaltic concrete surface | $a_1$ | 0.39 | 0.44 |
| Crushed stone base | $a_2$ | 0.15 | 0.14 |
| Sand, gravel subbase | $a_3$ | 0.12 | 0.11 |

TABLE 4

COMPARISON OF PAVEMENT COEFFICIENTS

| Course | Thickness (in.) | Stress in Course (psi) | | Stress Reduction in Layer | | |
|---|---|---|---|---|---|---|
| | | Top | Bottom | Total (psi) | Per In. (psi) | Comparative a[a] |
| Surface | 4 | 90 | 44 | 46 | 11.5 | 0.39 |
| Base | 3 | 44 | 27 | 17 | 5.7 | 0.19 |
| Subbase | 4 | 27 | 17 | 10 | 2.5 | 0.08 |

[a]Assuming surface = 0.39.

Although the AASHO Road Test report (1, p. 36) states that the weighted values indicate that an inch of surfacing ($a_1$ = 0.44) is about three times as effective as an inch of base ($a_2$ = 0.14) or four times as effective as an inch of subbase ($a_3$ = 0.11), this does not necessarily mean that these materials have support qualities or load-spreading abilities in the same ratio. For example, one design of the AASHO Loop 4, where the 18-kip axle load was employed, consisted of the layers shown in Table 4. If the vertical stresses are computed at the top and bottom of each layer using the Boussinesq equation (which applies to a semi-infinite homogeneous isotropic elastic mass), they will be seen to be less at the bottom of each successive course, as shown in the table. The stress reduction in each layer and stress reduction per inch of layer are also shown. The comparison indicates that the first layer is 2 times more effective than the second and 4.6 times more effective than the third. The effectiveness of each layer in terms of the pavement coefficient is tabulated. The resulting values are remarkably similar to the Road Test values for a. Therefore, it must be concluded that the relative values of $a_1$, $a_2$, and $a_3$ do not only reflect the load-spreading or supporting qualities of the pavement materials but also their relative position with respect to the pavement surface.

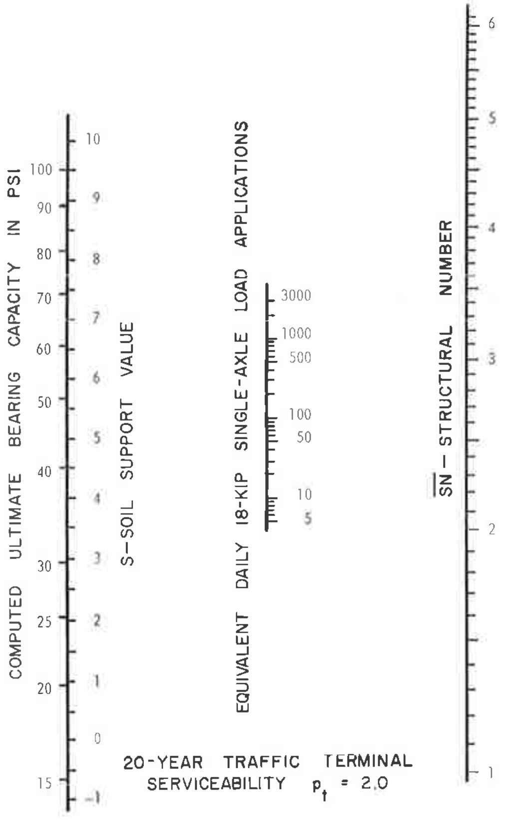



Figure 7. AASHO interim guide to design of flexible pavement (2) adapted to Georgia subgrades.

The Road Test correlation does not include any terms reflecting the subgrade soil support because it was assumed that this was constant and uniform. However, a possible "second" subgrade soil value was inferred from the pavement section with such thick crushed stone bases that the base, in effect, was the subgrade; however, this inference was not checked by any rational procedure. Arbitrary soil support values, S, were assigned to the subgrade and the thick stone base of 3 and 10, respectively. Of course, these values do not necessarily reflect relative support but instead are points of reference.

Nomographic design charts were constructed for the Road Test correlation utilizing an axle load of 18 kips and for terminal serviceabilities of 2 and 2.5. The former is reproduced in Figure 7, from the AASHO Interim Guide for the Design of Flexible Pavements. The same chart has been used by Liddle (2, 11).

Design Constants, Georgia Pavements

The use of the AASHO design charts requires calibration of the soil support scale in terms of some quantitative index or measure of the appropriate property of the subgrade soil for which the pavement is designed. Whereas the AASHO test results do not directly point to the mechanism of pavement deterioration and failure, clues are given by the results of the trenching program. Trenches were cut into pavement sections that had deteriorated to the point of removal from test in 1959. An extensive trench program was undertaken in 1960 when 39 pavement sections were investigated. In each, the transverse profile of the boundary between each of the pavement components and the densities of each layer in and beyond the wheelpath were obtained accurately. These tests indicated that about 25 percent of the thickness change of the pavement layers could be attributed to densification or consolidation of the layers. The remaining change, therefore, must be shear displacement. Such shear displacements can be clearly seen in the transverse profiles of the subgrade surface. Therefore, because it is shown that the major part of the subgrade's contribution to the deterioration of the pavement surface is shear failure, it appears reasonable to presume that the subgrade bearing capacity (its resistance to shear displacement) would be a valid index to the subgrade support of S. On this basis, the AASHO support value would represent an ultimate bearing capacity of 99 psi, based on tests of the samples secured in 1963 some time after the critical period of spring softening. This value probably does not represent the bearing capacity during the periods of most rapid deterioration. This is confirmed by the plot of safety factor vs traffic for the AASHO Road Test (Fig. 6). The lowest curve, which represents the more valid traffic-related deterioration, shows a safety factor of 3.6 required under conditions of very little traffic. The corresponding Georgia data gave a safety factor of about 1 for the same low traffic. Therefore, it is concluded that the actual bearing capacity of the AASHO subgrade was appreciably less than 99 psi during the critical periods of the Road Test.

A plot of the required Georgia subgrade safety factors for the same level of serviceability (1.5), based on Figure 4, is shown for comparison in Figure 6. The Georgia values everywhere are 1/3.2 or 31 percent of the indicated AASHO values. On this basis, the Georgia ultimate bearing capacity equivalent to the Road Test subgrade bearing capacity would be 0.31 × 99 or 31 psi. This value is recommended for use of pavement design in Georgia as the equivalent of the subgrade support value of 3 (2, 11).

Other values on the subgrade support value scale were established from this key bearing capacity utilizing the AASHO pavement thickness relation for an 18-kip axle load and 100 equivalent axle loads per day. The required safety factors for Georgia pavements for different amounts of traffic and different serviceabilities at the end of the pavement life were found from Figure 4 and plotted in Figure 8. One hundred axles per day for 20 years is a total of 730,000 axle loads. For a serviceability limit of 2.0, the required Georgia safety factor is 4. The safe limit of stress for the correct design would be $\frac{1}{4} \times 31$ or 7.75 psi. The total pavement thickness (3-in. surface plus soil-macadam base) required to maintain the stress at this level, from Figure 1, is 19.5 in. From the AASHO chart for a serviceability of 2, an S of 3 and 100 axles per day require a pavement $\overline{SN}$ of 3.58. The weighted average a for the Georgia pavement, therefore, must be 3.58/19.5 or 0.184.

A second point on the support scale can be found by utilizing a different assumed Georgia ultimate bearing capacity and the computed weighted average a for the Georgia pavement. For example, if the subgrade has an ultimate bearing capacity of 50 psi, the actual stress on the subgrade would be limited to 50/4 = 12.5 psi. This corresponds in Figure 1 to a total thickness of 13.7 in. Utilizing the previously computed weighted average a, the $\overline{SN}$ would be 13.7 × 0.184 = 2.52. From the AASHO chart, the soil support number corresponding to $\overline{SN}$ = 1.52 and 100 axle loads daily would be 5.7. Therefore, the S value corresponding to a bearing capacity of 50 psi would be 5.7.

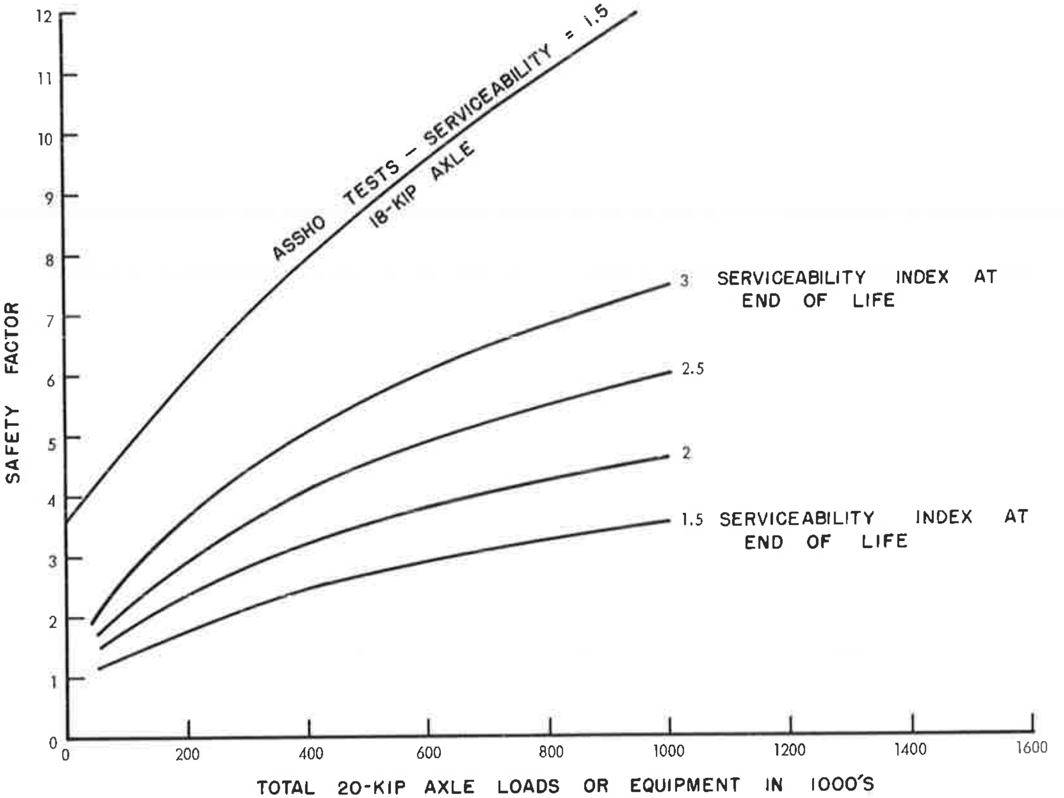



Figure 8. Required safety factors for Georgia subgrades for different numbers of 20-kip axle loads.

By this process the bearing capacities corresponding to various S values were found and are given in Table 5. These values apply to the 100 axle loads daily and the performance rating of 2.0 at the end of the pavement life. However, their applicability to other conditions is probably as valid as the other assumptions made in developing the design method.

In utilizing these values for design, consideration must be given to the test method on which the Georgia evaluations were based. The samples were secured in actual subgrades during the winter and the time of greatest soil moisture and lowest strength. Until data are available on the variations in soil moisture with the season, it is safe only to assume that the limit of capillary saturation is the limiting moisture corresponding to the Georgia test data. It is recommended that the bearing capacity be found from c and φ values determined as follows:

TABLE 5

RELATION OF S TO COMPUTED ULTIMATE BEARING CAPACITY[a]

| Bearing Capacity (psi) | S | Bearing Capacity (psi) | S |
|---|---|---|---|
| 15 | - 0.7 | 60 | 6.6 |
| 20 | + 0.9 | 70 | 7.4 |
| 30 | 2.9 | 80 | 8.2 |
| 40 | 4.5 | 90 | 8.9 |
| 50 | 5.7 | 100 | 9.5 |

[a]For using AASHO tentative design chart with subgrade bearing capacities computed from undrained triaxial shear tests of Georgia subgrade soils computed and inundated.

1. Compact two specimens of the subgrade to the lowest density and highest moisture permitted by the construction specifications.

2. Confine each in a triaxial chamber, one at a confinement of 10 psi and the other at 20 psi.

3. Subject each to a head of 1 ft of water from the bottom and allow to saturate until

no more water is absorbed (period of saturation to be found by experiment).

4. Load axially until failure occurs, without further change in moisture. For design utilize a seasonal weighting factor of 1 throughout, as in Figure 7.

Determination of the a coefficients for use in the AASHO design method is more difficult because there is little on which to base a correlation. The AASHO values are 0.44 for the asphaltic concrete surface and 0.14 for the crushed stone base. The weighted a for a typical Georgia pavement of 3 in. surface and 8 in. soil-bound macadam would be $3/11 \times 0.44 + 8/11 \times 0.14 = 0.22$. This compares reasonably well with 0.18, indirectly computed from the equivalent thicknesses utilizing the AASHO design chart as previously described.

Based on the subgrade stress studies of the author, the value of 0.44 for the surface appears large. Considering the stress-spreading value of the layers alone, a value of 0.35 is suggested for the asphaltic surface and 0.14 for the stone base. The weighted average of these is 0.197, which is closer to that of the value computed from the bearing study. These values have a ratio of 2.5 to 1, which is close to that found on the basis of the Boussinesq distribution to be applicable to a flexible pavement system employing a granular base.

The a value of the soil-macadam-cement base can be found indirectly from Figure 1. This shows that an 8-in. soil-macadam-cement base (5 psi on subgrade) is equivalent to 22 in. of soil-bound macadam, etc., or would have an a of $22/8 \times 0.14 = 0.38$. The 6-in. soil-macadam-cement base (15 psi) is as effective as 9 in. of soil-bound macadam; it has an a of $9/6 \times 0.14 = 0.21$. At first glance the different a values for the same material might appear contradictory. However, stress theory indicates that the load-spreading ability of a layer capable of supporting tension is a nonlinear function of the layer thickness as well as the material rigidity. For an all-over design value, an a of 0.25 to 0.30 for a soil-macadam-cement base would appear reasonable. This is not greatly different from the value of 0.23 estimated for the AASHO pavements.

The 8-in. thick sand-asphalt base stressed the subgrade to 23 psi which is equivalent to a 5.5-in. thick soil-bound macadam base. The equivalent a value for the sand-asphalt, therefore, would be $5.5/8 \times 0.14 = 0.10$. This is considerably less than the 0.25 estimated from the AASHO test results. (Of course, the AASHO tests did not include a sand-asphalt base; the value was only a guess.) The great difference between the AASHO a values and the values inferred from the Georgia stress tests is possibly the result of the higher Georgia temperatures and resulting lower rigidity, as well as the slower rate of loading.

## Alternate Georgia Design Method

An alternate design procedure can be evolved from the Georgia pavement evaluation data:

1. Determine the c and $\varphi$ of the soil as outlined previously.
2. Compute the ultimate bearing capacity, using an assumed tentative pavement thickness D.
3. Find the appropriate safety factor from Figure 8.
4. Compute the safe bearing capacity by dividing the ultimate bearing capacity, Step 3.
5. Find the total pavement thickness from Figure 1 utilizing the appropriate curve for the type of base course to be employed.

This procedure is no more complicated than that of the AASHO interim guide. It makes use of the AASHO serviceability concept and the traffic-serviceability decline principle. It is based on Georgia performance and on the stress spreading ability of the Georgia base courses. Finally, it is a more nearly rational approach to design than is the AASHO method.

## Recommendations for Further Study

Test sections of pavement should be constructed specifically for serviceability-performance studies. These should be a part of the highway system so as to reflect the use and traffic patterns of real highways. They should be placed on typical subgrades in each geologic region and should be constructed with varying pavement thicknesses and Georgia bases. They should be accompanied by a traffic count station where periodic Loadometer studies can be made to determine the distribution of the heavier axle loads. The soil moisture variation should be measured periodically, and the bearing capacity determined by laboratory tests of samples secured so as to reflect the typical range of moistures, particularly the highest. Pavement serviceability should be determined accurately by profile studies.

Subgrade moisture studies should be undertaken to define the range in moisture content changes for the typical subgrades in each geologic region and each different drainage regime.

Triaxial tests should be made on typical subgrade materials utilizing the procedure outlined in this paper for "saturating" the soils, or when more realistic subgrade moisture data become available, by making the tests at those moistures. The bearing capacities and deflections of these materials should be computed from the appropriate theories and correlated with the geology, soil classification, and region for use in preliminary design.

Finally more realistic theories should be developed for the bearing capacity and deflection of the subgrade and each of the pavement components.

## CONCLUSIONS

1. There are numerous causes or factors involved in the failure of a pavement to perform adequately. Of the Georgia pavements studies, however, most were load related.
2. The study of the performance of Georgia pavements disclosed a correlation among the serviceability rating or index, traffic, and computed deflection and bearing capacity of the subgrade.
3. The study of the AASHO data disclosed a similar correlation among these factors.
4. The AASHO subgrade had an ultimate bearing capacity of 99 psi. For comparable load, traffic and performance, Georgia roads required an ultimate bearing capacity of 31 psi. The difference is probably the result of differences in environment. The 31-psi required bearing capacity for Georgia subgrades corresponds to the Soil Support Number 3 of the AASHO design.
5. Georgia pavement thicknesses can also be designed by the use of triaxial tests on subgrade soils tested under field moisture conditions. The required safety factor against a bearing capacity (shear) failure is the required thickness to reduce the stress to that necessary to provide the safe bearing. This factor can be found from graphs of the test data and stress distribution below a pavement.

## ACKNOWLEDGMENTS

This work has been sponsored by the Georgia State Highway Department and the U. S. Bureau of Public Roads, through the Engineering Experiment Station of the Georgia Institute of Technology. The author is indebted to T. S. Wallace, T. K. Brasher and D. Wheeless, graduate research assistants; to Dr. A. Schwartz who directed the field studies of the pavements and the laboratory testing; to W. F. Abercrombie, Engineer of Materials and Tests and T. Moreland, Soil Engineer, the Georgia State Highway Department; and to the late W. E. Chastain, Sr., Engineer of Physical Research, Illinois Division of Highways.

## REFERENCES

1. The AASHO Road Test, Report 5: Pavement Research. Highway Research Board Spec. Rept. 61E, 1962.
2. Liddle, W. J. Application of AASHO Road Test Results to the Design of Flexible Pavements. Preprint Vol. Supple., Int. Conf. on the Structural Design of Asphalt Pavements, Univ. of Mich., Ann Arbor, 1962.

3. Carey, W. N., Jr., and Irick, P. E. The Pavement Serviceability-Performance Concept. Highway Research Board Bull. 250, pp. 40-58, 1960.
4. Sowers, G. F., and Vesic, A. B. Vertical Stresses in Subgrades Beneath Statically Loaded Flexible Pavements. Highway Research Board Bull. 342, pp. 90-123, 1962.
5. Terzaghi, K. Theoretical Soil Mechanics. New York, John Wiley and Sons, 1941.
6. Hveem, F. N. Types and Causes of Pavement Distress. Highway Research Board Bull. 187, pp. 1-52, 1958.
7. The AASHO Road Test, Report 2: Materials and Construction. Highway Research Board Spec. Rept. 61B, 1962.
8. Shook, J. F., and Fang, A. Y. Cooperative Materials Testing Program at the AASHO Road Test. Highway Research Board Spec. Rept. 66, 1961.
9. Sowers, G. F., and Sowers, G. B. Introductory Soil Mechanics and Foundations. New York, Macmillan Co., 2nd Ed. 1961.
10. Eiland, E. Alabama Highway Departments Use of the AASHO Data. Proc. 6th Alabama Joint Highway Conf., Auburn, April 1963.
11. Langsner, G., Huff, T. S., and Liddle, W. J. Use of Road Test Findings by the AASHO Design Committe. Highway Research Board Spec. Rept. 73, pp. 399-414, 1962.

# Low Modulus Pavement on Elastic Foundation

JON HAGSTROM, RICHARD E. CHAMBERS, and EGONS TONS
Respectively, Research Assistant, Senior Research Engineer, and Assistant Professor of Civil Engineering, Department of Civil Engineering, Massachusetts Institute of Technology

The Westergaard theory for a plate on a "heavy liquid foundation" has been applied to the analysis of stresses and deflections in the asphalt-bound layer(s) of a flexible pavement system. A computer program has been developed to give a solution for the symmetrical loading case. Plots of the bending stresses, shear stresses and vertical deflection vs pavement thickness have been made to determine the influence of the various parameters involved in the analysis. It was found that the bending and shear stresses depend on the ratio of the pavement stiffness to the support stiffness, whereas the deflection is determined by the stiffness ratio and the magnitude of the modulus of support reaction.

A limiting stiffness ratio of $E/k \geq 100$ was chosen as a criterion for plate-type behavior in the asphalt-bound layer. This limit, however, was not rigorously determined. Certain limitations as to thickness for given values of the other variables were discovered. Relatively thin asphalt-bound layers give little load spreading action because of their high flexibility. The load is transferred directly to the support, causing large vertical deformations in the support layer. The behavior of very thick asphalt-bound layers approaches that of a single homogeneous layer. Stresses are dependent only on load, and plate theory no longer applies. A computer program for stress-deflection calculations is applied.

