TRANSPORTATION RESEARCH RECORD

No. 1451

Public Transit

Public Transportation 1994: Current Research in Planning and Development; Management, Marketing, and Fare Policy; and Technology

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1994

Transportation Research Record 1451
ISSN 0361-1981
ISBN 0-309-06059-1
Price: $26.00

Subscriber Category
VI public transit

Printed in the United States of America

Sponsorship of Transportation Research Record 1451

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Chairman: Thomas F. Humphrey, Massachusetts Institute of Technology

Public Transportation Section
Chairman: Subhash R. Mundle, Mundle & Associates, Inc.

Committee on Public Transportation Planning and Development
Chairman: Patricia V. McLaughlin, Los Angeles County Metropolitan Transit Authority
Secretary: David R. Miller, Parsons Brinckerhoff et al.
Paul J. Ballard, Edward A. Beimborn, Chester E. Colby, Sally Hill Cooper, Stephen P. Gordon, George Edward Gray, S. Olof Gunnarsson, Brendon Hemily, Jay A. Higle, Nathan L. Jaschik, Hermann Knoflacher, Perry J. Maull, Marianne A. Payne, Patti Post, James P. Redeker, Jacques P. Roulet, G. Scott Rutherford, Dennis H. Ryan, George M. Smerk, Katherine F. Turnbull, Samuel L. Zimmerman

Committee on Rail Transit Systems
Chairman: S. David Phraner, Port Authority of New York and New Jersey
Secretary: John P. Aurelius, New Jersey Transit Corporation
John A. Bailey, Jack W. Boorse, Thomas J. Carmichael, Peter Fahrenwald, Alex E. Friedlander, Thomas R. Hickey, Wendy J. Hoyt, Robert J. Landgraf, Thomas F. Larwin, Herbert S. Levinson, Richard S. Marshment, R. David Minister, David J. Mitchell, Marianne A. Payne, Steven E. Polzin, Lee H. Rogers, P. Takis Salpeas, Peter J. Schmidt, John W. Schumann, John D. Wilkins, Nigel H. M. Wilson

Committee on Public Transportation Marketing and Fare Policy
Chairman: William R. Loudon, JHK & Associates
Carol Pederson Ambruso, Jitendra N. Bajpai, Beth F. Beach, A. Jeff Becker, Daniel K. Boyle, Rita Brogan, Roslyn A. Carter-Phillips, Daniel M. Fleishman, Thomas J. Higgins, J. David Jordan, Janet E. Kraus, Sarah J. LaBelle, Pierre Laconte, Robert R. Lockhart, Gail Murray, Thomas E. Parody, Marilyn M. Reynolds, Peter M. Schauer, Kenneth O. Stanley, Cy Ulberg, William S. Vickrey, Donna L. Vlasak, Linda M. Zemotel

Committee on New Transportation Systems and Technology
Chairman: Rolland D. King
Murthy V.A. Bondada, Daniel Brand, Salvatore Castronovo, Ovi Colavincenzo, Chester E. Colby, Vincent R. DeMarco, G. Bruce Douglas III, Lawrence J. Fabian, Franz K. Gimmler, Milton D. Harmelink, William Hurrell, Dennis Kershner, Shinya Kikuchi, Thomas J. McGean, Richard M. Michaels, Edward S. Neumann, William J. Sproule, Peggy Tadej, Katherine F. Turnbull, John P. Walsh

GROUP 2—DESIGN AND CONSTRUCTION OF TRANSPORTATION FACILITIES
Chairman: Charles T. Edson, Greenman Pederson, Inc.

Railway Systems Section
Chairman: Scott B. Harvey, Association of American Railroads

Committee on Rail Transit System Design
Chairman: Charles L. Stanford, Greater Cleveland Regional Transit Authority
Secretary: Donald A. Shoff, Daniel Mann Johnson Mendenhall
Robert C. Arnlund, Robert E. Clemons. Ronald H. Dunn, Michael E. Duval, George A. Figueredo, Earle M. Hughes, Bertie Jackson, Mohammad S. Longi, Vincent P. Mahon, Philip J. McQueen, James Tuman Nelson, Patrick A. Nowakowski, Don H. Painter, Robert Schumacher, Mahesh J. Sharma, Andrew Sluz, Roy T. Smith, Erland A. Tillman, John Insco Williams

Transportation Research Board Staff
Robert E. Spicher, Director, Technical Activities
Peter L. Shaw, Senior Program Officer, Public Transportation
Elaine King, Rail Transport Specialist
Nancy A. Ackerman, Director, Reports and Editorial Services
Marianna Rigamer, Oversight Editor

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1993.

Transportation Research Record 1451

Contents

Part 3—Technology

Foreword

The papers in this volume, which were presented at the 1994 TRB Annual Meeting, report new research in planning and management of public transportation. Each paper, in accordance with established TRB procedures, has been reviewed by peers (practitioner and academic) in the field of public transportation. New ideas are explored and improved practices discussed. Potential application is real and holds significant promise of utility and better customer service.

Part 1, Planning and Development, addresses six subjects. Urban decentralization has had significant impact on transit service and makes it difficult for transit to survive in the suburbs (Cervero). For traditional urban corridors, a new framework to determine bus transit coverage has been developed for American cities (Spasovic et al.) and proposed for improved quantitative analysis for Tel Aviv, Israel (Vovsha and Goodovitch). To pull it all together, a multiattribute utility theory for transit decisionmaking may help improve transit service design (Reed et al.). Expert systems are also reviewed for their value in deciding among transit technologies (Mackett). When the decision to build has been made, the civil/utilities drawings and information included by transit facility construction contract documents are needed (Berliner).

Part 2, Management, Marketing, and Fare Policy, focuses on three explorations of transit user information. A basic part of customer service is being on-time, thus the importance of a causal model (Henderson and Darapaneni). A comprehensive review of transit fare policies at large systems offers useful insights (Hinebaugh and Boyle). A study of human versus automated telephone information systems found that callers preferred ''live'' people (Hall et al.).

Part 3, Technology, considers three technological operational challenges. In Gothenburg, Sweden, personal rapid transit (PRT) appears to be a viable candidate for large-scale operation (Blide). What to do with empty PRT vehicles in Gothenburg and Gävle, Sweden, was studied (Andréasson). On an even larger level, a 11.6 kilometer/12 station automated people mover system opened in Taipei, Taiwan, and is designed to carry more than 27,340 passengers an hour (Shen and Lai).

The research discussed in this volume demonstrates the depth and breadth of transit research in the past few years and sets the stage for further advances in the future.

PART 1

Planning and Development

Making Transit Work in Suburbs

ROBERT CERVERO

Rapid decentralization of population and employment over the past several decades has chipped away at the U.S. transit industry's market share. The implications of decentralization on the ridership, operating performance, and fiscal health of the nation's largest transit operators are examined. On the basis of the results of a national survey, a number of service strategies that offer hope for reversing transit's decline are explored, including timed transfers, paratransit services, reverse commute and specialized runs, employer-sponsored van pools, and high-occupancy-vehicle and dedicated busway facilities. Land use options, like traditional neighborhood designs and transit-based housing, are also examined. A discussion of various institutional, pricing, and organizational considerations when implementing suburban-targeted service reforms and land use initiatives is also provided. Century-old models involving joint public-private development of communities and transit facilities, it is argued, also deserve reconsideration.

The ongoing decentralization of U.S. cities continues to plague the nation's transit industry. Today transit competes with the automobile in an environment of low densities, dispersed trip patterns, abundant free parking, cheap fuel prices, and inhospitable walking environs. It is losing the competition. From 26 billion passengers in 1946, U.S. transit patronage fell steadily for 30 years, reaching 8.8 billion in 1980. Through the 1980s the total number of transit riders remained roughly the same, but those numbers represented a smaller share of commute trips, from 6.4 percent in 1980 to 5.3 percent in 1990 (*1*).

This paper explores the challenges of making transit work in the suburbs—that is, making it viable, competitive, and sustainable. Performance statistics are used to compare suburban and urban transit operations in the United States. On the basis of the results of a national survey of suburban transit operations, the paper then turns to various service strategies that offer mass transit the most promise in competing with the private automobile in suburbia. The paper ends with a discussion of institutional, pricing, and land use considerations.

The challenge of making transit work in suburbia is not new. In the keynote address at the 1940 meeting of the American Transit Association, H. Bartholomew (*2*) warned, ''Can we not pause long enough in this headlong decentralization process to see where we are going? The mass transportation industry is caught in a strong tide which is sweeping this and many other businesses toward disaster.''

DECENTRALIZATION AND TRANSIT

Transit's falling fortunes in suburbia are an outcome of many factors. Traditional fixed-route services radially linked to downtowns are ill-suited for lateral suburb-to-suburb journeys, the most rapidly growing travel market (*3,4*). Also the densities and built environment of U.S. suburbs are generally not conducive to transit riding. A recent survey of several thousand office workers whose jobs were relocated from downtown San Francisco to the 560-acre Bishop Ranch Office Park found that transit's modal split plummeted from 58 percent before the move to under 3 percent after the move (*5*).

Demographics and institutions also work against transit in suburbia (*6*). Suburban residents and workers tend to be more affluent and own more cars than do their central-city counterparts. Suburbs also produce high rates of off-peak and weekend travel, when bus headways tend to be longest. Service coordination is also sometimes hampered by a multitude of competing suburban jurisdictions. In the San Francisco Bay Area, for instance, some two dozen separate transit agencies operate bus services outside of central cities.

Suburbanization and Transit Commuting

How has decentralization had an impact on transit? The following statistics were drawn to address this question for the nation's largest metropolitan areas (by using 1980 and 1990 census data from Summary Tape File 3A). Figure 1 shows that suburban population and employment grew rapidly in the four largest consolidated statistical areas (CSAs) in the United States. (For each CSA the suburbs are defined as areas outside the central city, using U.S. Bureau of the Census definitions of what constitutes a central city.) Suburbanization of jobs was the dominant trend, increasing on average 50 percent in the four CSAs compared with only 13 percent in their central cities.

The movement of jobs from the metropolitan core to the metropolitan periphery and beyond has been spurred by postindustrialization—the restructuring of the U.S. economy from a predominantly manufacturing base to a service and information processing economy. For example by 1990 New York City, Philadelphia, and Boston each had more employees in white-collar service industries—in which executives, managers, professionals, and clerical workers dominate—than in the manufacturing, construction, retail, and wholesale industries combined (*7*). Although many decentralized jobs have involved back-office support functions, corporate headquarters and entire companies in fields such as finance, retailing, and wholesaling are increasingly relocating to the suburbs (*8*). And where jobs and people go, so does retailing. New York's suburban ring now has 48 fully enclosed regional malls encompassing 49 million ft^2 of retail space (*9*).

Paralleling the rapid suburban growth has been a diminishing role for transit. Transit commutes actually fell by about 50,000 trips per day in the Chicago region during the 1980s and increased only slightly in the other three large metropolitan areas. In all four metropolitan areas, transit's modal share fell between 1980 and 1990; in the greater New York area this fall was by 10 percentage points (Figure 2). This trend was hardly limited to the biggest

---

Department of City and Regional Planning, University of California, Berkeley, 228 Wurster Hall, Berkeley, Calif. 94720.

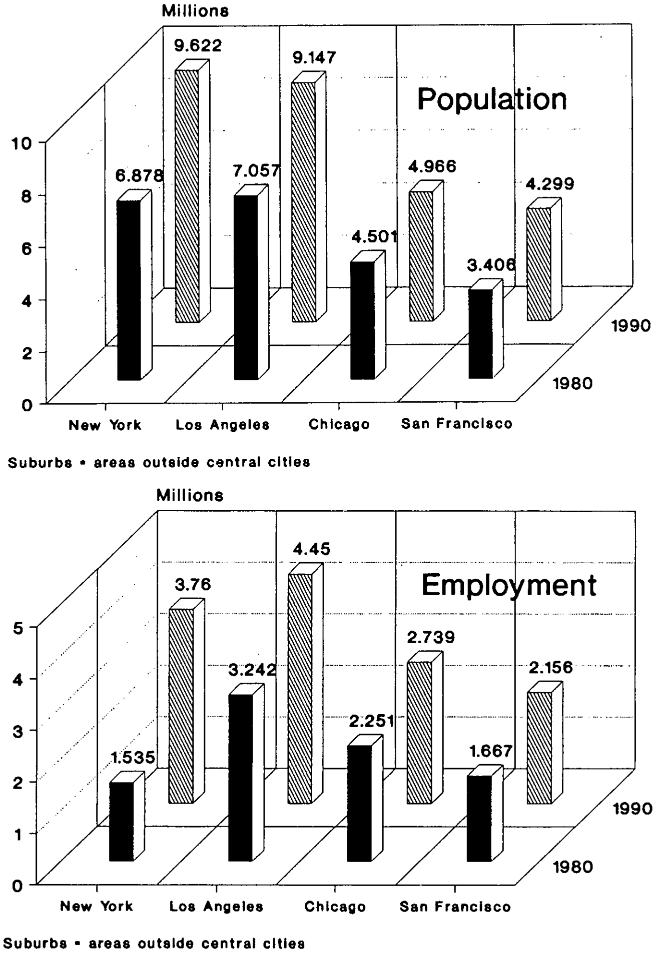


FIGURE 1 Suburban population and employment changes in four largest CSAs, 1980 and 1990.

areas—only 12 of the 75 largest U.S. metropolitan areas registered an absolute increase in transit journeys to work during the 1980s (mostly from the Sun Belt and western regions), and in only 4 of these (Houston-Galveston, Orlando, Dallas–Fort Worth, and San Diego) did transit's market share of work trips increase (*10*).

Trends Among Suburban Residents

Transit's falling fortunes are more alarming among suburban residents. Figure 3 shows that there were actually about 130,000 fewer daily transit work trips made by the suburban residents of the four largest metropolitan areas in 1990 than in 1980. This is despite the 6.2 million residents who were added to the suburbs of these four metropolises during the 1980s. The net result was a sharper decline in transit's market share commute trips of suburbanites than the metropolitan averages (Figure 4).

Trends in the New York metropolitan area were particularly pronounced. From 1980 to 1990 Manhattan added 54 million ft^2 of office space. The suburban ring, including Long Island, northeast New Jersey, and Westchester County, added 173 million ft^2 (equal to the entire Chicago metropolitan office market). Thus suburban counties captured two-thirds of the region's office growth during the 1980s. The impact on transit commuting was unequivocal. In 1980 about one of four suburbanites rode buses and trains to jobs, many of which were in Manhattan; by 1990 fewer than one of 10 suburbanites commuted by transit, many choosing to drive to suburban office parks and other outlying work destinations.

Performance Comparisons

Comparing the performance of urban and suburban transit operations is fraught with difficulties, in part because operating statistics within metropolitan areas are not usually broken down to match the census definitions of the core cities and the suburbs. A second-best approach is to compare operations for those metropolitan areas that have set up different transit properties to serve

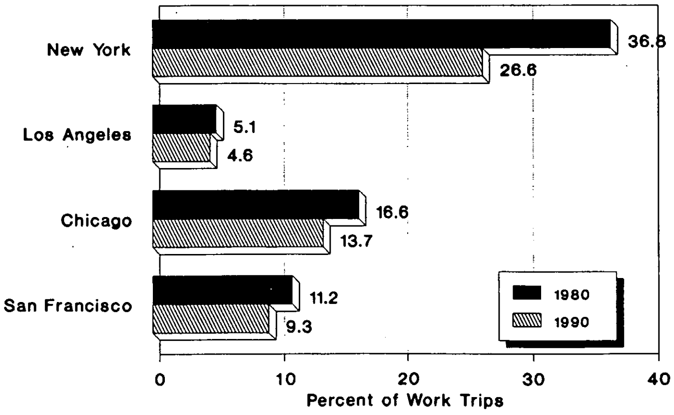
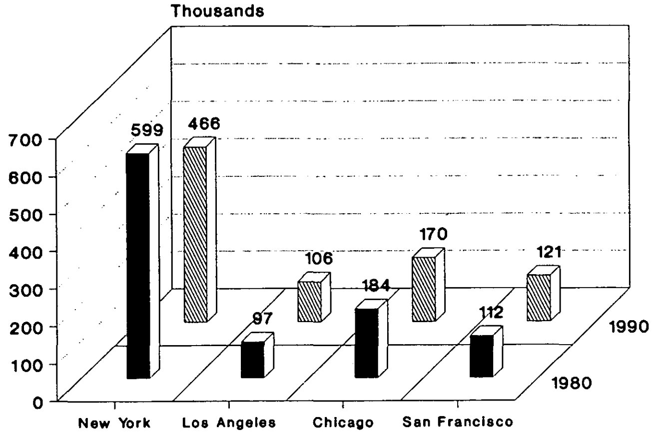


FIGURE 2 Changes in transit modal splits for work trips in four largest CSAs, 1980 and 1990.



Suburbs • areas outside central cities

FIGURE 3 Changes in daily commute trips by suburban residents, large CSAs, 1980 and 1990.

central-city and suburban markets. The best example of this is metropolitan Chicago, wherein the Regional Transportation Authority has divided administrative and operating authority for transit in the region into two groups: CTA, which is in charge of rail and bus services in the city of Chicago (as well as portions of suburban Cook County), and the operators in charge of suburban commuter rail (Metra) and bus (Pace) services.

Figure 5 gives performance statistics for suburban operators as a share of regional totals for four large metropolitan areas for which suburban operators could be reasonably distinguished from urban operators. (See footnotes b to e of Table 1 for transit operations that were defined as urban versus those that were defined as suburban.) Statistics for metropolitan San Diego instead of the San Francisco–Oakland–San Jose Bay area were used in this analysis mainly because the San Diego region has two operators that operate almost exclusively in the suburbs (North San Diego County Transit and San Diego Regional Transportation Service) and two that operate mainly in the central city (San Diego Transit and San Diego Trolley). On the other hand, many of the Bay Area's largest operators, Alameda–Contra Costa County (AC Transit) and Santa Clara County Transit, operate in both central cities (Oakland and San Jose) and suburban areas. The data in Figure 5 are from the 1991 Section 15 report on transit operating performance.

Figure 5 shows that relative to ridership and service output suburban transit services in the four metropolitan areas for which data are shown were far more dependent on public operating assistance than their urban counterparts (except in the New York region, where many suburban operations are either private or contracted). This was mainly because of their low passenger volumes relative to their costs (Table 1). (On a revenue mile basis, however, suburban services cost less than urban ones in three of the four metropolitan areas.) In the Chicago region the operating assistance per passenger for suburban services was more than four times that for urban services ($1.89 versus $0.84); on a revenue mile basis they were twice as high ($5.60 versus $2.85). To the extent that transit's customer base shifts to suburbia, funding allocations should be responsive to these shifts. Currently funding

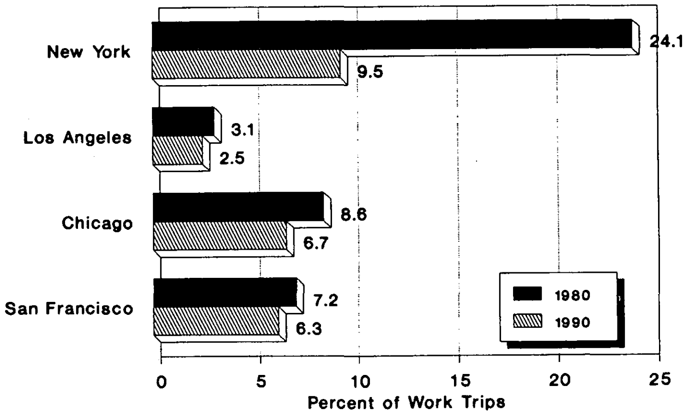
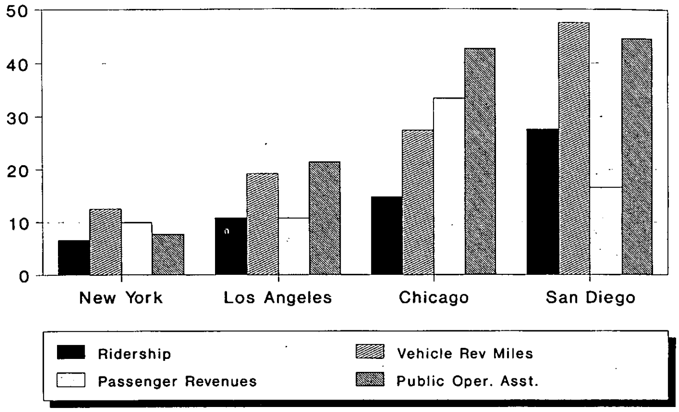


FIGURE 4 Changes in transit share of work trips by suburban residents, large CSAs, 1980 and 1990.



Source: 1990 Section 15 Data

FIGURE 5 Suburban transit as percentage of regional totals for four large metropolitan areas, 1991.

in all four metropolitan areas favors higher-cost suburban services. If economic efficiency is to be rewarded, any redistribution of funding should be based on output (e.g., ridership) instead of input (e.g., service delivery) measures, balanced by some recognition of the harder task of cost-effectively serving suburban markets.

A second comparison was carried out. That comparison examined urban versus suburban performance for a larger set of metropolitan areas; however, data only for the largest suburban versus urban bus operators in each metropolitan area were used. Table 2 summarizes the findings drawn from 1991 Section 15 statistics for (urban followed by suburban) operations in the following areas: New York (New York City Transit Authority and Metropolitan Suburban Bus Authority), Los Angeles (Southern California Rapid Transit District, now renamed Metropolitan Transit Au-

TABLE 1 Operating Cost Comparisons Between Urban and Suburban Services for Four Large Metropolitan Areas, 1991

	Operating Cost per Passenger[a]		Operating Cost per Revenue Mile[a]	
	Urban Services	Suburban Services	Urban Services	Suburban Services
New York[b]	$1.94	$2.43	$8.18	$5.08
Los Angeles[c]	1.39	1.98	5.76	4.20
Chicago[d]	1.26	3.49	5.53	7.06
San Diego[e]	1.19	1.80	4.92	2.25

[a] Statistics are for both bus and rail transit operations in the New York, Los Angeles, Chicago, and San Diego regions, based on 1991 Section 15 data. Data are exclusive of non-surface transit (e.g., ferries) and specialized services like dial-a-ride.

[b] Urban: New York Metropolitan Transit Authority (NYCTA, Metro-North, Long Island Rail Road, SIRTOA), PATH (rail only), Queens Surface Corporation, New Jersey Transit (non-contract and urban division services), and Command Bus Company; Suburb: NYMTA Metropolitan Suburban Bus Authority, New Jersey transit (all contract services and Suburban Transit Corporation), Westchester County Bus, Jamaica Buses, Hudson Bus Transportation, Green Bus Lines, Liberty Lines Express, New York Bus Tours, Putnam County Transit, Rockland Coaches, Suffolk Transit, Triboro Coach, and municipal service for Rockland, Clarkstown, Long Beach, and Spring Valley.

[c] Urban: Southern California Rapid Transit District, Los Angeles County (LACTCT) Motor Bus, and municipal services for Santa Monica, Montebello, Long Beach, Commerce, Gardena, Torrance, and Culver City; Suburban: Orange County Transit District, Omnitrans, Riverside Transit Agency, and municipal services for Laguna Beach, Arcadia, Corona, and Riverside.

[d] Urban: Chicago Transit Authority (including contract services, but excluding suburban Cook County bus runs); Suburban: Metra (including contract services), Pace (including contract services), and municipal services for Niles and Willmette.

[e] Urban: San Diego Transit Corporation and San Diego Trolley; Suburban: North San Diego County Transit and San Diego Regional Transportation Services.

TABLE 2 Summary Comparison of Performance Measures, Suburban Versus Urban Operators for Six Metropolitan Areas

	Farebox Recovery Ratio (%)	Operating Cost per Vehicle ($): Hour	Operating Cost per Vehicle ($): Mile	Passengers per Vehicle: Hour	Passengers per Vehicle: Mile	Operating Cost per ($): Trip	Operating Cost per ($): Pass. Mile
Average for Urban Operators	38.9	82.94	8.73	56.2	5.8	1.44	0.59
Average for Suburban Operators	30.4	72.81	5.24	38.8	2.9	2.06	0.42

thority, and Orange County Transit District), Chicago (Chicago Transit Authority and Pace Suburban Bus Division), Detroit (Detroit Department of Transportation and Suburban Michigan Area Regional Transit), San Francisco (San Francisco Municipal Railway and Transit and Santa Clara County Transit Authority), and San Diego (San Diego Transit Corporation and North San Diego County Transit Development).

Table 2 illustrates that on average urban operators outperformed their suburban counterparts in terms of fare box recovery rates and service effectiveness (in terms of passengers per mile by a factor of two). Of course the unit cost per mile or hour of urban services was substantially higher than that of suburban services; however, the costs per passenger were about 30 percent less. Because of the longer average trip distance suburban services cost less on a per-passenger-mile basis. However because most bus operations charge flat fares, fare revenues per passenger-mile for suburban operators tend to be proportionally less than those for urban operators, resulting in a higher deficit per passenger.

POLICY RESPONSES: ADAPT TRANSIT SERVICES

Transit's shrinking market share in suburbia, its relatively poor fiscal and operating performance, and continuing restraints on government spending underscore the need to overhaul how suburban services are delivered. During the 1980s the chief policy response to rising transit deficits was to competitively contract out services with an eye toward lowering input costs, particularly labor. Although this indeed slowed the deficit growth, it did not substantially change the service features of most suburban operations. Transit is continuing to lose market share to the automobile. To effectively compete radical surgery in how transit services are designed and delivered will be necessary.

At the simplest level policy makers can respond to the challenges posed by decentralization by (a) adapting transit services, making them more flexible, demand responsive, and responsive and suitable for serving dispersed origins and destinations and (b) adapting land uses to make them more supportive of transit—for example, greater densities and mixtures of uses. These of course are not mutually exclusive approaches, although pursuit of the first policy complicates efforts to achieve the second.

Adapting transit to a landscape of spread out and automobile-oriented development means, in many ways, making it more automobilelike. Similar to telephone networks, for transit to compete in suburbia it must cast a larger net to allow more patrons to get from anywhere to everywhere. Strategies that make transit more flexible, interconnected, and ubiquitous include initiating timed-transfer services, paratransit, reverse commute and special services, employer van pools, transitways, and advanced technologies, such as automated vehicle locator systems. This section summarizes some of the recent developments with these service strategies, drawing on a recent national survey of 88 U.S. transit properties.

The self-administered survey was sent to all U.S. transit properties with 50 or more vehicles during February and March 1993. In all, 88 of the 192 surveys were returned, providing information on types of service strategies, impacts on ridership and operations, and attitudes toward service changes. For the most part survey respondents were planners or analysts within an agency who were familiar with specific suburban-targeted strategies that had been introduced.

Timed Transfers and Transit Centers

The timed meeting of buses at transit centers improves inter-suburban services, especially those with long headways, by reducing wait times. The national survey found that 68 percent of U.S. transit properties have some form of timed-transfer and transit center services; among properties with more than 350 vehicles, almost 90 percent used timed transfers. Comparisons of ridership 1 year after introducing timed transfers showed systemwide ridership increases of 3.2 percent in Dayton, Ohio (between 1990 and 1991), and 40 percent in Painsville, Ohio (between 1989 and 1990), even though ridership was falling for most other Ohio transit properties in the same period. AC Transit, serving the Oakland, California, area has begun phasing in timed transfers, with promising results to date. AC Transit's ridership began falling in the mid-1980s as more and more jobs were locating in suburban areas away from its traditional routes. AC Transit planners initiated a multidestinational transit centers program in early 1989. Table 3 shows that ridership has risen noticeably in the two subdistricts where gridlike, interconnected services operating on a pulse schedule have been introduced. On the other hand patronage on the rest of the AC Transit's service area where traditional radial services remain has continued to fall off.

Tidewater, Virginia [Tidewater Regional Transit (TRT)] converted to a timed-transfer network in 1991. The network was designed by the same transit planners who first introduced timed transfers in Edmonton, Alberta, Canada, in the 1970s. Although TRT's ridership has fallen in recent years because of the local

TABLE 3 Ridership Trends Associated with Phase-in by AC Transit of Multidestinational, Timed-Transfer System

Subdistrict	Average Weekday Ridership December 1989	Average Weekday Ridership December 1991	% Change
West Contra Costa County[a]	12,488	28,329	+32
Oakland-Berkeley-Alameda[b]	146,386	156,987	+7
Remainder of AC Transit Service Area	58,671	49,357	-16
SYSTEM TOTAL	226,545	234,673	+4

[a] Grid and Timed-Transfer System introduced in September 1990

[b] Grid and Timed-Transfer System introduced in April 1991

recession, patronage has increased at four large employment centers in Virginia Beach served by buses operating in sync. A recent survey, moreover, revealed that three-quarters of TRT's customers prefer timed transfers to previous services (*11*).

Paratransit

Paratransit services, like shared-ride taxis and minibuses, are particularly suited to suburbia because of their flexible routing and curb-to-curb service features. From the national survey, 43 percent of U.S. transit properties were found to operate some form of demand-responsive service that is available to the general public (instead of exclusively for the elderly or other targeted groups); smaller agencies relied most heavily on paratransit. In the case of Broward County, Florida, five fixed-route services were converted to contract route-deviation dial-a-ride services in 1991—1 year later ridership increased from 15,000 to 27,000/month; this was accompanied by a 47 percent decline in operating costs.

Private jitneys have been part of greater Miami's transportation scene for many years, serving a number of inner-city neighborhoods unserved by public transportation. In 1992 Miami's jitneys carried nearly 50,000 riders per weekday, or about one-quarter of Miami Metrobus's ridership (*12*). Surveys show that Miami's jitneys have developed a market of their own instead of merely siphoning off riders from Metrobus. Jitneys were also mobilized to provide cross-country services in the wake of Hurricane Andrew, which left thousands of south Florida residents without vehicles and homes and displaced many businesses to temporary sites in northern Dade County.

One promising marriage is paratransit and automated vehicle locator (AVL) technologies. Satellite vehicle tracking systems enable vehicles equipped with sensors to be located and promptly dispatched to customers to minimize waits, detours, and deadheading. In Germany paratransit vehicles with on-board terminals are linked to central computers, allowing flexible-route buses, shared-ride taxis, and minibuses to be dispatched to customers waiting at suburban rail stations and rural areas. Ridership on these "call-a-bus" services has increased between 36 and 80 percent above those on the fixed-route bus services that they replaced in several German metropolises (*13*).

The biggest barriers to successful paratransit in the suburbs are restrictive regulations, subsidized bus fares, and free parking. Attempts to operate jitneys in Los Angeles as well as suburban-targeted, on-call shuttle buses (e.g., airport shuttles) in the 1980s were scrapped because the private operators could not compete with cheaper public buses and win over commuters who enjoyed free parking (*14,15*). One of the primary reasons regional shuttle services such as Supershuttle focus almost exclusively on airports is that commercial rates are charged for airport parking, whereas at most other locales parking is free or heavily subsidized. At airports shuttles are cost-competitive; at most other destinations they are not.

Reverse Commutes and Specialized Runs

Special reverse commute and rail station feeder runs are incorporated by about 38 percent of the U.S. transit properties surveyed, most of which are large operators. Most reverse commute services introduced in the 1970s and 1980s as "poverty abatement transportation programs" folded over time because of high attrition. A reverse commute program initiated in the mid-1980s in greater Washington, D.C., that connected inner-city residents to jobs in Fairfax County, Virginia, found that only 18 percent of the 255 original participants who got jobs still had their jobs 2 years later (*16*). In general many of these specialized programs overestimated the extent of suburban vacancies matched to the skills of inner-city residents, the willingness of suburban employers to hire and train inner-city residents, and the willingness of inner-city residents to endure long commutes for low-paying, often dead-end service-sector jobs.

The success of reverse commute services should not be gauged in transit ridership terms however. A study of another program in the Washington, D.C., area found that many of the original passengers either had earned enough money to buy a car to drive to work or had met coworkers and formed car pools (*17*). The ultimate success of reverse commute services lies in helping urban residents find jobs with some growth potential. Surveys by Pace of two reverse commute runs from south Chicago to job centers in DuPage County revealed that the services influenced the decision of 60 to 66 percent of surveyed passengers to take and retain the jobs (*18*). Moreover surveys found that about 30 percent of Pace's reverse commuters formerly drove alone to work.

Employer-Sponsored Van Pools and Subscription Services

Employer-sponsored van pools and subscription services are suited mainly for highly dispersed suburban markets, such as office parks in the exurbs. Particularly where fixed-route schedules cannot be justified, van can serve the commuting needs of clusters of workers. They are most economical when employees operate the vehicles. Pace's subscription van services, wherein employers and Pace share van purchase and operating expenses and rely on employee drivers, enjoy an 83 percent cost-recovery rate (*19*). More than half of Pace's 75 vans serve the new Sears center in Hoffman Estates. The program has been very successful, with about 30 percent of Sear's 5,000 suburban workers commuting by some form of mass transit (*20*). When these workers were in downtown Chicago, 92 percent of them commuted by mass transit, so part of this success is no doubt attributable to workers' ingrained habits of patronizing transit. Pace capitalized on the situation by designing an ambitious market development program that approached all employees about their individual commuting needs and delivered a rich mix of transit options (subscription bus runs, fixed-route services, and car pools in addition to employer-sponsored van pools). In the case of Sears and others, guaranteed ride home programs and on-site retail and other mixed-use activities have encouraged workers to join van pools.

HOV Lanes and Dedicated Busways

Dedicated busways and high-occupancy-vehicle (HOV) facilities improve suburban services because, unlike rail systems, vehicles can leave guideways and filter into low-density neighborhoods, reducing the need for a transfer. About 12 percent of the U.S. properties surveyed have some form of HOV or contraflow lanes for suburb-to-suburb runs in addition to the more traditional radial services. The 30-km busway in Ottawa, Ontario, Canada, captures as much as one-third of all trips to several large shopping plazas and work centers outside the core (*3*). Houston's transitway, slated to extend to 95 mi by 1995, is already the world's largest, a seemingly perfect technology for a region that is spread out but that features a dozen or more large-scale activity centers. Despite strong economic growth, Houston's average freeway speeds and transit patronage have increased faster and arterial congestion levels have fallen more than those of any large U.S. city in the past 5 years (*10,21*). Presently more than 6 percent of commuters from the Woodlands, an affluent community about 50 mi north of downtown Houston, patronize the Woodlands Express bus services that operate via the I-45 Transitway to downtown Houston, the Medical Center, and Greenway Plaza.

LAND USE INITIATIVES

A criticism of suburban-targeted strategies is that they reinforce the low-density, automobile-reliant development patterns that they attempt to serve. Some observers argue that regions should be restructured so that more people will ride transit. Transit works best when it connects relatively dense nodes along radial axes (*22*). The presence of mixtures of apartments-condominiums, office towers, and other activities is also needed for balanced, two-way flows. Greater Stockholm, Sweden, has such a built environment and operates a world-class rail system that handles 60 percent of all suburban work trip origins and destinations (*23*).

Traditional Neighborhoods

Transit-oriented and neotraditional developments have gained popularity in recent years as design motifs that reduce dependency on the automobile and create attractive environments for walking and using transit. Neotraditionalist designers borrow many of the successful elements of traditional turn-of-the-century transit villages: commercial cores within walking distance of most residents, well-connected (typically grid) street patterns, various densities of housing, and mixed land uses. It is still not known whether designing such places in the 1990s will lure many people from their cars. A Montgomery County, Maryland, study found that workers in ''transit and pedestrian friendly neighborhoods'' use transit 8 to 45 percent more often than workers from neighborhoods conducive to automobile use (e.g., with curvilinear roads and no retail shops). All neighborhoods in the study were about the same distance from transit facilities (*24*). Another recent study of ''streetcar'' neighborhoods (ones that at one time were served by a streetcar and have inherited higher densities, gridded streets, and mixed uses) and relatively close by ''automobile'' neighborhoods (postwar, typical suburban neighborhoods) reveals some degree of elasticity between urban design and travel behavior (*23*). A comparison of San Francisco Bay area neighborhoods matched in terms of comparable average household incomes and levels of bus service intensities showed that the denser, mixed-use streetcar neighborhoods average 2.5 to 5.5 percent more work trips by transit and 1.2 to 13.2 percent more work trips by walking or cycling.

In recognition of the need to build communities more easily served by transit, about 30 U.S. transit properties have prepared site and urban design guidelines in the past decade (*23*). These guidelines are meant to encourage developers to incorporate public transportation considerations into their project designs. Although none of the design guidelines have yet to be codified into local ordinances, eight of the transit properties with guidelines have prepared checklists that local planners use in evaluating the degree to which a proposed project encourages transit and pedestrian access.

Transit-Based Housing

In some suburban area with rail services transit-based housing is being actively promoted. In the San Francisco Bay Area Bay Area Rapid Transit (BART) officials have entered into joint development agreements with private home builders at several stations that will convert portions of park-and-ride lots to housing projects, using lease revenues to help finance replacement parking. Besides boosting ridership, planners hope that the placement of new housing near rail stations will allow more riders to walk or ride bikes to the station, yielding important air quality benefits. Short automobile trips currently account for about 60 percent of access trips to suburban BART stations; high levels of pollutants are emitted from automobiles during these trips as a result of the impacts of cold starts.

Recent research shows that 32 percent of residents living within 1,500 ft of a suburban BART station patronize transit to work, compared with only about 5 percent of the region's suburbanites who live more than 1,500 ft away (*25*). These market shares are smaller than those found in studies or ridership by proximity in suburban Toronto (*26*) and Washington, D.C. (*27*). Trip destination and parking policies at the workplace were the major determinants of whether those living near stations ride BART. More than 95 percent of suburban residents commuted by BART if they worked in downtown San Francisco and paid for parking. If they worked in downtown Oakland, Berkeley, or Walnut Creek and paid for parking, about 65 percent commuted by BART. For most other destinations (where employees typically park for free), BART's share was between 3 and 12 percent. As jobs continue to suburbanize, the ability of transit-based housing to serve work trips will be jeopardized. Thus successful transit-based housing programs will need to be matched by initiatives that target more employment growth around major suburban transit stops as well as policies (such as free parking) that eliminate subsidies to commute alone.

Land Use Dilemma

Other land use initiatives that have been suggested as a means of reducing automobile dependence and ostensibly increasing the regional role of mass transit include jobs-housing balancing, urban growth limits, and urban reinvestment. All of these initiatives are politically unpopular, however, because they interfere with market forces and in the minds of most Americans involve excessive government regulation (*28*). In general land use initiatives as a response to transportation problems suffer from the lack of common vision on the ideal metropolis (i.e., how a region should be planned) and not-in-my-backyard (NIMBY) resistance. They also receive lackluster political support because they typically yield mobility dividends only over the long run, well beyond existing politicians' terms of office.

INSTITUTIONAL, FISCAL, AND PRICING CONSIDERATIONS

Suburbanization also calls for creative institutional responses. New regional alliances are one option. A successful model in Germany has been transit federations. In greater Munich, Hamburg, and Essen-Dortmund regional federations have been formed to reverse the fragmentation of transit enterprises. These federations set fares, decide on route changes, and coordinate timetables to improve integration and avoid duplication. The concept is basic: a single organization should be managing services for the entire "commuteshed" of a region. Day-to-day operations of the urban, suburban, and inter-city carriers are run by individual transit companies. Managers of these companies sit on the boards of the transit federations. The federations collect all revenues and redistribute them so that each operation averages the same cost recovery rate, currently about 65 percent. Fares are totally integrated—a ticket purchased for U-Bahn (urban rail) services lets one transfer free to an S-Bahn (suburban rail), bus, or tram.

From a fare policy standpoint rapid suburbanization means that costs will likely vary increasingly more among individual trips depending on travel distance and perhaps even time of day. Areas experiencing rapid suburban growth should address whether zonal, peak surcharge, or other differentiated fares are needed. Of the seven U.S. transit properties that in 1989 charged a flat fare within the region's main city and a zonal charge for crossing into the suburbs, the average cost recovery rate was 4 percent (*29*). This compared with a 25 percent recovery rate for properties serving comparably sized metropolitan areas that had flat fares. For three U.S. transit agencies that had peak and off-peak fare differentials, on average, 39 percent of the operating costs were covered by fare receipts. More differentiated pricing is correlated with higher fare box recovery rates.

Rapid suburbanization will also invariably create political tensions between city and suburban agencies competing for the same shrinking share of public operating assistance. This battle is being played out in nearly all large metropolitan areas, including Chicago, Los Angeles, and San Francisco–Oakland, where multiple transit agencies vie for dedicated sales tax receipts that are returned to a regional transportation commission. Two principles should be considered when setting fiscal allocation policies. First, agencies should be rewarded with public assistance by doing something that benefits the region—such as achieving higher ridership and controlling costs. Such criteria are essential for stimulating innovation. Second, funding policies should be more people oriented than place oriented. Targeting public monies to places, whether in the form of transit subsidies or enterprise zones, will yield few societal benefits if the people in those places do not gain. Perhaps the most promising people-oriented fiscal policy in the transit arena would be to convert most subsidies from the provider side to the user side. Placing funds in the hands of the intended beneficiaries of most subsidies—those who are poor and disadvantaged—would, along with regulatory reforms, encourage sorely needed transit service innovations among competing transit operators. Everyone, inner-city and suburban residents alike, would benefit from the increased diversity in travel options.

BACK TO THE FUTURE

Fixed-route, fixed-schedule transit services will have a difficult time competing and surviving in the suburbs. Today transit's market shares are rapidly eroding nearly everywhere. Major policy reforms are needed. We are well advised to borrow from yesteryear as we look to the future. Early streetcar suburbs were successful in part because private entrepreneurs were allowed to link transit investments and land development, producing moderately dense, mixed-use land patterns (*30*). Well over half of suburban rail services in greater Tokyo are privately built, typically by large consortiums that link transit investments to new town development. In California private tollway franchises are building four different tollways throughout the state with the hope of reaping a nice profit, perhaps less from toll revenues than from selling land at key interchanges that the franchisers own; possibilities for franchising rail line extensions, however, have largely been ignored. Resurrecting the jitney services found three-quarters of a century ago in most U.S. cities might also be considered. Given the freedom to operate, door-to-door van and jitney services, similar to regional airport shuttles, would likely emerge in many suburban settings, tapping new market niches such as suburban mall and office complexes, sports stadia, and recreational theme parks.

The model of publicly led transit and privately led land development has been tried in the past 50 years with generally disappointing results. Another option deserves consideration: allowing developers to link transit and real estate projects and entrepreneurs to carve out new transit market niches in suburbia—with the hope that they will create more transit-oriented communities in the process.

Although the private sector is probably better suited to responding to many of the needs of suburban travelers, there will always be a role for the public sector: assembling rights-of-way for dedicated busways, providing start-up funds for smart transit technologies, and zoning for moderate-density housing around major transit stops. In combination profit-seeking entrepreneurs and community-minded governments can create the kinds of built environments and service innovations that within a decade or two could allow transit to compete successfully with the automobile in suburbia.

ACKNOWLEDGMENT

This research was supported by a grant from the University of California Transportation Center. Mark Dunzo was a research assistant on the project.

REFERENCES

1. American Public Transit Association. *Transit Fact Book.* Washington, D.C., 1991.
2. Bartholomew, H. Effects of Urban Decentralization Upon Transit Operations and Policies. *Proc., American Transit Association.* Washington, D.C., 1940.
3. Cervero, R. *Suburban Gridlock.* Center for Urban Policy Research, New Brunswick, N.J., 1986.
4. Pisarski, A. *New Perspectives in Commuting.* FHWA, U.S. Department of Transportation, 1992.
5. Cervero, R., and J. Landis. Suburbanization of Jobs and the Journey to Work. *Journal of Advanced Transportation,* Vol. 26, No. 3, 1992, pp. 275–297.
6. Rosenbloom, S. *Developing a Comprehensive Service Strategies to Meet a Range of Suburban Travel Needs.* Urban Mass Transportation Administration, U.S. Department of Transportation, 1990.
7. Kasarda, J. Economics and Technology. In *Land Use in Transition.* Urban Land Institute, Washington, D.C., 1993.
8. Stanback, T. *The New Suburbanization: Challenge to the Central City.* Westview Press, Boulder, Colo., 1991.
9. Hughes, M. Regional Economics and Edge Cities. In *Edge Cities and ISTEA—Examining the Transportation Implications of Suburban Development Patterns.* FHWA, U.S. Department of Transportation, 1992.
10. Cervero, R. Surviving in the Suburbs: Transit's Untapped Frontier. *Access,* Vol. 2, 1993, pp. 29–34.
11. Tidewater Regional Transit. *TRT On-Board Ridership Survey.* Norfolk, Va., 1992.
12. Urban Mobility Corporation. *The Miami Jitneys.* Federal Transit Administration, U.S. Department of Transportation, 1992.
13. Benke, R. *German "Smart Bus" System: Potential for Application in Portland, Oregon.* FTA, U.S. Department of Transportation, 1993.
14. Teal, R. Transportation Privatization: Experiences and Issues. *Urban Resources,* Vol. 4, No. 1, 1986, pp. 7–12.
15. Cervero, R. *Fostering Commercial Transit: Alternatives in Greater Los Angeles.* The Reason Foundation, Los Angeles, Calif., 1992.
16. Rosenbloom, S. *Reverse Commute Transportation: Emerging Provider Roles.* FTA, U.S. Department of Transportation, 1992.
17. Lemov, P. Impossible Commute. *Governing,* June 1993, pp. 32–35.
18. Paquet, J., B. Reynolds, and A. Murphy. *Reverse Commuting: How Do You Measure Success*? Pace Transit, Chicago, Ill., 1991.
19. *1993 Operating and Capita Program.* Pace Transit, Chicago, 1992.
20. Grzesiakowski, T. Pace's Experiences with Sears' Move to Hoffman Estates. *Proc., Second Annual Metropolitan Conference on Public Transportation Research,* Chicago, Ill., 1993.
21. Hanks, J., and T. Lomax. Roadway Congestion in Major Urban Areas: 1982–1988. In *Transportation Research Record 1305,* TRB, National Research Council, Washington, D.C., 1990, pp. 177–189.
22. Smith, W., Mass Transit for High-Rise, High-Density Living. *Journal of Transportation Engineering,* Vol. 110, No. 6, 1984, pp. 521–535.
23. Cervero, R. *Transit-Supportive Development in the United States: Experiences and Prospects.* FTA, U.S. Department of Transportation, 1994.
24. *Transit and Pedestrian Oriented Neighborhoods.* Maryland National Capital Park and Planning Commission, Silver Spring, Md., 1992.
25. Cervero, R. *Ridership Impacts of Transit-Focused Development in California.* Institute for Urban and Regional Development, University of California, Berkeley, 1993.
26. Stringham, M. Travel Behavior Associated with Land Uses Adjacent to Rapid Transit Stations. *ITE Journal,* Vol. 52, No. 4, pp. 112–116.
27. *Metrorail Ridership Survey.* JHK & Associates, Washington Metropolitan Area Transportation Agency, Washington, D.C., 1989.
28. Downs, A. *Stuck in Traffic: Coping with Peak-Hour Traffic Congestion.* The Brookings Institution, Washington, D.C., 1992.
29. Cervero, R. Transit Applications of Pricing and Market-Based Approaches. *The Role of Pricing and Market-Based Strategies: Summary of Proceedings.* UCLA Extension Public Policy Program, Los Angeles, Calif., 1992.
30. Warner, S. *Streetcar Suburbs.* Harvard University Press, Cambridge, Mass., 1962.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Bus Transit Service Coverage for Maximum Profit and Social Welfare

LAZAR N. SPASOVIC, MARIA P. BOILE, AND ATHANASSIOS K. BLADIKAS

A framework for finding the optimal bus transit service coverage in an urban corridor is presented. The service variables considered are a combination of route length, route spacing, headway, and fare. The criterion for optimality is either operator profit or social welfare maximization. The social welfare, a sum of user and operator surplus, is optimized with both unconstrained subsidy and breakeven constraints. The equations for the optimal design variables that maximize operator profit and social welfare are derived analytically for a rectangular transit corridor with elastic demand, uniformly distributed passenger trip density, and many-to-one travel patterns. The equations provide considerable insight into the optimality conditions and interrelations among variables. These equations are also incorporated within an efficient algorithm that computes optimal values for the decision variables for a more realistic model with vehicle capacity constraints. The numerical results show that at the optimum the operator profit and welfare functions are rather shallow, thus facilitating the tailoring of design variables to the actual street network and particular operating schedule without substantial decreases in profit or welfare. The social welfare function is relatively flat near the optimum for a relatively large range of subsidies. This result implies that for a given set of input data the breakeven constraint may be an economically preferable objective because it eliminates subsidy, whereas it reduces social welfare only marginally. The sensitivities of the design variables to some important exogenous factors are also presented. The presented methodology is also applicable to the problem of optimal service coverage of feeder bus systems serving rapid rail line stations.

The basic elements that must be determined in planning bus transit service in an area are route lengths, route spacing (or density), headways, and fares. Determining how far outward to extend transit routes from the central business district (CBD) is particularly important. The general trade-off is between the cost of service to the operator and the cost of travel to users. Operators prefer short routes to minimize costs. Passengers, especially those from the outer suburbs, prefer longer routes to minimize their access impedance. When the demand for transit service is elastic [i.e., passengers are sensitive to the level of service (LOS) characteristics and the fare], shorter routes and thus higher access impedance will decrease the attractiveness of the service and cause potential travelers to switch to other modes. Because the route length has a significant impact on both operator costs and passenger impedance, its value should be carefully selected.

The purpose of this paper is to develop a method for optimizing the lengths of bus transit routes that extend radially outward from the CBD or those of a feeder bus system serving rapid rail line stations. However this problem may not be considered independently of route location and service scheduling. Therefore the problem considered here is that of finding an optimal combination of route length, route spacing, headway, and fare that maximizes operator profit and social welfare for a rectangular-shaped urban corridor with uniformly distributed passenger trip densities.

Demand is considered to be elastic. Service characteristics affect ridership, which in turn has an impact on revenue. Ridership also affects service characteristics, and thus operator cost. The method proposed in this paper recognizes these interactions between demand and supply (operator cost) and calculates equilibrium LOS characteristics and fare that optimize transit service coverage under several design objectives, which are (a) maximization of operator profit, (b) maximization of social welfare with unconstrained subsidy, and (c) maximization of social welfare with a breakeven constraint.

BACKGROUND

Several previous studies sought to optimize various elements of transit service and network design by using calculus and, to a lesser extent, mathematical programming methods (*1–23*). An extensive review of optimization models can be found in Chang and Schonfeld (*20*). A summary of pertinent analytical models classified according to the design variables optimized is presented in Table 1. In most studies travel demand was inelastic and uniformly distributed over the service area. The usual travel pattern was many to one, whereas the most common objective function was the minimization of the sum of operator cost and user time cost. The assumptions of inelastic demand precluded the models from analyzing the impacts of pricing policies and subsidies.

Kocur and Hendrickson (*12*) developed an analytical model with elastic demand and derived closed-form solutions for optimal route spacing, headway, and fare but not route length for different design objectives. Morlok and Viton (*21*) and Viton (*22*) developed a similar model to evaluate the profitability of bus transit service.

A literature review revealed only two published papers (*15,16*) that dealt with the optimization of a radial transit route length in an urban transportation corridor, which is the focus of this paper. Wirasinghe and Seneviratne (*15*) developed closed-form solutions for the optimal rail transit line length for sectorial and rectangular corridors with inelastic demand and uniformly distributed passenger trip density. The objective function to be minimized included the total rail fleet cost, rail and feeder bus operating cost, and passenger time cost. Spasovic and Schonfeld (*16*) presented a model for optimal service coverage for rectangular and sectorial urban corridors with uniform and linearly decreasing density func-

---

L. N. Spasovic, School of Industrial Management and National Center for Transportation and Industrial Productivity, New Jersey Institute of Technology, Newark, N.J. 07012. M. P. Boile, Transportation Program, New Jersey Institute of Technology, Newark, N.J. 07012. A. K. Bladikas, Industrial and Management Engineering Division, and National Center for Transportation and Industrial Productivity, New Jersey Institute of Technology, Newark, N.J. 07012.