•A TYPICAL flexible pavement cross-section may include the natural in-place subgrade, a compacted subgrade, a compacted subbase, a compacted base of treated or plain gravel or crushed rock, and a surface of one or more layers of asphaltic concrete. This vertical variation in the material composition of the highway structure, coupled with the complex nature of the behavior under load of the individual materials, has hindered the development of rational analysis for the stresses and displacements produced by traffic loads.

In 1943, Burmister (1) introduced a theory of stresses and displacements in layered systems based on the assumption that the materials of each layer are ideally elastic. Burmister's analysis provides an exact solution for a given surface loading. The equations are rather cumbersome to work with in practice; however, computer solutions have been developed for a large range of applications (5).

Attempts to correlate theoretical stresses and deflections with actual soil behavior by using Burmister's theory have met with little success. Sowers and Vesic (6) found that the reduction in subgrade stresses predicted by Burmister's theory occurs only when the stiffer top layers have the ability to develop tensile stresses. In general, investigators have concluded that soil properties cannot be accurately described in terms of a single Young's modulus and Poisson's ratio as is assumed in elastic theory.

Paper sponsored by Committee on Flexible Pavement Design and presented at the 44th Annual Meeting.

When the surface layer is fairly stiff in comparison with the support and the deflections are small, the tensile stresses produced in the support will be negligible. Assuming that these conditions are satisfied and that the behavior of the surface layer is consistent with elastic theory, Burmister's equations can be used for the analysis of stresses and deflections in the asphalt-bound layer.

By making certain simplifying assumptions concerning the behavior of the asphalt-bound and supporting layers, the equations for the stresses and displacements can be reduced to a manageable form. This simplified theory was first suggested by Westergaard (7) for use in the design of concrete pavements. It is the purpose of this paper to explore Westergaard's theory as it applies to the analysis of stresses and deflections in the surface layer of a flexible pavement system. The term "surface layer" used in this context is meant to include the asphalt-bound layer(s) only. The treatment in this report is mathematical in nature. It is intended as background material for future research which will include a treatment of asphaltic concrete properties for various time-temperature and loading conditions.

## ELASTIC APPROACH

The behavior of asphaltic concrete is not consistent with several of the assumptions of elastic theory. The relation between stress and strain is not, in general, linear; moreover, it is time dependent. However, under certain conditions a flexible pavement will exhibit nearly complete rebound on removal of load, the recovery occurring over some time interval. The amount of deflection under load and the time for recovery on removing the load will depend on the temperature of the pavement and the duration of the load. Baker and Papazian (10) have pointed out that the effect of choosing a particular temperature and time of loading to determine the elastic modulus of the asphalt-bound material is equivalent to selecting a secant modulus. The complexity of the mathematics for rigorous treatment of the viscoelastic theory makes the use of this secant modulus artifice and elastic theory necessary at this time.

## APPLICATION OF WESTERGAARD THEORY

### Development

Westergaard treats the asphalt-bound layer as a circular plate of infinite extent. The flexural rigidity of the plate can be expressed as:

$$D = \frac{EI}{1 - \mu^2} \tag{1}$$

where

$E$ = modulus of elasticity for plate (psi),
$I$ = moment of inertia for plate (in.$^4$), and
$\mu$ = Poisson's ratio.

For a plate of constant thickness, this equation becomes:

$$D = \frac{Eh^3}{12\,(1 - \mu^2)} \tag{2}$$

where h is the thickness of plate in inches. The deflection of a pavement depends not only on its flexural rigidity but also on the stiffness of the support beneath. Westergaard has expressed the stiffness of the support as a modulus of support reaction, k. The modulus is defined mathematically as follows:

$$k = \frac{p}{w} = \frac{\text{reaction of support}}{\text{deflection of support surface}} \tag{3}$$

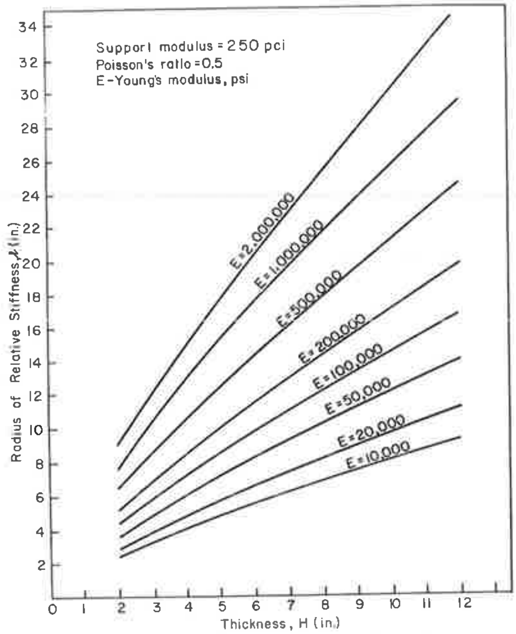



Figure 1. Radius of relative stiffness vs thickness of asphalt-bound layer.

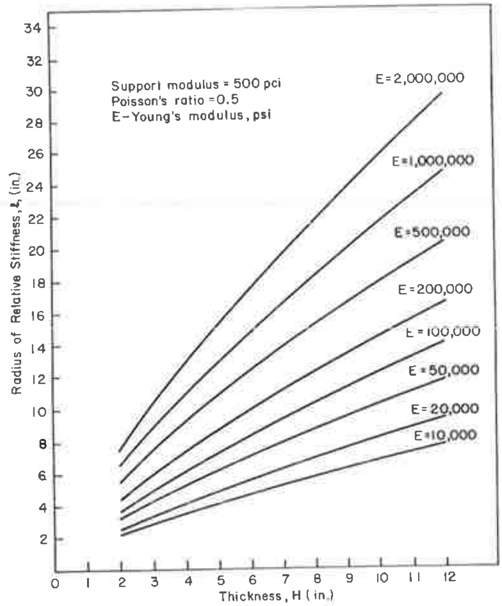



Figure 2. Radius of relative stiffness vs thickness of asphalt-bound layer.

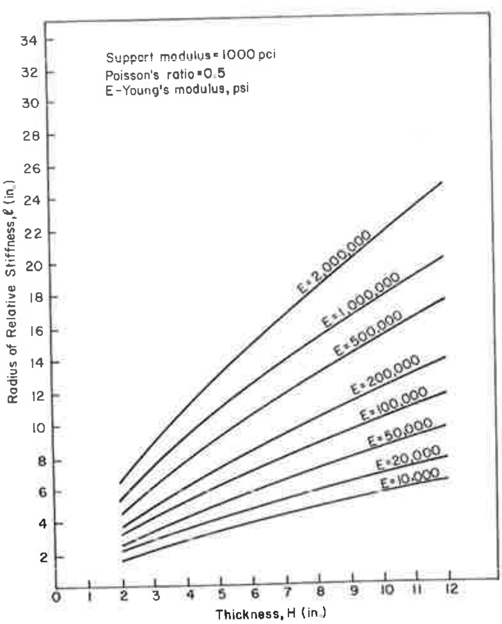



Figure 3. Radius of relative stiffness vs thickness of asphalt-bound layer.

To facilitate the mathematical treatment, Westergaard introduced the term "radius of relative stiffness," denoted by the symbol $\ell$ and defined mathematically as:

$$\ell^4 = \frac{D}{k} \tag{4}$$

For a pavement of constant thickness this equation becomes:

$$\ell^4 = \frac{Eh^3}{12\,(1 - \mu^2)k} \tag{5}$$

Plots of radii of relative thickness vs thicknesses of asphalt-bound layers for varying support moduli are shown in Figures 1 through 3.

To apply this mathematical treatment in an analysis of the forces and deflections due to an interior load on a layered pavement system, the following assumptions must be made:

1. A pavement loaded some distance from the edge can be represented by a circular plate of infinite extent.
2. The reactions of the support are vertical only, are proportional to the deflection of the asphalt-bound layer, and

are independent of the loaded area. The latter portion of this assumption is equivalent to assuming that there is no transfer of stress through shear resistance in the support, analogous to the behavior of a "heavy liquid."

3. The asphalt-bound layer is at every point in contact with the support.

4. Volumetric changes, variations in support properties, temperature, horizontal components of support reaction, and dynamic effects can be neglected or accounted for otherwise.

## Support

Because the support, consisting of particulate materials, is neither a perfectly elastic solid nor a heavy liquid, it is necessary to evaluate the support modulus, k, by some sort of plate bearing test. This necessitates the adoption of some standard plate size and testing procedure. Work by the U. S. Bureau of Public Roads (11) indicates that for circular plates with diameters between 26 and 84 in., there is little variation in the pressure required to produce a given deflection. Furthermore, Terzaghi (12) has indicated that a linear relationship between deflection and pressure does exist for some soils up to one-half the bearing capacity of the soil. This justifies the assumption of linearity between reaction and deflection, and negligible transfer of stress by shear resistance in the soil within the limits indicated.

## Asphalt-Bound Layer

Westergaard represents the asphalt-bound layer as a circular plate of infinite extent. He assumes that the properties of the asphalt-bound material are consistent with the assumptions of elastic theory (13); that is, the material must be continuous, homogeneous, isotropic, and governed by Hooke's law.

By requiring equilibrium of forces and moments in the three directions of space, a system of six equations with six unknowns can be derived. These six equations can be reduced to a single sixth-order differential equation describing the behavior of an ideally elastic body under load. For a body whose dimension in the vertical direction is much smaller than in the two other directions, the governing sixth-order differential equation can be reduced to the familiar fourth-order differential equation of ordinary plate theory by making several simplifying assumptions:

1. The middle plane of the plate remains unstrained under load.
2. Plane sections remain plane under load.
3. The direct stress in the vertical direction is small in comparison with the other stress components and can be neglected in the stress-strain relations.
4. A plane section normal to the middle plane before loading remains normal under load.

The last assumption is equivalent to stating that the deformation due to vertical shearing stresses is very small and can be neglected.

If the behavior of the asphalt-bound layer under load is to be consistent with plate theory, the asphalt-bound layer must be relatively rigid in comparison with the support layer. However, it is difficult to define the critical E/k value because of the many assumptions involved.

A critical ratio of E/k = 100 is proposed in this paper with the following reasoning:

1. Plate theory requires that the thickness of the plate be small in comparison with the dimensions in the other two directions. For circular plates, it is commonly required that the thickness-radius ratio be greater than 10 (26).
2. The radius of influence of a concentrated load is approximately $4\ell$, where $\ell$ depends on the E/k ratio and the asphalt-bound layer thickness (14).
3. For values of E/k less than 100, the radius of influence of a concentrated load is at best seven times greater than the asphalt-bound layer thickness.

Because the critical ratio has not been rigorously determined, it should be used only as a general guide. However, it is important to realize that the Westergaard theory becomes increasingly inaccurate as the stiffness of the asphalt-bound layer approaches that of the support. Thus, the application of the theory is limited to conditions which lead to small deflections, the vertical stress at the pavement support interface should not exceed one-half the ultimate bearing stress of the support, and the E/k ratio should be 100 or greater.

Development of Equations

The fourth-order differential equation describing the behavior of a plate can be expressed as follows:

$$\nabla^2 (\nabla^2 w) = \frac{p}{D} \tag{6}$$

where w represents the vertical deflection of the plate, p represents the vertical load, D represents the flexural rigidity of the plate, and $\nabla^2$ represents the Laplace operator which is defined in Cartesian coordinates as $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. For a plate on a "heavy liquid" foundation where the reaction of the subgrade is vertical only and is proportional to the deflection, w, the differential equation becomes:

$$\nabla^2 (\nabla^2 w) = \frac{p - kw}{D} \tag{7}$$

By representing a tire load on a pavement as a circular load of uniform intensity, the intensity being equal to the tire pressure, Putnam (14) has derived equations for the bending moments in the asphalt-bound layer and for the deflection of the surface of the pavement (Appendix B). Because of the symmetry of the assumed load representation, the moments and deflections are expressed in terms of Bessel functions. Putnam's equations for deflection and bending moment under the center of the circular load are as follows:

$$w_o = \frac{P\ell^2}{\pi D c^2} \left[ 1 + c\,\mathrm{ker}'c \right] \text{ in.} \tag{8}$$

and

$$m_{max} = \frac{P(1 + \mu)}{2\pi} \left[ \frac{1}{c} \mathrm{kei}'c \right] \text{ in.-lb/in.} \tag{9}$$

where

- $P$ = total load (lb);
- $\mu$ = Poisson's ratio;
- $\ell$ = radius of relative stiffness (in.);
- $D$ = flexural rigidity of the plate (lb-in.$^2$);
- $c$ = radius of relative load distribution, defined as radius of applied circular load/$\ell$;
- $\mathrm{ker}'x$ = Bessel function with real argument; and
- $\mathrm{kei}'x$ = Bessel function with real argument.

Using tabulated values for the Bessel functions, these equations provide a relatively simple means of analysis for deflection and maximum bending stresses. These equations can also be adapted to computer analysis and the stresses and deflections can be computed for a range of asphalt-bound layer thicknesses and moduli, and moduli of support reaction.

## Moment and Deflection Curves

Figures 4 through 9 show plots of the maximum bending stress (maximum moment per unit length/section modulus per unit length) and the maximum deflection under the center of a 10-kip single-axle wheel load with a 70-psi tire contact pressure vs asphalt-bound layer thickness. The ordinate values have been divided by the total load because the magnitude of the stresses and deflections is a function of the ratio of total load to tire contact pressure (i.e., loaded area) and is not uniquely determined by the magnitude of the load.

The graphs indicate that the bending stresses depend on the ratio of the asphalt-bound layer stiffness to the modulus of support reaction, D/k, and are independent of the magnitude of these quantities. As the asphalt-bound layer decreases, it loses flexural rigidity, and at some point a decrease in the maximum bending stress under the center of the load occurs, as is indicated by the peaking in the curves of Figures 4, 5 and 6. In this range of asphalt-bound layer thickness, the radius of influence of the load is small, on the order of twice the radius of the loaded area. (For a concentrated load, the radius of influence is approximately $4\ell$.) Thus, the load is transferred directly into the support with little spreading by plate action.

For relatively thick asphalt-bound layers, Figures 4, 5 and 6 indicate that the bend-stress tends to become less dependent on the Young's modulus of the asphalt-bound layer and the support modulus. The pavement behavior approaches that of a Boussinesq half space and plate theory breaks down. An increase in asphalt-bound layer thickness beyond 10 to 11 affects only slightly the bending stresses of the structure. It is interesting to note that this 10- to 11-in. boundary for diminishing returns coincides with the optimum pavement thickness determined for a similar range of variables by McLeod (15) with his Burmister-type layered analysis.

The deflection plots indicate that the deflection depends not only on the ratio of plate rigidity to support stiffness but also is greatly affected by the modulus of support reaction, k. Using the deflection curves, it is possible to determine the magnitude of the vertical stress in the soil at the interface. The vertical stress is equal to the deflection multiplied by the modulus of subgrade reaction: p = kw.

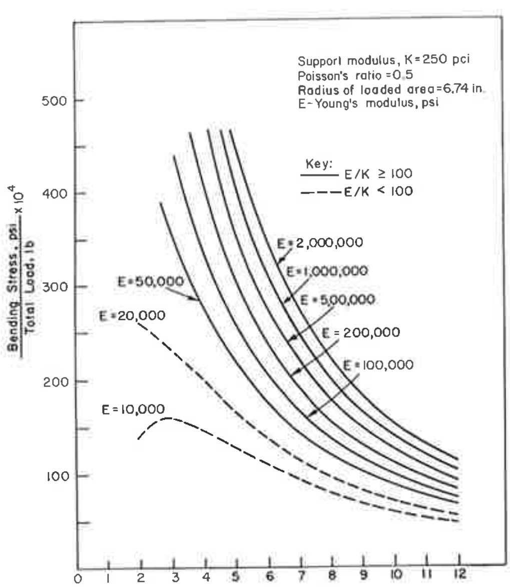



Figure 4. Bending stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

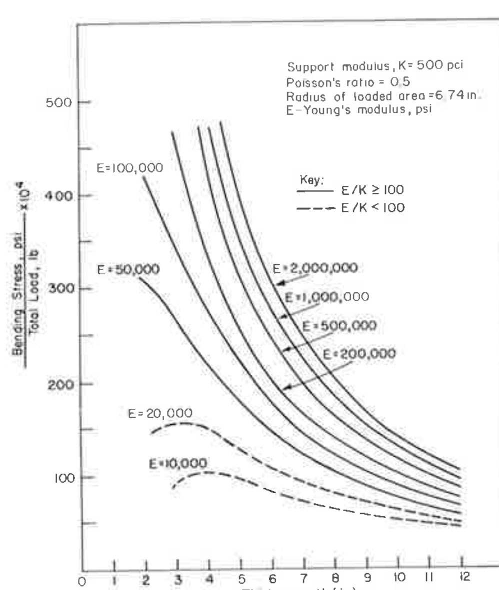



Figure 5. Bending stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

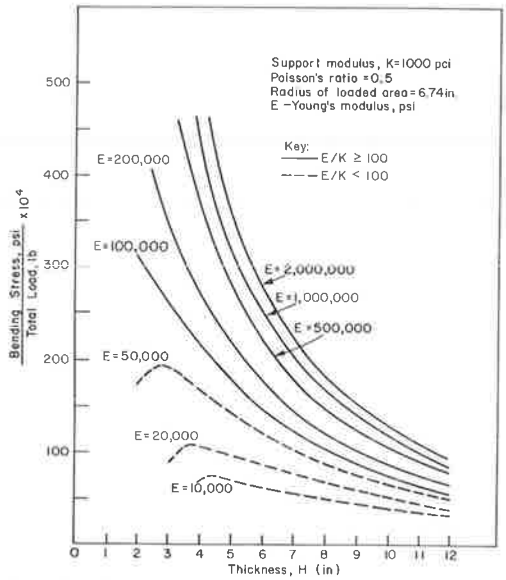



Figure 6. Bending stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

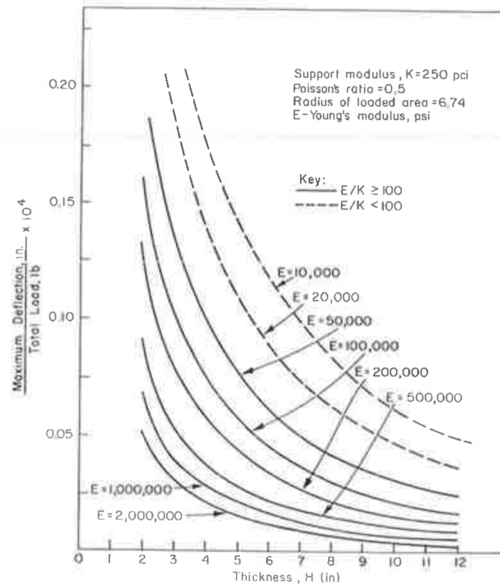



Figure 7. Maximum deflection vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

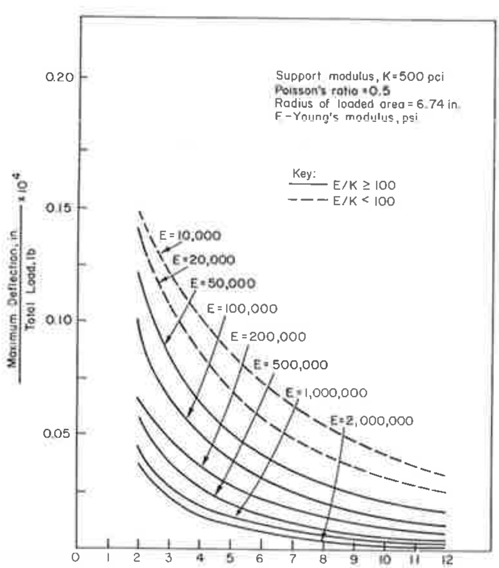



Figure 8. Maximum deflection vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

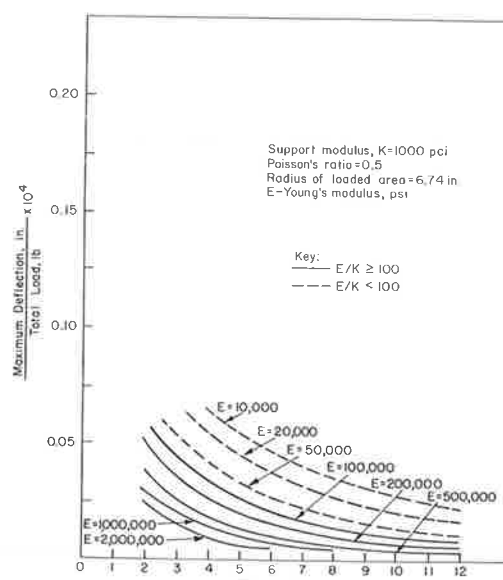



Figure 9. Maximum deflection vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

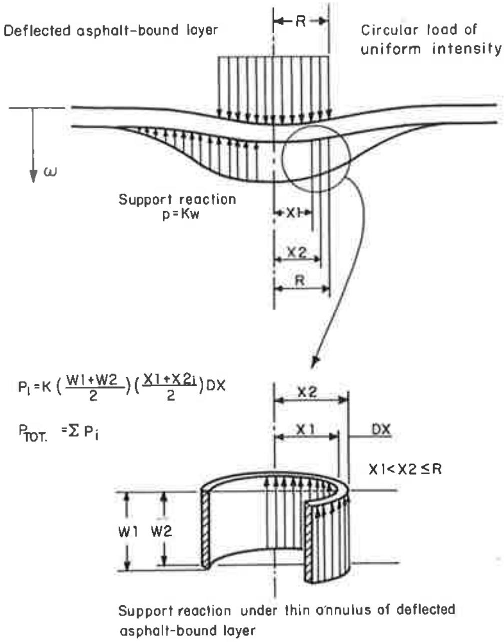



Figure 10. Computation of total support reaction within distance, R, from center of loaded area.

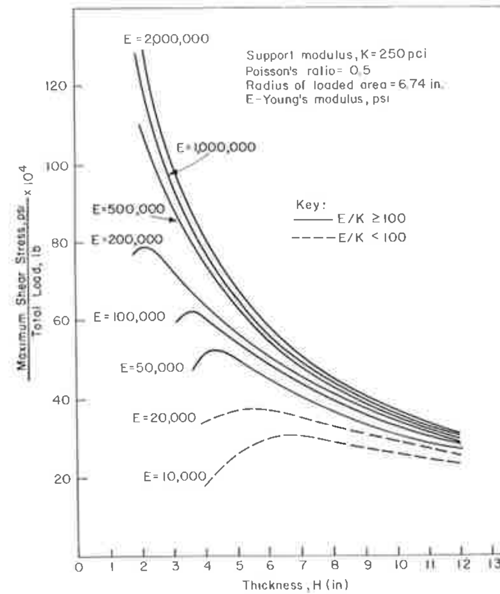



Figure 11. Maximum shear stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

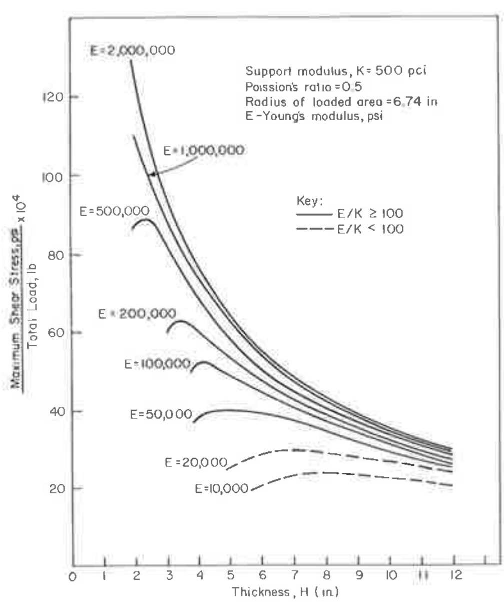



Figure 12. Maximum shear stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

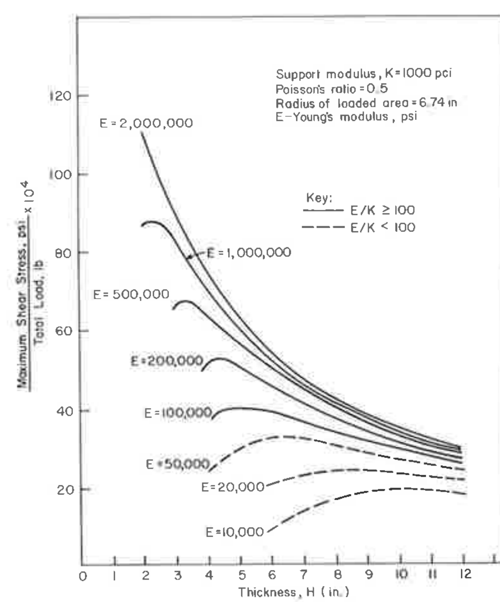



Figure 13. Maximum shear stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

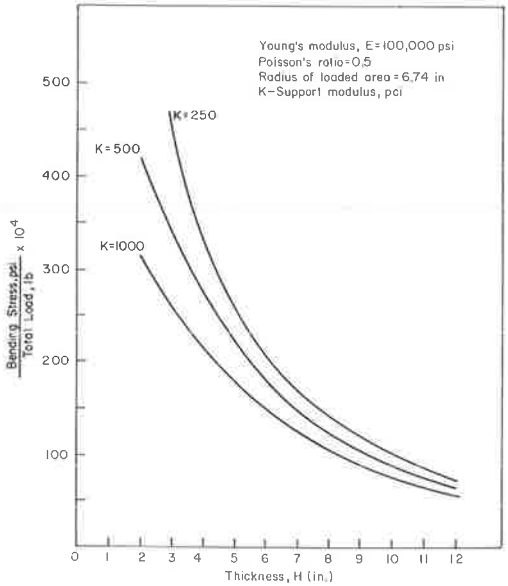



Figure 14. Bending stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

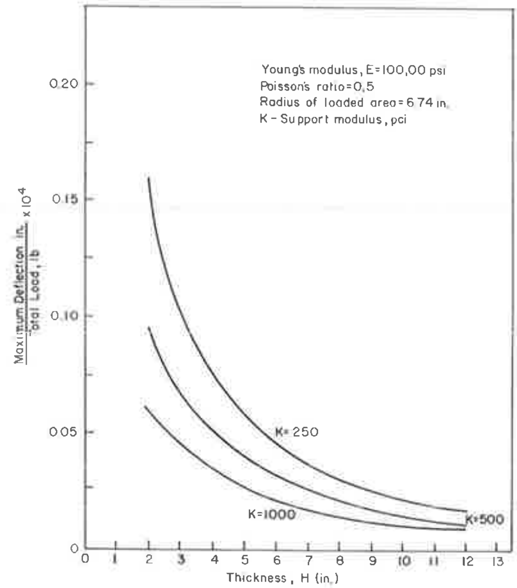



Figure 15. Maximum deflection vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

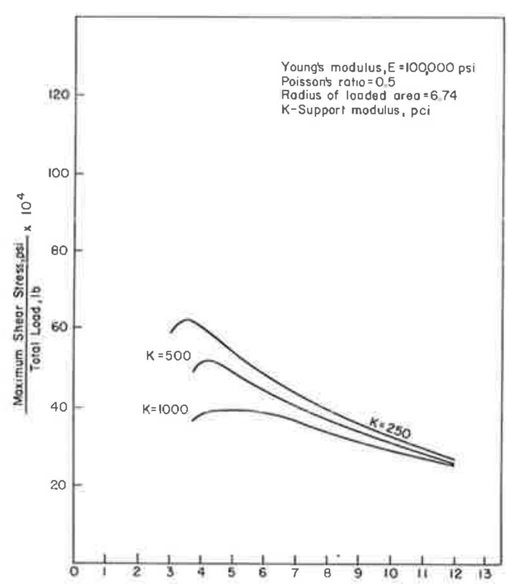



Figure 16. Maximum shear stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

## Vertical Shear

Because of the assumptions made in reducing the general sixth-order differential equation of elastic theory to the fourth-order equation of ordinary plate theory, it is impossible to solve directly for the vertical shear stresses in the asphalt-bound layer. However, the shear force acting over the thickness of the asphalt-bound layer at a point can be determined by evaluating the slope of the moment curve at that point. By assuming a distribution of the shear force over the thickness, a value for the maximum shear stress can be determined.

The shear force at any distance from the center of the applied load can also be determined by computing the magnitude of the total support reaction within the radius of the point being considered, subtracting this reaction from the magnitude of that portion of the applied load which lies within the radius of the point being considered and assuming that the difference in these forces is carried by a shear force distributed uniformly around the perimeter of the cylinder with the radius being considered. The magnitude of the support reaction can be computed by dividing the circular area within the radius being considered into many thin annuli, evaluating the average reaction for each annulus, multiplying the average reaction by the area of the annulus and summing the reactions for each annulus. This procedure is indicated schematically in Figure 10.

Figures 11, 12, and 13 are plots of the maximum shear stress at the edge of the applied load for a 10,000-lb wheel load with 70-psi tire contact pressure. In determining the shear stress, it was assumed that the shear force is distributed parabolically over the asphalt-bound layer thickness. As was the case for the bending stresses, the shear curves exhibit a peak value for relatively thin asphalt-bound layers and the shear stress tends to become independent of E and k for thicknesses greater than the 10- to 11-in. range.

Figures 14, 15, and 16 illustrate an alternative way of plotting the bending stresses, deflections and shear stresses. The Young's modulus value of the asphalt-bound layer is held constant and the ordinates are plotted vs thickness for various values of the support modulus.

## CONCLUSIONS

1. When the asphalt-bound layer is fairly rigid in comparison with the support ($E/k \geq 100$), an applied load is spread over the support by the plate action in the asphalt-bound layer. If, in addition, the deflection under load is small, the Westergaard theory will provide a useful analysis.
2. By representing the wheel load as a circular load of uniform intensity, the bending stress under the center of the load can be computed for a given E/k ratio in terms of Bessel functions.
3. Given the E/k ratio and the value of the support modulus, k, the pavement deflection and the vertical stress at the asphalt-bound layer-support interface can be computed.
4. By assuming a parabolic distribution for the shear force over the asphalt-bound layer thickness, it is possible to determine the maximum vertical shear stress acting under the edge of the loaded area.

## REFERENCES


1. Burmister, D. M. The Theory of Stresses and Displacements in Layered Systems and Applications to the Design of Airport Runways. Highway Research Board Proc., Vol. 23, pp. 126-148, 1943.
2. Burmister, D. M. The General Theory of Stresses and Displacements in Layered Systems. Jour. of Applied Physics, Vol. 16, Nos. 2, 3 and 5, 1945.
3. Burmister, D. M. Stress and Displacement Characteristics of a Two-Layer Rigid Base Soil System: Influence Diagrams and Practical Applications. Highway Research Board Proc., Vol. 35, pp. 773-814, 1956.
4. Burmister, D. M. Applications of Layered System Concepts and Principles to Interpretations and Evaluations of Asphalt Pavement Performances and to Design and Construction. Internat. Conf. on the Structural Design of Asphalt Pavements, Univ. of Mich., 1962.
5. Mehta, M. R., and Veletoses, A. S. Stresses and Displacements in Layered Systems. Univ. of Illinois, Urbana, June 1959.
6. Sowers, G. F., and Vesic, A. B. Vertical Stresses in Subgrades Beneath Statically Loaded Flexible Pavements. Highway Research Board Bull. 342, pp. 90-123, 1962.
7. Westergaard, H. M. Computation of Stresses in Concrete Roads. Highway Research Board Proc., Vol. 5, Part 1, pp. 90-112, 1925.
8. Westergaard, H. M. Stresses in Concrete Pavements Computed by Theoretical Analysis. Public Roads, Vol. 7, No. 2, April 1926.
9. Westergaard, H. M. Analytical Tools for Judging Results of Structural Tests of Concrete Pavements. Public Roads, Vol. 14, No. 10, Dec. 1933.
10. Baker, R. F., and Papazian, H. S. The Effect of Stiffness Ratio on Pavement Stress Analysis. Highway Research Board Proc., Vol. 39, pp. 61-85, 1960.
11. Teller, L. W., and Sutherland, E. C. The Structural Design of Concrete Pavements. V. An Experimental Study of the Westergaard Analysis of Stress Conditions in Concrete Pavement Slabs of Uniform Thickness. Public Roads, Vol. 23, No. 8, Oct. 1942.

12. Terzaghi, K. Evaluation of Coefficients of Subgrade Reaction. Geotechnique, Vol. 5, Dec. 1955.
13. Wang, C. T. Applied Elasticity. New York, McGraw-Hill, 1953.
14. Putnam, J. O. Theoretical Analysis of the Behavior of Rigid Airfield Pavements for the Case of Interior Loading. unpublished Doctoral Thesis, MIT, 1963.
15. McLeod, Norman W. Some Notes on Pavement Structural Design. Highway Research Record No. 13, pp. 66-141, 1963.
16. Yoder, E. J. Principles of Pavement Design. New York, John Wiley and Sons, 1959.
17. Taylor, D. W. Fundamentals of Soil Mechanics. New York, John Wiley and Sons, 1948.
18. Hogg, A. H. A. Equilibrium of a Thin Plate, Symmetrically Loaded, Resting on an Elastic Foundation of Infinite Depth. Philosophical Magazine, Vol. 25, 1938.
19. McLachlan, N. W. Bessel Functions for Engineers. 2nd Ed., Oxford, Clarendon Press, 1955.
20. Timoshenko, S., and Goodier, J. N. Theory of Elasticity. New York, McGraw-Hill, 1951.
21. Timoshenko, S., and Woinowsky-Krieger, S. Theory of Plates and Shells. 2nd Ed., New York, McGraw-Hill, 1959.
22. Investigations of Pressures and Deflections for Flexible Pavements. U. S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Rept. No. 1, Homogeneous Clayey Silt Test Section, Tech. Memo No. 3.323, March 1951.
23. Investigations of Pressures and Deflections for Flexible Pavements. U. S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Rept. No. 4, Homogeneous Sand Test Section, Tech. Memo No. 3.323, Dec. 1954.
24. Acum, W. E. A., and Fox, L. Stresses in a Three-Layer System. Great Britain, Bond Research Note No. R. H./1257/WEAA-LF, Nov. 1949.
25. McMahon, T. F., and Yoder, E. J. Design of a Pressure-Sensitive Cell and Model Studies of Pressures in a Flexible Pavement Subgrade. Highway Research Board Proc., Vol. 39, pp. 650-682, 1960.
26. Seely, F. B., and Smith, J. O. Advanced Mechanics of Materials. New York, John Wiley and Sons, 1957.
27. Odemark, N. Investigation as to the Elastic Properties of Soils and Design of Pavement According to the Theory of Elasticity. Statens Vaginstitut, Stockholm, Maddelande No. 77, 1949.

# *Appendix A*

## PROGRAM DESCRIPTION

The bending stresses, deflections, and shear stresses plotted in Figures 4 through 9 and 11 through 16 were evaluated using a FORTRAN source program developed from Putnam's equations. The input required by the program includes the design wheel load, the design tire pressure, Young's modulus for the pavement material, Poisson's ratio for the pavement material, and the support modulus. As output, the program lists the bending stress divided by total load, deflection divided by total load, maximum shear stress divided by total load and the ratio of the radius of the loaded area to the radius of relative stiffness of the pavement. These quantities are computed for values of pavement thicknesses ranging from 2 to 12 in. in 2-in. intervals. The input and output formats are consistent with any FORTRAN II processor and can be readily altered for use with any FORTRAN processor. The output format contains Hollerith fields, or H-type alphanumeric data fields, which identify the quantities computed. Figure 17 shows the actual output for a typical analysis.

The reasoning used in the development of the program can be seen in the flow diagram shown in Figure 18.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| H= | 2.00 | STRESS= | .0663 | DEF= | .00001637 | SHEAR= | .00475281 | C= | 1.5529 |
| H= | 4.00 | STRESS= | .0345 | DEF= | .00000750 | SHEAR= | .00614423 | C= | .9233 |
| H= | 6.00 | STRESS= | .0207 | DEF= | .00000443 | SHEAR= | .00487677 | C= | .6812 |
| H= | 8.00 | STRESS= | .0139 | DEF= | .00000300 | SHEAR= | .00391795 | C= | .5490 |
| H= | 10.00 | STRESS= | .0100 | DEF= | .00000219 | SHEAR= | .00324716 | C= | .4644 |
| H= | 12.00 | STRESS= | .0076 | DEF= | .00000169 | SHEAR= | .00276339 | C= | .4050 |

E = 100,000
K = 250
Poisson's ratio = 0.5

Figure 17. Typical program output.

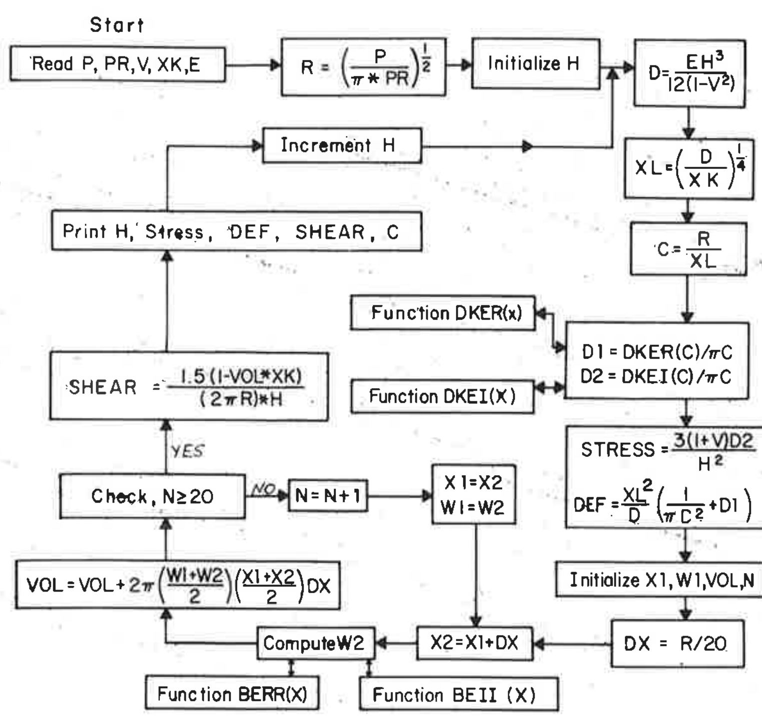



Figure 18. Flow diagram for computer program.

The definitions of the symbols used in the flow chart are as follows:

BEII (X) - bei (x), Bessel function with real argument;
BERR (X) - ber (x), Bessel function with real argument;
C - ratio of relative load distribution, R/XL;
D - flexural rigidity of pavement (lb-in.$^2$/in.);
D1 - function of C;
D2 - function of C;
DEF - deflection under center of load area, divided by the total load (in./lb);
DKEI (Z) - kie'(z), Bessel function with real argument;

DKER (Z) - ker'(z), Bessel function with real argument;

DX - width of annuli used in calculating reaction under loaded area (in.);

E - Young's modulus for pavement material (psi);

H - pavement thickness (in.);

N - number of annuli used to compute subgrade reaction under loaded area;

P - total load applied at surface (lb);

PR - tire contact pressure for applied load (psi);

R - radius of applied load (in.);

SHEAR - maximum vertical shear stress at edge of applied load divided by total load (psi/lb);

STRESS - maximum bending stress under center of applied load divided by total load (psi/lb);

V - Poisson's ratio;

VOL - volume of support reaction diagram divided by modulus of subgrade (cu in.);

W1 - deflection at inner radius of annulus (in.);

W2 - deflection at outer radius of annulus (in.);

XK - modulus of support reaction (pci);

XL - radius of relative load distribution;

X1 - inner radius of annulus; and

X2 - outer radius of annulus.

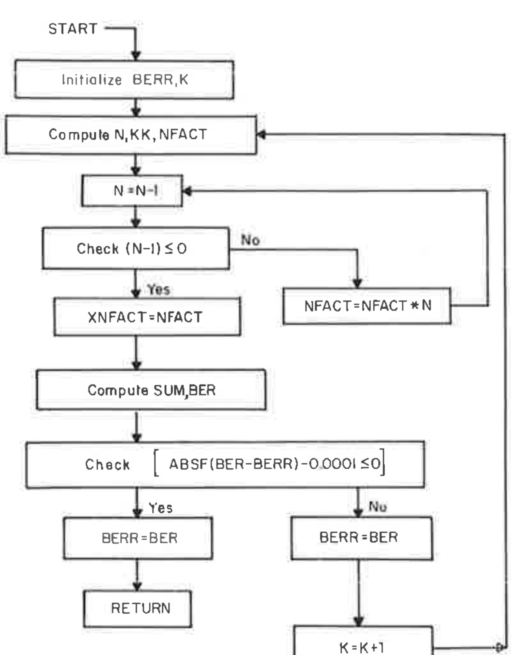



Figure 19. Flow diagram for computer subprogram, FUNCTION BERR (X).