TABLE 1 Pertinent Analytical Models for Transit Network Design

Decision Variables	Objective Function	Transit Mode	Street Network Geometry	Passenger Demand	Authors
Route Length, Spacing, Headway, Stop Spacing	Min. operator and user cost	bus	rectangular and sectorial grid	Uniform and Linear Decreasing, inelastic, many-to-one	Spasovic and Schonfeld (1993)
Route Length	Min. operator and user cost	rail	rectangular grid	General, inelastic, many-to-one	Wirasinghe and Seneviratne (1986)
Route Spacing, Zone Length, Headway	Min. operator and user cost	bus	rectangular grid	Uniform, inelastic, many-to-one	Chang and Schonfeld (1992)
Route Spacing, Lengths and Headway	Min. operator and user cost	bus and rail	rectangular grid	Uniform, inelastic, many-to-one	Byrne (1976)
Route Spacing	Min. operator and user cost	bus	rectangular grid	Uniform, inelastic, many-to-many	Holroyd (1967)
Route Spacing and Headway	Min. operator and user cost	bus	rectangular grid	Uniform, inelastic, many-to-one	Byrne and Vuchic (1972)
Route Density and Frequency	Min. operator and user cost	bus	rectangular grid	General linear, inelastic, many-to-one	Hurdle (1973)
Route Spacing, Headway and Fare	Max. operator profit, Max. user benefit, etc.	bus	rectangular grid	Uniform elastic, many-to-one	Kocur and Hendrickson (1982)
Route Spacing, Headway and Stop Spacing	Min. operator and user cost	feeder bus to rail	rectangular grid	General, inelastic, many-to-one	Kuah and Perl (1988)
Route Spacing, Headway and Fare	Max. profit, max. welfare, min. cost	bus	rectangular grid	Irregular, elastic, many-to-many, time dependent	Chang and Schonfeld (1989)
Route Spacing, Headway	Max. profit	bus	sectorial grid	Uniform, elastic, many-to-one	Morlok and Viton (1984)

tions that were inelastic. The model jointly optimized route length, headway, route, and stop spacing, and it also considered stations along the line and the associated access cost.

This paper extends the methodology of Spasovic and Schonfeld (*16*) to the case of a rectangular corridor with elastic demand. The assumption of elastic demand enables the model to analyze the impacts of pricing policies and subsidies on the system's design characteristics and service coverage.

EQUILIBRIUM FRAMEWORK

The framework for planning optimal bus transit service coverage in which the resources and costs of providing the service are related to its operating characteristics and the induced ridership is presented in Figure 1. In this process the values of the service characteristics such as route length, route spacing (or route density), headway (or its inverse, the frequency), and fare must be carefully selected to satisfy prespecified design objectives.

Because the demand is elastic the service characteristics chosen will have an impact on ridership, and thus system revenue. On the other hand ridership will have an impact on the service characteristics, and thus operator cost.

The LOS characteristics and fare could be optimized by using several objectives. For example the maximization of operator profit—the difference between the fare box revenue and operating cost—could be one objective. However most transit systems do not recover their operating cost from the fare box and need to be subsidized from additional external revenue sources.

As mentioned earlier, there is a conflict between the operator's and users' objectives. Users prefer to have short access to the route and short waiting time, whereas the operator would prefer to have a very long headway and shorter, sparsely located routes with few stops to minimize costs. To alleviate the perceived con-

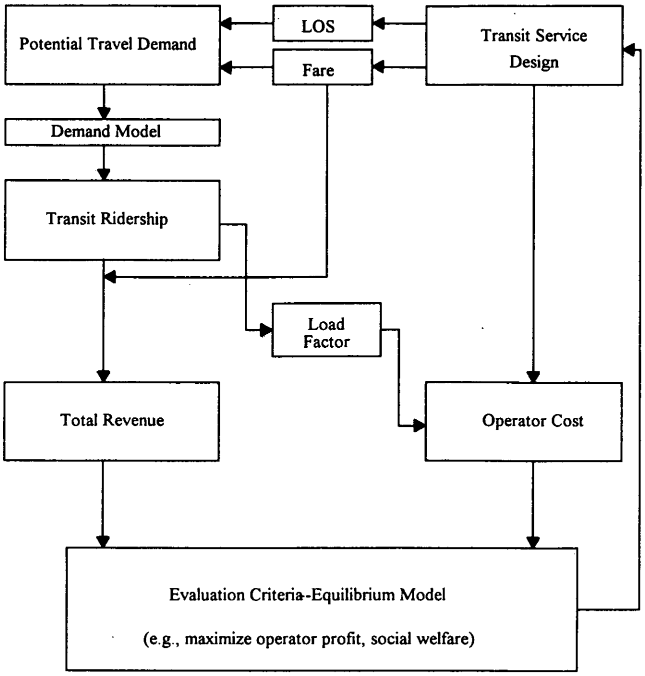


FIGURE 1 Equilibrium framework for optimal transit service coverage.

flict between user and operator objectives, the sum of operator and user time costs (i.e., access, waiting, and in-vehicle riding times multiplied by the value of user time) is often used as a suitable design criterion. In the case of elastic demand, with no requirements for minimum service provision, it is possible to find a set of LOS and fare that minimizes operator and user costs by effectively eliminating ridership. In this case the objective of minimizing the total system cost should be replaced with the maximization of social welfare (defined as the sum of consumer surplus and operator surplus, or profit) subject to a budget constraint.

In this paper the bus service coverage problem of Figure 1 is formulated as an optimization problem wherein the route length, route spacing, headway, and fare must be chosen to maximize either profit or social welfare. The optimization process yields optimal values for service characteristics taking into consideration the interaction of demand and operator cost.

STUDY APPROACH

The problem under consideration is to provide optimal transit service coverage with a simplified bus transit system in an urban corridor as illustrated in Figure 2. The corridor of length E and width Y is divided into two zones. Zone 1 is the area between the end of the corridor and the route terminus, and Zone 2 is the area between the CBD and the route terminus.

The basic approach of this paper is to formulate design objectives as functions of the decision variables. The optimal values of the decision variables are found by taking partial derivatives of the objective function with respect to all decision variables, setting them equal to zero, and solving them simultaneously. This approach, as will be seen later, resulted in a simple model that offered considerable insight into the optimality conditions and interrelations among variables. The equations obtained are incorporated within an efficient algorithm that optimizes service coverage for a more realistic model that includes a vehicle capacity constraint.

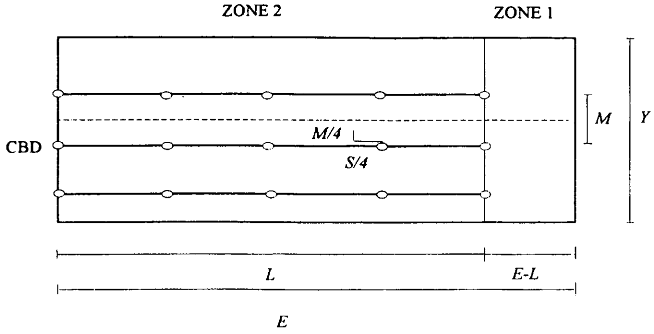


FIGURE 2 Urban corridor and transit network under study.

BUS SYSTEM CHARACTERISTICS AND DEMAND/SUPPLY FUNCTIONS

This section describes briefly the assumptions of the bus system's operating characteristics and presents the derivation of the system's passenger demand and cost functions.

Assumptions About Bus System Characteristics

1. An urban rectangular corridor is served by a bus transit system consisting of n parallel routes of uniform length L separated laterally by route spacing M.
2. The routes extend from the CBD outward.
3. The total transit demand is uniformly distributed along the entire corridor and over time and is sensitive to the quality of transit service and fare.
4. The commuter travel pattern consists of many-to-one or one-to-many trips focused on the CBD.
5. A dense rectangular grid street network allows passengers orthogonal access movements (i.e., access paths are parallel and perpendicular to the route).
6. Transit vehicles operate in local service (i.e., all vehicles serve all stations).
7. The average access speed is constant. Walking is the only access mode.
8. Average waiting time equals half the headway. The headway is uniform along the route and among all parallel routes.
9. Operator costs are limited to those for vehicles (i.e., the infrastructure is free).
10. There is no limit on vehicle fleet size.

Demand Functions

The urban corridor demand is assumed to be a linear function sensitive to price and various travel time components (waiting, access, and in-vehicle times). A conceptual form of the demand density function is as follows:

$$q = P[1 - e_w * \text{wait time} - e_a * \text{access time} - e_{iv} * \text{in-vehicle time} - e_p * \text{fare}] \qquad (1)$$

where

q = unit transit demand density (passengers/mi^2-hr),
P = potential travel demand density (passengers/mi^2-hr),
e_w = sensitivity factor for waiting time,
e_a = sensitivity factor for access time,
e_{iv} = sensitivity factor for in-vehicle time, and
e_p = sensitivity factor for fare.

The demand function is similar to the one suggested by Kocur and Hendrickson (*12*) and is almost identical to that of Chang and Schonfeld (*20*).

For the particular application presented in this paper total demand consists of the sum of Zone 1 and Zone 2 demands. It is obvious that access, waiting, and in-vehicle times will affect demand in both zones. Because the trip origins are uniformly distributed over the corridor an average passenger accessing the route walks perpendicularly one-quarter of the spacing between the two routes—an access distance of $M/4$. The access distance parallel to the route depends on whether the trip originated within Zone 1 or Zone 2. Passengers originating in Zone 1 must board vehicles at the terminus, thus having a total average access distance of $(E - L)/2 + M/4$. A passenger from Zone 2 walks along the route one-quarter of the local stop spacing S to reach a stop. The total access time for an average passenger in Zone 1 equals the average access distance divided by the access speed g [i.e., $(E - L)/2g + M/4g$]. For a passenger in Zone 2 the access time is $(M + S)/4g$.

The in-vehicle time is the actual riding time between the stop of origin and the CBD. The average in-vehicle time is obtained as the average distance traveled divided by the average transit speed V and is different for each zone. Passengers originating in Zone 1 travel the whole length of route L, whereas those from Zone 2 travel approximately an average distance of $L/2$. According to Assumption 8 passengers wait $H/2$.

The hourly transit demand in Zone 1 (in passengers per hour) is then given as

$$Q_1 = PY(E - L)\left[1 - e_w\frac{H}{2} - e_a\left(\frac{M}{4g} + \frac{E - L}{2g}\right) - e_{iv}\frac{L}{V} - e_p f\right] \qquad (2a)$$

where:

Q_1 = transit demand in Zone 1 (passengers/hr),
P = potential transit trip density (passengers/km^2-hr),
Y = corridor width (km),
E = corridor length (km),
L = length of transit route (km),
H = route headway (hr/vehicle),
M = route spacing (km/route),
g = access speed (km/hr), and
V = average transit speed (km/hr).

The hourly transit demand in Zone 2 (in passengers per hour) is as follows:

$$Q_2 = PYL\left[1 - e_w\frac{H}{2} - e_a\frac{M + S}{4g} - e_{iv}\frac{L}{2V} - e_p f\right] \qquad (2b)$$

where S is average stop spacing (km/stop).

The total hourly corridor demand, Q, is the sum of Q_1 and Q_2.

Operator Cost

The operator cost includes maintenance and overhead as well as the more direct cost of operation (driver wages, fuel, spare parts, etc.) and is represented by the all-inclusive hourly operating cost per vehicle, c. The total hourly operator cost is obtained by multiplying the active fleet size by the hourly operating cost per vehicle. Fleet size is the number of on-line vehicles required to provide service and is obtained by dividing the total round-trip time (running time and layover time) by the headway. The total round-trip time is the round-trip route length divided by the average speed. The total hourly operator cost is then

$$C = \frac{2cYL}{HMV} \qquad (3)$$

where

C = operator cost (\$/hr),
c = vehicle operating cost (\$/vehicle-hr),
Y = corridor width (km),
L = length of transit route (km/route),
H = route headway (hr/vehicle),
M = route spacing (km/route), and
V = average transit speed (km/hr).

TRANSIT SERVICE DESIGN OBJECTIVES

The two objectives considered in this paper are maximization of operator profit and maximization of social welfare. The analysis consists of optimizing service coverage under each objective, comparing the results, and deriving insights about the optimal coverage.

Maximizing Operator Profit

Operator profit (Π) is defined as a difference between the fare box revenue R and operator cost C

$$\Pi = R - C \qquad (4)$$

Revenue R is defined as the fare multiplied by ridership

$$R = PYE\left(1 - e_w\frac{H}{2} - e_a\frac{M}{4g} - e_p f\right)f + PY(E - L)\left(- e_a\frac{E - L}{2g} - e_{iv}\frac{L}{V}\right)f + PYL\left(- e_a\frac{S}{4g} - e_{iv}\frac{L}{2V}\right)f \qquad (5)$$

The hourly operator profit (Π) is the difference between the total operator revenue (Equation 5) and operator cost (Equation 3)

$$\Pi = PYE\left(1 - e_w\frac{H}{2} - e_a\frac{M}{4g} - e_pf\right)f + PY(E - L)\left(- e_a\frac{E - L}{2g} - e_{iv}\frac{L}{V}\right)f + PYL\left(- e_a\frac{S}{4g} - e_{iv}\frac{L}{2V}\right)f - \frac{2cYL}{HMV} \tag{6}$$

The operator profit function can be maximized by setting its partial derivatives with respect to the route length L, headway H, route spacing M, and fare f, to zero. When the resulting equations are solved independently, the following expressions for route length L, headway H, spacing M, and fare f are obtained:

$$L^* = E - \frac{2cg}{PHMf(e_aV - e_{iv}g)} - \frac{e_aSV}{4(e_aV - e_{iv}g)} \tag{7a}$$

$$H^* = \left(\frac{4cL}{e_wMPEVf}\right)^{1/2} \tag{7b}$$

$$M^* = \left(\frac{8cLg}{e_aHPEVf}\right)^{1/2} \tag{7c}$$

$$f^* = \frac{2 - e_wH}{4e_p} - \frac{e_a[ME + 2(E - L)^2 + SL]}{8e_pEg} - \frac{e_{iv}(2LE - L^2)}{4e_pEV} \tag{7d}$$

Solving Equations 7b and 7c simultaneously yields the following expressions for H and M:

$$H^* = \left(\frac{2cLe_a}{PEVfe_w^2g}\right)^{1/3} \tag{8a}$$

$$M^* = \left(\frac{16cLe_wg^2}{PEVfe_a^2}\right)^{1/3} \tag{8b}$$

When the route length, route spacing, headway, and fare are optimized independently of each other, their relation to the other decision variables can be read directly from Equations 7a to 7d. These equations provide the optimal value of one of the decision variables as a function of the other three. For exmaple Equation 7a can be used to find the optimal route length when the headway, route spacing, and fare are given. Equations 8a and 8b may be useful by themselves in cases in which the route length L or fare f cannot be modified.

Equations 7a to 7d also provide useful insights into the relationship between the decision variables and the various parameters. For example according to Equation 7a the optimal route length varies directly with the corridor length E, passenger density P, headway H, route spacing M, fare f, sensitivity factor for access time e_a, and transit speed V. It varies inversely with the vehicle operating cost c, access speed g, stop spacing S, and the sensitivity factor for in-vehicle time e_{iv}.

It should be noted that the simultaneous solution of Equations 8a and 8b produces an interesting result. Optimally the ratio of route spacing and headway is constant and has the following value:

$$\frac{M^*}{H^*} = 2\frac{e_wg}{e_a} \tag{9}$$

Unfortunately all four equations, Equations 7a to 7d, cannot be solved simultaneously by algebraic methods.

Maximizing Social Welfare

Social welfare (W) is defined as the sum of consumer surplus (T) and producer surplus or profit (Π)

$$W = T + \Pi \tag{10}$$

Consumer surplus (T) is the total social benefit minus the total cost that users actually pay. The total social benefits (also known as the users' willingness to pay) for each of the zones can be obtained by inverting the demand functions (Equations 2a and 2b) to find the fare as a function of demand and by integrating the inverted functions from zero to Q_1 and Q_2, respectively. Then the total consumer surplus (T) can be stated as

$$T = \frac{PY(E - L)}{2e_p}\left[1 - e_w\frac{H}{2} - e_a\left(\frac{M}{4g} - \frac{E - L}{2g}\right) - e_{iv}\frac{L}{V} - e_pf\right]^2 + \frac{PYL}{2e_p}\left(1 - e_w\frac{H}{2} - e_a\frac{M + S}{4g} - e_{iv}\frac{L}{2V} - e_pf\right)^2 \tag{11}$$

Therefore, the social welfare objective can be formulated as follows:

$$W = \frac{PY(E - L)}{2e_p}\left[1 - e_w\frac{H}{2} - e_a\left(\frac{M}{4g} + \frac{E - L}{2g}\right) - e_{iv}\frac{L}{V} - e_pf\right]^2 + \frac{PYL}{2e_p}\left(1 - e_w\frac{H}{2} - e_a\frac{M + S}{4g} - e_{iv}\frac{L}{2V} - e_pf\right)^2 + PY(E - L)\left[1 - e_w\frac{H}{2} - e_a\left(\frac{M}{4g} + \frac{E - L}{2g}\right) - e_{iv}\frac{L}{V} - e_pf\right]f + PYL\left(1 - e_w\frac{H}{2} - e_a\frac{M + S}{4g} - e_{iv}\frac{L}{2V} - e_pf\right)f - \frac{2cYL}{HMV} \tag{12}$$

In solving for the maximization of social welfare a deficit constraint is considered. This constraint states that the operator cost must be equal to the sum of the total revenue R and a prespecified acceptable level of subsidy K, namely

$$C = R + K \tag{13}$$

Therefore, the deficit constraint is as follows:

$$\frac{2cYL}{HMV} - PY(E - L)\left[1 - e_w\frac{H}{2} - e_a\left(\frac{M}{4g} + \frac{E - L}{2g}\right) - e_{iv}\frac{L}{V} - e_p f\right]f - PYL\left(1 - e_w\frac{H}{2} - e_a\frac{M + S}{4g} - e_{iv}\frac{L}{2V} - e_p f\right)f - K = 0 \qquad (14)$$

The breakeven constraint is introduced by eliminating subsidies (i.e., $K = 0$) from Equation 14.

Unconstrained Subsidy Results

If the subsidy is unconstrained the first-order conditions at optimum are

$$\frac{\partial W}{\partial L} = 0 \qquad (15)$$

$$\frac{\partial W}{\partial M} = 0 \qquad (16)$$

$$\frac{\partial W}{\partial H} = 0 \qquad (17)$$

$$\frac{\partial W}{\partial f} = 0 \qquad (18)$$

The optimized fare can be immediately obtained for Equation 18 and is $f^* = 0$.

This result is not surprising because the marginal operator cost is zero according to the assumptions made so far. The marginal cost, and thus the fare, would become positive if a vehicle capacity constraint is introduced, as will be shown later.

By substituting a zero fare back into Equations 15 to 17, the expressions for the optimal route length, spacing, and headway are obtained, and they are given in Appendix A.

Results with Breakeven Constraint

To solve the problem by using the breakeven constraint, the constraint was introduced into the objective function (Equation 12) with a multiplier, λ. The purpose of λ is to introduce a penalty for violating the constraint. In economic terms it is the "shadow price" of the subsidy (i.e., it indicates the change in welfare that will result from a $1 subsidy).

The expressions for the optimal route length, spacing, headway, and fare assuming a breakeven constraint (i.e., no subsidy) are also shown in Appendix A. A detailed derivation of these expressions can be found in Spasovic et al. (*23*).

Capacity Constrained Headway

The models for maximizing either operator profit or social welfare presented so far have not taken into account a vehicle capacity constraint. This constraint ensures that the total capacity provided on the routes satisfies the demand at some reasonable LOS by restricting the maximum allowable headway. The constraint is written as

$$PYE\left(1 - e_w\frac{H}{2} - e_a\frac{M}{4g} - e_p f\right) + PY(E - L)\left(- e_a\frac{E - L}{2g} - e_{iv}\frac{L}{V}\right) + PYL\left(- e_a\frac{S}{4g} - e_{iv}\frac{L}{2V}\right) \le k\frac{Y}{MH}l \qquad (19)$$

where k is the capacity of transit vehicle (in spaces), and l is the allowable peak load factor at the CBD. The expression for maximum allowable headway, derived from Equation 19, is used within an optimization algorithm that is described next.

OPTIMIZATION ALGORITHM

Although the models presented so far provided valuable insights into the relations among decision variables and exogenous parameters, they are too complex for simultaneously optimizing all of the decision variables algebraically. To solve the model an algorithm that sequentially used Equations 7a to 7d (or Equations a1 to a3 or Equations b1 to b5 in the Appendix, depending on the objective to be optimized) was developed to advance from an initial feasible solution toward the optimal solution. The algorithm starts with a trivial feasible solution and in each step improves the value of the objective function by computing an optimal value of one decision variable while keeping the others at their feasible levels. In computing the optimal values of decision variables, the algorithm computes sequentially the route length, route spacing, headway, and finally fare. In each step the value of a newly computed variable is recorded and used in the next step for computing the optimal values of the other decision variables. The algorithm keeps improving the objective function until it converges to an optimal solution. It terminates when the values of the objective functions from two successive iterations are sufficiently close and no significant further improvement can be expected. The objectives turned out to be relatively flat (shallow, four-dimensional, U-shaped) functions. Thus small deviations from the optimal decision variables result in even smaller relative changes in the values of the objectives.

It is quite possible that buses may overload if no capacity constraint is introduced. Instead of formulating a model as a constrained optimization problem with a nonlinear objective function and a linear constraint and solving it by using a penalty method, the following modification of the algorithm is made to incorporate the vehicle capacity constraint:

1. Examine whether the newly obtained optimal headway satisfies the capacity constraint, by computing the optimal busload and checking whether the busload exceeds capacity.
2. If the busload is smaller than the available capacity there is no need for capacity-constrained results.
3. Otherwise set the optimal headway equal to the maximum

allowable headway (obtained by solving Equation 19), which is as follows:

$$H^* = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \tag{20}$$

where

$$B = -(E - L)^2 \frac{e_a}{2g} + E\left(1 - e_a \frac{M}{4g} - e_p f\right) - e_{iv} \frac{(E - L)L}{V} - L\left(e_a \frac{S}{4g} + e_{iv} \frac{L}{2V}\right),$$

$$A = -e_w \frac{E}{2}, \text{ and}$$

$$C = -\frac{k}{PM}.$$

Then calculate the set of decision variables that satisfies the capacity constraint. This is considered to be the optimal solution.

NUMERICAL EXAMPLE

A numerical example is developed to demonstrate how the model optimizes transit service coverage.

Table 2 gives results from the maximization of operator profit and social welfare (with both the unconstrained subsidy and breakeven constraint) objectives. The results include optimal route length, route spacing, headway, fare, operator profit, social welfare, and consumer surplus for a rectangular corridor of 8.045 × 4.824 km (5 × 3 mi) with a potential demand density of 77.35 passengers/km^2-hr (200 passengers/m^2-hr). The hourly operating cost of the bus is assumed to be $40/vehicle, the average transit speed is assumed to be 16.09 km/hr, and the average access speed is assumed to be 4.02 km/hr. The transit vehicle capacity is 50 seats/vehicle, the allowable peak load factor is 1, the stop spacing is 0.402 km/stop, and the sensitivity factors for waiting time, access time, in-vehicle time, and fare are 0.7, 0.7, 0.35, and 0.5, respectively.

Under the profit maximization objective the optimal route length is 5.3 km (3.296 mi), route spacing is 1.614 km (1.004 mi), headway is 0.201 hr, and the fare is $0.88. The induced hourly ridership is 743 passengers, yielding a $264 profit.

Under the welfare maximization objective with unconstrained subsidy the optimal service design has a route length of 4.560 km (2.834 mi), route spacing of 1.120 km (0.699 mi), headway of 0.140 hr, and a $0.36 fare. The induced hourly ridership is 1,536 passengers, the social welfare is $719, and the subsidy is $150. The introduction of the breakeven constraint results in an optimal service design with a route length of 4.57 km (2.838 mi), route spacing of 1.19 km (0.739 mi), headway of 0.148 hr, and a $0.45 fare. These service variables induce an hourly ridership of 1,372 passengers, yielding a social welfare of $710.

A comparison of the welfare maximization results with the unconstrained subsidy and breakeven constraints reveals that when the breakeven constraint is removed the welfare increases slightly (by $8.50 or approximately 1.2 percent), whereas the deficit increases much more (from $0 to $150). The welfare function appears to be relatively flat near the optimum. This indicates that minor deviations from the optimum will not decrease welfare significantly. This result is similar to the one found by Chang and Schonfeld (*20*). This implies that for a given set of input data the welfare objective with a breakeven constraint seems quite reasonable and far more desirable from an economical standpoint than the welfare objective with unconstrained subsidy.

The shadow price in Table 2 implies that relaxing the breakeven constraint and thus increasing the operator deficit from $0 to $1/hr (i.e., to a $1/hr subsidy) will result in a $0.128/hr increase in welfare.

The equilibrium demand is strongly influenced by the level of service and the fares optimized under different objectives. The total hourly demand level is 24.8 percent of the potential demand under profit maximization. It is 51.2 and 45.76 percent of the potential demand under welfare maximization for the unconstrained subsidy and breakeven conditions, respectively.