The required Bessel functions are computed using Function Subprograms, otherwise known as FORTRAN functions. The definition for FUNCTION BERR (X) is as follows:

BERR (X) = ber(x) =

$$\sum_{k=0}^{\infty} (-1)^K \frac{\left(\frac{X}{2}\right)^{4K}}{\left[(2K)!\right]^2} \qquad (10)$$

Figure 19 shows the flow diagram for this function. BEII (X) is defined as follows:

BEII (X) = bei(x) =

$$\sum_{k=0}^{\infty} (-1)^K \frac{\left(\frac{X}{2}\right)^{4K+2}}{\left[(2K+1)!\right]^2} \qquad (11)$$

```
C       WESTERGAARD ANALYSIS
      1 FORMAT (5F12.2)
      2 FORMAT (3H H=F5.2,4X,7HSTRESS=F8.4,4X,4HDEF=F12.8,4X,
       16HSHEAR=F12.8,4X,2HC=F8.4)
      3 READ 1, P, PR, V, XK, E
        R=(P/(3.1416*PR))**.5
        DO 7 IH=2,12,2
        H=IH
        D=(E*(H**3))/(12.*(1.-V*V))
        XL=(D/XK)**0.25
        C=R/XL
        D1=DKER(C)/(3.1416*C)
        D2=DKEI(C)/(3.1416*C)
        STRESS=(D2*3.*(1.+V))/(H**2)
        DEF=(XL**2)*((1./(3.1416*(C**2)))+D1)/D
        X1=0.0
        W1=DEF
        DX=R/20.
        VOL=0.0
        N=1
      4 X2=X1+DX
        W2=((XL**2)/D)*((1./(3.1416*(C**2)))+(D1*BERR(X2/XL))
       1+(D2*BEII(X2/XL)))
        VOL=VOL+(1.5708*(W1+W2)*(X1+X2))*DX
        IF(20-N)6,6,5
      5 X1=X2
        W1=W2
        N=N+1
        GO TO 4
      6 SHEAR=(1.5*(1.-(VOL*XK)))/(6.2832*R*H)
      7 PRINT 2, H, STRESS, DEF, SHEAR, C
        GO TO 3
        END


        FUNCTION DKEI(Z)
        A=0.1159-LOGF(Z)
        X=Z/2.
        DKEI      =(A+.5)*X+3.1417*(X**3)/8.-(A+5./3.)*(X**5)/12.
       1-3.1417*(X**7)/576.
        RETURN
        END


        FUNCTION DKER(Z)
        A=0.1159-LOGF(Z)
        X=Z/2.
        DKER      =-.5/X+(3.1417*X/4.)-(.5*(A+1.25)*(X**3))
       1-(3.1417*(X**5)/48.)+((A+(47./24.))*(X**7)/144.)
        RETURN
        END
```

Figure 20.

```
      FUNCTION BEII(X)
      BEII=(.5*X)**2
      K=1
    6 N=2*K+1
      KK=4*K+2
      NFACT=N
    3 N=N-1
      IF(N-1)1,1,2
    2 NFACT=NFACT*N
      GO TO 3
    1 XNFACT=NFACT
      SUM=((-1.)**K)*((.5*X)**KK)/(XNFACT**2)
      BEI=BEII+SUM
      IF(ABSF(BEI-BEII)-.0001)4,4,5
    5 BEII=BEI
      K=K+1
      GO TO 6
    4 BEII=BEI
      RETURN
      END
```

```
      FUNCTION BERR(X)
      BERR=1.
      K=1
    6 N=2*K
      KK=4*K
      NFACT=N
    3 N=N-1
      IF(N-1)1,1,2
    2 NFACT=NFACT*N
      GO TO 3
    1 XNFACT=NFACT
      SUM=((-1.)**K)*((.5*X)**KK)/(XNFACT**2)
      BER=BERR+SUM
      IF(ABSF(BER-BERR)-.0001)4,4,5
    5 BERR=BER
      K=K+1
      GO TO 6
    4 BERR=BER
      RETURN
      END
*     DATA
```

Figure 20. (Cont'd.)

The flow chart for FUNCTION BEII (X) is similar to that of FUNCTION BERR (X) and, therefore, will not be illustrated.

The subprograms used to evaluate DKEI (Z) and DKER (Z) are relatively simple; therefore, they will not be diagramed. These two subprograms are exact for values of the argument less than or equal to one. For values of the argument between 1 and 2, the error is negligible (less than 1 percent), but for values greater than 2 the error increases rapidly. It is for this reason that the quantity C, which is the argument used in this program, is included in the output of the program. An inspection of the program output will reveal the limits of the pavement thickness for which the computed stresses and deflections are exact. For values of $E/k \geq 100$, the solutions for a 10-kip load with a 70-psi tire pressure are accurate for pavement thicknesses greater than 2 in. Bessel functions available as Library Functions in the software of a particular data processing unit can readily be incorporated into the program and the subprograms can be omitted. The log function, natural log, is the only Library Function required by the program as presented here. Figure 20 shows a complete listing of the main FORTRAN program and the required FORTRAN functions.

# *Appendix B*

## DERIVATION OF EQUATIONS FOR MOMENT AND DEFLECTION*

The differential equation describing the behavior of a plate on a "heavy liquid" foundation (Eq. 7) becomes in cylindrical coordinates and for a symmetrical loading:

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}\right)\left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r}\right) = \frac{p - kw}{D} \qquad (12)$$

where

$p$ = unit load on plate,
$r$ = radial distance from center of loaded area, and
$w$ = vertical displacement.

For the case of a concentrated load, p is equal to zero except at the center of the plate; therefore, by setting p equal to zero and substituting $\ell^4$ for $D/k$, the governing differential equation becomes:

$$\ell^4\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}\right)\left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r}\right) + w = 0 \qquad (13)$$

The following dimensionless quantities are introduced: $z = w/\ell$ and $x = r/\ell$. Eq. 13 becomes

$$\left(\frac{\partial^2}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x}\right)\left(\frac{\partial^2 z}{\partial x^2} + \frac{1}{x}\frac{\partial z}{\partial x}\right) + z = 0 \qquad (14)$$

By substituting $\nabla$ for $\frac{\partial}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x}$, Eq. 14 becomes

$$\nabla\nabla z - z = 0 \qquad (15)$$

*This material is taken from the work of Putnam (14).

The variable $\epsilon = x\sqrt{i}$ is introduced and Eq. 15 becomes

$$\nabla'\nabla' z - z = 0 \tag{16}$$

where the symbol $\nabla'$ stands for $\frac{\partial}{\partial \epsilon^2} + \frac{1}{\epsilon}\frac{\partial}{\partial \epsilon}$. Eq. 15 is equivalent to

$$\nabla' (\nabla' z + z) - (\nabla' z + z) = 0 \tag{17}$$

and

$$\nabla' (\nabla' z - z) + (\nabla' z - z) = 0 \tag{18}$$

Thus, it is satisfied by the solution of the Bessel differential equation:

$$\nabla' z + z = \frac{\partial^2 z}{\partial \epsilon^2} + \frac{1}{\epsilon}\frac{\partial z}{\partial \epsilon} + z = 0 \tag{19}$$

and

$$\nabla' z - z = \frac{\partial^2 z}{\partial \epsilon^2} + \frac{1}{\epsilon}\frac{\partial z}{\partial \epsilon} - z = 0 \tag{20}$$

The combined solution of these two equations can be written as:

$$z = B_1 I_o (x\sqrt{i}) + B_2 I_o (xi\sqrt{i}) + B_3 K_o (x\sqrt{i}) + B_4 K_o (xi\sqrt{i}) \tag{21}$$

where $I_o$ and $K_o$ are Bessel functions of the first and second kind and of imaginary argument.

The argument x is real and all functions contained in the solution appear in complex form. The real part of the solution is determined by the introduction of four other functions: $I_o (x \pm i) = \text{ber } x \pm \text{bei } x$, $K_o (x \pm i) = \text{ker } x \pm \text{kei } x$, and setting $B_1 + B_2 = C_1\ell$, $B_1 - B_2 = -C_2 i\ell$, $B_3 + B_4 = C_4\ell$, and $B_3 - B_4 = -C_3 i\ell$, where the constants $C_1$, $C_2$, $C_3$ and $C_4$ are real. The following expression is obtained for the deflection of the plate:

$$w = C_1 \text{ ber } x + C_2 \text{ bei } x + C_3 \text{ kei } x + C_4 \text{ ker } x \tag{22}$$

All functions contained in Eq. 22 are real for real values of the argument. The series expressions which are represented by the Kelvin function, for small values of the argument are:

$$\text{ber } x = 1 - \frac{x^4}{4(16)} + \frac{x^8}{4(16)(36)(64)} \ldots,$$

$$\text{bei } x = \frac{x^2}{4} - \frac{x^6}{(4)(16)(36)},$$

$$\text{ker } x = (\ln 2 - j - \ln x) + \frac{\pi x^2}{16} - \frac{x^4}{64}\left(\ln 2 - j - \ln x + \frac{3}{2}\right) \ldots, \text{ and}$$

$$\text{kei } x = \frac{-\pi}{4} + (\ln 2 - j - \ln x + 1)\frac{x^2}{4} + \frac{x^4}{256},$$

in which j is 0.57722, Euler's constant; and ln 2 - j = 0.11593. The values of the Kelvin functions and their first derivatives are available in tabulated form for the pertinent values of the argument (19).

The general solution can be used for the analysis of any problem of symmetrical bending of a circular plate resting on a dense liquid foundation. The four constants must be determined for each particular case. For the case of loading under consideration (a concentrated load, P, applied at a point, x = 0, of an infinitely extended plate), the deflection of the plate at some distance, x, must be equal to zero. The functions ber x and bei x increase indefinitely with increasing values of x; therefore, the constants $C_1$ and $C_2$ must assume a value of zero to satisfy the boundary condition. The value of ker x becomes infinitely large at the origin, x = 0.

By setting $C_1 = C_2 = C_4 = 0$, the deflection equation (Eq. 22) reduces to

$$w = C_3 \text{ kei } x \tag{23}$$

The constant, $C_3$, is evaluated by considering the shearing forces. The shearing force, per unit length, is written as:

$$Q_r = -\frac{D}{\ell^3}\frac{d}{dx}\frac{d^2w}{dx^2} + \frac{1}{x}\frac{dw}{dx} \tag{24}$$

By substituting the first terms of the series expression for kei x in Eq. 23 and taking the derivative, the following expression is obtained:

$$Q_r = -\frac{C_3 D}{3}\frac{d}{dx}\left(-\ln x + \frac{\pi x^2}{16}\ldots\right) \tag{25a}$$

$$Q_r = \frac{-C_3 D}{\ell^3}\left(\frac{1}{x} + \frac{\pi x}{8}\ldots\right) \tag{25b}$$

As the value of x decreases, the value of $Q_r$ approaches $\frac{C_3 D}{\ell^3}\frac{1}{x}$.

$$Q_r = \frac{C_3 D}{r\ell^2} \tag{26}$$

The shearing force may also be obtained by distributing the load, P, uniformly over a circumference of radius, r:

$$Q_r = -\frac{P}{2\pi r} \tag{27}$$

Setting Eq. 23 equal to Eq. 27 yields:

$$C_3 = -\frac{P\ell^2}{2\pi D} \tag{28}$$

and

$$w = -\frac{P\ell^2}{2\pi D} \text{ kei } x \tag{29}$$

The value of D can be substituted into Eq. 29 to obtain:

$$w = -\frac{P}{2\pi k \ell^2} \text{ kei } x \tag{30}$$

To obtain the deflection at the center of a uniform circular load, consider an element of the loaded area:

$$dA = r d\theta \, dr \tag{31}$$

and the increment of load on the element of area:

$$dP = pdA = \frac{P}{\pi a^2} \, r \, dr \, d\theta \tag{32}$$

The increment of deflection which is caused at the origin (r = 0) by the load on the element of area is a function of the distance of the load from the origin (r) and is dependent on the direction of the load from the origin ($\theta$). Therefore, by use of the influence function for the concentrated load, Eq. 29, the increment of deflection at the origin caused by the load, dP, at a distance, r, is:

$$dw_o = -\frac{P}{\pi a^2} \, r \, dr \, d\theta \left(\frac{\ell^2}{2\pi D} \text{ kei } x\right) \tag{33}$$

Because Eq. 33 represents the increment of deflection caused at the center of the loaded area by the load on each element of the area, the total deflection at the origin will be the summation of the increments of deflection:

$$w_o = -\frac{P\ell^2}{\pi a^2 \, (2\pi D)} \sum_{r=0}^{r=a} \sum_{\theta=0}^{\theta=2\pi} r \, dr \text{ kei } x \tag{34}$$

In the limit Eq. 34 becomes the integral expression for deflection under the center of a distributed load; thus

$$w_o = -\frac{P\ell^2}{\pi a^2 \, (2\pi D)} \int_0^a \int_0^{2\pi} r \, dr \, d\theta \text{ kei } x \tag{35}$$

By integrating between the limits, $\theta = 0$ and $\theta = 2\pi$, and substituting $x = r/\ell$, Eq. 35 becomes:

$$w_o = -\frac{P\ell^2}{\pi D a^2} \int_0^a r \text{ kei } \left(\frac{r}{\ell}\right) dr \tag{36}$$

A solution to Eq. 36 is written as

$$w_0 = \frac{P\ell^2}{\pi Da^2}\left[r\ell \ker'\left(\frac{r}{\ell}\right)\right]_0^a \tag{37}$$

For a value of r = 0, the expression in the brackets of Eq. 37 takes a value of $-\ell^2$; therefore, by substitution of the limits, r = 0 and r = a, Eq. 37 may be written:

$$w_0 = \frac{P\ell^2}{\pi a^2 D}\left[\ell^2 + a\ell \ker'\left(\frac{a}{\ell}\right)\right] \tag{38a}$$

or

$$w_0 = \frac{P^2}{\pi D}\left[\left(\frac{\ell}{a}\right)^2 + \frac{\ell}{a}\ker'\left(\frac{a}{\ell}\right)\right] \tag{38b}$$

and by substituting the dimensionless parameter, $c = a/\ell$ (the radius of relative load distribution):

$$w_0 = \frac{P\ell^2}{\pi D}\left[\frac{1}{c^2} + \frac{1}{c}\ker' c\right] \tag{39a}$$

or

$$w_0 = \frac{P\ell^2}{\pi Dc^2}\left[1 + c\ker' c\right] \tag{39b}$$

The increment of moment which is produced at the origin by the load on an element of area is a function of the distance from the origin, r, and the direction, θ. However, for the special case of symmetrical loading, represented by the circular area of load distribution, the moment which is produced at the origin may be obtained by integration over the surface of loading by considering the equation of deflection for a concentrated load and the equations of bending for a symmetrically loaded plate:

$$m_r = -D\left(\frac{d^2w}{dr^2} + \frac{\mu}{r}\frac{dw}{dr}\right) \tag{40}$$

and

$$m_t = -D\left(\frac{1}{r}\frac{dw}{dr} + \mu\frac{d^2w}{dr^2}\right) \tag{41}$$

The radial and tangential bending moments are equal at the origin for a symmetrical loading condition, therefore,

$$m_{r=0} = \frac{m_r + m_t}{2} = -D\left(\frac{1+\mu}{2}\right)\left(\frac{d^2w}{dr^2} + \frac{1}{r}\frac{dw}{dr}\right) \tag{42}$$

By considering Eq. 30 for deflection of a concentrated load, and the derivatives:

$$\frac{dw}{dr} = -\frac{P\ell^2}{2\pi D}\operatorname{kei}'\frac{r}{\ell} \tag{43a}$$

and

$$\frac{d^2w}{dr^2} = -\frac{P\ell^2}{2\pi D} \text{kei}'' \frac{r}{\ell} \tag{43b}$$

and by adding:

$$\frac{d^2w}{dr^2} + \frac{1}{r}\frac{dw}{dr} = -\frac{P\ell^2}{2\pi D}\left(\text{kei}'' \frac{r}{\ell} + \frac{1}{r}\text{kei}' \frac{r}{\ell}\right)$$

and by substituting the identity relationship:

$$\text{kei}'' \frac{r}{\ell} + \frac{1}{r}\text{kei}' \frac{r}{\ell} = \frac{1}{\ell^2}\text{ker} \frac{r}{\ell}$$

we may obtain

$$\frac{d^2w}{dr^2} + \frac{1}{r}\frac{dw}{dr} = -\frac{P}{2\pi D}\text{ker} \frac{r}{\ell} \tag{44}$$

Thus, the increment of total bending moment produced at the origin by the load on an element of the loaded area is given by:

$$dm = \frac{(1 + \mu)}{2} \frac{P}{\pi a^2} \frac{1}{2\pi} r \, dr \, d\theta \, \text{ker}\left(\frac{r}{\ell}\right) \tag{45}$$

and, at the limit, the total bending moment produced at the origin by loading over the circular area is:

$$m = \frac{(1 + \mu)P}{2\pi a^2} \int_0^a r \, \text{ker}\left(\frac{r}{\ell}\right) dr \tag{46}$$

and integration of Eq. 46 between the limits of r = 0 and r = a yields:

$$m_{max} = \frac{P(1 + \mu)}{2\pi a^2}\left[a\ell \, \text{kei}'\left(\frac{a}{\ell}\right)\right] \tag{47a}$$

or

$$m_{max} = \frac{P(1 + \mu)}{2\pi}\left[\frac{1}{c}\text{kei}' \, c\right] \tag{47b}$$

Eq. 47b for the maximum moment and Eq. 39b for the deflection under the center of the loads are the equations used in the computer program to compute STRESS and DEF, respectively.

# Layered Pavement Design Method for Massachusetts

EGONS TONS, RICHARD E. CHAMBERS, and MICHAEL A. KAMIN
Respectively, Assistant Professor of Civil Engineering, Senior Research Engineer, Principal Civil Engineer, Massachusetts Department of Public Works

A method for the design of flexible pavements has been developed for Massachusetts. The method is simple, rational and practical, and can be applied immediately on a routine basis in the design office, yet is flexible enough to permit modification as indicated by future research. Data and analyses that evolved from the AASHO Road Test experiments were used as a guide in the development of the design method.

Straightforward conventional procedures have been selected to provide a value for soil support. Preliminary soils data are obtained from geological and soil maps, and borings and samples for test are taken as needed. Laboratory Bearing Ratio and other tests are performed. Test results and boring data are then used as a basis for a Design Bearing Ratio. Traffic factors are computed from anticipated traffic conditions and existing traffic and Loadometer data. The 18-K Daily Equivalent Axle Load is used directly as defined in AASHO procedures.

The Regional Factor as recommended by the AASHO guide was not adopted in its present form. This approach was deemed too specifically related to materials, soils and environment of the test site. To account for possible moisture, frost, traffic, time (aging) and other effects, a blanket increase of 15 percent in Structural Number was introduced. This increase probably approximates a reasonable Regional Factor. More study is needed in this area.

There are an infinite variety of layered pavement systems that will satisfy the strength requirements as dictated by a given soil and traffic. AASHO coefficients of relative strength of various materials for surface, base and subbase were adopted where possible. A new coefficient was derived for the penetrated base. The general design chart correlating DBR values with traffic and structural number permits the use of any type and thickness of surface, base, and subbase, provided the strength coefficient for each material is known. The pavement thicknesses obtained by using the derived design chart are reasonable when compared to past experiences in Massachusetts and other states.

•FIELD EXPERIENCE has formed the basis for proportioning of flexible pavements in Massachusetts as well as in many other states. In an attempt to place pavement design in Massachusetts on a more quantitative basis, a committee was formed of members of the Massachusetts Department of Public Works and of the Department of Civil Engineering of the Massachusetts Institute of Technology; their mission was to review and evaluate available approaches and recommend a practical design method that could

Paper sponsored by Committee on Flexible Pavement Design and presented at the 44th Annual Meeting.

be used immediately in the design office. It was, of course, recognized that any approach adopted would have to be augmented with subsequent research, the direction of which would also be recommended by the committee. This paper covers the first phase of the study, i.e., the background and development of the design method.

There are a variety of sources that can be used as a basis for the design of flexible pavements. There are empirical methods that become so modified with time that the reasons for their existence are vague and often not documented. At the other extreme are theoretical approaches that are oversimplified in assumptions but complicated in application, demanding development of data and techniques not now available. Therefore, it was decided that a new approach, that offered by the AASHO guide, would be a desirable path to follow. Although the AASHO concept is controversial and certainly not conclusive in many respects, it does offer an extensive, well-documented, and well-controlled fund of data as its background. There is also the hope that more information will be developed to strengthen further the AASHO approach.

In essence, the AASHO approach considers: (a) soil supporting capacity, (b) traffic factors, (c) regional variables, and (d) structural capabilities of pavement materials. These factors form the basis for proportioning a pavement section that will serve the riding public for a predictable period of time. Certainly the critical parameters are recognized and the criterion for performance is realistic. A design method based on the AASHO approach should, therefore, be reasonable.

## DEFINITIONS OF TERMS

Axle Load.—Load transmitted by a single axle having two single- or dual-tired wheels.

Base Course.—Layer of specified material of designed thickness placed on a subbase or subgrade to support a surface course.

Bearing Ratio (BR).—Stress required to produce a certain penetration, using a standard piston, in a given soil relative to a standard reference stress.

Black Base.—A plant-mixed graded bituminous mixture used under the surface layer.

Design Bearing Ratio (DBR).—That bearing ratio selected as being typical for design purposes of the section under consideration.

18-Kip Equivalence Factor.—Number of applications of an 18,000-lb single-axle load that will have the same effect on the serviceability of a pavement as a single application of a given load.

Equivalent Daily 18-Kip Load.—Average number of equivalent 18-kip load applications that will be applied to the pavement structure in one day.

Freezing Index (FI).—An index of the severity of a winter which takes into account the temperature drop below freezing and the number of days in which this occurs.

Frost Heave.—Increase in elevation of the pavement surface due to ice lens formation in the underlying soils or materials.

Frost Susceptible Material.—A soil in which significant detrimental ice segregation can occur if conditions of moisture and temperature are favorable.

Future Average Daily Traffic (Future ADT).—Estimated average daily traffic at the time that the Terminal Serviceability Index is reached.

Initial Serviceability Index (ISI).—Serviceability index of a newly constructed road before the commencement of traffic.

Layered (Flexible) Pavement.—A pavement structure which maintains intimate contact with and distributes loads to the subgrade and depends on aggregate interlock, particle friction, and cohesion for stability.

Longitudinal Profile.—Contour of the surface grade in the direction of traffic.

Pavement Structure.—Combination of subbase, base course, and surface course placed on a subgrade to support the traffic load and distribute it to the subgrade.

Penetrated Base.—An open-graded crushed rock layer penetrated in the field with asphalt, used under the surface layer.

Present Average Daily Traffic (Present ADT).—Average daily traffic expected to occur immediately after the highway is opened to traffic.

Present Serviceability Index (PSI).—Serviceability index at time of observation.

Regional Factor.—A factor used to adjust the structural number for climatic and environmental conditions.

Serviceability.—Ability of a pavement to serve traffic that it was meant to serve.

Serviceability Index (SI).—A number that estimates ability of a pavement to serve traffic that it was meant to serve. A measure of the roughness, rutting and degree of cracking and patching found in a pavement.

Soil Support Value (S).—An index of the relative ability of a subgrade material to support traffic loads imposed on it by the pavement structure.

Strength Coefficient.—A number indicating relative effectiveness of pavement materials as they contribute to the performance of the pavement structure.

Structural Number.—Product of thickness and strength coefficient of a given layer of the pavement structure. Sum of the structural numbers for all layers is the structural number of the pavement and is an index of the performance of the section.

Subbase.—Layer or layers of specified or selected material of design thickness placed on a subgrade to support a base course.

Subgrade Line.—Level above which the pavement structure and shoulders are constructed.

Subgrade Material.—Material below subgrade line in cuts and embankments and in embankment foundations, extending to such depth as affects the support of the pavement structure.

Subgrade Weakening.—Loss in supporting capacity of subgrade soils due to increase in water content.

Surface (Wearing) Course.—Top layer(s) of the pavement structure that resists the direct stress applied by traffic loads. It provides a smooth riding surface, resists skidding, abrasion and climatic effects, and protects the underlying layers from moisture.

Terminal Serviceability Index (TSI).—Serviceability index at which the pavement is deemed unable to serve the traffic that it was meant to serve; the point at which major resurfacing of the pavement is necessary.

Traffic Analysis (Design) Period.—Number of years that the highway will be in service before the Terminal Serviceability Index is reached.

Transverse Profile.—Contour of surface grade across the roadway.

Truck Weight (Loadometer) Study.—A field survey made for the purpose of obtaining information on trends in weight dimensions, axle spacings, types, loads, etc., of freight vehicles actually using the road.

## FACTORS IN PAVEMENT DESIGN

The purpose of a pavement is to provide a smooth riding support for various loaded vehicles. This is achieved by designing and constructing a layered densified composite consisting of various sizes of rock particles, often bound together by some organic or inorganic cementing agent. The function of each layer is as follows (Fig. 1):

1. Surface Layer.—This must resist the high vertical stresses applied under the tire. As it acts as a plate under dynamic conditions, it develops substantial bending and shear stresses as well. The surface must also act as a protective layer to shield the layers beneath from water, provide a smooth wearing surface to afford a satisfactory ride to the user, and reduce stresses on the layers below it.

2. Base Layer.—This must resist high stresses. Although the surface layer acts to reduce them to some degree, the stresses within the base are still high. This is particularly true for bases which underlie thin surface layers. The base must also serve to reduce stresses on the weaker layers below it, and provide a smooth support on which the surface can be laid.

3. Subbase.—This must provide drainage for water that may penetrate the surface layers and also must allow drainage for the water that percolates from below. The drainage function is particularly critical because of the frost conditions found in Massachusetts. The subbase must also resist stress applied from above, and reduce stresses on the underlying soil layer.

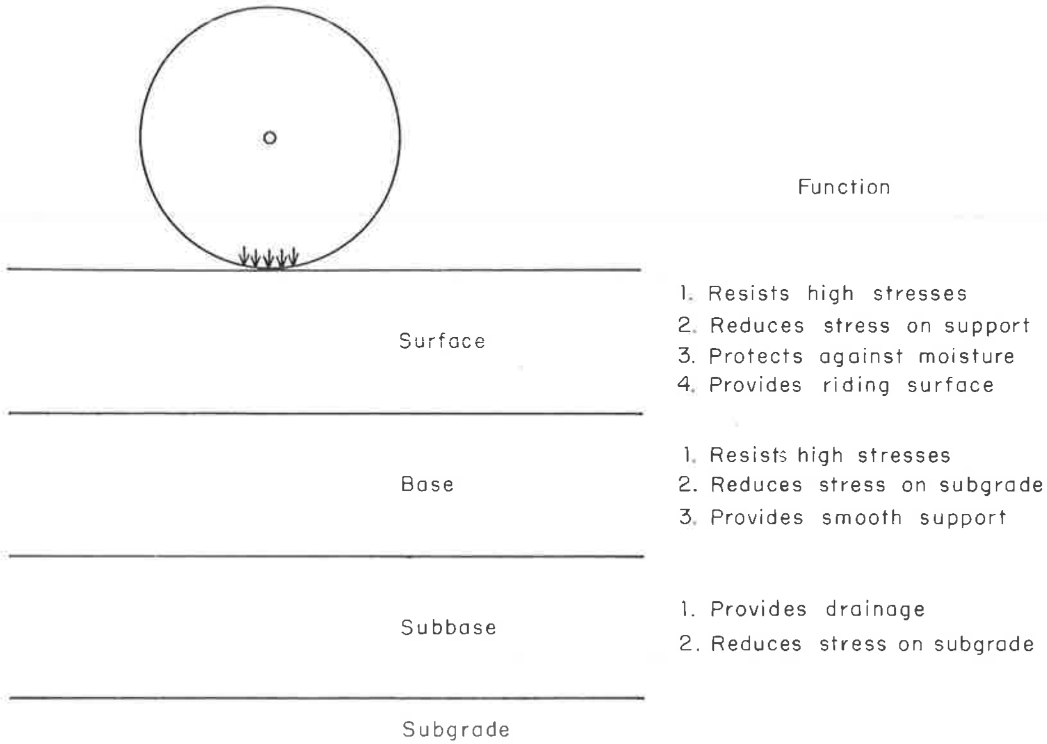



Figure 1. Primary functions of pavement layers.

## Static vs Moving Load

The load applied to the layered pavement composite can be static, such as at stop lights, or moving, as encountered in rural and many urban highways. In the methods of design developed in the past, the static load was considered the most critical; therefore, static test values were often used in design. Most of the mileage of the interstate highways and expressways is subjected exclusively to moving loads during its service. Furthermore, the data developed at the AASHO Road Test are based on performance under moving loads(1).

There are materials, such as crushed rock or gravel, which may not have properties influenced by time of load duration. On the other hand, materials like bituminous concrete and some soils can be greatly affected by the speed of load application and by temperature. Therefore, it is necessary for the purposes of design to decide what kind of loading the designed pavement will predominantly serve.

The speed of traffic used in the AASHO Road Test was about 30 mph. On primary highways and most secondary roads, 30 mph and above are frequently encountered, although it is possible that trucks on grades may slow down to below the 30-mph speed.

In this design approach, using the AASHO Road Test findings as background, it was assumed that a moving load would be applied to a given point on the wheelpath for very short durations, usually lasting much less than one second. Figure 2 shows the relationship between speed of given vehicle and the approximate time of load duration at a point the wheel is traversing.

## Present Serviceability Index

One of the major contributions towards quantitative measurement of ridability of the pavement is the Present Serviceability Index (PSI) developed during the AASHO Road Test experiments (2). The index can vary between 0 and 5 (4 to 5 = very good; 3 to 4 =

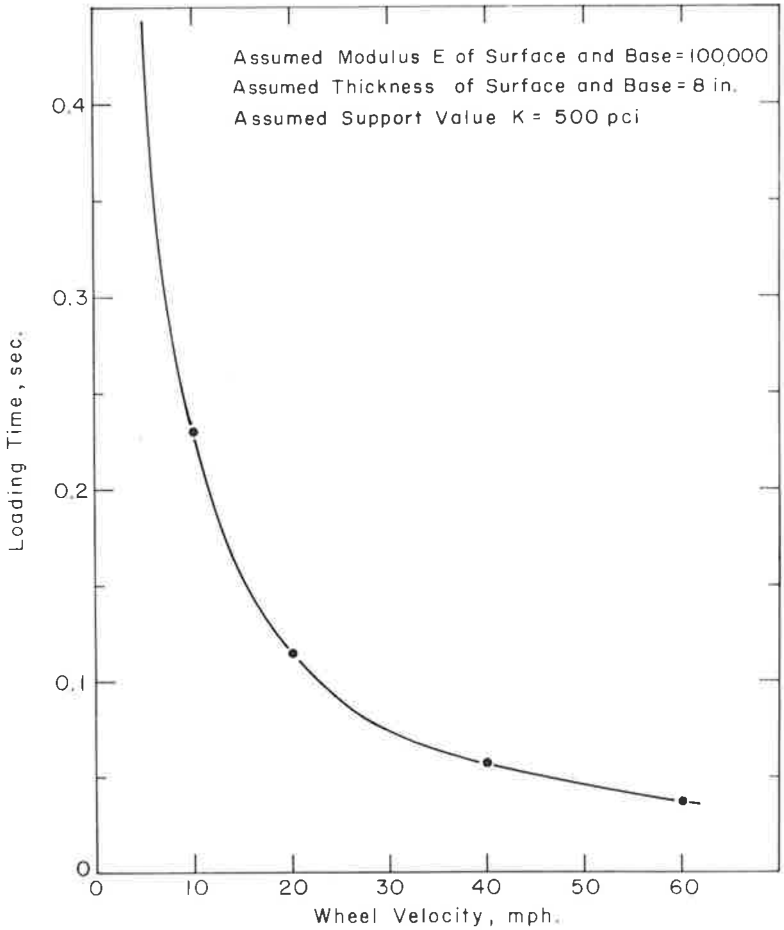



Figure 2. Example of pavement loading time as affected by vehicle speed.

good; 2 to 3 = fair; 1 to 2 = poor; 0 to 1 = very poor). This numbering system was obtained by first sending out various representatives of highway users on designated road sections and asking them to give their evaluation of the condition of the road as to the ability to carry traffic at that time. They were asked to indicate their decision by assigning a number for each section.

The human survey was followed by a mechanical survey, measuring washboarding, rutting, cracking, patching, etc., to determine the relative influence of each of these deficiencies in assigning a PSI for a given section by an "average" user of the road. The most important factors that were found to affect the ridability of the road are: (a) longitudinal profile of the road (washboarding, bumps, etc.), (b) transverse profile of the road (rutting), (c) amount of cracking, and (d) amount of patching. From the AASHO Road Test, the following equation relating these variables was derived (2):

$$PSI = 5.03 - 1.91 \log (1 + SV) - 1.38\, RD^2 - 0.01 \sqrt{C + P} \qquad (1)$$

where

SV = mean of the slope variance in the two wheelpaths (longitudinal variation),
C + P = measure of cracking and patching in pavement surface, and
RD = measure of rutting in wheelpaths.

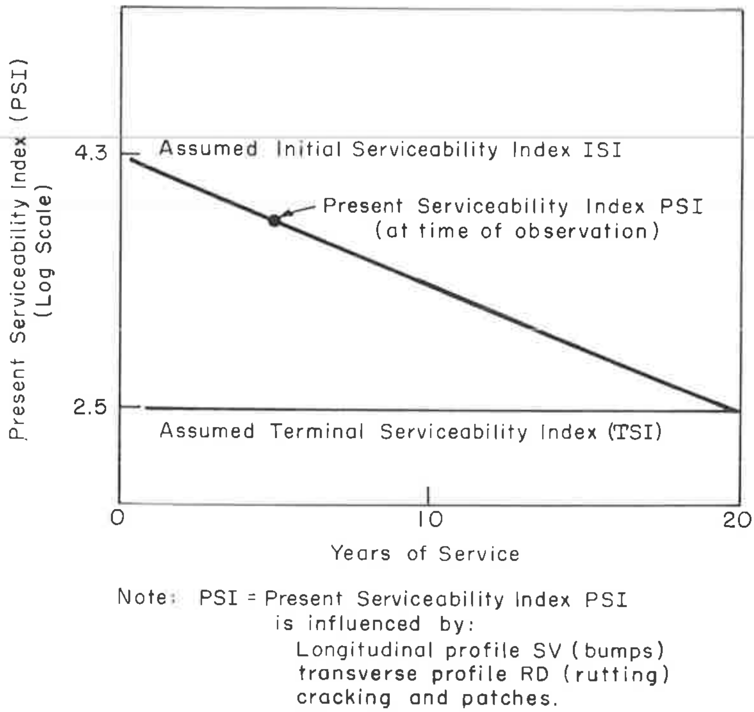



Figure 3. Serviceability index decline with time.

All these variables can be measured on a road and the PSI values can be calculated at any time desired. For primary roads with a high volume of high-speed traffic, a PSI value of 2.5 is often assumed to be the low point, a time when some kind of resurfacing work should be considered. If the initial PSI value is known and the terminal value is assumed to be 2.5, a trend curve with age for PSI can be obtained for a given traffic, as illustrated in Figure 3. That is, the PSI concept can be used as a criterion for the design of a pavement to carry given traffic for, e.g., 20 yr before the PSI value drops to 2.5. At this time, major resurfacing work is needed to restore the riding comfort. This approach is the basis of the design method presented here.

Major Factors Affecting PSI

The amount of roughness, cracking, and patching of a road with time in service will be dictated by many factors. The most important are the following: (a) subgrade, (b) materials in the pavement layers and their arrangement, (c) quality of construction, and (d) traffic. All these factors may not have equal weight under given conditions, but they are all important in general considerations of pavement design. If any of them are neglected, a pavement may have only a fraction of the useful life intended. The PSI defines primarily the smoothness of a road surface, which can be affected by any of the factors mentioned.

## SURVEY OF SUBGRADE SOILS

One of the major parameters considered in the AASHO guide approach to pavement design is the supporting capacity offered by the subgrade soil. Because the final cross-section and grades of the highway will depend on the supporting capacity of the soils encountered, a detailed survey of the soils existing along the proposed route must be undertaken early in the design phase. A study of available information on soils and soils types existing in Massachusetts was undertaken as a background for preliminary

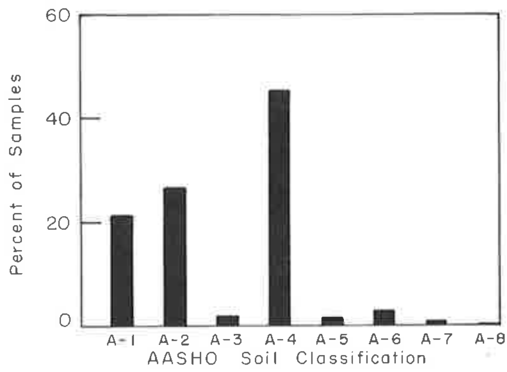



Figure 4. Frequency of soil types from records on existing roads (Mass. Department of Public Works, Materials Division).

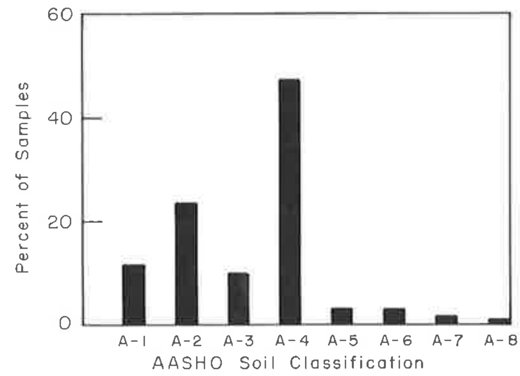



Figure 5. Frequency of soil types from records of Soil Conservation Service (Amherst, Mass.).

surveys along proposed routes. An analysis was made of data available on embankment material on existing roads, and of data developed by the U.S. Soil Conservation Service. Figures 4 and 5 show the distribution of soils types as determined from representative data from these sources. The majority of soils fall into AASHO Classifications A-1 through A-4, with the A-4 materials being predominant.

Maps of value for soils evaluation are currently being prepared by the U.S. Soil Conservation Service and the U. S. Geological Survey; however, it will be several years before either of these sources has information covering all areas of Massachusetts. Data from these agencies should aid in initial route planning, as well as in selection of locations for soil sampling along proposed routes. Topological maps, prepared as a matter of course during the route-planning phase of highway design, should also prove helpful.

## Route Survey

A field survey must also be undertaken to determine the locations at which samples for test should be removed. A specialist in the area of soils and geology should accompany the highway designer in an on-site inspection of the proposed route. Observation of characteristics of the terrain such as surface water, rock, outcroppings, variations in soil types, and condition of existing cuts and fills in the area, if present, will provide some index as to the variations to be expected in embankment materials and identify possible problem areas that may need special treatment.

## Criteria for Location of Soil Test Samples

The subgrade on which the pavement structure will be placed will seldom, if ever, coincide with the surface of the existing terrain (Fig. 6). To determine the type and strength of the soil on which the pavement will be constructed, soil sampling and testing is necessary. Samples may be obtained either by digging or by boring if the depth required is great. It is important that borings be made in both cuts and fills. The location and spacing of the borings should be determined from the preliminary soil surveys.

In cuts, the pavement is placed on top of soils existing along the predetermined subgrade line. There may be instances where these soils vary greatly in their ability to support the loaded pavement. Especially capricious are so-called transition areas from a cut to a fill. To predict the support abilities of the soils, soil samples for laboratory evaluation are needed from the subgrade level. Borings in cuts should be taken to at least 3 ft below the prescribed subgrade line. Soil profiles should also be plotted from the boring data. Knowledge of the soil properties in cuts is very important

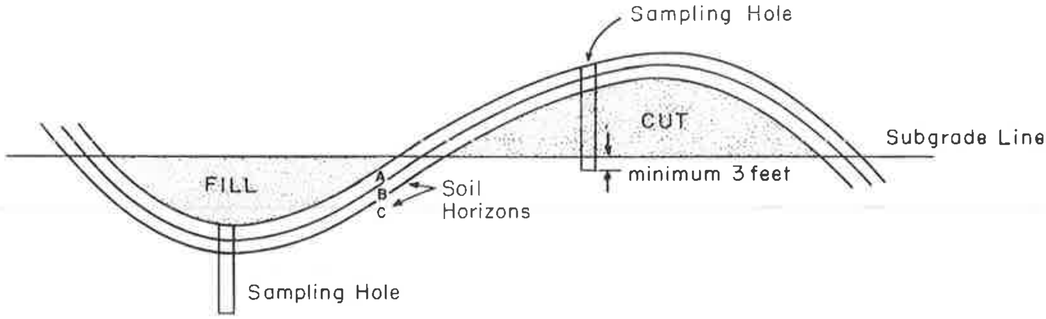



Figure 6. Terrain necessitating soil sampling in cuts and fills.

in this design procedure because the pavement thickness is based primarily on the soil conditions in the cut.

The purpose of taking borings and obtaining soil information under fill areas is different from that of cuts. The excess material from cuts is usually used in fill and the subgrade material will be more or less blended soils from the cut. Therefore, the concern here is the ability of the existing soils to support the fill material rather than the pavement and the traffic load. The depth of borings under fills will depend on circumstances but in most cases should be about equal to the height of the fill material.

## SOIL TESTING AND IDENTIFICATION

### Strength Test

In most pavement design methods, a "strength" value of the embankment soil forms one basis for proportioning of the pavement structure. In essence, the AASHO approach permits any measure of soil supporting capacity to be used, provided the strength of the AASHO Road Test subgrade is known under the test. (This is a tentative situation; subsequent testing must be performed to determine the extent of correlation that exists with any test method that may be selected.) Thus, virtually complete freedom was allowed in the selection of test for soil supporting capacity.

Strength Test Values Shown in AASHO Guide.—The AASHO Road Test sections were built on one type of soil, A-6, and, therefore, only one point on a soil support scale is available (3). In Figure 7, taken from the AASHO guide, a support value of 3 has been assigned for the A-6 soil. The highest support value on the scale was obtained by analysis of performance of various sections with a thick crushed rock base. Thus, Point 10 represents soils having characteristics of the crushed rock base materials. The support values have been compared with several known test procedures. These comparisons are approximate and were not adopted for the basis of Massachusetts design for the following reasons:

1. Kentucky CBR.—The Kentucky CBR scales given in the AASHO guide (Fig. 7) are basically not Kentucky curves because that state does not use dynamic compaction in preparing specimens. Furthermore, the guide calls for the Modified Proctor compaction (AASHO Method T-180-57) which Massachusetts is not using at the present time.
2. Stabilometer (R) Curves.—The stabilometer scales given in the AASHO guide were studied. The compaction again is not that of the Standard Proctor; also, the R values of Washington State differ from those of California, indicating that the test has not been generally standardized.