A comparison of the optimal route length for different objectives indicates that profit maximization yields longer routes of 5.3

TABLE 2 Optimal Objectives and Design Variables

	Objective Functions		
	Profit	Social Welfare	
		Unconstrained	Break-Even
Route Length (km)	5.3 (3.296 mi)	4.56 (2.834 mi)	4.57 (2.838 mi)
Route Spacing (km)	1.61 (1.004 mi)	1.12 (0.699 mi)	1.19 (0.739 mi)
Headway (hr)	0.201	0.140	0.148
Fare ($)	0.88	0.36	0.45
Ridership (pass/hr)	744	1536	1372
Operator Cost ($/hr)	392	696	623
Revenue ($/hr)	656	546	623
Profit ($/hr)	264	-150	0
Consumer Surplus ($/hr)	2367	869	710
Welfare ($/hr)	501	719	710
Bus Load (pass/bus)	50	50	49.267
Fleet Size (buses)	3.28	4.05	3.83
Shadow Price of Subsidy Increase ($/hr)	N/A	N/A	0.128

km (3.29 mi) than welfare maximization [4.56 km (2.834 mi)]. Also the optimal design for profit maximization has longer headway and route spacing than that of welfare maximization [i.e., 0.201 hr and 1.61 km (1.004 mi) versus 0.14 hr and 1.120 km (0.699 mi)]. This can be explained by the presence of the vehicle capacity constraint and the customers' higher value for waiting and access times than for in-vehicle time (as indicated by the values of e_w, e_a, and e_{iv}) that replace the transit system with one with denser routes and more frequent service so that welfare can be maximized.

A sensitivity analysis was performed to show how changes in the more important exogenous parameters given in the numerical example affect the values of the decision variables and objective functions. The changes in design variables, namely route length, spacing, headway, and fare with respect to the corridor length, passenger density, transit and access speed, operator cost, sensitivity factors, and fare, are shown in Table 3. The two values used for each parameter are between 10 and 20 percent above and below those that were used to generate the basic results of Table 2.

Table 3(*A*) illustrates the effects that changes in parameters have on the optimal design variables under profit maximization. For example if corridor length is increased by 10 percent (from 8.045 to 8.8495 km) the route length is increased by 10.4 percent. This implies that the optimal route length L is elastic (i.e., the absolute

TABLE 3 Sensitivity Analysis

A) FOR PROFIT MAXIMIZATION

		Design Variables and Objective Function					
Parameters		Length (km)	Spacing (km)	Headway (hr/veh)	Fare ($/pass)	Profit ($/hr)	Demand (pass/hr)
Corridor Length (km)	7.24	4.761	1.636	0.203	0.87	286.33	725
	8.85	5.845	1.610	0.200	0.89	234.96	749
Density (pas/km^2-hour)	69.53	5.303	1.71	0.213	0.87	227.66	664
	84.98	5.303	1.535	0.191	0.89	301.39	824
Transit Speed (km/hr)	14.48	4.962	1.684	0.209	0.86	211.68	684
	17.7	5.573	1.562	0.194	0.90	313.02	795
Access Speed (km/hr)	3.62	5.601	1.556	0.215	0.89	239.27	722
	4.425	4.997	1.672	0.189	0.87	287.96	764
Stop Spacing (km)	0.362	5.314	1.614	0.201	0.89	265.96	746
	0.442	5.292	1.617	0.201	0.88	262.53	742
Operator Cost ($/hr)	36	5.565	1.570	0.195	0.89	305.61	787
	44	5.041	1.664	0.207	0.87	227.06	702
(e_w, e_a)	0.6	4.788	1.574	0.196	0.88	318.01	784
	0.8	5.673	1.655	0.206	0.87	217.66	708
(e_{iv})	0.25	4.788	1.578	0.196	0.92	298.08	779
	0.45	5.673	1.655	0.206	0.84	232.34	708
(e_p)	0.4	5.828	1.527	0.190	1.11	439.37	832
	0.6	4.777	1.712	0.213	0.72	161.46	662
Optimal Results*		5.300	1.61	0.201	0.88	264.24	744

B) FOR WELFARE MAXIMIZATION WITH BREAK-EVEN CONSTRAINT

		Design Variables and Objective Function					
Parameters		Length (km)	Spacing (km)	Headway (hr/veh)	Fare ($/pass)	Welfare ($/hr)	Demand (pass/hr)
Corridor Length (km)	7.24	4.22	1.199	0.149	0.42	733.53	1353
	8.85	4.854	1.195	0.149	0.48	678.40	1358
Density (pas/km^2-hour)	69.53	4.563	1.263	0.157	0.45	623.86	1218
	84.98	4.568	1.128	0.140	0.45	797.96	1528
Transit Speed (km/hr)	14.48	4.149	1.250	0.155	0.45	618.77	1243
	17.7	4.927	1.144	0.142	0.44	798.07	1483
Access Speed (km/hr)	3.62	4.798	1.152	0.159	0.47	663.54	1315
	4.425	4.327	1.224	0.138	0.43	754.92	1424
Stop Spacing	0.362	4.574	1.187	0.148	0.46	714.47	1376
	0.442	4.557	1.190	0.148	0.45	706.60	1369
Operator Cost ($/hr)	36	4.894	1.150	0.143	0.44	785.33	1466
	44	4.256	1.231	0.153	0.46	644.46	1283
(e_w, e_a)	0.6	4.167	1.145	0.142	0.41	805.65	1479
	0.8	4.849	1.236	0.154	0.48	627.74	1272
(e_{iv})	0.25	4.761	1.157	0.144	0.47	768.93	1451
	0.45	4.37	1.224	0.152	0.43	657.04	1295
(e_p)	0.4	5.239	1.115	0.139	0.52	1087.39	1564
	0.6	3.966	1.274	0.158	0.39	488.42	1197
Optimal Results*		4.57	1.19	0.148	0.45	710.52	1372

* For the values of the exogenous parameters given in the numerical example

value of the elasticity exceeds 1.0) with respect to the corridor length *E*. The reason for this is that, as the length of the corridor *E* is increased, the length of the area between the terminus and the end of the corridor ($E - L$) is increased very slowly, thus increasing *L* faster than *E*. This result is consistent with those obtained by Spasovic and Schonfeld (*16*) for fixed-demand systems. Also if the passenger density is increased by 10 percent the headway will be reduced by 5 percent. This result confirms that headway varies inversely with the cube root (approximately) of the passenger density. Table 3(*A*) also shows that the route length would decrease by 10 percent if the sensitivity factor for fare is increased by 20 percent (from 0.5 to 0.6).

Table 3(*B*) shows the effect that changes in parameters have on the optimal design variables under welfare maximization with a breakeven constraint.

The effect of the route length on profit and on welfare is shown in Figure 3. For a given route length the system design variables have been reoptimized, yielding the optimal profit or welfare. The profit and welfare functions are relatively flat near the optimum. A practical application of this result is that, for a given set of data, the optimal design variables can be tailored to the actual street network without substantially reducing the optimal profit or welfare.

The effect of subsidy on welfare and consumer surplus is shown in Figure 4. For a given subsidy level, the system design variables have been reoptimized, yielding the optimal welfare. The consumer surplus increases with subsidy. For no subsidy the breakeven constraint holds and the social welfare equals consumer surplus. The net effect of the profit and consumer surplus interactions is that the welfare function is relatively flat near the optimum for a relatively large range of subsidies. A practical implication of this result is that for a given set of data the breakeven constraint may be economically and politically preferable because it eliminates subsidy and marginally reduces social welfare. Furthermore Figure 4 shows that a negative subsidy (profit) can be obtained by marginally decreasing welfare.

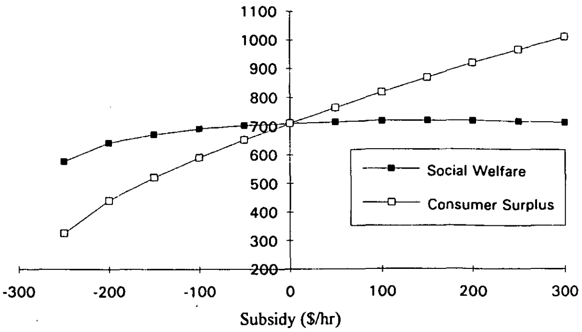


FIGURE 4 Social welfare and consumer surplus for various subsidy levels.

CONCLUSIONS

The paper presented a model of optimal bus transit service coverage that was optimized to maximize profit and social welfare with unconstrained subsidy and a breakeven constraint. The model provides simple guidelines for optimizing the extent of transit routes and other major operating characteristics such as route spacing, headway, and fare. Equations 7a to 7d can be used to optimize route length, route spacing, headway, and stop spacing separately. They provide insights into the interrelationships among the optimized variables. For example the cube root in Equations 8b and 8c indicates that optimal solutions for headway and route spacing are relatively insensitive to changes in system parameters.

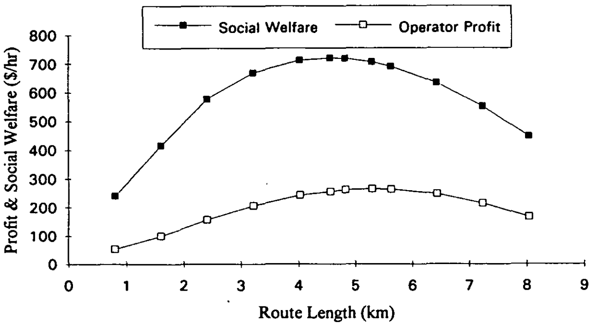


FIGURE 3 Impact of route length on design objectives.

The optimality of a constant ratio between route spacing and headway, which has been found in previous studies for various bus network and demand conditions (*12.20*), is also found to be maintained in the present study, which optimized the route length as well. The route spacing and headway that optimize profit, welfare with unconstrained subsidy, and welfare with a breakeven constraint closely maintain a ratio of 5.00, irrespective of the values of the other parameters such as potential demand density, sensitivity factors, or speed.

The profit and social welfare functions are relatively flat near the optimum. For practical applications this implies that a near-optimal profit or welfare can be attained while fitting the transit network to the particular street network or modifying its operating schedule.

The results of maximization of social welfare for different subsidy levels indicate that the welfare function is relatively flat near the optimum. A practical application of this result is that for a given set of data the subsidy can be reduced (or eliminated) by providing passengers a service with marginally worse quality. Therefore the welfare objective under a breakeven constraint seems reasonable and more desirable from an economical standpoint than the welfare objective with unconstrained subsidy. Furthermore for a given set of input data in the numerical example a negative subsidy or profit can be obtained for a marginal decrease in social welfare.

FUTURE RESEARCH POSSIBILITIES

Several simplifying assumptions could be relaxed in future models. The linear demand function may be replaced by a nonlinear function that more precisely reflects traveler behavior. More realistic and irregular distributions of temporal and spacial demand (e.g., nonuniform lateral distributions) could be used. The model could be improved to handle non-CBD trips (e.g., many-to-many travel pattern) and access modes other than walking. A modified model could handle sectorial service areas with possible overlaps in service coverage among the routes, and the assumption that average stop spacing is constant could be relaxed. The impacts of passengers boarding and alighting on bus dwell time, cruising speeds, and the cost of operations may also be included.

ACKNOWLEDGMENT

The authors appreciate the partial support provided by the National Center for Transportation and Industrial Productivity at the New Jersey Institute of Technology funded by the University Transportation Centers Program.

APPENDIX A
Optimal Bus Transit Service Design Variables

ROUTE LENGTH

(*a*) With unconstrained subsidy

$$L^* = R - {}^1\!/_2\,(S + T)^{1/2}/(C^2 - A^2)\sqrt{HMVP} \qquad \text{(a-1)}$$

(*b*) With breakeven constraint

$$\begin{aligned} L^* = R - {}^1\!/_2\{&S(1 + \lambda) + T \\ &+ fHMVPe_p 2(A + B)[B(A + 3C) \\ &+ D(3A + C)](1 + \lambda) \\ &+ fEHMVPe_p 2A(A - 3C)(A + C) \\ &+ f^2HMVPe_p^2 4(A + C)^2(1 + \lambda)^2\}^1\!/_2 \\ &\div (C^2 - A^2)\sqrt{HMVP} \end{aligned} \qquad \text{(b1)}$$

where, for Equations a1 and b1

$$A = \frac{e_a}{2g} - \frac{e_{iv}}{V},$$

$$B = 1 - e_w\frac{H}{2} - e_a\frac{M}{4g} - e_a\frac{E}{2g},$$

$$C = \frac{e_{iv}}{2V},$$

$$D = 1 - e_w\frac{H}{2} - e_a\frac{M + S}{4g},$$

$$R = \frac{-[2(AB + CD) - A^2E]}{3(A^2 - C^2)},$$

$$S = 12ce_p(A^2 - C^2), \text{ and}$$

$$\begin{aligned} T = {}&HMVP[(AB + CD)^2 + 3(AD + BC)^2] \\ &+ EHMVPA(A^3E + 2A^2B - 4ACD - 6BC^2). \end{aligned}$$

HEADWAY

(*a*) With unconstrained subsidy

$$H^* = \frac{\sqrt{J}}{A + B(1 - C - D + F - 2G) + L(1 - G - I)} \qquad \text{(a2)}$$

(*b*) With breakeven constraint

$$H^* = \frac{\sqrt{J(1 + \lambda)}}{\begin{aligned} A + B(1 - C - D + F - 2G) + L(1 - G - I) \\ + epf[-B + (E - L) + E\lambda] \end{aligned}} \qquad \text{(b2)}$$

ROUTE SPACING

(*a*) With unconstrained subsidy

$$M^* = \frac{\sqrt{J'}}{A' + B(1 - C' - D + F - 2G) + L(1 - G - I')} \qquad \text{(a3)}$$

(*b*) With breakeven constraint

$$M^* = \frac{\sqrt{J'(1 + \lambda)}}{\begin{aligned} A' + B(1 - C' - D + F - 2G) + L(1 - G - I') \\ + e_pf[-B + (E - L) + E\lambda] \end{aligned}} \qquad \text{(b3)}$$

where, for Equations a2, b2, a3, and b3

$A = -Ee_w/2,$
$A' = -Ee_a/2,$
$B = E - L,$
$C = e_aM/4g,$
$C' = e_wH/2,$
$D = e_aE/2g,$
$F = e_aL/2g,$
$G = e_{iv}L/2V,$
$I = e_a(M + S)/4g,$
$I' = e_a(2Hg + S)/4g,$
$J = (4ce_pL)/(MVPe_w)$, and
$J' = (8cge_pL)/HVPe_a.$

FARE

(*a*) With breakeven constraint

$$f^* = \frac{\begin{aligned} 2g\lambda Le_{iv}(-2E + L) - 2e_a\lambda V(E - L)^2 \\ + e_a\lambda V(LS - EM) + 2Eg\lambda V(2 - He_w) \end{aligned}}{4Ege_pV(1 + 2\lambda)} \qquad \text{(b4)}$$

(*b*) Shadow price for breakeven constraint:

$$\lambda^* = \frac{-X + \sqrt{X^2 - 4(A - B)(X + ICDJ + IFGJ - Z)}}{2(X + ICDJ + IFGJ - Z)} \qquad \text{(b5)}$$

where, for Equation b5

$X = 4A - 4B + ICDJ + IFGJ$,
$Z = I^2e_p(C + F)$, and
$A = 2cLY/HMV$,
$C = (E - L)PY$,
$D = 1 - e_a(E - L)/2g - e_aM/4g - Le_{iv}/V - He_w/2$,
$F = LPY$,
$G = 1 - e_a(M + S)/4g - Le_{iv}/2V - He_w/2$,
$J = -4ELge_{iv} + 2L^2ge_{iv} - 2E^2Ve_a + 4EVg + 4ELVe_a - 2e_aL^2V - e_aEMV - e_aLSV - 2EgHVe_w$, and
$I = 4EgPV$.

REFERENCES

1. Holroyd, E. M. The Optimal Bus Service: A Theoretical Model for a Large Uniform Urban Area. *Vehicular Traffic Science* (L. C. Edie, ed.), American Elsevier, New York, 1967, pp. 309–328.
2. Vuchic, V. R., and G. F. Newell. Rapid Transit Interstation Spacing for Minimum Travel Time. *Transportation Science*, Vol. 2, No. 4, 1968, pp. 359–374.
3. Vuchic, V. R. Rapid Transit Interstation Spacing for Maximum Number of Passengers. *Transportation Science*, Vol. 3, No. 4, 1969, pp. 214–232.
4. Newell, G. F. Dispatching Policies for a Transportation Route. *Transportation Science*, Vol. 5, No. 1, 1971, pp. 91–105.
5. Byrne, B. F., and V. R. Vuchic. Public Transportation Line Position and Headway for Minimum Cost. *Traffic Flow and Transportation* (G. F. Newell, ed.), American Elsevier, New York, 1972, pp. 347–360.
6. Hurdle, V. F. Minimum Cost Locations for Parallel Public Transit Lines. *Transportation Science*, Vol. 7, No. 4, 1973, pp. 340–350.
7. Byrne, B. F. Public Transportation Line Positions and Headways for Minimum User and System Cost in a Radial Case. *Transportation Research*, Vol. 9, 1975, pp. 97–102.
8. Byrne, B. F. Cost Minimizing Positions, Lengths and Headways for Parallel Public Transit Lines Having Different Speeds. *Transportation Research*, Vol. 10, 1976, pp. 209–214.
9. Hurdle, V. F., and S. C. Wirasinghe. Location of Rail Stations for Many to One Travel Demand and Several Feeder Modes. *Journal of Advanced Transportation*, Vol. 14, No. 1, 1980, pp. 24–45.
10. Wirasinghe, S. C. Nearly Optimal Parameters for a Rail/Feeder-Bus System on a Rectangular Grid. *Transportation Research-A*, Vol. 14, No. 1, 1980, pp. 33–40.
11. Wirasinghe, S. C., and N. S. Ghoneim. Spacing of Bus-Stops for Many to Many Travel Demand. *Transportation Science*, Vol. 15, No. 3, 1981, pp. 210–246.
12. Kocur, G., and C. Hendrickson. Design of Local Bus Service with Demand Equilibration. *Transportation Science*, Vol. 16, No. 2, 1982, pp. 149–170.
13. Tsao, S. M., and P. M. Schonfeld. Optimization of Zonal Transit Service. *Journal of Transportation Engineering*, ASCE, Vol. 109, No. 2, 1983, pp. 257–272.
14. Tsao, S. M., and P. M. Schonfeld. Branched Transit Service: An Analysis. *Journal of Transportation Engineering*, ASCE, Vol. 110, No. 1, 1984, pp. 112–128.
15. Wirasinghe, S. C., and P. N. Seneviratne. Rail Line Length in an Urban Transportation Corridor. *Transportation Science*, Vol. 29, No. 4, 1986, pp. 237–245.
16. Spasovic, L. N., and P. M. Schonfeld. A Method for Optimizing Transit Service Coverage. In *Transportation Research Record 1402*, TRB, National Research Council, Washington, D.C., 1993, pp. 28–39.
17. Kuah, G. K., and J. Perl. Optimization of Feeder Bus Routes and Bus Stop Spacings. *Journal of Transportation Engineering*, ASCE, Vol. 114, No. 3, 1988, pp. 341–354.
18. Chang, S. K., and P. M. Schonfeld. *Analytical Optimization of Parallel Bus Routes with Time Dependent and Elastic Demand.* Transportation Studies Center Working Paper 89-22. University of Maryland, College Park, 1989.
19. Chang, S. K., and P. M. Schonfeld. Optimal Dimensions of Bus Service Zones. *Journal of Transportation Engineering*, ASCE, Vol. 119, No. 4, 1993, pp. 567–585.
20. Chang, S. K., and P. M. Schonfeld. Welfare Maximization with Financial Constraints for Bus Transit System. In *Transportation Research Record 1395*, TRB, National Research Council, Washington, D.C., 1993, pp. 48–57.
21. Morlok, E. K., and P. A. Viton. Feasibility of Profitable Transit Service in Bus Transport. In *Transportation Research Record 980*, TRB, National Research Council, Washington, D.C., 1984, pp. 46–54.
22. Viton, P. A. The Possibility of Profitable Bus Service. *Journal of Transport Economics and Policy*, Vol. 16, No. 1, 1980, pp. 85–94.
23. Spasovic, L. N., M. P. Boile, and A. K. Bladikas. *A Methodological Framework for Optimizing Bus Transit Service Coverage.* National Center for Transportation and Industrial Productivity Working Paper 93-01. New Jersey Institute of Technology, Newark, 1993.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Novel Methodological Approach to Transit Network Analysis: Application to Tel Aviv Metropolitan Area

PETER VOVSHA AND TOMER GOODOVITCH

Analysis of changes in spatial structure of transit lines represents a serious mathematical and computational problem, one that has not been resolved by existing models and transportation software. An original methodology to enable quantitative analysis of spatial characteristics of transit networks is presented. A set of criteria is developed, and practical experience of the technique is presented as developed for the Tel Aviv metropolitan area.

The Tel Aviv metropolitan area is a contiguous urbanized region with a transit system based on two large, privately owned and operated bus cooperatives with over 200 lines. In 1993 this system was serving a metropolis of 2 million people, 400,000 of whom resided in the core city of Tel Aviv. Historical allocation of bus routes preserved for many years a segregation of the intrametropolitan "Dan" company from the intercity operator "Eged." However, the urbanization process overflowed to the fringes of the region, the spread of the built-up area has blurred the border between intercity and intracity transit. Development of the transit network (TN) has not followed an organized pattern with systematic planning, and this has led to inefficient route structure, with many lines running along the same major corridors and most ending at the central bus station. Today Tel Aviv's TN is characterized by line duplication and levels of service that do not attend to the public's needs.

The contracts that each company has with the Transport and Finance ministries determine the service level, fares, and subsequent government subsidy; in most instances, the municipalities are not involved in transportation planning. Reorganization of the TN generally would be initiated by the bus companies themselves but would require official approval from the Ministry of Transport before going into effect. However, the government has little information and lacks adequate tools to evaluate proposed changes. It has operated on an intuitive and experimental instead of quantitative basis. This situation worsens when service modifications by both companies are introduced for the same area. In such a case, the transit authority also has to choose which proposition to approve. Present policy for route allocation forces compensatory balancing between the companies even though that means unnecessary duplication of Tel Aviv's TN.

Any change to the current situation is bound by a regulatory process and public scrutiny and encounters political pressure to preserve current economic conditions and government subsidies to the operators. Government regulation is based on maintaining a fine economic balance between the cooperatives. Structural change to TN, although economically and socially justifiable, could disturb this stability. Therefore, only marginal changes and not a system improvement have been acceptable. The novel method of TN analysis presented in this paper is based on research undertaken for the Ministry of Transport to facilitate the decision making in TN planning and development.

METHODS FOR TRANSIT NETWORK ANALYSIS

Existing methods for TN analysis are diverse, ranging from rule of thumb guidelines to computer packages. All such methods fall into one of two groups according to Nes et al. (*1*): evaluation methods for a predefined TN and construction methods for new TN design. Evaluation is a necessary element of any construction procedure, so construction methods also employ appropriate evaluation subroutines, see Baaj and Mahmassani (*2*). If planning a local improvement or comparing a limited number of alternatives it is sufficient for planners to perform the evaluation stage only. Evaluation and construction methods can be divided into three subgroups: (*a*) descriptive (based on a simplified calculation of performance measures and a good portion of intuition for TN design), (*b*) heuristic (usually allows user intervention at crucial points of TN design, but performs computerized evaluation of performance and marginal TN improvements), and (*c*) formal (mathematical optimization of TN).

Axhausen and Smith (*3*) have presented a comprehensive review of TN optimization algorithms. Practically all of them consist of three main steps: (*a*) initial TN construction (or input), (*b*) route development (by link addition and deletion), and (*c*) selection of optimum route set. Optimization methods for TN design are classified according to whether both main variables, routes and frequencies, are to be included (*1*). A new algorithm was developed that includes three stages: (*a*) route generation, (*b*) analysis procedure, and (*c*) route improvement (*2*). Summarizing the more than 20-year history of TN optimization, the work by Baaj and Mahmassani (*2*) lists difficulty of formulation, nonlinearity and nonconvexity [see also (*4*)], combinatorial explosion for a large TN, multiobjective nature [see also (*5*)], and difficulty of formalization of spatial layout of routes as problematic areas.

Combinatorial explosion is predetermined by the nature of a transit line that can be represented as a combination of links whose number is multiplied with an increasing number of nodes. Taken together with nonlinearity and nonconvexity, this practically prohibits application of optimization methodology for TN design in large cities. All of existing effective algorithms (*1,2,6–14*) have been applied successfully for relatively small TNs made

---

The Israel Institute of Transportation Planning & Research, 7 Nahal Ayalon Street, Tel Aviv, Israel 61090.

up of not more than 100–250 nodes. A fairly schemative representation of TN in Tel Aviv metropolitan area requires about 2,000 nodes.

The problem of size is closely connected to the formalization of the spatial layout of routes. Frequency allocation to a given set of routes presents a particular problem that can be successfully resolved by using optimization techniques (*15–17*). It is spatial configuration of transit lines (layout of routes) that generates the most cumbersome set of decision-making variables and complicates both evaluation and construction stages of TN analysis.

In a descriptive approach, the planner visually defines the form of TN development. This can be strengthened by use of any evaluation tool. Certain planning guidelines are suggested, based on interviews with transit agencies over a broad spectrum of U.S. and Canadian cities (*18*). Schneider and Smith (*19*) proposed a complete concept for redesigning urban TN (a transit-center-based approach). Alternative route configurations determined by practical considerations were evaluated (*20*). Criteria for TN design suggested by Giannopoulos (*21*) include minimal demand, straightness of lines, and avoidance of overlapping. These practical guidelines give valuable information on basic structural features of an analyzed (designed) TN that, transformed into quantitative measures, can be used as input into an optimization procedure, ensuring integration of formal and intuitive planning techniques.

LEVELS OF TRANSIT NETWORK ANALYSIS

Several levels of detail for TN analysis are possible (Table 1). The first level deals with spatial configuration of TN. A major weak-

TABLE 1 Levels of Transit Network Analysis

Level	Scope	Core of Evaluation	Core of Construction
Line Spatial Configuration	Route Layout, Stop Spacing, Service type, Frequency, Traffic Constraints, Transit Facilities (Vehicle type, Number), Speed	Transit Operation Simulation[a], then Transit Performance Indicators Calculation (including Transit System Cost)[a]	Alternative Networks Development under Coverage or other Service Standard Conditions[b], Systemwide Network Optimization
Trip Distribution	Travel Demand for Transit (+ as above)	Transit Assignment with Fixed Demand[a], Transit Performance Indicators Calculation (including Transit System Cost and Level of Service)[a]	Alternative Networks then Development under Demand with Service Standard Conditions[b], Systemwide Network Optimization
Mode Interaction	Total Travel Demand, Alternative Modes (+ as above)	Modal Split[b] (Multimodal Assignment[a]), Diverted Demand Definition, then Transportation Performance Indicators Calculation (including Transport System Cost and Level of Service)[a]	Alternative Networks Development by Different Scenarios[c]
Linkage with Land-Use	Population Employment and Activity System (+ as above)	Trip Generation and Distribution Sensitive to Transit Accessibility[b], Derivative Demand Definition, then Modal Split[b] (Multimodal Assignment[a]), then Transportation Performance Indicators Calculation (including Transport System Cost, Level of Service and other Social Benefits[c])	Alternative Networks Development by Different Scenarios

a - Implemented Applicapable Software
b - Prospective Software
c - First Attempts to Formulation

ness of such an approach is obvious: TN will not directly follow the demand.

For this reason, routing based on road network calculations (e.g., the shortest path between predetermined terminals) is recognized as ineffective, even for initial route skeleton construction (*3*). In later proposed algorithms, special assignments (''expanded'' or ''first and second'' shortest path) of the demand matrix on the road network (*2*) were applied. Also, account must be taken of the approximate character of the demand matrix. If the demand matrix has been obtained by a trip survey, it will be biased toward existing transit corridors. If demand has been estimated by means of modeling (trip generation and distribution), it could likely be abstract and not sensitive to TN alternatives.

Spatial configuration can be thought of as a self-defining basis of a TN. A basic framework for TN development that cannot be completely formalized is most naturally expressed at this level. Such analysis, which can be performed without a data-taking demand component, is capable of reasonable TN evaluation and can be useful, especially in a preliminary stage in which a large number of solutions is to be reduced to a limited alternative pool. An unknown demand matrix can to a certain extent be substituted by coverage indicators on the basis of service standards per capita or employee. A number of packages can be applied for this task (*22,23*), but existing approaches have concentrated more on passive performance estimators than on TN weakness identification.

The second level is based on a known transit demand matrix. The core of the evaluation is a transit assignment procedure that allows trip distribution by lines and further estimation of transit system cost and components of level of service. The algorithm implemented in the IANO package (*13*) uses an assignment procedure at the last stage of line recombination. The EMME/2 package has a transit assignment module that can simulate the operation of a large TN. Another variant of transit assignment has been proposed (*2*). Transit assignment today is the most accessible tool for planning practice that is useful for comparison of TN alternatives under a fixed demand condition. Trip distribution is placed second to spatial configuration as a planning tool. The capability of existing software for TN construction is extremely limited by TN size.

The third level gives a new facet for substantiation of planning decisions, namely, possible diverted demand from private car to public transport as a result of level of service improvement (assuming total travel demand is fixed). The VOLVO package (*12*) suggested estimation of TN improvement's impact on demand with an independent mode choice model. A deterrence function was borrowed from the simultaneous distribution–modal split model (*24*), which describes the relation between supply and demand (*1*). The conclusion was reached that an external modal split model should be applied iteratively with a TN optimization algorithm to ensure supply-demand equilibrium (*3*). The EMME/2 package has a macro-option of multimodal assignment as a combination of successive automobile and transit assignments, with a demand function incorporating mode choice attributes. The difficulty of running such a technique limits its practical use today.

The fourth level is achievable only in the long-term masterplanning framework in which transport system development relates to regional land use. Such large-scale transit projects as a new LRT or subway, which can radically change the accessibility of an area, will inevitably influence activity patterns, including residential and employment choice, and will generate new (derivative) travel demand. However, for practical transit planning this approach is impossible.

PRINCIPLES AND CRITERIA FOR SPATIAL CONFIGURATION ANALYSIS

Table 2 presents criteria used in various approaches at the first and second levels of TN analysis. Among the most common are operational cost, fleet size, and route directness, which serves for initial skeleton construction in almost all TN optimization algorithms. Level-of-service estimators at spatial configuration level were often not formulated, leaving this to a trip distribution stage when travel time or the number of transfers can be estimated by transit assignment. The lack of measures of spatial configuration from the passengers' point of view precludes TN construction subroutines from producing a good starting TN in most of the algorithms (*3*).

A number of criteria have been introduced in descriptive approaches and recent formal models. Among these are area coverage and direct connection possibility. Two spatial characteristics have been formulated (*19*): location of focal points (transfer centers) and service type (e.g., express or collector), but they are in need of quantification. Route duplication is noted as significant for TN design quality (*18* and *21*). An empirical maximum acceptable level of route overlapping was estimated as 50 percent (*21*). Duplication was introduced into the route joining subroutine of the TN improvement algorithm (*2*).

Ridership measures (passenger-miles, passenger per length) are the most commonly used indicator of system effectiveness, after trip distribution by lines. Every planner wants first to ensure that a proposed line will get reasonable ridership. Two level-of-service indicators have been established (*3*): demand density (ratio of the number of trips with 0,1,2 transfers to the number of origin-destination pairs connected with 0,1,2 transfers) and time deformation (difference in travel time between an optimum TN and actual TN). Demand slices by number of transfers are also considered (*2*).

To incorporate different criteria on a uniform methodological platform, systemwide TN characteristics should be stressed. The criteria list must be completed by indicators of TN connectivity and supplementation of each line to others. In real TNs, where great importance is placed on transfers, searching for the most effective way to deliver passengers on all origin-destination pairs in the entire TN is suggested. The proposed approach is composed of previously applied criteria but concentrates more on TN spatial configuration. It is based on three chief criteria: (*a*) line duplication, (*b*) area coverage and transit accessibility, and (*c*) network integration.

Line duplication allows recognition of a similarity of lines. Duplication represents an undesirable factor that negatively affects bus occupancy and adds nothing to the level of service. Significant duplication among lines reveals poor TN design. Practically, a systemwide duplication check can be reduced to a sequential check of all pairs of lines in the TN. For this reason, only pair duplication is described.

Line duplication (D) is composed of three components: (*a*) route overlapping (DR), (*b*) service identity (DS), and (*c*) frequency surplus (DF). D is calculated as

$$D(\%) = \frac{DR(\%) \cdot DS(\%) \cdot DF(\%)}{100(\%) \cdot 100(\%)} \tag{1}$$

TABLE 2 Criteria for Spatial Configuration Analysis (Existing Approaches)

	Spatial Configuration Only										Trip Distribution by Lines						
	System Cost					Level of Service					System Effect		Level of Service				
	1. Operational costs 2. Fleet size 3. Route Directness 4. Route Length 5. Route Duplication					6. Service standards 7. Area coverage 8. Direct connection 9. Focal points 10.Service Type					11.Line/link Ridership 12.Vehicle Occupancy		13. Travel Time 14. Waiting Time 15. No. of Transfers 16. Demand Density 17. Time Deformation				
Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Formal Methods																	
Axhausen,Smith			+	+			+				+		+		+	+	+
Baaj,Mahmassani		+			+	+	+	+			+	+	+	+	+	+	
Billheimer,Gray	+																
Dubois,Bell,Llibre		+	+								+		+				
Furth,Wilson											+			+			
Hasselstroem											+	+			+		
Hsu,Surti			+								+						
Jun,Schnaider	+	+									+			+			
Lampkin,Saalmans		+	+								+		+				
Mandl			+								+		+	+	+		
Marwah	+	+											+		+		
Nebelung			+					+									
Nes, Hamerslag,Immers		+					+		+		+				+		
Rea			+								+			+			
Rosello	+		+	+							+						
Scheele		+									+		+	+			
Sahling			+								+						
Sharp	+												+	+			
Silman, Barzily,Passy			+								+						
Sonntag								+			+						+
Descriptive Methods																	
Chua,Silcock							+										
Giannopoulos	+		+	+	+						+						
NCHRP69			+	+	+	+	+										
Schnaider,Smith		+		+		+	+		+	+	+		+	+			
Thelen,Chatterjee, Wegmann	+						+						+				

All components are scaled from 0 through 100 percent. Zero means full divergence and 100 percent indicates complete identity of lines. Route overlapping is calculated first by comparing lines with regard to mutual segments. If a sufficient level of overlapping is identified (more than 40 percent), service identity and frequency surplus are checked. Criterion (*D*) becomes 0 in a case of lines' divergence by one component at least, but tends to 100 percent only when all components do so.

Route overlapping (*DR*) is calculated as a weighted sum of line segment proximity coefficients (*DD*).

$$DR(\%) = \frac{\sum_i DD_{i,j(i)} \cdot W_i \cdot W_{j(i)}}{\sum_i W_i^2} \qquad (2)$$

where

i = segments of given line,
$j(i)$ = segment of original line most closed to segment i, and
$W(i)$ = segment weight.

Coefficients and consequent route overlapping vary from 0 to 100 percent. The proximity coefficient for a particular segment of the original line gets the value 100 percent if an identical segment exists in the given line. Route overlapping is 100 percent only in the case of complete identity of all segments of given line with appropriate segments of the original line. Details of segment proximity (*DD*) and weight (*W*) estimation vary depending on objectives and data available. Comprehensive discussion on this issue can be found in a work by Cohen et al. (*25*).

In some cases, route overlapping derives from the road network structure. This occurs in southern Tel Aviv, where few possible entries predetermine overlapping for more than 70 routes connected CBD with the southern sector of the city. There are two possibilities for treatment of such enforced overlapping. The first suggests an additional multiplying coefficient for segment weight. The second assumes a permissible overlapping level. A permissible overlapping level for Tel Aviv has been set at 40 percent. Besides the CBD entrance, it reflects enforced overlapping in the central bus-station neighborhood.

Service identity (*DS*) is specified with regard to service type pairs. There are three main types of bus service: (*a*) direct (from origin to destination without intermediate stops), (*b*) express (intermediate stops only at key points), and (*c*) collector (all stops along line route are available). The classification is conventional, and it is possible to further differentiate types of service. The type of line service is specified as follows:

- 100 percent—full service identity (the same type),
- 50 percent—partial service similarity (direct or collector with express), and
- 10 percent—distinct kind of services (direct against collector).

The lines of service reflect a ridership that may "migrate" from one line to another. Detailed description of stop spacing could be a substitute for a service identity estimation. In that case, route overlapping includes service pecularities. Criterion *DS* is of importance for aggregate spatial description of routes and gives supplementary information on route overlapping. For example, an express line passing along a collector line should not be treated as completely duplicating.

Frequency surplus (*DF*) is defined as total frequency over a given maximum level divided by frequency of a given line. When ridership is unknown, the maximum level of frequency is defined a priori, for example, by a 5-min headway during the peak hours that yields (*DF*) values. The criterion depends largely on the maximum level defined. Taking a 3-min headway, one would obtain a concordance for lines both having 6-min headway as opposed to 80 percent surplus. It can also be calculated separately for peak and off-peak periods. Frequency surplus distinguishes between supplementary duplication of lines in the particular segment that needs frequency enhancement and real duplication that means useless bus trips. If transit assignment results are available, maximum reasonable frequency is calculated on the basis of passenger flows assigned but should not be less than a policy headway (15 to 30 min).

Area coverage and transit accessibility (Figure 1) reflect service scope and level for a given line or the TN as a whole. Coverage or accessibility improvement represents an advantage for a given line. Usually accessibility improvement is a main argument for a new line in an existing TN. Coverage reflects an aggregative relationship, "area-service," as accessibility represents a disaggregative relationship, "population-service," within the area unit. If the level of spatial aggregation is low (area unit size is comparable with walking time), coverage and accessibility indicate the same. If aggregation level is high (area unit size is much greater than walking time), accessibility is not defined. An intermediate level suggests a supplementary role: coverage represents aggregative parameters of area transit supply, whereas accessibility reflects internal parameters for each zone.

Coverage is based on a comparison of total parameters of TN with total parameters of the service area. There are two methods derived from two main quantitative characteristics of transit service: total frequency and number of lines (or total length). Total frequency is more usual for transportation planning. Line coverage is more suitable for land use or geographical study.

The frequency group is in turn divided into two types of coverage: nonoriented and oriented. For nonoriented coverage, a total frequency of all lines crossing the area unit is compared with unit travel demand factors (e.g., population and employees). This method is simple and practically available but not sensitive to trip directions. To take directionality into account, oriented coverage should be applied when total frequency and demand are compared for origin-destination pairs. Frequency is summed for all lines connecting a given origin and destination. Demand can be given by a trip matrix or estimated by travel potential (for instance, population and employees). Oriented coverage does not include transfers and is therefore also approximate. Line coverage is usually calculated as total line length/km^2 (transit density), but number of lines/km^2 is also of interest because it characterizes service multitude.

Accessibility includes two parameters: walking time or distance (formal accessibility) and total time, including walking, waiting, and riding to the most closed focal point in the TN (real accessibility to a wide service range). Transit planning usually is oriented to simple formal accessibility, assuming that any line fits travel demand direction. This appears reasonable for a radial or grid TN in which directions are obvious. A complex TN presents a problem for accessibility definition, because the closest line in origin can be useless for the passenger depending on destination. In such cases, the time to get to the closest focal point is more informative. A focal point can be defined as a node where at least

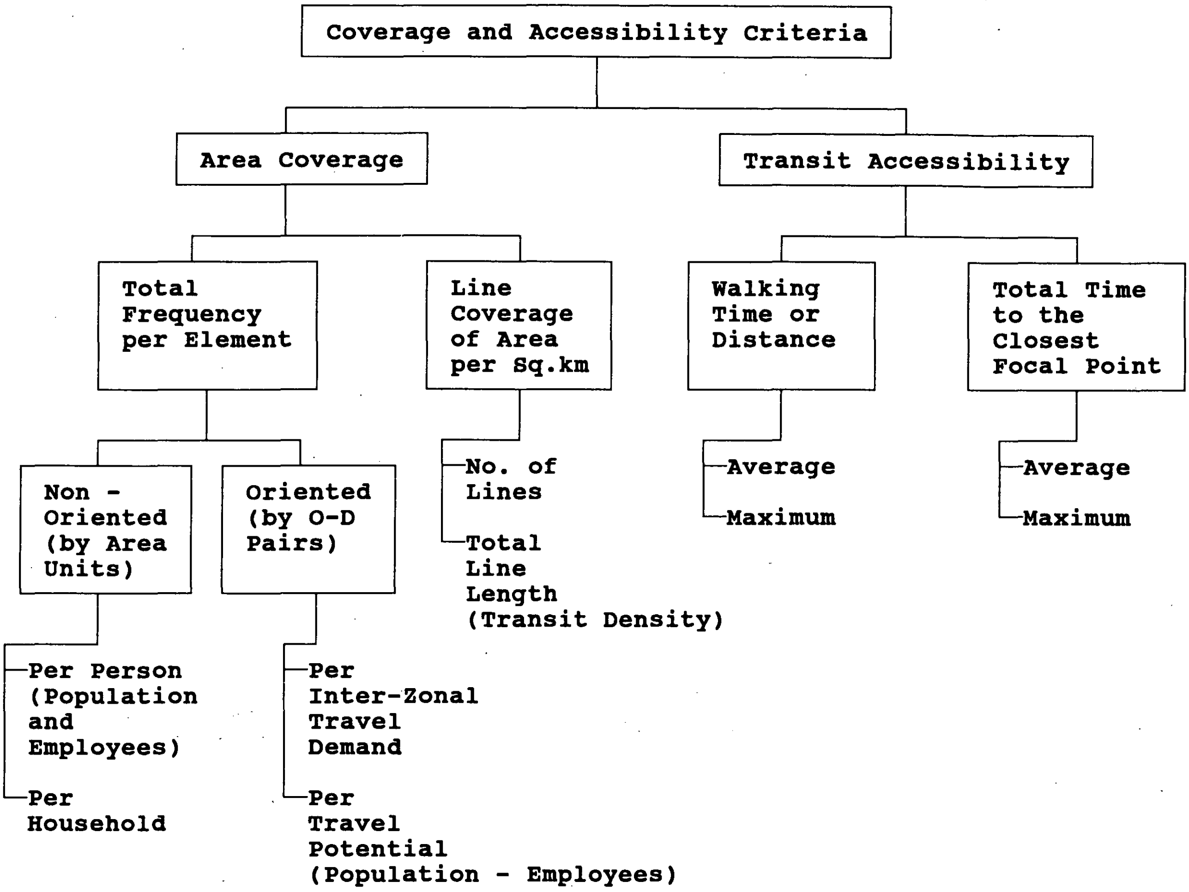


FIGURE 1 Classification of coverage and accessibility criteria.

three nonduplicative lines cross. Both parameters are estimated for each area unit. Later, two main summary principles can be involved: "average," which characterizes accessibility in a given region as a whole, and "maximum," which reveals the most problematic area units.

For total frequency per person, quantitative boundaries can be established in terms of low, standard, and surplus level of service (*25*). Improvement of coverage means reduction in the number of area units with low level of service. A growth in number of units with surplus service is not desirable.

Network integration can be viewed as different lines' consistency and allows recognition of the role of a given line in TN. It positively affects trip structure and level of service and balances between two contradictory objectives of TN design: direct connection and transfer convenience. Integration evaluation has two stages (Figure 2): (*a*) evaluation of given line characteristics and (*b*) evaluation of TN systemwide characteristics.

In the first stage, two subsets are involved: line crossing and line "exclusive addition to the TN integration." Crossing characterizes transfers to other lines. There are three pairs of parameters in the subset: (*a*) number of lines crossing the given line (*CL*), and the same for unduplicative lines only (*UCL*); (*b*) number of transfer points in the given line (*TP*), and the same for unduplicative lines only (*TPU*); and (*c*) minimum number of transfer points in the given line, which enables access to all crossing lines (*MTP*), and the same for unduplicative lines (*MTPU*) only.

Figure 3 presents a TN fragment. There are two compared lines (*A,B*) and five others (1–5). Number of crossing lines reveals an integration advantage for line *A*: *LC*(*A*) = 6; *LC*(*B*) = 2. Nevertheless, some of the lines crossing line *A* (1,5) duplicate it. A duplicative line usually presents no interest to transfer, because it has no (or few) additional destinations over the original line. To take this into account, the number of unduplicative crossing lines is of help. According to this parameter, the advantage of line *A* is not so appreciable: *UCL*(*A*) = 4 (without lines 1,5); *UCL*(*B*) = 2.

The second pair of parameters deals with transfer-point allocation. They are distinct from the number of crossing lines for two reasons. First, more than two lines can cross at one point (*b* for lines *A*,2,3,4). Second, two lines can cross each other several times. Usually, this occurs in duplicative lines, but not necessarily. Duplicative lines *A* and 1 have a list of transfer points *g*, *f*, . . ; unduplicative lines *A* and 2 also have transfer points *b* and *c*. The number of transfer points shows an advantage to line *A*: *TP*(*A*) = 12, *TP*(*B*) = 2. Most transfer points of line *A* are related to the duplicative lines *1* and *5*. Without them, the advantage of line *A* is less appreciable: *TPU*(*A*) = 3, *TPU*(*B*) = 2.

The third pair of parameters reflects transfer-point distinction. Two transfer points in the line are distinct if they have different crossing lines, as opposed to having the same lines. Practically, each line has a limited number of focal points that allow access to all crossing lines. This parameter is significant to line integra-

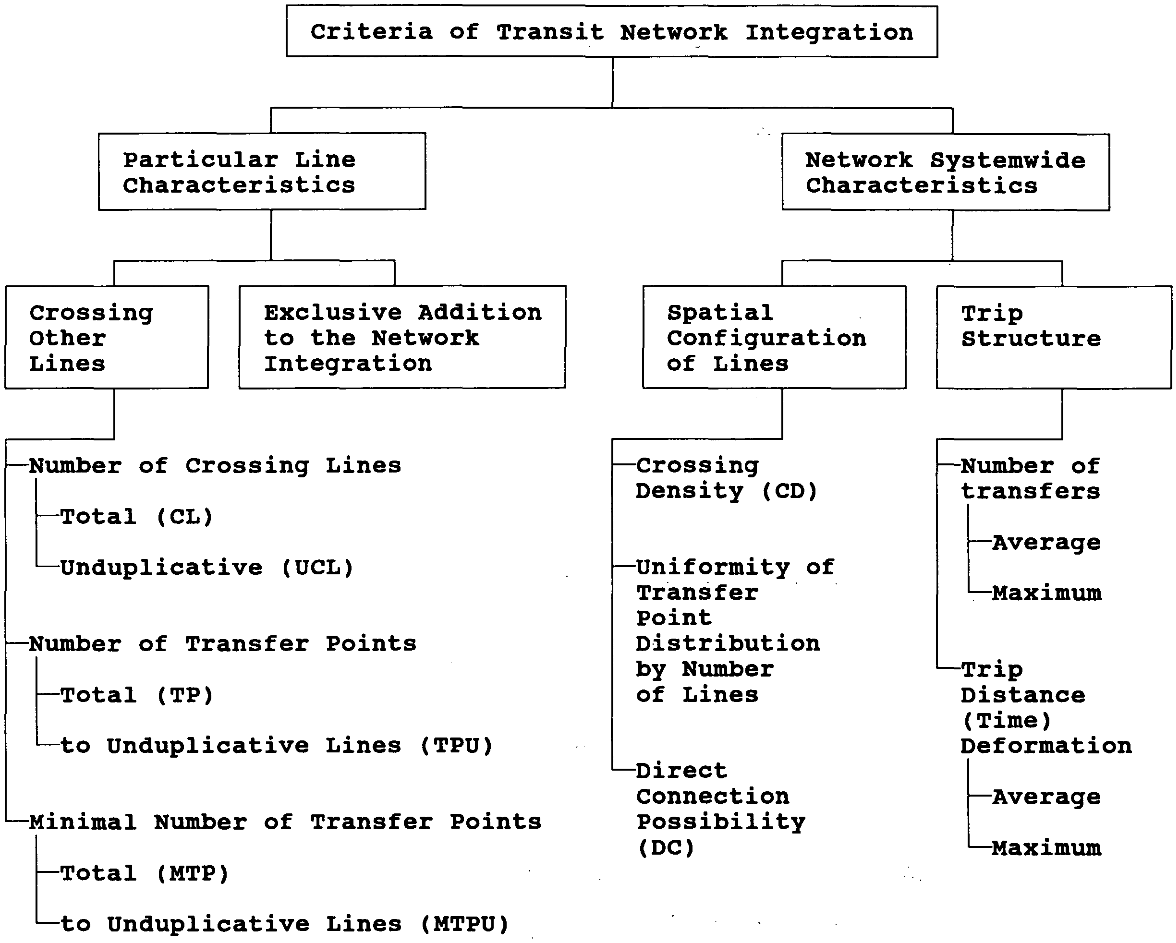


FIGURE 2 Classification of network integration criteria.

tion. It is similar for both given lines: $MTP(A) = 3$ (a,b,i), $MTP(B) = 2$ (a,m). Without duplication, equality of lines A,B is revealed: $MTPU(A) = 2$ (a,b), $MTPU(B) = 2$ (a,m).

In sum, two parameters (*UCL* and *MTPU*) can be viewed as most important in the subset. Simply by crossing parameters (*CL*, *TP*), line *A* has a significant advantage. Nevertheless, most of the advantages of line *A* should not be treated as real because they derive from the duplicative lines. More accurate analysis shows that line *B* is comparatively successful.

Exclusive addition to the TN integration, *ANI*, represents a given line's contribution to transfer-point variety. First, interim calculations must be made, including the total number of TN transfer points (*TPN*) and focal points (*TPNF*), and the same without the given line (TPN^0,$TPNF^0$). Then line-exclusive addition to the number of transfer points (ΔTP) and focal points (ΔTPF) are calculated:

$$\Delta TP = TPN - TPN^0; \quad \Delta TPF = TPNF - TPNF^0 \qquad (3)$$

After that the final calculation should be made.

$$ANI = \frac{(\Delta TP + \Delta TPF)}{TPN} \qquad (4)$$

A line is considered significant for TN integration if it generates original transfer points or new focal points. If a line provides neither additional transfer points nor focal transfer points, it means that the line crosses others only in existing focal points. Normally, line-exclusive addition to the TN integration should be comparable with the total number of lines. For a TN including 100 lines, an average level of *ANI* can be defined as 1 percent. A line that has more than 1 percent addition can be viewed as especially important.

At the second stage, two subsets are involved: spatial configuration and trip structure indicators. A TN configuration is related to the spatial structure of the region. There are three criteria in the subset: crossing density (*CD*), uniformity of transfer point distribution, and direct connection (*DC*).