Other Laboratory Strength Tests.—Besides the California Bearing Ratio and the stabilometer tests, there are a variety of laboratory procedures, with triaxial tests being quite prominent. Several states use a triaxial test value in their pavement design. The Texas Highway Department was contacted to find out whether any correlation

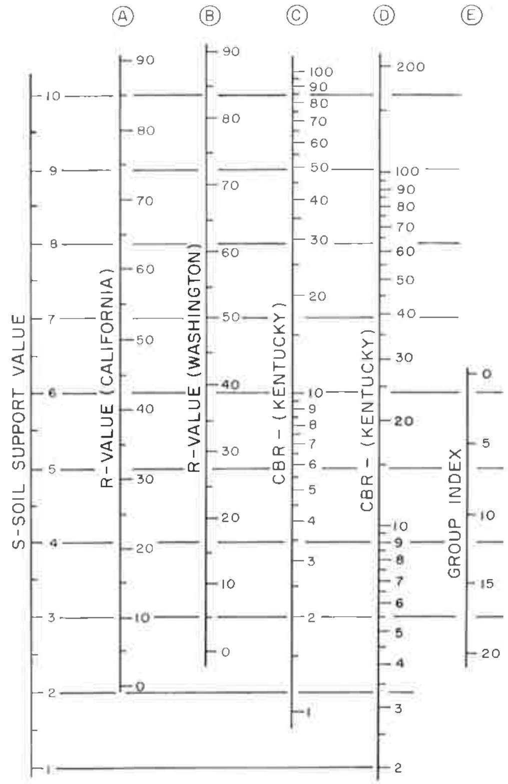



Figure 7. AASHO correlation chart for estimating soil support value (S).

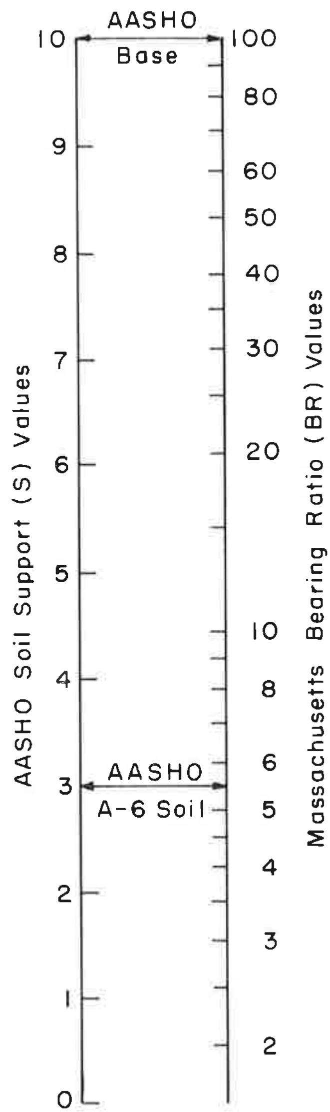



Figure 8. Assumed correlation between Massachusetts bearing ratio (BR) and AASHO soil support value (S).

has been established between the triaxial and the soil support values (S) in the AASHO guide. Apparently such correlation does not exist.

Bearing Ratio for Massachusetts.—It was decided that the Massachusetts test for soil supporting capacity should fit the following requirements:

1. The test should be relatable to the S values of AASHO guide.
2. The test should be well established, so that as much background data as possible could be obtained as needed.
3. Samples should be prepared using Standard Proctor compaction.

It was judged that a bearing ratio test would satisfy these requirements. Therefore, a literature search was undertaken to establish a bearing ratio scale for Massachusetts, tied in with the hypothetical soil support values in the AASHO guide. The most helpful publication in this respect was a paper by Shook and Fang (4), which summarizes all cooperative tests done by various agencies on the AASHO test road materials.

The Massachusetts scale was established by averaging appropriate CBR values for the AASHO test road soil compacted by the Standard Proctor (5.5-lb hammer, 12-in. drop). This bearing ratio value was approximately 5.5 and was assumed to be equal to the soil support value, S = 3. Also, the appropriate bearing ratio values for the crushed rock base material, compacted according to the Standard Proctor method, were averaged. This value was about 100 and it was assumed to be equal to 10 on the soil support scale. Then a logarithmic relationship was assumed between the two established points which resulted in a bearing ratio scale for Massachusetts as shown in Figure 8. Further research and improvement of this scale may be necessary. At the same time, the scale is reasonable if compared to the other attempts made along the same lines. The test basically calls for AASHO T-99-57 compaction, and a 4-day soaking before test. Detailed procedure is given in ASTM D1883-61T.

### Classification Tests

Standard soil classification tests should be performed on the subgrade materials. These include: (a) sieve analysis (AASHO Designation T-88-57), (b) hydrometer analysis (AASHO Designation T-88-57), (c) liquid limit (AASHO Designation T-89-60), and (d) plastic limit (AASHO Designation T-90-56). Although these tests are not directly involved in structural design calculations, they should be performed for purposes of identification and determination of the degree of frost susceptibility of the soil.

### Desirable Additional Tests

In addition to tests needed for design purposes, three other tests are suggested: (a) volume change, (b) permeability, and (c) unconfined compression. The first two tests are fundamental in describing engineering behavior of soils. A serious thought should be given to incorporating these two tests in design considerations. To achieve this, accumulation of measurements and research is needed. The unconfined compression test can be used only with clays. This simple test, in comparison to the bearing ratio values, should add to the understanding of the strength behavior of cohesive soils in subgrades.

## SUBGRADE SUPPORT FOR DESIGN

### Selection of Design Bearing Ratio

One of the problems facing the design engineer is the selection of a Design Bearing Ratio value (DBR) for a given section of a roadway. The soils vary widely from place to place, as is illustrated in Figure 9. If the design is based on different bearing ratio values, the cross-section of the pavement may have to be changed every 500 to 1,000 ft. If the lowest bearing ratio value is used, overdesign in most sections will result. The economics of the project and minimum pavement thickness requirements should be the guiding factors for a decision to change a cross-section.

A practical criterion for the selection of the DBR results if normal construction practice is considered. Usually in cut areas practice is to utilize the soils in situ as the subgrade for the pavement. Fill areas, however, are built up from either cut or borrow soils. It is reasonable to expect that the fill, if made up from the cut material, will be at least as strong as the weakest material in the cut. In cases where fill is built up from borrow, the borrow material can be specified as having a given strength. It is clear that the governing factor in the choice of a DBR is the strength of the soil in the cuts; the properties of the soils in fill areas will usually exceed those in the cut area either by nature or by control.

The choice of the DBR within the cut will depend on the variations of strength of the soils. Figure 10 shows three examples of cases that may arise:

1. If the bearing ratio values in the cut do not vary greatly, the DBR should be close to the minimum bearing ratio found. If fill is to be borrow, the borrow must be specified to have a bearing ratio equal to the DBR.

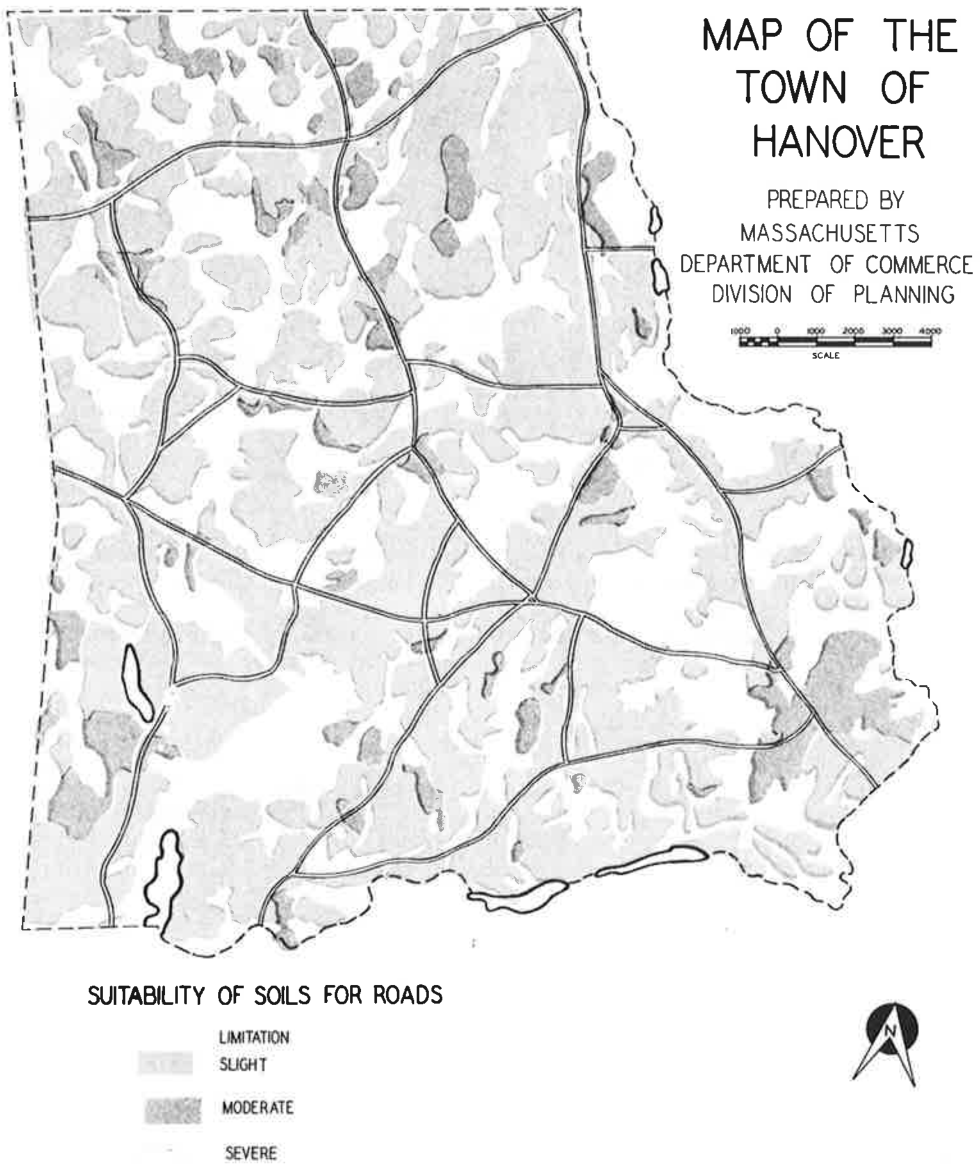



Figure 9. Interpretive map developed by U. S. Soil Conservation Service, showing frequent soil variation in road construction.

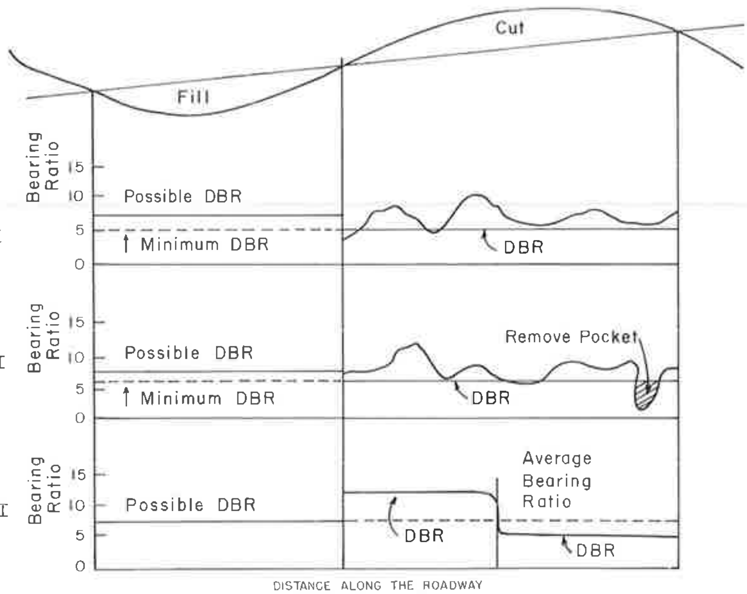



Figure 10. Three examples in selection of Design Bearing Ratio (DBR).

2. If there is a "soft spot" for only a short distance (Fig. 10), it may be more economical to treat this area (dig out or cover up) rather than use the lowest bearing ratio value for a design.

3. If there are changes in soils and two (or more) types of soils exist for relatively long distance (e.g., half a mile or more), a possibility of using two (or more) thicknesses of pavement should be considered.

## Soil Support as Affected by Frost and Moisture

Depth of frost penetration is determined by several parameters: (a) freezing index, an index of the severity of a wind, which takes into account the temperature drop below freezing and the number of days in which this occurs; (b) thermal conductivity, which depends on the soil or layer type, density and moisture content; and (c) the volumetric latent heat of fusion, which depends on moisture content and dry density of the soil or layer. Several formulas are available to determine the frost penetration using these parameters.

The U. S. Army Corps of Engineers (5), using the modified Berggren formula derived by Aldrich (6), has developed simple graphical solutions for frost depth determinations under pavements. From these, Figure 11 was plotted for frost depths expected to occur in Massachusetts under average conditions in average soils. The wide variation of penetration is to be expected because of the spread in freezing index from, for example, Cape Cod to the Berkshires. There are, of course, yearly variations in penetration due to climactic changes. The criterion used in the development of Figure 11 was the average freezing index of the three coldest years occurring in a 30-yr interval. It is certain that frost permeates the pavement and well into the supporting soils everywhere in Massachusetts. Frost and its effects on the pavement structure and performance is, therefore, an important consideration in this design study.

Concepts in Frost Design.—Under certain conditions frost heaving can occur, causing severe and dangerous changes in elevation of the pavement surface. Heaving is possible if a certain combination of factors are present: (a) freezing temperatures within the soil, (b) a source of water, and (c) a frost-susceptible material. The depressed temperature starts to freeze the water in the soil. If the soil has a high cap-

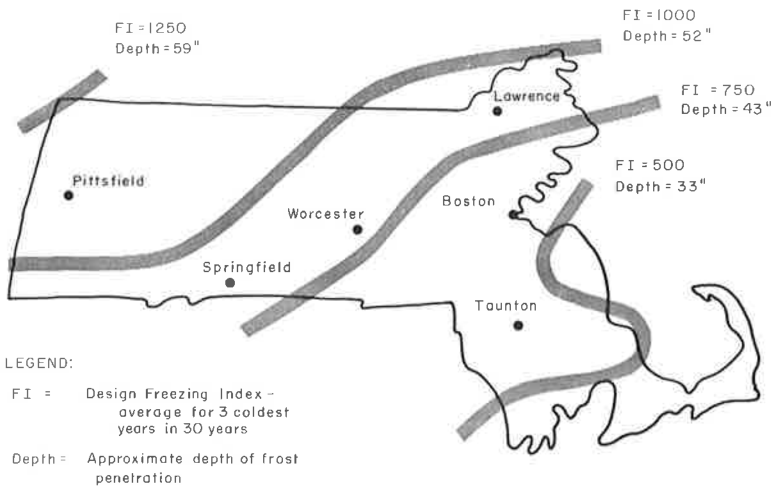



Figure 11. Variation in design freezing index and depth of frost penetration in Massachusetts.

illary tendency, it will act to siphon water from the source. As the new water freezes, lenses are formed which can cause severe changes in elevation at the surface.

The types of soils that possess capillarity always contain fines, but to be frost susceptible they also must be permeable. Thus, for example, sands with relatively large particle size are not considered frost susceptible; such soils as gravels having a high percentage of fines, silts, and varved clays are considered frost susceptible. Criteria for frost susceptibility, which are said to be 85 percent sure, have been developed by the U. S. Army Corps of Engineers (5). These are based on the percentage of fines passing 0.02-mm mesh for given soil types. On this basis, frost susceptibility has been divided into four categories ranging from low susceptibility (F-1) through high susceptibility (F-4). Table 1 gives the criteria for determining the degree of susceptibility for different materials.

The effects of frost heaving vary with local conditions. If the soils and water conditions are uniform over a reasonable distance, the heave will be uniform, and probably unnoticeable, to the motorist. The problem of heaving may be serious if conditions are variable in localized areas. In this case, differential heaving is possible, and uncomfortable or dangerous bumps may result. It is principally the horizontal variability of soils or water conditions that cause serious effects from frost heaves.

Horizontal variations are taken into account in some agencies. The Corps of

TABLE 1

SOIL CLASSIFICATION FOR FROST DESIGN[a]

| Frost Group | Soil Type | Finer than 0.02 mm (% by wt) |
|---|---|---|
| F-1 | Gravelly | 3-10 |
| F-2 | Gravelly | 10-20 |
| | Sands | 3-15 |
| F-3 | Gravelly | >20 |
| | Sands except very fine silty sands | >15 |
| | Clays with plasticity indexes >12 | |
| F-4 | All silts | |
| | Very fine silty sands | >15 |
| | Clays with plasticity indexes <12 | |
| | Varved clays and other fine-grained banded sediments | |

[a]U. S. Army Corps of Engineers (5).

Engineers, for example, uses these as the basic consideration that determines the approach to frost effects in airfields, where maintenance of uniform grade is critical. If variations are high, removal of most or all of the susceptible material is recommended. Such a step is not deemed economically feasible for highway pavements except in unusual circumstances. Michigan, too, considers horizontal variations to be serious. In their study of the problem (7), they found that 98 percent of all heaves occurred in cuts where the soil pattern is variable. In fills, few heaves were observed because, by nature, sharp variations are minimized. Blending of the top 12 in. of the subgrade may be a helpful step in minimizing horizontal variations, and hence frost heave.

Once the soil and select materials of a pavement are frozen, beneficial effects occur. Freezing essentially stabilizes the particulate materials and increases their resistance to deflection. At the AASHO test road a marked decrease in loss of serviceability was observed in the winter months, indicating improved performance.

When frost leaves, significant weakening of the soil and select pavement materials can result. If there are frozen lenses within the soil, thawing essentially leaves a pocket of liquid. This liquid then saturates the surrounding soil. Because the material below remains frozen, the path of escape for the water is upwards. According to the AASHO test results, this weakens the subbase as well as the subgrade soil. Thus, while the gravel acts as a draining layer, it is also weakened in strength.

The Corps of Engineers recommends designing highways for the reduced subgrade strength encountered during spring breakup. They recommend increasing the design traffic number in their design method as an adjustment to, in effect, arrive at a stronger pavement section than would normally be used. The size of the increase depends on the degree of frost susceptibility of the soil under consideration. The Corps of Engineers' approach cannot be applied directly to the AASHO data because the criteria used for measuring performance by the two methods differ.

Massachusetts' Experience with Frost Problems.—A survey was taken of all the district offices of the Massachusetts Department of Public Works regarding their experience with frost problems on major roadways. There were a few isolated cases where serious heaving was reported, for example, in places where springs were found close to the surface. But problems were considered the exception rather than the rule. The Design Committee on its several survey trips made special note of the frequency of heaves. None were apparent, although one area had been marked as such on Route 128.

On the basis of experience, frost heaves do not appear to be a major problem in Massachusetts, although certainly conditions are favorable for them to develop. There are apparently several reasons for this. Massachusetts has been particularly careful to provide good surface and subsurface drainage in their pavements. For example, practice has been to keep the road surface at least 5 ft above free water to allow for drainage. Also, a gravel subbase layer extending to 20 in. below the pavement surface has been provided on major highways. These precautions have, in most cases, probably reduced the amount of free water available to create frost heaving.

The effects of subgrade weakening during spring breakup have not been documented. But these effects, although damaging, are more subtle and would probably not be noticed unless specific study were given to the problem.

AASHO Approach—Regional Factor.—At the AASHO test road, the rate of decay of serviceability was found to vary significantly with season. Indeed, a major conclusion of the research effort was that 80 percent of the sections failed during spring breakup. To account for these variations in performance, deflection data were analyzed. Deflections found in plate bearing tests on many of the test sections were found to vary with season. They were low in the winter, high in spring, and intermediate in summer. By statistical manipulation a weighting function was derived that related serviceability loss and deflection variations (Fig. 12).