Crossing density indicates TN connectivity. It is calculated as a number of TN transfer points divided by a number of regional area units (*AU*).

$$CD = \frac{TPN}{AU} \qquad (5)$$

Figure 4 presents typical TN patterns. Crossing density (*CD*) reveals an advantage of grid and shortcoming of radial and polycentric form, which allows limited possibility of transfer. How-

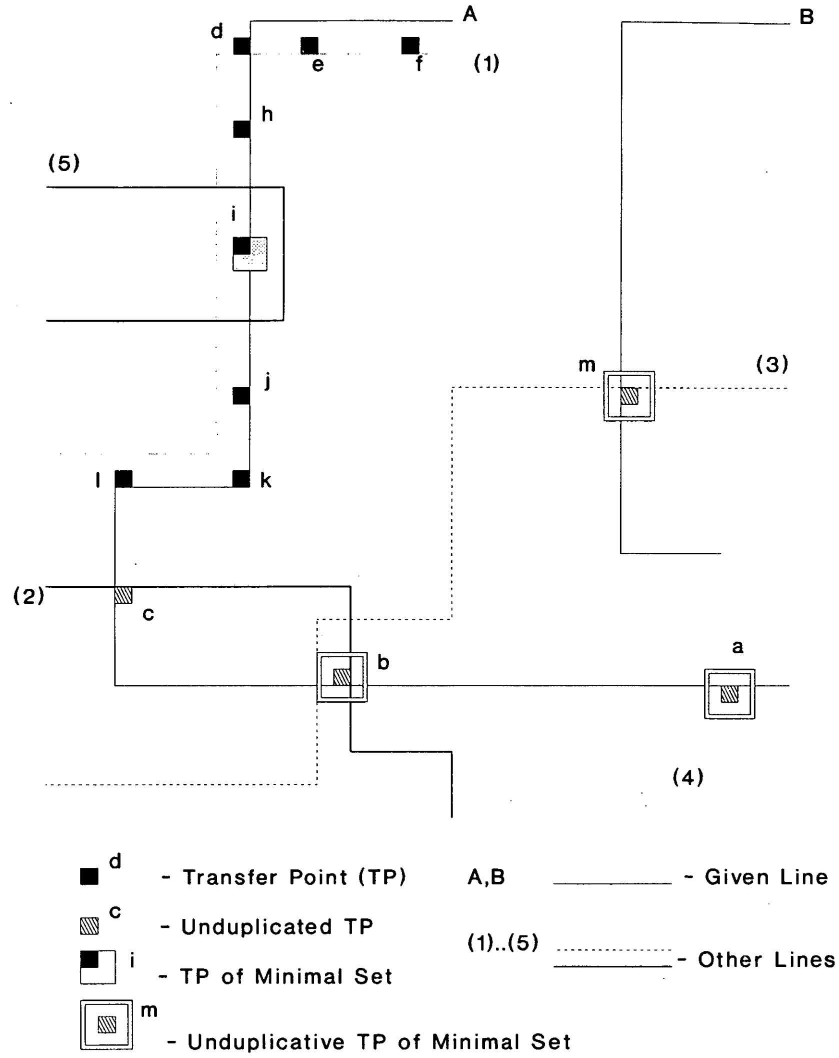


FIGURE 3 Line crossing in transit network.

ever, they are recognized as the most economical. A directionally oriented form is attractive in a case of oriented demand, but at the same time it is the most expensive. The TN in Tel Aviv is mixed, but all forms can be revealed. The Eged subnetwork is mostly radial, centered at the Tel Aviv central bus station. The Dan subnetwork in the northern part of Tel Aviv has a grid form, but in the southern part it is close to a directionally oriented form. Crossing density is of help for TN connectivity estimation as a whole. Approximate values of *CD* are 40 percent and more for tied TN (grid), 20 to 40 percent for mixed, and less than 20 percent for radial.

Because crossing density does not differentiate between ordinary and focal transfer points, radial TN (with wide center including a few area units) and a directionally oriented TN may have the same CD value. To distinguish between them, uniformity of transfer point distribution is of help (Figure 4). A Grid TN has the most uniform distribution (no focal points), and a radial TN has one dominant central point. Uniformity can be seen as a positive factor of TN integration.

Direct connection (*DC*) indicates the possibility to travel to a destination without transfer. It is calculated as the number of area unit pairs directly connected by at least one transit line (*DSP*) divided by the total number of pairs. The latter can be calculated as $AU \cdot (AU - 1)/2$, thus

$$DC = \frac{2 \cdot DSP}{AU \cdot (AU - 1)} \tag{6}$$

Crossing density and direct connection are contradictory—the first suggests transfers, but the second avoids them. That reflects the contradictory objectives of TN design: to improve level of service and reduce costs. Direct connection reveals an advantage of a directionally oriented TN and a shortcoming of a radial or polycentric TN. Normally, the cost criterion yields the opposite result. The directionally oriented form is effective from different points of view when demand is oriented. For example, in southern Tel Aviv, 80 percent of trips are south-north oriented as a result of the spatial structure of population and employment.

Trip structure summarizes the effectiveness of TN design. Two criteria should be noted: the number of transfers and trip distance or time deformation. Deformation is calculated as trip distance (time) on TN divided by trip distance (time) by shortest path

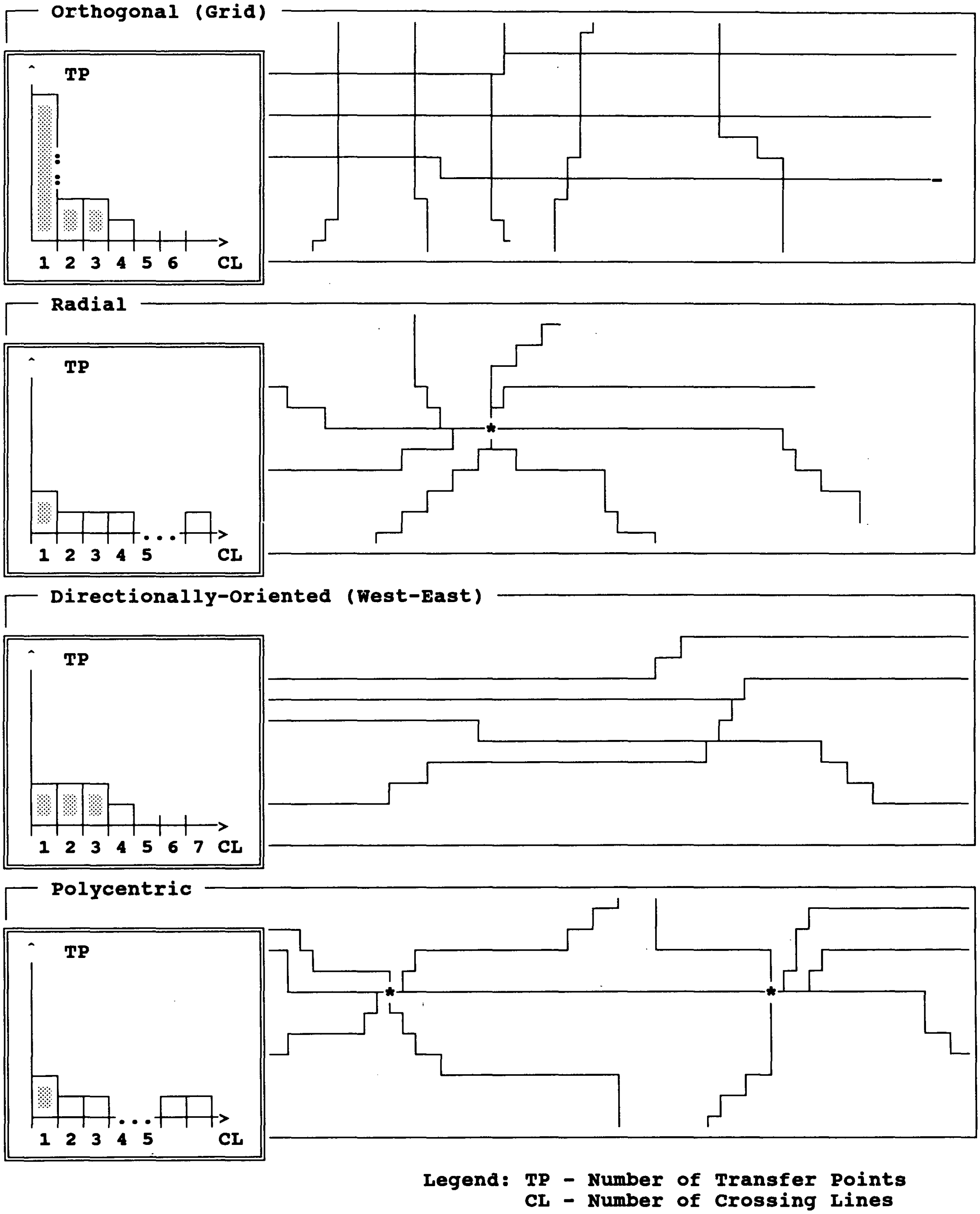


FIGURE 4 Typical transit networks and transfer point distribution.

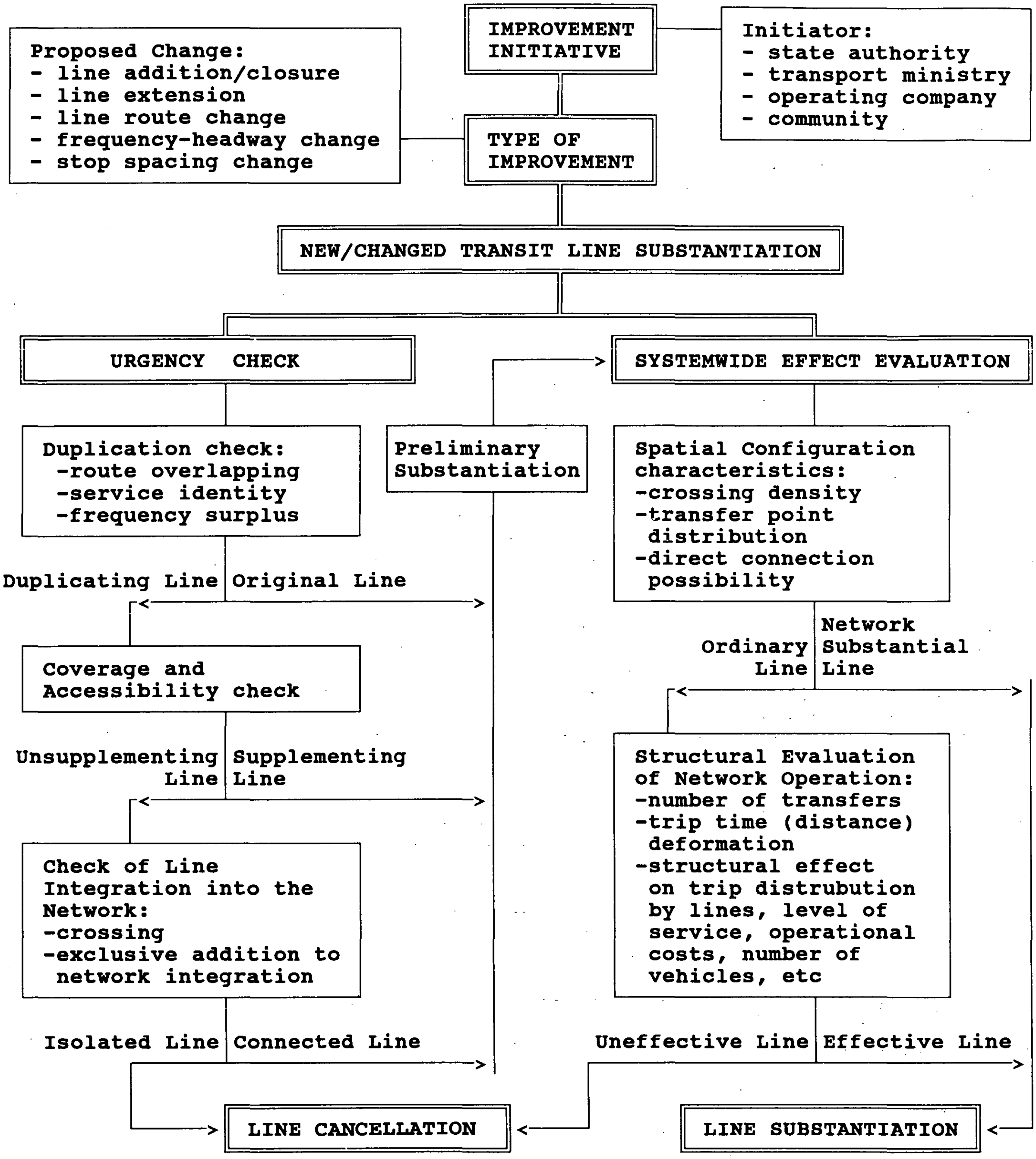


FIGURE 5 Decision-making framework for transit network improvement.

(minimum time) on the road network. These two criteria inevitably contradict under real constraints. Commonly, TN needs a sufficient level of transfers to supply directness of trips. Direct connection for most origin-destination pairs is too expensive. Any proposal that improves one of these criteria without worsening the other is of interest.

Both criteria should be checked as an average TN value and maximum value for particular origin-destination pair. Average parameters are useful to transit planning; however, a decision that is good systemwide may prove extremely inconvenient for particular passengers. This problem is usual for regions combining urban settlements with rural areas. Direct connection for a pair of rural zones will not be effective systemwide, and feeder lines to appropriate urban settlements yield at least two transfers and trip distance (time) deformation for these passengers.

A TN for a metropolitan area can be recognized as acceptable if the averge number of transfers is no more than 40 percent (20 percent shows an excellent pattern) and less than 5 percent of passengers experience two or more transfers. Both distance and time deformation are of interest. Time can be thought of as an ultimate service indicator, but it is affected by road congestion. Thus, trip distance deformation is a more direct indicator of TN design.

TABLE 3 Summary Report for Transit Line Substantiation (Extension of Line 8)

Evaluation Stage/Step	Scope	Criterion	Value			Conclusion (ª - before,º - after)
			Before	After	%	
1.Substantiation	28 lines "Eged" 20 lines "Dan"					Unsubstantial lineª Additional Substantiationº
1.1.Duplication over 40%	- || -	No. of duplicative lines	4	4	0%	Duplicative lineª Reduced Duplication with some lines but increased Duplication with some othersº
	Line 18 "Dan"	Route overlapping	59%	42%	-29%	
		Service identity	100%	100%	0%	
		Frequency surplus	100%	100%	0%	
	Line 42 "Dan"	Route overlapping	51%	-	-100%	
		Service identity	100%	-	-100%	
		Frequency surplus	0%	-	0%	
	Line 85A "Eged"	Route overlapping	49%	55%	12%	
		Service identity	100%	100%	100%	
		Frequency surplus	0%	0%	0%	
	Line 85 "Eged"	Route overlapping	48%	53%	10%	
		Service identity	100%	100%	0%	
		Frequency surplus	0%	0%	0%	
	Line 83A "Eged"	Route overlapping	-	43%	+100%	
		Service identity	-	100%	+100%	
		Frequency surplus	-	0%	0%	
1.2.Coverage	28 lines "Eged" 20 lines "Dan" 59 traffic zones	No. of served zones uncluding:	59	59		Non-Supplementing lineª No additional coverageº
		- surplus level of service	4	4	0%	
		- standard level of service	45	45	0%	
		- low level of service	10	10	0%	
1.3.Integration	28 lines "Eged" 20 lines "Dan"					
- Line	- || -	No. of crossing lines	37	46	+24%	Connected lineª Additional connectionº
		No. of undupl. cross. lines	33	42	+27%	
		No. of transfer points	43	54	+26%	
		No. of tr.p. to undupl. lines	43	54	+26%	
		Min. no. of trans. points	5	6	+20%	
		Min. no. of tr.p. to und. l.	5	6	+20%	
		Addition to Integration	0.5%	0.5%	0%	
- Network Spatial Configuration	- || -	Crossing density	51%	53%	+4%	Network substantial lineª Additional substantialityº
		Transfer point distribution				
		- 2 crossing lines	32	35	+9%	
		- 3 crossing lines	24	26	+8%	
		- 4 crossing lines	14	13	-7%	
		- 5 and more crossing lines	16	17	+6%	
		Direct connection possibility	41%	46%	+12%	

(Continued on next page)

TABLE 3 *Continued*

Evaluation Stage/Step	Scope	Criterion	Value			Conclusion (ª - before, º - after)
			Before	After	%	
- Trip Structure	- || -	Average no. of transfers	1.32	1.30	-2%	Uneffective lineª
		Maximum no. of transfers	2	2	0%	Additional effectº
		Average distance deformation	1.24	1.23	-1%	
		Maximum distance deformation	2.68	2.66	-1%	
		Average time deformation	1.18	1.17	-1%	
		Maximum time deformation	2.60	2.55	-2%	
		Trips distrubution by lines:				
	Line 8	No. of boardings	263	711	+170%	
		Average occupancy ratio	0.12	0.24	+100%	
		Maximum occupancy ratio	0.26	0.53	+104%	
	Line 83A	No. of boardings	633	398	-37%	
		Average occupancy ratio	0.35	0.20	-43%	
		Maximum occupancy ratio	0.86	0.45	-48%	
2. Allocation to Companies						"Dan" lineª Line open to competitionº
2.1. Integration within Subnetworks	28 lines "Eged" (+line 8)	No. of crossing lines	14	21	+50%	Line more connected
		No. of undupl. cross. lines	12	19	+58%	with subnetwork "Dan"ª
		No. of transfer points	38	48	+26%	Line equally connected with
		No. of tr.p. to undupl. lines	38	48	+26%	subnetworks "Eged" and "Dan"º
		Min. no. of trans. points	3	4	+33%	
		Min. no. of tr.p. to und. l.	3	4	+33%	
		Addition to Integration	5.4%	8.1%	+50%	
	20 lines "Dan"	No. of crossing lines	19	19	0%	
		No. of undupl. cross. lines	16	16	0%	
		No. of transfer points	43	53	+23%	
		No. of tr.p. to undupl. lines	43	49	+14%	
		Min. no. of trans. points	3	3	0%	
		Min. no. of tr.p. to und. l.	3	3	0%	
		Addition to Integration	2.7%	9.2%	+241%	
2.2. Subnetworks Structure	28 lines "Eged" (+ line 8)	Crossing density	42%	43%	+2%	Ordinary line for
		Transfer point distribution:				"Dan" subnetworkª
		- 2 crossing lines	40	43	+8%	Equally substantial line for
		- 3 crossing lines	19	16	-16%	"Eged" and "Dan" subnetworksº
		- 4 crossing lines	2	5	+150%	
		- 5 and more crossing lines	2	2	0%	
		Direct connection	29%	35%	+21%	
	20 lines "Dan"	Crossing density	25%	30%	+20%	
		Transfer point distribution:				
		- 2 crossing lines	13	18	+38%	
		- 3 crossing lines	5	7	+40%	
		- 4 crossing lines	6	6	0%	
		- 5 and more crossing lines	5	5	0%	
		Direct connection	21%	24%	+14%	

DECISION MAKING FRAMEWORK FOR TRANSIT NETWORK DEVELOPMENT

Figure 5 presents a decision-making framework for TN improvement in a real planning environment. The starting point is an initiative for change from one party. Potential changes may range from line opening or closure to change of line characteristics. Any change can be expressed ultimately as inserting new lines or (if necessary) discontinuing service on existing lines. There are two stages of substantiation: (*a*) urgency check and (*b*) systemwide effect evaluation.

Urgency check enables line estimation without the cumbersome procedure of trip structure analysis. It will become increasingly relevant in a longer term when travel demand is unknown and can only be approximated. Urgency check involves three sequential steps: duplication, coverage and accessibility, and line integration into TN. As a result of first step, the line can be defined either as original (its route has no mutual segments with existing lines, its type of service is different from service provided by other lines, its frequency gives reasonable addition to a total frequency of existing lines) or as duplicating. After the second step, the line can be recognized either as supplementing (if it serves new area previously unserved or improves transit accessibility in a particular area) or as unsupplementing. At the third step, the line can be qualified either as connected (if it has a number of TN-important transfer points) or as isolated.

The line is considered preliminarily substantiated in the case of a positive result from at least one of the checks. A preliminary substantiated line is subjected to systemwide effect evaluation. There are two steps here: (*a*) spatial configuration, and (*b*) structural evaluation of TN operation. As a result of the first step, the given line gets a status either as a network substantial line (significant for TN operation for a strong relationship with other lines) or as an ordinary line (limited by its own operational sphere). The second step is the final one in the decision-making process. A transit assignment with a defined demand matrix should be used. At this stage such TN performance indicators as trip redistribution by lines, level of service, and operational cost are available. As a result, the line can be defined either as effective (it attracts sufficient ridership, its introduction leads to a level of service improvement within an acceptable cost range) or as ineffective.

A line is substantiated in the case of a positive result at one of the stages: either it is substantial for TN or effective in itself. A line is otherwise canceled. For practical evaluation, a programming package has been developed by the Israel Institute of Transportation Planning and Research (*25*). Transit assignment is performed by the EMME/2 package, which is compatible with developed programs.

SUMMARY OF RESULTS

Final results for the evaluation indicate that the proposed modification will improve TN integration (Table 3). At present, line 8 is marked as over 40 percent duplication of four different lines. In addition, line 8 is not supplementary to transit coverage of the area under study. Therefore, the change would be beneficial to passengers and operator. Extension of a line to zones that are served by other lines results in a 24 percent increase in the number of crossing lines (from 37 to 46). Crossing density, however, does not increase much, showing only a 9 percent increase in the number of crossing lines with two or more lines. Direct connection possibility has increased by 12 percent as more destinations can be reached along the extended route of line 8. All these results indicate the value of the change to greater passenger mobility and accessibility.

By using a transit assignment, an increase of 170 percent in the number of boardings was observed, and the average occupancy ratio doubled. The findings demonstrate the attractiveness of the line to passengers. In a closed system with the fixed demand matrix that the authors use, any increased passenger boardings in one line will worsen other lines' levels of occupancy. It is beyond the scope of this research to present the effect of service quality on demand. Nonetheless, lines suffering as a result of service improvements in another line will usually have some level of duplication. In this case, most line boardings decreased slightly, with the exception of line 83A, which suffered a 37 percent decline in ridership. Although, it is not the line with the highest level of duplication, it suffered most. One of the reasons is that, whereas other lines had a relatively high level of duplication with line 8 before the change, line 83A had no duplication at all. As a result of the modification, lines 8 and 83A were competing on passenger ridership in the same areas, demand was split between them, and thus 83A was affected significantly.

The final analysis phase considers line allocation to companies. The basic rationale for picking one transit company over the other is a better line integration in existing companies' subnetworks and enhancement of their structure. In such case, passengers may benefit from improved service and the operator will benefit from improved operating costs. The area under study was specifically picked to test the model behavior in a case where both Dan and Eged are pushing for additional network expansion and neither is dominating. The result showed that although at present line 8 is run by Dan for justifiable reasons, after extension, TN can be allocated equally to both companies. This was the major reason Dan was pushing the Ministry to accept the proposed change and to give them a foot in the door to better market share. This is an example of how a policy-oriented approach can favor one operator.

REFERENCES

1. Nes, R. V., R. Hamerslag, and B. H. Immers. Design of Public Transport Networks. In *Transportation Research Record 1202*, TRB, National Research Council, Washington, D.C., 1988, pp. 74–83.
2. Baaj, M. H., and H. S. Mahmassani. An AI-Based Approach for Transit Route System Planning and Design. *Journal of Advanced Transportation*, Vol. 25, No. 2, 1991, pp. 187–210.
3. Axhausen, K. W., and R. L. Smith. Evaluation of Heuristic Transit Network Optimization Algorithms. In *Transportation Research Record 976*, TRB, National Research Council, Washington, D.C., 1984, pp. 7–20.
4. Newell, G. Some Issues Related to the Optimal Design of Bus Routes. *Transportation Science*, Vol. 13, 1979, pp. 20–35.
5. Jun, M., and J. B. Schneider. Designing Personal Rapid Transit (PRT) Networks. *Journal of Advanced Transportation*. Vol. 25, No. 3, 1991, pp. 247–268.
6. Lampkin, W., and P. D. Saalmans. The Design of Routes, Service Frequencies and Schedules for a Municipal Bus Undertaking: A Case Study. *Operations Research Quarterly*, Vol. 18, No. 4, 1967, pp. 375–397.
7. Silman, L. A., Z. Barzily, and U. Passy. Planning the Route System for Urban Buses. *Computers and Operations Research*, Vol. 1, 1974, pp. 201–211.
8. Hoidn, H. P. Busnetz von Aarrau. *Neukonzeption der Linien.* Institut fuer Operations Research der ETH Zuerich, Zuerich, Switzerland, 1975.

9. Rosello, X. An Heuristic Algorithm to Generate an Urban Bus Network. *Advances in Operations Research: Proceedings*. Elsevier, Amsterdam, the Netherlands, 1976.
10. Sonntag, H. Ein Heuristisches Verfahren zum Entwurf Nachfrageorientierter Linienfuerhrung im Oeffentlichen Personennahverkehr. *Zeitschrift fuer Operations Research*, Vol. 23, 1979, pp. B15–B23.
11. Dubois, D., G. Bell, and M. Llibre. A Set of Methods in Transportation Network Synthesis and Analysis. *Journal of the Operations Research Society*, Vol. 30, No. 9, 1979, pp. 797–808.
12. Hasselstroem, D. *A Method for Optimization of Urban Bus Route Networks*. Report 88-DH-03. Volvo Bus Corporation, Gothenburg, Sweden, 1979.
13. Mandl, C. *User Reference and Implementer Manual IANO-System*. Institutsarbeit 122. Institute for Advanced Studies, Vienna, Austria, 1979.
14. *Sommer Semester 1981: Linienplanung*. Seminar Verkehrswesen, Institutsnotiz.
15. Furth, P. G., and N. H. M. Wilson. Setting Frequencies on Bus Routes: Theory and Practice. In *Transportation Research Record 818*, TRB, National Research Council, Washington, D.C., 1981, pp. 1–7.
16. Hagberg, B., and D. Hasselstrom. *A Method to Assign Frequencies and Vehicle Types to Fixed Route Urban Public Transport Systems*. Volvo Transportation Systems, Gothenburg, Sweden, 1980.
17. Scheele, S. O. A Supply Model for Public Transit Services. *Transportation Research*, Vol. 14B, No. 3, 1980, pp. 133–146.
18. *NCHRP Synthesis of Highway Practice 69: Bus Route and Schedule Planning Guidelines*. TRB, National Research Council, Washington, D.C., 1980.
19. Schneider, J. B., and S. P. Smith. Redesigning Urban Transit Systems: A Transit-Center-Based Approach. In *Transportation Research Record 798*, TRB, National Research Council, Washington, D.C., 1981, pp. 56–65.
20. Thelen, K. M., A. Chatterjee, and F. J. Wegmann. Evaluation of Alternative Transit Routing Configurations in a Hypothetical Low-Density Area. In *Transportation Research Record 761*, TRB, National Research Council, Washington, D.C., 1980, pp. 53–56.
21. Giannopoulos, G. A. *Bus Planning and Operation in Urban Areas: A Practical Guide*. Avebury, 1989.
22. Rapp, M., and C. D. Gehner. Transfer Optimization in an Interactive Graphics System for Transit Planning. *Transportation Research Record 619*, TRB, National Research Council, Washington, D.C., 1976, pp. 27–33.
23. Douglas, G. B. Applications of Interactive Graphics in Urban Transit Analysis. In *Transportation Research Record 866*, TRB, National Research Council, Washington, D.C., 1982, pp. 18–24.
24. Wilson, A. G. The Use of Entropy Maximizing Models in the Theory of Trip Distribution, Modal Split and Route Split. *Journal of Transport Economics and Policy*, Vol. 2, No. 1, 1969, pp. 108–126.
25. Cohen, A., P. Vovsha, and T. Goodovitch. *Decision Support System for Planning and Development of Public Transport: An Application to Tel Aviv Metropolitan Area* (in Hebrew). The Israel Institute of Transportation Planning and Research, Tel Aviv, Israel, 1993.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Application of Multiattribute Utility Theory to Public Transit System Design

THOMAS B. REED, CHELSEA C. WHITE III, MICHAEL P. BOLTON, AND WILLIAM D. HILLER

Advanced Public Transportation Systems (APTSs) represent a promising concept that applies existing and emerging technologies and techniques to enhance and expand transit services. However the rapid emergence of a large variety of technologies and techniques might increase the possibility that new designs for transit systems will focus heavily on available products, neglect functional issues, and thus lead to partial and fragmented systems. Transit authorities need a mechanism they can use to avoid this negative situation and to develop transit systems that are as comprehensive, effective, and efficient as possible. Multiattribute Utility Theory (MAUT) is a normative model of decision making. Transit system design is a context in which the application of MAUT may be useful. Specifically MAUT provides tools for systematically evaluating, priority ranking, and integrating desired transit functionalities and APTS capabilities. The application of a MAUT-based framework for the process of transit system design is described and illustrated. It is hoped that the framework will both aid transit authorities in systems-design efforts and stimulate discussions that might influence the development of a nationwide specification for APTSs and lead to a standard, open transit system design.