The approach recommended by the AASHO guide regarding seasonal variations is to use a regional factor. Although no quantitative means of determining the regional factor is given, consideration is recommended of such factors as finished grade elevation above water, drainage, depth of frost penetration and number of freeze-thaw cycles. Furthermore, other considerations are lumped in with this factor, such as steepness of grade and areas of concentrated stopping or turning. The regional factor, which can

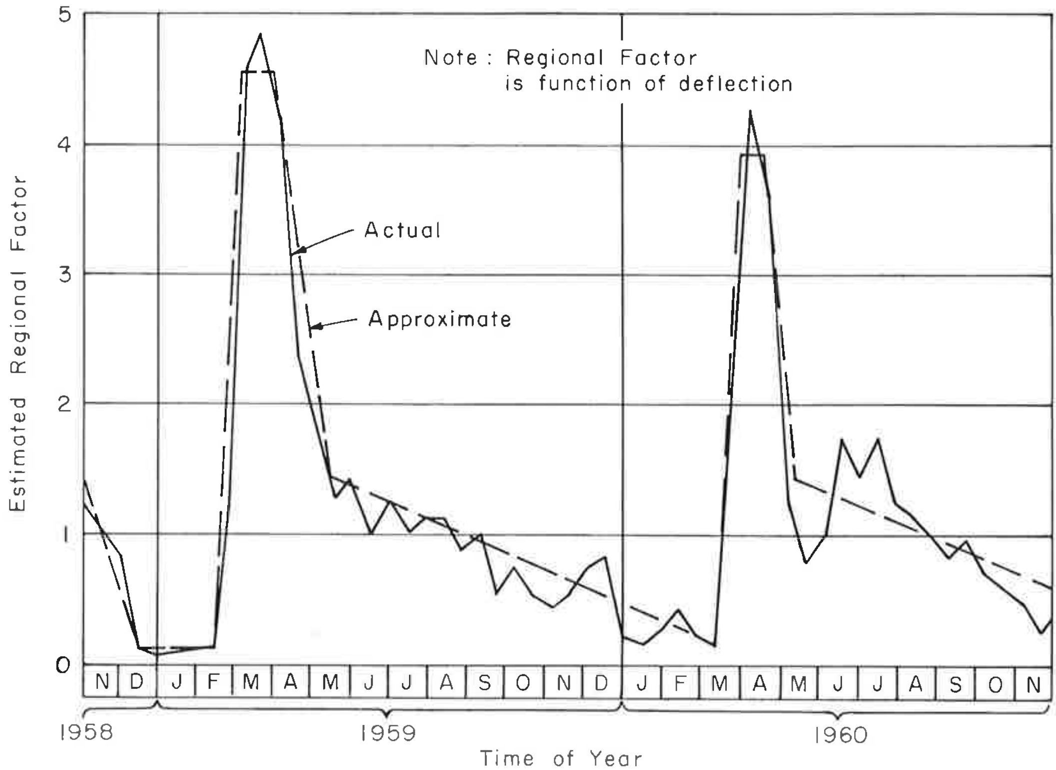



Figure 12. Monthly variations in estimated regional factor for the AASHO test road.

vary from 0.5 to 5.0, can then be applied to adjust the structural proportions of the pavement to account for these effects.

The regional factor should also account for the relative weakening that can occur in soils and materials, depending on their frost susceptibility. Further study of this area should be undertaken to account for seasonal variations in properties of soils and materials.

Recommendations.—Soils and materials should be checked for frost susceptibility according to the Corps of Engineers criteria (Table 1). The F-4 soils, as designated by the Corps of Engineers are highly susceptible and, therefore, offer the possibility of creating significant differential heaves and weakening. Special attention should be given to surface drainage and water table when these materials are encountered. If good drainage is impossible to achieve or if ground or surface water conditions exist near F-4 soils, they should be removed and replaced with less susceptible materials, preferably of the same type as is adjacent to the problem area. The depth of removal should be at least 80 percent of the depth of frost penetration in the area (Fig. 11). Economics and practicality of the specific situation should govern the exact quantity of material to be removed.

Variations in soils cause differential frost heaves and weakening effects. In fills, where the material is randomly deposited and intermixed in the grading operation, the chances of local variations in soils are small. In cuts, soils will be of a more variable nature; therefore, consideration should be given to blending at least the top 12 in. of subgrade in cut areas. To reduce the potential hazards of frost heave and weakening, the present practice of holding surface grade at least 5 ft above free or groundwater should be continued.

Non-frost-susceptible materials should be used in the select materials of all pavement. In all cases, a minimum depth below pavement surface of 20 in. of non-frost-susceptible material should be provided. Presently, insufficient data are available to determine quantitatively the regional factor. If the regional factor is ignored, the resulting design may not be adequate. Thus, a factor should be introduced to account for: (a) frost weakening effects, (b) Massachusetts construction practice differing from the super-controlled procedures used at the AASHO Road Test, (c) unknown differences in materials properties from those used at the AASHO Road Test, (d) effects of mixed traffic, (e) effects of traffic application over longer periods of time than was possible at the AASHO Road Test, and (f) other unknown factors. It is recommended that a blanket increase of 15 percent be applied to the structural requirements of the pavement. In fact, this is equivalent to a regional factor of 3.

## PAVEMENT AS A STRUCTURE

### Structural Approach

The design method evolved here is, in essence, a structural design. It differs significantly in its approach from those normally used by the civil engineer in designing bridges, buildings or other structures. Designers of the latter structures usually can estimate the loadings, the stresses within the structure, and the behavior of the materials that they will use. The immensity of the variables that accompany a flexible pavement design have precluded a full rational design treatment, and so the AASHO Road Test and resulting design guide depend heavily on empirical results. Although the state of the art in design theory and materials understanding cannot yet replace empirical methods, it can supplement the AASHO approach and contribute to an understanding of the behavior of the layered pavement as a structural system.

### Applicable Theories

There are several theoretical solutions for the stresses and deflections existing in a pavement section when subjected to wheel loads. One simplified solution is that developed by Boussinesq (8) using the theory of elasticity. As it is applied to pavement sections, it must be assumed that the materials in the section have the same elastic properties, i.e., modulus of elasticity and Poisson's ratio. It also must be assumed that the material is elastic. Whereas neither of the assumptions are true for particulate layered systems, the Boussinesq solution provides a first approximation of pavement stresses.

A solution that is more refined than the Boussinesq solution has been offered by Burmister (9). The Burmister solution theoretically is more applicable to the pavement system because it can account for the differing properties in each layer. The solution still assumes that the materials are elastic and, therefore, it suffers some of the weaknesses of the Boussinesq solution.

A third solution extended from the Westergaard (10) theory by our Materials Research Laboratory (11) makes significantly different assumptions than were made by either Boussinesq or Burmister. This theory assumes that an elastic plate rests on a dense liquid foundation. It maintains that the surface layer, the plate, behaves elastically and that the foundation is made up of a bed of independent springs. Thus far, this theory has been developed to consider stresses in the asphalt-bound layer only and the vertical stress applied to the layer below.

### Stresses in Pavement Layers

Vertical Stress.—These three theories involve different stiffness parameters and, therefore, it is difficult to compare rigorously values obtained with them. By using some approximations, the plots of variations in vertical stress with depth were obtained (Fig. 13). The Boussinesq theory yields results that are reasonably valid for the static case only, whereas the Burmister and Westergaard theories can be applied to the dynamic or moving load case because they can account for differing properties of the pavement layers.

The plots in Figure 13 clearly illustrate the important structural functions of the surface layer. First, it must support a high stress at the surface; most soils cannot

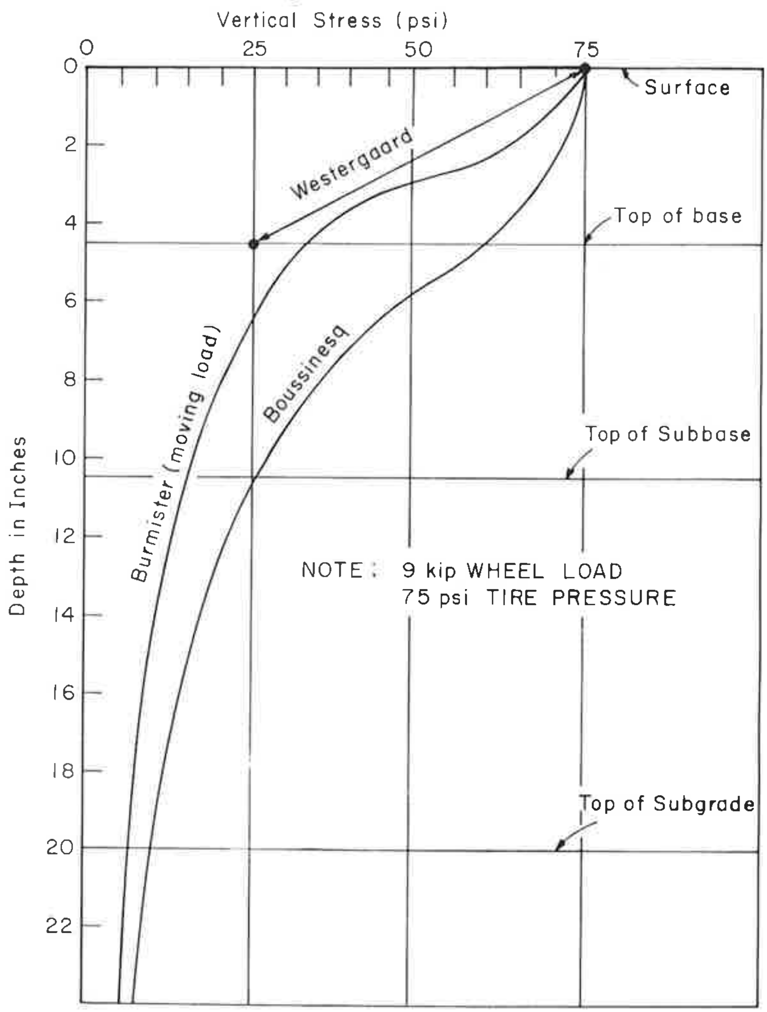



Figure 13. Variations in vertical stress with depth according to various theories.

support such stresses without failure or excessive deformation. Second, for the dynamic moving load case, the surface layer absorbs a significant amount of vertical stress, thereby reducing the stress on the layers below. For the dynamic case, the stress on the top of this particular base layer is around 40 percent of that exerted by the tire at the surface. A reduction of only 20 percent is expected for the static case. If the surface layer is thinner than the $4\frac{1}{2}$ in. shown, the stress reduction will be less, whereas thicker layers will produce the converse result. Thus, the surface layer serves a dual structural function. It carries the high concentrated surface stress and, if thick enough, is capable of reducing the vertical stress applied to the layers below.

The structural role of the base is also illustrated in Figure 13. Despite the stress reduction provided by the surface layer, the top of the base layer is exposed to a substantial stress. For the illustration shown, this stress is about 30 psi. The depth of the base layer is caused for further stress reduction so that the stresses are reduced by about one-half for the 6-in. depth. As with the surface layer, the base resists high stress at its surface and reduces stress on the layers below.

The structural role of the subbase follows that of the base. It still must resist a reasonably high stress of 15 psi which reduces by about one-half over the depth of the

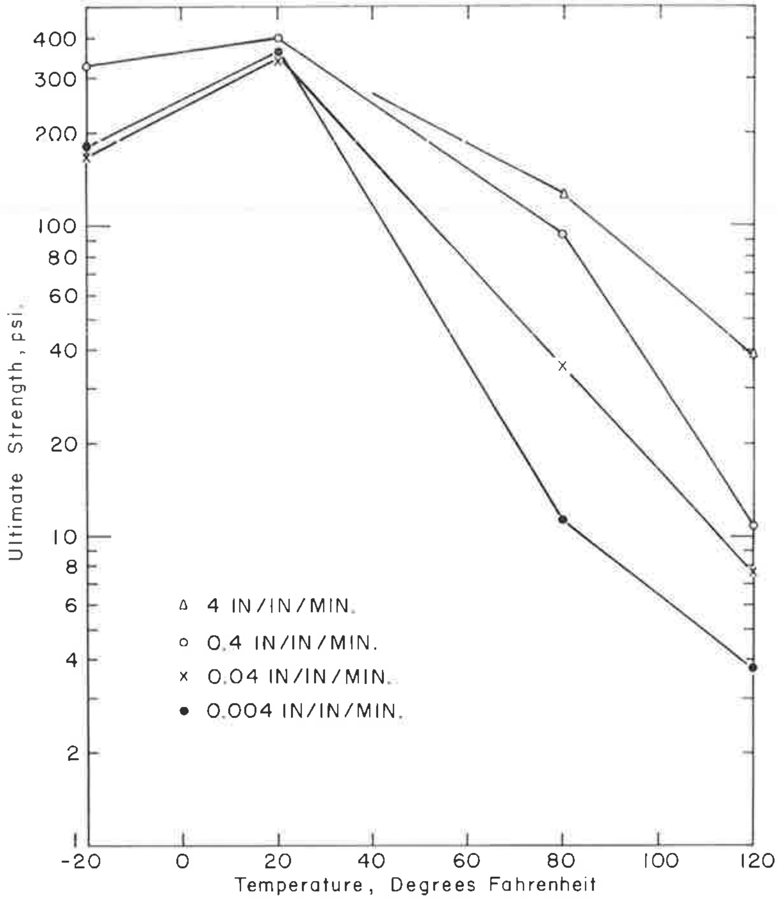



Figure 14. Tensile strength of Massachusetts Class I bituminous concrete top course at various temperatures and rates of loading.

layer. This layer, which is usually gravel in Massachusetts, is in a zone of reasonably high stress; it is important to maintain its quality at a high level because it must act as a drainage layer to select its gradation so that it is not weakened by water.

The subgrade, of course, is the ultimate support for the structure. The stresses on it are a direct function of the action of the layers above. The stresses that are applied to the subgrade surface eventually die out at great depths. The important consideration here is that the materials above distribute the stress so that this layer is subjected to tolerable stress levels.

Each layer in the pavement structure has a dual role to fulfill. First, the layer must be strong enough to sustain the load from above without failure or excessive permanent deformation. Second, it must be stiff and/or deep enough to keep the stress on the layer below at a tolerable level. Although these functions are qualitatively clear from the preceding analysis, the exact parameters that govern these functions have not yet been fully developed.

Stresses in Surface Layer.—The theoretical analysis just discussed dealt with the variations of vertical stress with depth and materials of the pavement section. Theory has been developed to explore further the stresses in the critical surface layer, particularly as it acts as a plate. This plate action occurs when the moving load condition is coupled with the materials properties under dynamic loads. The action is one of stiffening so that the applied load tends to spread over a large area of the underlying layers. In so doing, flexural stresses are brought into play. To provide an estimate

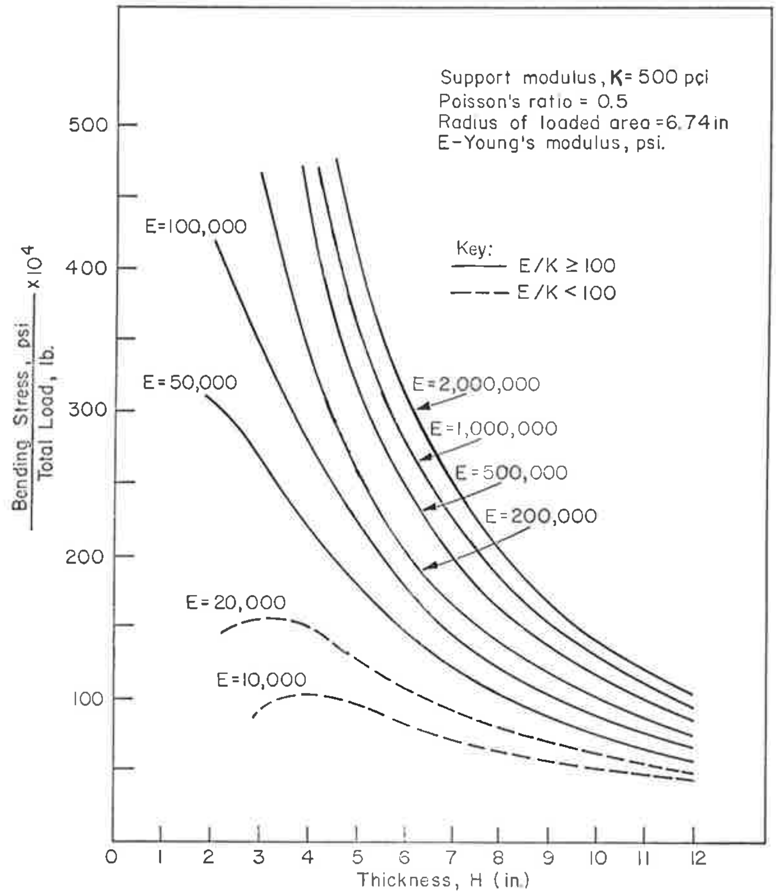



Figure 15. Bending stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

of the severity of stress in the surface layer, the dynamic properties of the bituminous surface are used in conjunction with layered pavement analysis.

Bituminous concrete, being a viscoelastic material, displays time- and temperature-dependent properties. Ideally, from viscoelastic theory, the effects of time and temperature are superimposable. That is, the modulus of a viscoelastic material may be altered from a given value by changing either temperature or the rate of loading. Thus, a material will offer more resistance to deformation if the rate of loading is increased or the temperature of the material is depressed. Clearly, no single value of modulus can be representative of bituminous concrete. A reasonable value for normal temperature and vehicle velocity can be taken as 100,000 psi for purposes of the ensuing discussion.

The tensile strength of bituminous concrete will also vary with temperature and rate of loading. Figure 14 is a plot of tensile strength of bituminous concrete with rate and temperature of test (12). At temperatures below the glass transition point of asphalt, the tensile strength is relatively unaffected by test rate and has a value of a few hundred psi. At temperatures above the glass transition point, test rate has significant influence on the tensile strength. Loading times at even slow vehicle speeds are much less than a second, representing several orders of magnitude faster loading than those shown in Figure 14. The limiting strength under such high loading rates would approach the strength observed at or below the glass transition point. A reasonable estimate of this value from the plot is 300 psi for Massachusetts-type bituminous surface course.

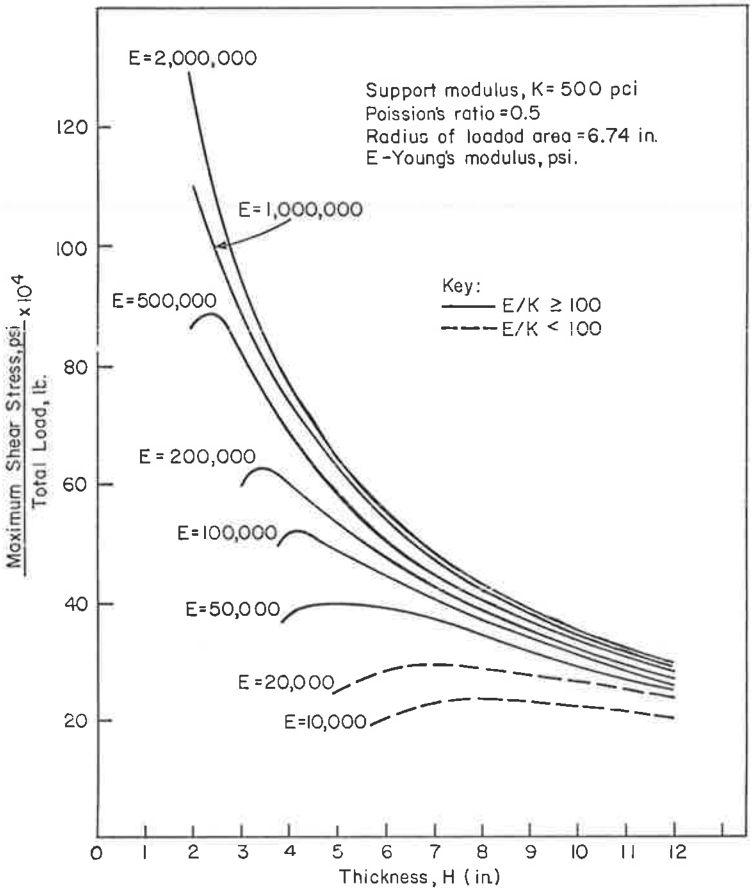



Figure 16. Maximum shear stress vs thickness of asphalt-bound layer for 10-kip load, 70-psi tire pressure.

Westergaard's Analysis Applied to Bituminous Concrete.—The Westergaard analysis was originally developed for use in concrete pavement design. It remains as the principal method by which airfield pavement is proportioned. Putnam (13) extended the Westergaard analysis to cover a broader range of problems than was considered in the original work. This enabled Hagstrom et al. (11) to study bituminous concrete using the Westergaard theory.

The accuracy of the Westergaard approach depends on certain conditions of relative stiffness. If the stiffness of the plate or top layer is substantially higher than that of the supporting foundation, the theory will be reasonably accurate. Such is the case if the dynamic properties of bituminous concrete are used in conjunction with reasonable values for the supporting media.

Figure 15 is a plot of the variation in bending stress with pavement thickness as derived by Hagstrom. Constant reasonable values are assumed for the support modulus (500 pci) and loading (a 10,000-lb single load simulating two 5,000-lb single wheels). The range of moduli shown represents the variation of the modulus of bituminous concrete that may occur under reasonable rates of loading and temperature conditions. The dashed curves represent areas where application of the theory is questionable because the stiffness of the surface layer is not sufficiently large when compared to the foundation stiffness. The curves in Figure 15 illustrate the strong influence that depth of section has in reducing bending stress within the top layer. For example, with the conditions discussed earlier, a modulus of 100,000 psi and tensile strength of 300 psi, the depth required so that the section would just fail under the loading is $3\frac{1}{2}$ in. If fatigue effects are considered, the allowable working stress might be about one-half of the test value, or 150 psi. Thus, to sustain repeated loadings, a depth of 7 in. would be required. Of course, tensile failure of the flexural section implies cracking on the under surface that will eventually reflect through to the surface. Clearly, to reduce cracking, the bituminous layers should be as thick as economically possible.