Public transit faces a simultaneous decline in customer base and a rise in costs (*1–3*). As a result transit authorities are more than ever faced with the unenviable task of developing means of increasing the share of the market held by transit and at the same time reducing costs and increasing revenues. Along with these systems-design objectives transit authorities must respond to numerous legislative mandates, for example, the Clean Air Act Amendments of 1990, the Americans with Disabilities Act, and acts related to energy use and must improve transit safety and security (*4*).

The recent growth of Advanced Public Transportation Systems (APTSs), a promising concept that applies existing and emerging technologies and techniques to enhance and expand transit services, would appear to provide a means of attaining the aforementioned systems-design objectives. However the rapid emergence of a large variety of technologies and techniques might increase the possibility that new designs for transit systems will focus heavily on available products, neglect functional issues, and thus lead to partial and fragmented systems.

Furthermore integration of available resources into a single unified system is difficult because of the involved and interrelated nature of the subsystems and the social and political issues that the designer must accommodate. Moreover forecasts of effects of interventions are speculative by nature, and there is no all-encompassing evaluative framework by which to judge the efficacy, justice, and cost of any action. The complexity inherent in this situation often overwhelms designers and promotes retreat to simplistic solutions.

Therefore there is a pressing need to approach the design of transit systems in a more organized manner. Designers need a clarifying mechanism that increases the quality of the design process through avoiding potential piecemeal and haphazard designs while developing comprehensive and coherent transit systems that are effective and efficient at meeting transit objectives.

METHODOLOGY

The cognitive psychology literature shows that unaided decision makers perform poorly in comparison with decision makers who use normative models (*5*). This is true for the quantitative fit between actual decision-making behavior and optimal behavior, that is, rational behavior as prescribed by a normative model, and for the qualitative character of decision makers' actions, evidenced by persistent violation of fundamental and self-evident axioms of rational behavior.

Studies have shown that decision makers are potentially amenable to support. In many problem-solving situations an aided decision maker can outperform an unaided decision maker. For example a study comparing the diagnostic capabilities of aided engineering graduate students against the baseline capabilities of unaided general practitioners found that, in diagnosing common ambulatory care complaints, the students were able to reduce diagnostic costs by 32 percent and the number of major diagnostic errors by a factor of 4 (*6*).

As implied earlier the difficulty of decision making in situations analogous to a systems-design process is intuitively largely because of the number, complexity, and interrelated nature of the issues faced. Research supports this intuition by demonstrating that people can maintain only a finite and relatively small number of issues in mind at any one time. Specifically studies have shown that people can adequately identify only 7 ± 2 levels of any unidimensional stimulus such as sound or sight (*7*). People are more adept at distinguishing multidimensional stimuli and, for example, can recognize hundreds or more categories when judging between faces or phonemes of human speech. Unfortunately the level of experience that leads to this capability appears to be lacking in most other decision-making environments. The implication of this human inadequacy is that issues need to be "chunked" or put into higher-level terms if far-reaching or intricate processes are to be handled with any degree of sophistication.

T. B. Reed, Urban, Technological, and Environmental Planning, University of Michigan, 4117 EECS, Ann Arbor, Mich. 48109-2122. C. C. White III, Department of Industrial and Operations Engineering, University of Michigan, 274 IOE, Ann Arbor, Mich. 48109-2117. M. P. Bolton and W. D. Hiller, Ann Arbor Transportation Authority, 2700 South Industrial Highway, Ann Arbor, Mich. 48104.

Multiattribute Utility Theory (MAUT) is a normative model of decision making (*8,9*) that designers can use to code the enormous amount of disparate information necessary for systems design into manageable chunks. In other words problem-solving methodologies based on MAUT can provide decision-making support for the task of selecting a single alternative or a set of alternatives in a risky or uncertain environment involving multiple, noncommensurate, and conflicting objectives. Such methods are also useful for inverse decision aiding or policy capture, which can provide insight into the reasons why a decision maker might prefer an alternative. Simply stated, MAUT-based methods provide a framework for decomposing a highly complex problem-solving or design task into a collection of simpler issues that, when resolved and recomposed, leads to a solution of the original problem.

As a result MAUT-based methods are useful for improving the capability of a decision maker or designer through aiding in the identification of contradictory facts and preferences and through enhancing the quality of the alternatives selected. These methods are also useful in improving the acceptability of decisions. This is possible because the basis of decisions is material provided by the user, not the method or a consultant. This user perspective also facilitates data acquisition because the information needed is often readily accessible to the designer, who sometimes is the knowledge source. Furthermore a common and major impediment to project evaluation is lack of clear objectives. MAUT-based methods avoid this by requiring a priori explication of design objectives. Such explication also provides a clear basis for arguing the merits of a project before and justifying project expenditures in the eyes of the relevant agencies and the general public.

Transit system design is a context in which the application of MAUT may be useful. Specifically MAUT provides tools for systematically evaluating, priority ranking, and integrating desired transit functionalities and APTS capabilities. MAUT can thus facilitate a comprehensive approach to the process of transit system design, in sharp contrast to a design approach based solely on implementation of technologies and techniques individually available in the marketplace. The work reported here framed a methodology as the 10-step process illustrated in Figure 1. Note that the designer can iterate the process if needed. Figure 1 shows only the most significant iterative loop; however, the process permits a return to any previous step at any point.

Step 1 in Figure 1 requires the determination of a hierarchy of objectives such as that shown in Figure 2. The objectives, which represent considerations that the designer thinks necessary and sufficient for a satisfactory system design, are instrumental in guiding and evaluating the design process. In developing the objectives the designer must remember that a good transit design will appeal to a number of potential target audiences, including transit personnel (general managers, operations managers, maintenance managers, motor coach operators, and maintenance and office personnel), suppliers, governmental authorities, and customers and taxpayers.

As illustrated in Figure 2, a given objective in a hierarchy, O_j, is often associated with several levels of subordinate or superordinate objectives. In many cases different objectives are associated with different numbers of levels as well. The design process can focus on any level of objectives or can even include objectives from a combination of levels, subject to the prerequisite that the analysis must include one and only one objective between the highest and lowest levels of each branch of the hierarchy, as illustrated by the shading in Figure 2. Use of the lowest-level objective for each branch would provide the greatest specificity; however, in actual practice, uncertainty and resource constraints often force the analysis to higher levels in the hierarchy. This is not a great hindrance.

The results of the objectives-determination process are inherently subjective. However the process outlined in Figure 1 enjoins the designer to explicate intentions, seek out necessary data, and then make decisions in a rational manner.

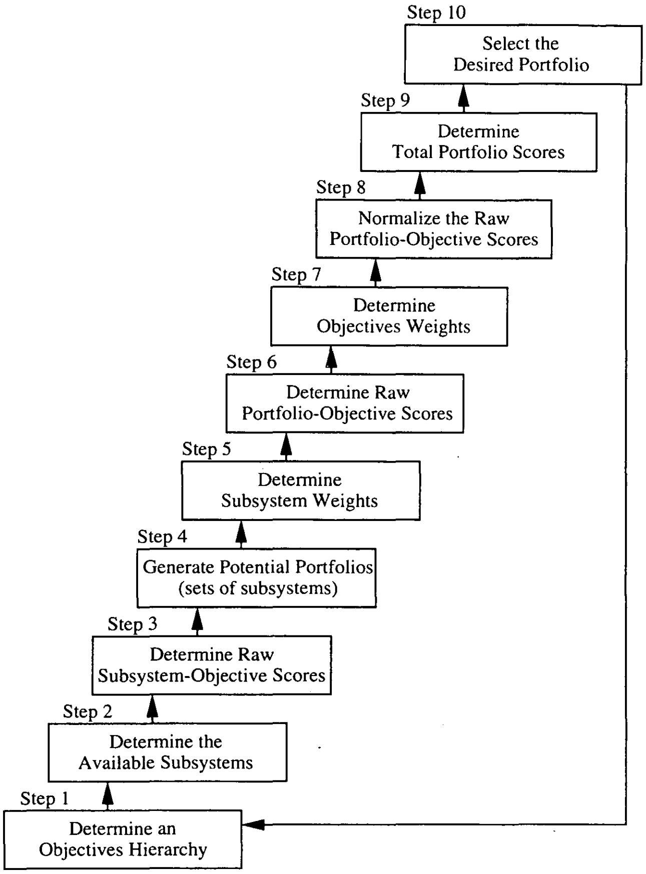


FIGURE 1 Transit system design methodology.

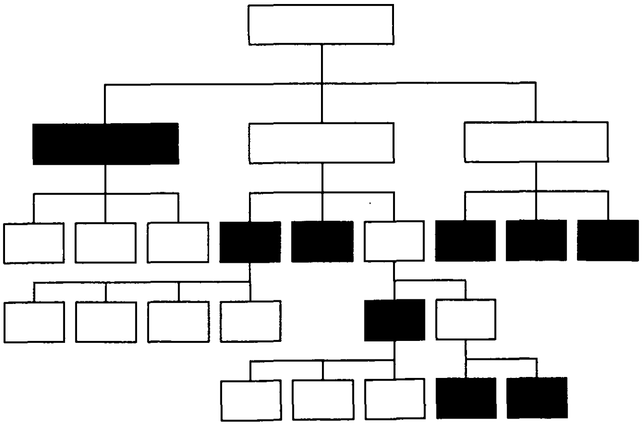


FIGURE 2 Format of objectives hierarchy.

After the designer has determined the objectives hierarchy, Step 2 requires determination of the available subsystems S_k. A subsystem can be either a technology (such as Automatic Vehicle Location) or technique (such as traffic regulation) that might enable attainment of the objectives.

At this point, Step 3, the designer must evaluate the ability of each subsystem to meet each of the objectives through determination of a raw (nonnormalized) score, called S_kO_j, for each subsystem-objective pairing, that is, for each (S_k, O_j) pair. The process of determining these raw subsystem-objective scores is equivalent to putting an S_kO_j score into each of the (S_k, O_j) cells of a matrix composed of J objectives in the columns and K subsystems in the rows. The designer can express the raw subsystem-objective scores in absolute terms. However in many designs it may be more helpful to express scores as changes from current conditions. Explicating each subsystem in this way will often highlight key subsystems and show the benefit of eliminating other subsystems because of the inability to meet design objectives or excessive cost.

Ideally each score or measure of effectiveness would be quantitative, such as fuel efficiency in kilometers per liter, service reliability in percentage of on-time operation, customer service in time to respond to telephone inquiries in seconds, or subsystem setup and maintenance costs in dollars per vehicle. Unfortunately many of the most important measures of effectiveness, such as customer perception of the quantitative measures and customer satisfaction, are qualitative and are thus somewhat more difficult to evaluate. However the designer may assign a utility to each level of a qualitative measure, utility being a method of subjectively valuing individual preference (*8,9*). In some cases the designer might wish to assign a utility to the quantitative measures as well. Other than possibly being preferentially inaccurate, the use of utility has no effect on the MAUT methodology. The designer will also have to use subjective judgment if the process includes evaluation of emerging technologies and concepts, because these offer little if any actual data and the designer will likely need to assign subjective values, even for quantitative measures.

After identifying and scoring the subsystems, then, as Step 4, the designer generates a number of portfolios, P_i. Each portfolio represents a different combination of the essential subsystems that constitute a system. The simplest method of generating portfolios is to arrange the K available subsystems in each of the possible 2^K include or exclude combinations to arrive at a total of I portfolios. Take, for example, a design problem in which two subsystems, S_1 and S_2, are available; that is, K is equal to 2. S_1 and S_2 can form four portfolios ($I = 4$), P_4 (both S_1 and S_2), P_3 (only S_1), P_2 (only S_2), and P_1 (neither S_1 nor S_2). Note the inclusion of P_1, the "do-nothing" option, that is, a portfolio with no (new) subsystems, which represents the status quo. Although P_1 could conceivably be a competitor for the best system design, the designer need not explicitly include it in the analysis.

This method quickly becomes impractical as the number of subsystems increases. A modest system with only 12 subsystems, for example, would have 2^{12} (4,096) possible portfolios. For a large or detailed study the number of portfolios could easily exceed this number by several orders of magnitude. Three means of handling this situation are available. The designer could make an early move to cull subsystems from consideration through review or revision of the results of Steps 1 to 3. Such action narrows the analysis. Alternatively the designer could concede to grouping the subsystems into supersubsystems, $S_{k'}$, and perform the analysis on these. Doing so limits the depth of the analysis in a manner analogous to performing the analysis at a higher level in the objectives hierarchy. Finally the designer could separately analyze the supersubsystems in detail and then treat the best subportfolios as inputs to a second analysis. This method does not reduce the size of the analysis as much as the former two methods do, but it does not constrain the level of detail as much either.

Often many of the possible combinations that make up portfolios are infeasible. This might be because, say, S_1 also requires the presence of S_3 if it is implemented in conjunction with S_2, and so portfolios with S_1 but not S_3 are infeasible. Similarly S_2 might not be workable in the presence of S_4. When designers know this in advance they can save some effort by excluding such combinations from the portfolio generation process. If not noted ahead of time, however, the designer can eliminate infeasible portfolios from consideration at this point.

During Step 5 the designer ascertains the subsystem weights, SW_{ijk}. The subsystem weights represent the relative importance of each subsystem to each portfolio-objective combination and need to be as objective as possible to avoid introducing ambiguity. Differences in weights might arise from different levels of importance of the subsystems, might be due to synergy or dissonance among the chosen subsystems, and so on. The weights must sum to 1 across subsystems for each portfolio-objective pairing, with the exception of the cost objective, for which the designer commonly simply adds the scores, that is, gives a weight of 1.0 to each subsystem. If economies of scale or economies owing to synergisms are possible, then the designer can reduce the cost weights accordingly.

During Step 6 the designer must determine raw (nonnormalized) scores, called P_iO_j, to evaluate the ability of the remaining feasible portfolios to meet each of the design objectives, that is, for each portfolio-objective or (P_i, O_j) pair. The process of determining these raw portfolio-objective scores is equivalent to putting a P_iO_j score into each of the (P_i, O_j) cells of a matrix composed of J objectives in the columns and I portfolios in the rows. The designer calculates each raw portfolio-objective score as a weighted sum of the raw subsystem-objective scores, S_kO_j, from Step 3. The appropriate weights are the subsystem weights, SW_{ijk}, from Step 5, with a factor termed PS_k, which reflects the composition of the portfolio under consideration; that is, PS_k is 1 if the portfolio has subsystem S_k and zero if not. In short the designer uses Equation 1 to determine P_iO_j. The designer should eliminate portfolios that fall short of, or exceed in the case of cost, design requirements.

$$P_iO_j = \Sigma(SW_{ijk})(S_kO_j)(PS_k) \qquad (1)$$

In Step 7 the designer must determine the weights of the objectives, OW_j. It would be best if the designer were able to assign absolute weights. However in many situations decisions about the relative importance of the objectives are difficult to make either because the designer has no basis for such decisions or because it is easier to attain consensus if the designer does not specify detailed weights. To circumvent the need for absolute weights the designer may determine the weights of the objectives on the basis of the difference in relative importance of the portfolios for the various objectives. If the designer views the difference in raw portfolio-objective scores between portfolios for a specific objective as negligible; that is, if the designer thinks all portfolios have essentially the same impact on the given objective, then the weight for that objective should be low. That is, because all portfolios are roughly equally fitted for meeting this objective, then with respect to this objective it does not matter which portfolio is selected and so the objective should not be considered in the decision-making process. If the portfolios have substantially different impacts on the objective under consideration, the designer should assign a high weight to the objective. The weights of the objectives, including cost, should sum to 1. If the designer uses this method, then the weights of the objectives are also known as *trade-off weights*, because the method involves "trading" the difference in portfolio-objective scores associated with one objective for the difference in scores associated with another.

In Step 8 the designer normalizes the raw portfolio-objective scores to make the comparison of totals meaningful. The normalization process results in a standardized distribution for each set of portfolio-objective scores. Each distribution ranges from a low score of zero to a high score of 1, 10, 100, and so on, if desired. If the designer uses expected value in decision making, calculation of the normalized portfolio-objective scores, called NP_iO_j, is by a linear function, such as Equation 2. Note that the normalization of cost scores requires subtraction of the NP_iQ_j calculated by using Equation 2 from the maximum normalized score, which is 10 in this case. This is because for cost less is better.

$$NP_iO_j = \frac{(10)[P_iO_j - \min(P_1O_j{:}P_lO_j)]}{[\max(P_1O_j{:}P_lO_j) - \min(P_1O_j{:}P_lO_j)]} \qquad (2)$$

Many large organizations judge decisions by expected value; that is, they are risk neutral toward any single project. This is possible because diversification enables the organization to spread the risk over a number of projects. In contrast the designer might be risk averse or risk seeking (*8,9*). For example the designer might be risk averse in system design if there is no fallback position; the designer might play it safe and keep sufficient funds in reserve to counteract any potential large-scale failure. On the other hand the designer might be risk seeking in system design if the goal is exploratory research and the budget is relatively unrestricted; the designer might take risks in seeking a potential breakthrough in system design. Regardless of the reason, if the designer is not risk neutral, some form of nonlinear function must replace Equation 2.

In Step 9 the designer determines a total score, called TP_i, for each portfolio. The designer calculates each TP_i as a weighted sum of the normalized portfolio-objective scores, NP_iO_j, from Step 8. The appropriate weights are the trade-off weights of the objectives, OW_j, from Step 7. In short the designer uses Equation 3 to determine TP_i.

$$TP_i = \Sigma(OW_j)(NP_iO_j) \qquad (3)$$

At Step 10 the designer selects a single portfolio for implementation. If no portfolio is clearly superior or no portfolio meets expectations, the designer can iterate the design process. Iteration could include obtaining more information about the subsystems, better elicitation of the designer's preferences, and perhaps better explication or even modification of the objectives.

A final word on the method is appropriate. If it is difficult to generate raw scores or weights, the designer can substitute natural language statements such as "approach *Q* is preferable to approach *R*." This type of substitution greatly complicates the analysis, however, because the designer must use inequalities in the analysis. Therefore the design process should handle only a very few items in this way.

ILLUSTRATION

Any method composed of 10 steps may understandably be difficult to grasp at first. This section illustrates the process to enable the reader to better understand the details. A fictional designer is followed through the process of planning an upgrade for a transit (bus) fleet of 100 vehicles. While reading through the illustration the reader may find it useful to occasionally refer to the preceding section.

Step 1, determination of an objectives hierarchy, is crucial to the ultimate usefulness of the design. After careful consideration the designer concluded that the hierarchy must incorporate three questions basic to transit system design. First, does the system as designed satisfy transit customers? That is, is the system "satisfying"? Second, does the transit authority have, or have access to through subcontract, the technical, systems integration, business, and management skills required to successfully implement and maintain the system as designed? That is, is the system "doable"? Third, does the transit authority have the financial resources to implement the system as designed? That is, is the system affordable?

Other objectives are also possible. For example a fourth question is important to transit system design. Does the system as designed adequately address policy issues such as mobility equity, energy conservation, and so on? For simplicity this question was left out of the illustration. Moreover political issues are important. However many designers might conclude, as in this example, that it is not politic to explicitly include politics as an objective. After further thinking the designer drew the rudimentary objectives hierarchy of Figure 3. If resources had been available the designer clearly could develop the hierarchy and carry out the design in greater detail.

In Step 2, determination of available subsystems, the designer included 13 subsystems. They were vehicle area network, vehicle self-diagnostics, automatic vehicle location, pacing, collision warning, smart card fare payment, digital voice and data radio, telephone-based itinerary selection assistance, information kiosks at major locations, computer integration of operations, transfer coordination, flexible routing, and automatic operator check-in.

Table 1 shows the matrix resulting from the designer's efforts in Step 3, determination of raw scores for the subsystem-objective pairings. Because the scores shown are for illustrative purposes only, the reader should not interpret any score as representing an actual or estimated value. Moreover inasmuch as the analysis is taking place at a high level in the objectives hierarchy, the designer had to score qualitatively.

Note that the designer focused on the objectives called *doable, satisfying*, and *affordable*. However in evaluating the affordability objective, the designer focused on set-up cost to the exclusion of operating cost and resources. The designer needs to consider these other objectives for the design to be complete; recall that the analysis must include one and only one objective between the highest and lowest levels of each branch of the hierarchy. However the designer estimated that future operating costs would be no greater than, and it is hoped would be significantly less than, current operating costs, and so continuation of the existing fare revenue

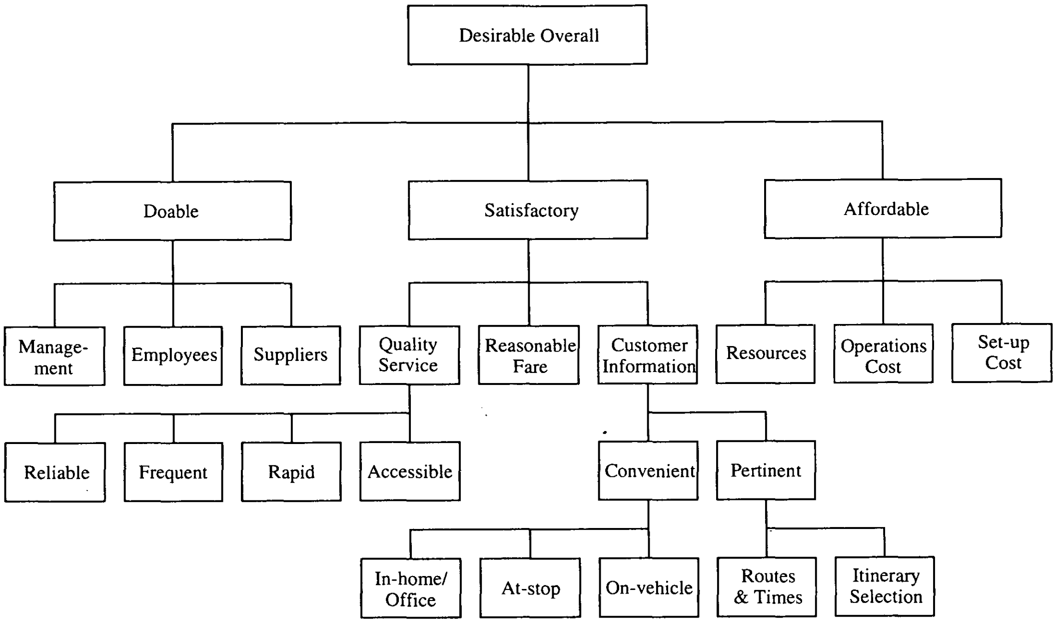


FIGURE 3 Example objectives hierarchy for transit system design.

and government subsidy could cover foreseeable needs. Therefore the designer temporarily set the issue of operating costs aside. The designer knew that the capital improvement budget was already available and fixed at $2,250,000, and so the resources were not included in the table either.

At this point the designer excluded the collision warning subsystem from consideration because of excessive cost. After elimination of collision warning, 12 subsystems were available, and so 2^{12} (4,096) portfolios resulted from the designer's initial attempt at Step 4. To circumvent this untenable situation the designer grouped the subsystems into three super subsystems, $S_{k'}$, of four subsystems each and continued with the design process. The first supersubsystem, which the designer called the vehicle-based supersubsystem, consisted of the vehicle area network, vehicle self-diagnostics, automatic vehicle location, and pacing subsystems. The second, or interface-based (vehicle-to-operations or customer-to-system), supersubsystem consisted of the subsystems providing smart card fare payment, digital voice and data radio, telephone-based itinerary selection assistance, and information kiosks at major locations. The third, or operations-based, supersubsystem, consisted of the subsystems providing computer integration of operations, transfer coordination, flexible routing, and automatic operator check-in.