The previous analysis probably relates directly to the studies of black bases at the AASHO Road Test. It was found that black bases offered significant improvement in pavement performance. The well-graded cohesive base layers served to add depth and, hence, flexural capacity to the surface layer. This would then reflect itself in retention of serviceability.

There are a number of assumptions necessary to develop the preceding discussion. But changing the modulus of the bituminous layer by a factor of two alters the stress by perhaps 10 percent. Varying the foundation modulus by a factor of two changes the bending stress by a maximum of 10 percent at reasonable depths. More recent work by Hagstrom shows that a more exact analysis for dual wheels yields results about 20 percent lower than assumed. On the other hand, wheel loads may in practice exceed those assumed here. Significant latitude in assumptions is possible without altering the general conclusion regarding the analysis.

Hagstrom also obtained values for shear stress existing in the bituminous layer. Figure 16 shows the variation in shear stress with pavement thickness for the same conditions assumed in the bending analysis. For the practical range of moduli and thickness of the bituminous layer, the shear stress is usually less than 100 psi. Although dynamic values of the shear strength of bituminous concrete are not available, it would be expected that the shear strength is of the same order as the tensile strength. Thus, shear strength does not appear to be a criterion governing the top layer of pavement.

## COEFFICIENTS OF RELATIVE STRENGTH FOR SUBBASE, BASE AND SURFACE

The preceding theoretical discussion indicated that conventional theory and test values could be applied to the surface layer to provide a reasonable estimate of failure conditions. However, as subsequent layers are treated, the analysis becomes more complex and the important parameters remain unknown. At present, there are severe limits to theory as it applies to the complex action of these non-cohesive layers.

In the absence of theoretical treatment of the layered pavement system, empirical methods were used at the AASHO Road Test. A statistical factorial experiment was designed in which several thicknesses and combinations of materials were placed in a number of test sections. By monitoring the traffic and serviceability, the relative contribution of each material to performance could be obtained. This relative contribution to performance can be assigned to the pavement materials in the form of coefficients.

### Coefficients of Various Layers

Table 2 gives the coefficients for various materials as provided by the AASHO guide. Of interest is the relative contribution that each material makes to the performance of a section. For example, 1-in. plant-mixed bituminous concrete (coefficient = 0.44) is equivalent to 3 in. of crushed stone (coefficient = 0.14) or 4 in. of gravel (coefficient = 0.11). This is the first time that quantitative relative values for the performance of materials have been made available to the designer. Although there is as yet no theoretical grasp or means of measuring the coefficients of materials that differ from the AASHO materials, a valuable incentive has been established.

Current Massachusetts practice is to use a penetrated crushed stone base. Because the AASHO guide did not give coefficients for this material, a value had to be estimated for design purposes. The coefficient for crushed stone in the base layer is 0.14, and the coefficient for a black base is 0.34. The strength of a penetrated base should be greater than crushed stone because it is bound by asphalt. Yet the penetrated base cannot be as effective as a well-graded black base because of its open gradation and incomplete penetration of the asphalt. It was decided on the basis of these considerations

TABLE 2

MATERIALS COEFFICIENTS

| Pavement Component | Coefficient | | |
|---|---|---|---|
| | $a_1$ | $a_2$ | $a_3$ |
| **(a) Surface Course** | | | |
| Road-mix (low stability) | 0.20 | - | - |
| Plant-mix (high stability) | 0.44[a] | - | - |
| Sand asphalt | 0.40 | - | - |
| **(b) Base Course** | | | |
| Sandy gravel | - | 0.07 | - |
| Crushed stone | - | 0.14[a] | - |
| Cement-treated (no-soil-cement): | | | |
| ≥650 psi[b] | - | 0.23[a] | - |
| 400-650 psi | - | 0.20 | - |
| ≤400 psi | - | 0.15 | - |
| Bituminous-treated: | | | |
| Coarse-graded | - | 0.34[a] | - |
| Penetrated stone | - | 0.24[c] | - |
| Sand asphalt | - | 0.30 | - |
| Lime-treated | - | 0.15-0.30 | - |
| **(c) Subbase** | | | |
| Sandy gravel | - | - | 0.11[a] |
| Sand or sandy clay | - | - | 0.05-0.10 |

[a]Based on results of AASHO Road Test, all other coefficients estimated.
[b]Compressive strength at 7 days.
[c]Estimated by Massachusetts Design Committee.

that a simple average of the coefficients for crushed stone and black base would be a reasonable approximation. Thus, a value of 0.24 was selected as the coefficient for penetrated stone.

Structural Number Concept

The coefficients as derived can be employed directly in the design of the pavement. The anticipated traffic, environmental and soil conditions can be combined to yield a required structural number (SN). The SN can be derived from any arrangement of materials for which the coefficients are known. The thickness of each material multiplied by its coefficient yields the materials contribution to the total SN. Obviously, the types and thicknesses of materials can be manipulated, within limits, to produce a section to meet economic or other criteria.

## EQUIVALENT DAILY 18-KIP AXLE LOADING

In addition to a knowledge of the characteristics of soils and paving materials, it is necessary for the adequate design of pavement elements to have certain traffic data and a criterion for the quality of service that is expected for a specified length of time.

Most of the traffic data required for geometric design will be used for the structural design of pavements. These include present average daily traffic (present ADT), future average daily traffic (future ADT), and the percentage of trucks (T). For structural design purposes, the traffic is assumed to be equally divided between the two directions. In addition to these traffic parameters, the daily overall truck traffic and lane distribution of trucks must be ascertained.

An analysis of the relationship between T for the entire day and ADT on Massachusetts highways shows that the overall daily truck traffic is three or more times greater than the peak-hour truck percentage. Because of this, a multiplying factor of 3 will be used to convert the T given for geometric design purposes to obtain the overall daily truck percentage.

Because there are no local data available relative to the distribution of truck traffic by lanes, the recommendations of two recognized organizations will be used. The Thickness Design Manual Series No. 1 (Ms-1), 7th edition, September 1963, by the Asphalt Institute and the Manual of Instructions for Pavement Evaluation Survey, August 1962, by AASHO suggest the following, in effect, identical distribution percentages for the most heavily traveled lanes:

1. If there are four traffic lanes (two in each direction), the percentage of trucks using the design lane is 90 percent of the total number of trucks in one direction.
2. If there are six or more lanes (three or more in each direction), the percentage of trucks using the design lane is 80 percent of the total number of trucks in one direction.

These lanes will hereafter be designated as the design lanes, i.e., the thickness of all other lanes will be the same as that of the design lane.

The AASHO method of determining the relative effects of different axle loadings on pavement performance is used in this design procedure. The AASHO concept relates the destructive effects of mixed axle loads to equivalent 18-kip single-axle loads. The distribution and magnitude of the various axle loads are indicated in Tables W-4A, W-4B and W-4C of the Massachusetts Truck Weight Study. (Truck Weight Study and Loadometer Study are used interchangeably; they refer to the same data.) The "All Stations, All Systems" tabulation of the Truck Weight Study is the source of these data.

The axle-load intervals used in the Truck Weight Study differ from those suggested in the AASHO guide. This required that new equivalence factors be established to fit the intervals of the Truck Weight Study. This was done by plotting the equivalence factors for small intervals (3, Appendix F), and then taking the average value of the interval of interest from the plot. The choice of Truck Weight Study axle-load intervals makes it possible to take advantage of the existing Massachusetts Traffic Planning Department computer program and eliminate the regrouping of axle load.

When the traffic data and truck weight study data are applied as shown in the Appendix to the AASHO guide, a value $T_{18}$, i.e., equivalent daily 18-kip applications, is obtained. The value of $T_{18}$ is applied to the design chart in combination with soils data (DBR) to obtain a SN from which pavement layer thicknesses can be determined.

## STRUCTURAL DESIGN CHARTS

There are two basic structural design charts included herein: one for a Terminal Serviceability Index (TSI) of 2.5 for high-volume high-speed roads (Fig. 17), and the other for a TSI of 2.0 for less traveled roads (Fig. 18). Any proposed combinations of pavement materials may be explored using these charts as a basis.

### Variables

The charts were obtained using the AASHO guide, correlating the DBR scale with the AASHO S scale and increasing the SN value by 15 percent for the various traffic volumes.

The design numbers needed before the SN can be obtained are: (a) DBR of the soil, (b) equivalent 18-kip daily loadings, and (c) materials and their strength coefficients used in the pavement layers.

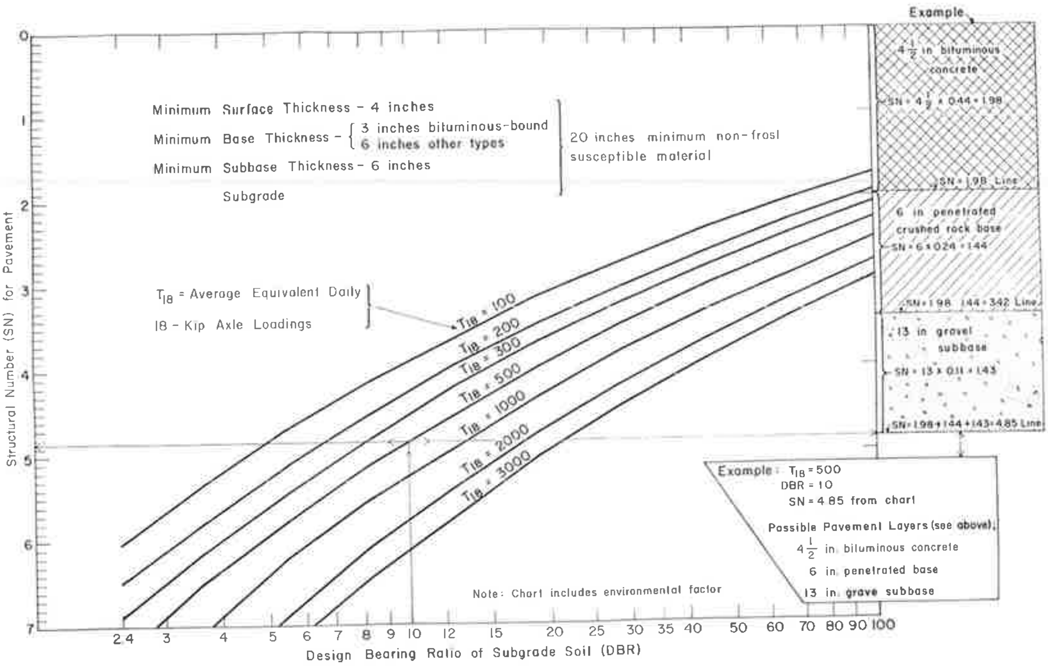



Figure 17. Structural design chart for pavements with TSI = 2.5 and 20-yr traffic analysis (for interstate, federal-aid-primary, and major state highways).

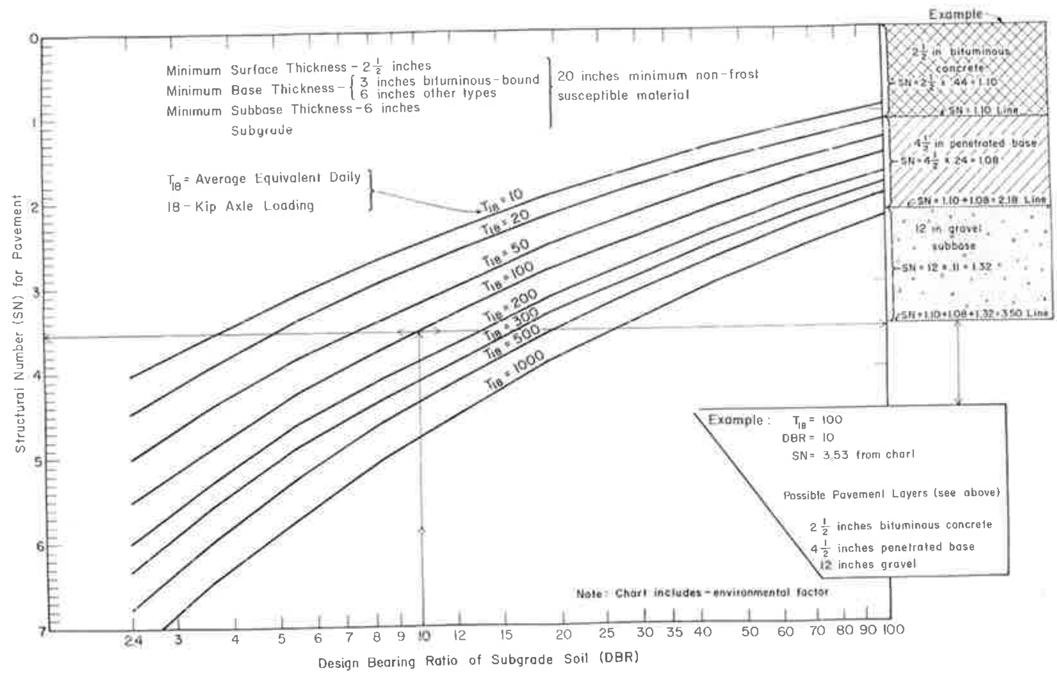



Figure 18. Structural design chart for pavements with TSI = 2.0 and 20-yr traffic analysis (for roads other than interstate, federal-aid-primary, and major state highways).

Design Using Various Materials

The two basic design charts (Figs. 17 and 18) can be used to design with various types of materials, provided their structural coefficients are known or can be estimated. An example for the design is given on the charts. It must be emphasized that the charts give a required SN and not pavement thickness for a given design DBR and traffic combination. Because each material in the layered system has a different strength coefficient, the total thickness of a pavement will vary for a given DBR and traffic number (DBR and $T_{18}$).

The charts can be used for any road. One can assume certain thickness of asphaltic concrete for the surface, multiply the strength coefficient of this layer by the thickness, and add necessary rock or gravel layers for a foundation to meet the required SN.

Bituminous Concrete/Penetrated Base/Gravel.—Current Massachusetts practice in flexible pavement design is to use a bituminous concrete surface supported by penetrated crushed rock base and gravel. If the designer wants to use these particular materials and knows the desired thicknesses of the surface and the base, a specific design chart can be easily prepared. For instance, a popular cross-section may have 4½-in. bituminous surface, 6-in. penetrated base, and variable gravel thickness. The thickness of each layer is multiplied by the appropriate strength coefficient and a SN for each material can be represented on the chart. The main variable is the gravel layer, except where the DBR's are high or the traffic number, $T_{18}$, is low.

Again it must be emphasized that the charts are based on SN rather than pavement thickness. For illustration, Figure 19 shows the result of replacing the SN with thickness for a specific combination of materials.

Bituminous Concrete/Black Base/Gravel.—A similar chart can be prepared for pavement with a bituminous base (black base). The 6-in. penetrated crushed rock base is approximately equal to 4 in. of black base.

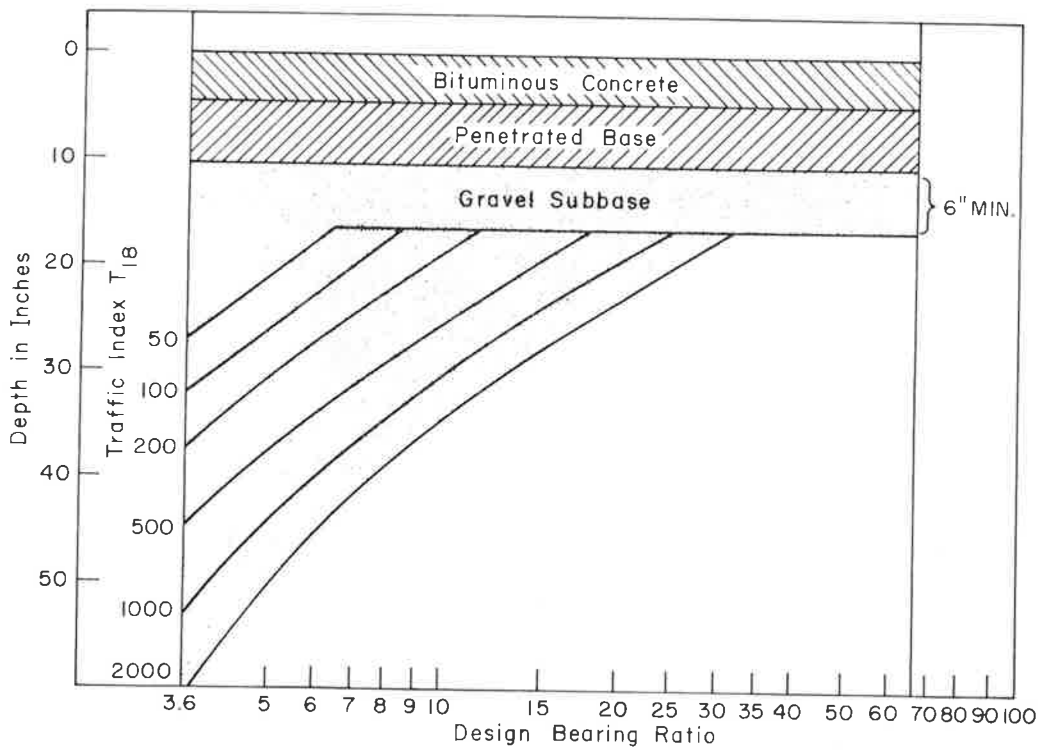



Figure 19. Thickness of pavements for a given combination of materials (not to be used for design purposes); TSI = 2.5.

## CONCLUDING REMARKS

The layered pavement design method developed for Massachusetts recognizes characteristics of soils, traffic, and materials as the parameters of major influence in the retention of serviceability of a pavement. Where possible, the results of the AASHO Road Test were used to assess quantitatively the influence of the major parameters. In addition, the special problems related to frost penetration into pavements, frost heaving and spring breakup have been analyzed and recommendations have been made concerning detection, control and treatment of frost-susceptible materials. The pavement structure was analyzed using layered theory to demonstrate the critical action of each layer. The importance of using deep surface layers to reduce cracking was particularly evident from this analysis. The procedure has been developed in such a way that it can be integrated into the sequence of present Massachusetts practice in highway planning, design, and construction.

It was recognized at the outset of this study that although the design method to be developed should represent an improvement over existing constant section design, there would be a necessity for further research in areas where knowledge was found to be particularly deficient. Future research should be directed toward the evaluation of the structural support offered by soils and pavement materials, and variations in this support with environmental conditions. Results of this work should provide a better means for assessment and control of the structural capacity of the pavement components than is currently available.

## ACKNOWLEDGMENTS

The work described herein was carried out under the sponsorship of the U. S. Bureau of Public Roads and the Massachusetts Department of Public Works. In the course of this work, the authors called on members of many state highway and federal agencies and drew on the special talents of many in the Massachusetts Department of Public Works and at the Massachusetts Institute of Technology. The number of contributors is so great that it is not possible to give individual credits; the help of all has been invaluable.

## REFERENCES

1. The AASHO Road Test, Report 3: Traffic Operations and Pavement Maintenance. Highway Research Board Spec. Rept. 61C, 1962.
2. Carey, W. N., Jr., and Irick, P. E. The Pavement Serviceability-Performance Concept. Highway Research Board Bull. 250, pp. 40-58, 1960.
3. AASHO Interim Guide for the Design of Flexible Pavement Structures. Oct. 12, 1962.
4. Shook, J. F., and Fang, H. Y. Cooperative Materials Testing Program at the AASHO Road Test. Highway Research Board Spec. Rept. 66, 1961.
5. Engineering and Design, Pavement Design for Frost Conditions. U. S. Army, Corps of Engineers, Rept. EM 1110-1-306, May 1962.
6. Aldrich, H. P. Frost Penetration Below Highway and Airfield Pavements. Highway Research Board Bull. 135, pp. 124-149, 1956.
7. Stokstad, O. L. Frost Action in Michigan. Symposium on Frost Heave and Frost Action on Soils. Highway Research Board Spec. Publ., 1951.
8. Boussinesq, J. Application of Potential Theory in a Study of Equilibrium and Motion of Elastic Solids. Paris, 1885.
9. Burmister, D. M. The General Theory of Stresses and Displacements in Layered Soil Systems. Jour. of Applied Physics, Vol. 16, 1945.
10. Westergaard, H. M. Theory of Concrete Pavement Design. Highway Research Board Proc., Vol. 7, Pt. 1, pp. 175-181, 1927.
11. Hagstrom, J., Chambers, R. E., and Tons, E. Low Modulus Pavement on Elastic Foundation. MIT, Dept. of Civil Engineering, Res. Rept. R64-11, April 1964.
12. Tons, E., and Krokosky, E. M. Tensile Properties of Dense Bituminous Concrete. Proc. 38th Annual Meeting AAPT.

13. Putnam, J. O. Theoretical Analysis of the Behavior of Rigid Airfield Pavements for the Case of Interior Loading. Unpublished Doctoral Thesis, MIT, 1963.
14. Helmer, R. A. Nuclear Moisture—Density Research Project. Oklahoma Highway Dept., 1963.
15. The AASHO Road Test. Highway Research Board Spec. Repts. 61A-G, 1962.
16. Proc. Int. Conf. on the Structural Design of Asphalt Pavements, Univ. of Michigan, Aug. 1962.
17. The AASHO Road Test. Highway Research Board Spec. Rept. 73, 1962.