TABLE 1 Raw Subsystem-Objective Scores (S_kO_j)[a]

		Objective j		
		O1	O2	O3
Subsystem k		Doable [b]	Satisfying [c]	Affordable [d]
S_1	Vehicle Area Network	3	1	350,000
S_2	Vehicle Self-Diagnostics	3	3	400,000
S_3	Vehicle Location	4	4	200,000
S_4	Pacing	4	5	50,000
S_5	Collision Warning	1	1	Unknown; high
S_6	Smart Card Fare Payment	3	5	450,000
S_7	Digital Voice/Data Radio	4	2	400,000
S_8	Customer Telephone Aid	3	4	100,000
S_9	Information Kiosks	3	4	250,000
S_{10}	Integrated Operations	3	2	50,000
S_{11}	Transfer Coordination	3	4	10,000
S_{12}	Flexible Routing	1	5	100,000
S_{13}	Operator Auto-Check-in	4	0	10,000

[a] Scores given are for illustrative purposes only.
[b] Score represents designer-perceived probability of successful system implementation (scaled 1 to 5).
[c] Score represents customer-perceived service quality (scaled 1 to 5).
[d] Score represents dollars needed to equip a transit system of 100 vehicles.

Because three supersubsystems were available for inclusion in a portfolio, 2^3 (8) portfolios were possible, as shown in Table 2. The designer assumed that implementing the interface-based supersubsystem without the operations-based supersubsystem would seriously degrade the effectiveness of the former. Therefore the designer eliminated those portfolios with the interface-based supersubsystem but not the operations-based supersubsystem, that is, Portfolios 3 and 7. The designer also put Portfolio 1, the status quo, in the background.

In Step 5 the designer used knowledge of the system and the design environment to determine an appropriate set of supersubsystem weights. The designer thought that the interface-based supersubsystem would influence the success of implementation of the system design more than the vehicle-based and operations-based supersubsystems and believed that the latter were of equal importance to this objective. Therefore the weights given to the

TABLE 2 Portfolio Descriptions

	Super-Subsystem k'		
Portfolio i	S_1'	S_2'	S_3'
(System i)	Vehicle	Interface	Operations
P_1	no	no	no
P_2	no	no	yes
P_3	no	yes	no
P_4	no	yes	yes
P_5	yes	no	no
P_6	yes	no	yes
P_7	yes	yes	no
P_8	yes	yes	yes

three components for the doable objective were 0.4, 0.3, and 0.3, respectively. Furthermore the designer thought that the interface-based component was the most important to travelers; this was followed first by the vehicle-based component and then by the operations-based component. Thus the weights given to the three components for the satisfying objective were 0.5, 0.3, and 0.2, respectively. Finally the designer thought the three super-subsystems were relatively independent from the vantage point of implementation; that is, the designer did not foresee any cost break for joint implementation of the various components. As a result the designer did not reduce any of the weights for the cost objective from the initial values of 1.0, 1.0, and 1.0, respectively.

Note that in this illustration the subsystem weights did not vary across portfolios. However for analyses done at a greater level of detail the weights might well vary across portfolios because of synergism among the subsystems for example. Furthermore many designers might find it difficult to provide weights of the form given in the illustration if the analysis requires more than a few weights. However most designers are able to rate each subsystem on a scale of, say, 1 to 100. These intermediate ratings are easy to normalize into the desired form.

The results from Step 6, determination of the raw portfolio-objective scores, are shown in Table 3. The designer calculated these scores using Equation 1, the subsystem weights from Step 5, and the raw subsystem-objective scores from Table 1. To get the supersubsystem scores from Table 1 the designer averaged the appropriate subsystem scores for the doable and satisfying objectives and summed the values for the affordable objective. Portfolio 8 clearly cost too much and the designer eliminated it.

In Step 7, determination of the objectives trade-off weights, the designer calculated the low/high difference in raw portfolio-objective scores for the doable, satisfying, and affordable objectives to be 1.3, 1.875, and $1,200,000, respectively. According to the perceived relative significance of these differences, the designer assigned weights of 0.3, 0.3, and 0.4, respectively, to these objectives.

In Step 8 the designer normalized the raw portfolio-objective scores from Table 3 using Equation 2. The result of such action is shown in Table 4, along with the total portfolio scores, which are described next.

To calculate the total portfolio scores, Step 9, the designer combined the normalized portfolio-objective scores from Step 8 using Equation 3 and the associated weights of the objectives from Step 7. Table 4 contains the weighted sums, which represent each total portfolio score, in addition to the normalized portfolio-objective scores from Step 8.

TABLE 3 Raw Portfolio-Objective Scores (P_iO_j)[a]

	Objective j		
Portfolio i (System i)	O1 Doable [b]	O2 Satisfying [c]	O3 Affordable [d]
P2	0.825	0.550	170,000
P4	2.125	2.425	1,370,000
P5	1.050	0.975	1,000,000
P6	1.875	1.525	1,170,000
P8	3.175	3.400	2,370,000

[a] $P_iO_j = \Sigma\,(SW_{ijk'})\,(S_{k'}O_j)\,(PS_{k'})$; the $SW_{ijk'}$ are from Step 5 (SW_{i11}=0.3, SW_{i12}=0.4, SW_{i13}=0.3; SW_{i21}=0.3, SW_{i22}=0.5, SW_{i23}=0.2; SW_{i31}=1.0, SW_{i32}=1.0, SW_{i33}=1.0), the $S_{k'}O_j$ are derived from Table 1 by averaging the appropriate subsystem scores, S_kO_j, for the doable and satisfying objectives and summing the scores for the affordable objective, and $PS_{k'}$ is 1 if the portfolio has super-subsystem $S_{k'}$ and 0 if not.

[b] Score represents designer-perceived probability of successful system implementation (scaled 1 to 5).

[c] Score represents customer-perceived service quality (scaled 1 to 5).

[d] Score represents dollars needed to equip a transit system of 100 vehicles.

TABLE 4 Normalized Portfolio-Objective Scores (NP_iO_j) and Total Portfolio Scores (TP_i)

	Normalized Portfolio-Objective Scores [a]			Total
Portfolio i (System i)	O1 Doable	O2 Satisfying	O3 Affordable	Portfolio Scores [b]
P2	0.0	0.0	10.0	4.0
P4	10.0	10.0	0.0	6.0
P5	1.7	2.3	3.1	2.4
P6	8.1	5.2	1.7	4.6

[a] $NP_iO_j = (10)[P_iO_j - \min(P_1O_j{:}P_IO_j)] / [\max(P_1O_j{:}P_IO_j) - \min(P_1O_j{:}P_IO_j)]$.

[b] $TP_i = \Sigma\,(OW_j)\,(NP_iO_j)$; the OW_j are from Step 7 (OW_1=0.3, OW_2=0.3, OW_3=0.4).

At this point the designer found Step 10, selecting the desired portfolio, to be straightforward. Portfolio 4, composed of the interface- and operations-based supersubsystems and costing $1,370,000, had the highest total score, so the designer chose it as the systems design.

DISCUSSION OF RESULTS

In the illustration a set amount of money was available and the designer needed to determine what level of sophistication the transit authority could implement. The reverse case is also possible; the designer can use the method both to determine the budget needed for capital improvement and to justify the inevitable appeal to governmental authorities (through a request for capital improvement funds), to the private sector (through issue of bonds or stocks), or to the general public (through proposal of a tax to support transit) to attain the funds necessary to implement the design. In other words the designer can develop a shopping list to show the community that at certain levels of funding, certain transit system functions are available.

The designer can also use the analysis to determine what percentage of the transit fleet the transit authority can afford to equip and how much less a given subsystem would have to cost to become feasible. For example if the designer had kept Portfolio 8 in the illustration and maintained the same objective weights, further analysis would show that reducing the cost of the vehicle-based supersubsystem by 15 percent, or equipping only 85 percent of the fleet, would have brought Portfolio 8 within budget and made it the best choice. The same would be true if the designer could increase the budget slightly.

The MAUT approach is also quite useful as a tool for developing requests for proposals to upgrade capital and operations and for evaluating systems designs submitted in response to those requests. The Ann Arbor Transportation Authority (AATA) in Ann Arbor, Michigan, is applying the method in this manner. Note that in systems evaluation, as opposed to systems design, relatively few portfolios exist because the number of portfolios equals the number of systems submitted in the competition, and the number of bids is commonly low. This greatly simplifies the evaluation at the expense of reducing the potential options.

Because the MAUT framework presented here represents an explicit and rational process, it is a good mechanism for drawing out needs, identifying solutions, and justifying decisions. Therefore the method should prove useful whether it is used to design systems or evaluate proposals.

The authors hope that this paper and the AATA experience (the results of which will be available soon) will stimulate discussion that might influence the development of a nationwide specification for APTSs. It is hoped that such a specification will detail a standard, open transit system design that addresses issues concerning systems architecture, technologies, and services from functional as well as product aspects. An appropriate APTS specification will also incorporate the current system and yet be highly amenable to modular expansion and upgrades in the future. For universal appeal any specification must be applicable in both large and small transit systems.

To develop a transit system capable of achieving desired transit objectives it will likely prove to be necessary to coordinate the entire transportation system by means of the process known as *mobility management.* Intermodal transportation linked by intermodal information may prove to be essential to future tansit competitiveness. Furthermore expansion of the system design process to encompass the larger community will be essential. Specifically the transit authority will need to gain the collaboration of authorities responsible for planning and oversight of roadways, land use zoning, and travel and parking regulations (*4*) as well as the cooperation of numerous community and special interest groups. Implicit in any transit growth scheme are cooperation and coordination regarding multijurisdictional, multiparty issues. Management of this larger problem might require broader methodologies such as social decision analysis (*10*) or policy exercise (*11*).

ACKNOWLEDGMENT

The research in this paper was partially funded by a grant from FTA (Project MI-26-0003).

REFERENCES

1. Wachs, M. U.S. Transit Subsidy Policy: In Need of Reform. *Science,* Vol. 244, June 30, 1989, pp. 1545–1549.
2. Wachs, M. Policy Implications of Recent Behavioral Research in Transportation Demand Management. *Journal of Planning Literature,* Vol. 5, No. 4, May 1991, pp. 333–341.
3. Rose, F. Empty Seats; Despite Huge Outlays, Transit Systems Fail to Lure Back Riders. *The Wall Street Journal,* Midwest Edition, June 29, 1993.
4. *Report on the Advanced Public Transportation Systems Smart Systems Integration Workshop, San Francisco, California, August 18, 1992.* Prepared at the University of Michigan, in conjunction with the Ann Arbor Transportation Authority, for FTA, U.S. Department of Transportation, 1993.
5. *Decision Making: Selected Readings.* (Edwards, W., and A. Tversky, eds.). Penguin Books Ltd., Harmondsworth, England, 1967.
6. White, C. C., E. C. Wilson, and A. C. Weaver. Decision Aid Development for Use in Ambulatory Health Care Settings. *Operations Research,* Vol. 30, No. 3, May-June 1982, pp. 446–463.
7. Miller, G. A. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. In *The Psychology of Communication: Seven Essays.* Basic Books, New York, 1967.
8. Raiffa, H. *Decisions Analysis: Introductory Lectures on Choices Under Uncertainty.* Addison-Wesley, Reading, Mass., 1968.
9. Keeney, R. L., and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* John Wiley & Sons, Inc., New York, 1976.
10. Chen, K., and T. B. Reed. Social Decision Analysis for Intelligent Vehicle-Highway Systems. *Proceedings of the Third Annual Meeting of IVHS America.* IVHS America, Washington, D.C., 1993.
11. Duke, R. D. *Gaming: the Future's Language.* SAGE Publications, John Wiley & Sons, Inc., New York, 1974.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Determining Appropriate Public Transport System for a City

R. L. MACKETT

Car ownership is growing in cities. This is leading to more congestion and environmental damage. To attract motorists from their cars it is necessary to improve the quality of public transport. In many cities this means building new systems. A variety of technologies are available, so decisions must be made to determine which is the most appropriate for a particular city. It is argued that the building of new transport systems can increase patronage and that cities in continental Europe have a much more positive approach to public transport than cities in Britain. There is scope for the transfer of knowledge about such systems from countries such as France and Germany to cities in Britain. As part of this process it is important to consider how decisions about the type of transport technology have been made. The methodology for the use of expert systems, a form of artificial intelligence, is described. The methodology is used to encapsulate the knowledge of experts in cities in continental Europe and to transfer it to cities in Britain, where decisions are being made about the type of public transport technology to that should be adopted.

Increasing car ownership is causing increasing congestion and environmental damage in cities. Greater car ownership leads to more car use and so reduces demand for public transport. In the long run as public transport revenue decreases the quality of service deteriorates and the downward spiral of public transport accelerates. Furthermore the shift from the use of public transport to the use of a car increases the rate of suburbanization, which in turn tends to favor car use and make public transport even more difficult to operate financially.

It would be perfectly possible to let this process continue, so that all urban mobility is offered by the car and public transport finally disappears. However there are a number of reasons why this is a bad idea:

1. It is impossible to provide all the road capacity to meet the demand, and so congestion occurs; this is inefficient because it wastes time and causes uncertainty in planning journeys.
2. Cars produce a variety of pollutants; although technical innovation can reduce emittants significantly in new cars, there are still many older cars on the road, and these pollute.
3. Some people will never be able to drive, for example, some of the young and the old, so there is a need to provide for their mobility. Some poor people cannot afford to buy or run a car, and lack of a suitable alternative can add to their deprivation, possibly leading to social problems.

Growing awareness of these issues has led to a recognition of the need to encourage urban public transport. This means not only that existing systems must be improved but also in some cases that new systems must be introduced. However such systems are expensive, can take a long time to build, and will have an impact on the city. Consequently care needs to be taken in making such decisions. It is possible to use computer models to assess the impacts of various possible systems, but such models require the specification of the systems to be tested. There is a need for a methodology to generate the systems to be evaluated. This procedure is a mixture of quantitative techniques and judgment in a political framework. This paper is concerned with the development of such a methodology by using techniques from the field of artificial intelligence.

In the next section the need for better urban public transport and the range of options are discussed. The issues involved in determining the appropriate form of public transport system are also discussed. Then the potential for using artificial intelligence techniques to address this issue is considered, and work on a project that uses such methods is reported.

NEED FOR BETTER PUBLIC TRANSPORT

Some of the problems caused by increasing levels of car ownership and the need for better public transport have been discussed. The two issues are complementary: there is a need to make car use less attractive and public transport more attractive. Some motorists at least are willing to forsake the car. The *Lex Report on Motoring* (*1*) shows that 35 percent of motorists agree with the view, "I would use my car less if public transport were better." In London, where congestion is the worst in Britain, 49 percent agreed with the statement. Currently there is considerable interest in the potential for road pricing. This means charging drivers for the use of the road so that they are paying an amount that better reflects the costs that they impose in terms of congestion as well as environmental damage. It also puts the charge for car travel on a similar basis to that for public transport, because once a person has bought a car, the marginal cost of making a journey tends to be lower than the equivalent cost of making a journey by public transport, where there are usually no capital costs for the user, so that the marginal cost is higher. In Britain the Department of Transport has commissioned a $4.5 million research project into road pricing in London, probably involving some form of electronic charging system (*2*).

If people are to be discouraged from using their cars the alternative modes must not only be attractive but must also have sufficient capacity. In many cities this means investing in new systems, because the existing public transport is provided by bus and suburban heavy rail only. However buses suffer from the same congestion caused by cars, and suburban heavy rail tends to have poor spatial coverage because it is expensive to build and requires heavy flows of at least 10,000 passengers per hour to justify the investment (*3*).

Centre for Transport Studies, University College London, Gower Street, London WC1E 6BT, Great Britain.

Table 1 shows the characteristics of various urban public transport modes. The mode that can be introduced in the cheapest and quickest manner is the standard bus running in traffic because it needs little new infrastructure. However it is subject to delays because of congestion and tends to have a poor image and so does not attract motorists from their cars. A guided bus system, such as that in Essen, Germany, or Adelaide, Australia, permits high-speed running along radial corridors, thereby avoiding congestion, but it retains the flexibility of covering the suburbs by using ordinary roads. It is debatable whether such systems can overcome the prejudice against buses. Many cities in continental Europe have trams, which can provide efficient movement of passengers to the city center. However, running on streets means that trams are delayed by cars, so in some cities, such as Vienna and Prague, tram routes are being removed as metro lines are being opened. Segregated light rail is really a modern form of tram, but it runs in separate corridors. Such systems carry large numbers of people at high speeds. The disadvantage is the need to find land on which to build the system. In some places, such as Newcastle-upon-Tyne in the north of England, the system goes underground in the city center. This can increase the cost substantially, but it may be necessary to provide sufficient penetration of the city center to attract car users. Higher capacity can be provided by a full-scale metro running underground. This system completely segregates the passenger from the surface, so that road congestion has no effect. The disadvantages are the high capital cost and the length of time it takes to build the system. These factors tend to mean that areal coverage is poor, particularly when a new system is built. Suburban rail can also convey large numbers along corridors, but penetration into the city center is usually poor.

In practice a large city needs a combination of public transport modes, with buses in the suburbs where their flexibility can be exploited, a high-capacity rail-based system along the radial corridors, and an efficient distributor system in the city center.

It was argued above that cities need good public transport to attract people from their cars and that this may require a major investment in new infrastructure. A variety of technologies is available, and so decisions must be made on what is appropriate for a particular city. British cities need investment in public transport if the damaging effects of cars are going to be limited. The following are two key questions: How does one decide what is the appropriate form of public transport technology that should be adopted, and how can Britain draw on the positive experiences in cities in continental Europe? These questions are addressed in the next two sections.

DETERMINING TYPE OF PUBLIC TRANSPORT SYSTEM

If it is accepted that there is a need to invest in new public transport systems, it is necessary to be able to determine what type of system is appropriate. Many factors are important. Some features of each city are unique, but there are many common factors. A variety of modeling techniques is available to assess the impact of a new system characterized by its capacity, speed, route pattern, and so on. These techniques can be used as part of an evaluation framework. What is lacking is a systematic way of generating the alternatives to be considered. In fact such decisions are based on experience and judgment as much as formal modeling techniques,

TABLE 1 Costs and Other Characteristics of Public Transport Modes

	Maximum capacity (1000pph/direction)	Commercial speed (km/h)	Operating cost per km per annum ($ x 10^6)	Capital cost for twin lanes ($ x 10^6)	Total cost over 30 year life ($ x 10^6)	Cost per passenger-km in cents
Standard bus in traffic	7.2 - 9.6	15	0.5 - 0.7	0.4 - 0.5	5.7 - 8.1	0.8 - 0.9
Guided bus	19 - 29	15 - 25	1.2 - 2.1	1.1 - 2.6	14.7 - 26.7	0.8 - 0.9
Tram (street running)	9 - 25	15 - 25	0.3 - 0.9	6.7 - 13.3	10.7 - 23.3	0.7 - 1.9
Light rail (segregated)	9 - 25	30 - 40	0.3 - 0.7	3.3 - 6.7	6.7 - 14.0	0.5 - 1.1
Metro (underground)	35 - 70	30 - 40	0.7 - 1.3	20.0 - 43.0	26.7 - 60.0	0.5 - 1.3

Note:

It is assumed that system is operating at 50 per cent capacity for 18 hours a day, 363 days a year over 30 years. The total operating costs over the 30 year life have been annualised at 8 per cent a year. The figures have been converted from £ to $ at an exchange rate of £1 = $1.50

Source:

Modified from a table in a review of people mover systems and their potential roles in cities, by B H North, published in the Proceedings of the Institution of Civil Engineers. Transportation, Volume 100, pp 95-110 so to use the lessons from one city in another it is necessary to find a method of encapsulating the relevant knowledge to transfer it from one city to another.

Before considering this matter further, it is relevant to examine some examples of decisions made on this topic to understand the type of knowledge to be transferred.

Tyne and Wear Metro and Docklands Light Railway

The Tyne and Wear Metro in Newcastle-upon-Tyne in northern England was opened in 1980. It was planned and operated by the Tyne and Wear Passenger Transport Executive under the directorship of T. Ridley. The Docklands Light Railway was opened in 1987 as part of the regeneration scheme in London Docklands. It was planned and operated by London Underground Ltd., the managing director of which was also T. Ridley. During the time between his work on these two systems he was also responsible for the development of the Hong Kong Mass Transit Railway, so he has considerable experience in making the type of decision being discussed here. Ridley (T. Ridley, unpublished data) argues that the following factors are required to get a new public transport system built in British cities:

1. A local political consensus, that is, agreement between all shades of political opinion;
2. A good working relationship between central and local governments at various levels (technical, managerial and political);
3. A consultant's report to give credibility to the project and to focus attention on the complexities of the issues; and
4. Luck.

Clearly, this is not quite the same issue as deciding between different types of technology, but it is illustrative of the factors that influence decisions in this field.

Shidami Human Science Town

An example in which a choice between a guided busway and a rail-based system was made was in Shidami Human Science Town to the northeast of Nagoya, Japan (*4*). A high-quality public transport link to the city center of Nagoya, about 12 km away, was required. In this case the guided busway was chosen for several reasons.

1. Duel-mode vehicles could have direct access to both the suburbs and the city center in the conventional bus mode and use the elevated section linking Shidami to Nagoya to provide a high-speed, frequent service in the guided busway mode,
2. Construction costs for the guided busway were lower than those for a rail-based system, and
3. The proposed system provided sufficient capacity initially but could be upgraded to a rail-based system later as demand grew.

In a later section of this paper the decisions about the appropriate scheme for Manchester, England, will be discussed. In that case light rail was chosen over guided bus. In Manchester and Shidami the final choice was between a light-rail system and a guided busway, and different solutions were found to be appropriate. This is a crucial point because the type of public transport system built must be appropriate to the problem being addressed. Wachs (*5*) argues that investment in rail transit in Los Angeles is taking funds away from local bus services, which are already overcrowded, and that what is really needed is increased local bus services together with adaptive improvements to the street network such as bus lanes and traffic signal priority for buses.

A research project has been set up to examine how these types of decision are made and to use methods of transferring the experience between cities. It is described after the discussion on the use of artificial intelligence methods.

USE OF ARTIFICIAL INTELLIGENCE METHODS

The need to provide new urban public transport technology has been demonstrated. However such investment is very expensive and takes a long time to come to fruition. Experience shows that different types of technology will be appropriate in different cities, with heavy rail most likely to be suitable in very large cities and buses most likely to be suitable in small urban areas, with the various alternatives shown in Table 1 fitting in between. A variety of modeling techniques for assessing the impacts of the various types of technology is available. The use of such techniques might well involve the characteristics of speed, route coverage, capacity, and so on, of a set of alternative technologies. The effects on patronage and fare revenue plus the costs could then be used in some form of cost-benefit analysis. Environmental effects such as emissions could also be modeled. However although such methods can be used to assist in assessing the appropriateness of the technology they can not take into account all the relevant factors because a lot of judgment is required, and that can only come from experience. By the nature of the type of system being considered here such decisions will be made very infrequently, so that many transport planners may be involved in only one such decision in a lifetime. Each decision is taken from first principles. One way to help overcome such problems is to circulate the knowledge of the various experts who have made such decisions in the past under a variety of situations. This can be done by using artificial intelligence methods, in particular, expert systems. Essentially an expert system is a computer program that provides advice on solving a problem, for example the best way to design a system, using the knowledge of experts. As Ortolano and Perman (*6*) explain an expert system has the following elements:

1. *Domain*, which is the subject area;
2. *Knowledge base*, which is a collection of facts, definitions, rules of thumb, and computational procedures applied to the domain;
3. *Control mechanism*, which is a set of procedures for manipulating the information in the knowledge base; this may be in the form of logical deductions from a set of facts and rules of the form of "if (premises) then (consequences)"; and
4. *User interface*, which usually is a visual display unit and keyboard linked to the computer running the expert system.

In the case being considered, the domain is the decision about the type of technology for an urban public transport system. The knowledge base will contain information about the characteristics of various technologies (speed, weight, capacity, and so on), the different types of system used in different cities along with their characteristics, the costs of the various systems, and so on. The control mechanism will be based on information from experts and could be of the following forms:

1. "If the maximum money available is less than $1.5 million per kilometer, then you cannot afford to tunnel,"
2. If traffic congestion is a problem in the city center, then you must segregate the new system from cars, or
3. If atmospheric pollution is a serious problem in the city center you need to use electric traction.

Such knowledge could come from interviewing experts in person or from written documents.

These are very simple examples, but when combined together and linked to some more conventional modeling techniques, a very powerful tool can be produced. The conventional modeling technique might be used to calculate the effects of the most appropriate alternatives, which would then be fed back through the expert system to give an explanation of why the proposed solution is the most appropriate and why others have been rejected.

Ortolano and Perman (*6*) identify six conditions for deciding whether a particular task can be codified into an expert system.

1. Knowledge needed for task performance is specialized and narrowly focused;
2. True experts, that is, people who know more than novices, exist;
3. The task is neither trivial nor exceedingly different;
4. Conventional computer programs are inadequate for the task;
5. The potential payoff from an expert system is significant; and
6. An articulate expert is available and willing to make a long-term commitment to build the expert system.

The problem being addressed here appears to meet the first five criteria: few people have made such decisions, so it is a specialized task, but such people do exist; the task is not trivial, but it does not approach the impossible; although conventional computing techniques are useful they do not really address the crucial question of how to choose the appropriate system; and given the huge costs of such systems and the problems if the wrong solution is developed, the potential payoff is huge. Whether the final condition is met depends on the particular application.

Hence it appears that there is scope for the use of an expert system in the determination of the appropriate type of public transport system that should be selected. A project to do this is described in the next section.

UTOPIA PROJECT

To study the issues identified a project was set up at the Centre for Transport Studies at University College London with funding of about $190,000 from the U.K. Science and Engineering Research Council for 3 years starting in January 1993. The project is known as UTOPIA (Urban Transport Operations and Planning Using Intelligent Analysis) and has the following objectives:

1. To help produce more civilized cities by improving transport operations and planning,
2. To transfer between cities experience of decision making about appropriate transport technology, and
3. To use artificial intelligence techniques to improve decision making in the field of transport.

The core of the UTOPIA work will be the use of expert systems to import knowledge to Britain from experts in cities in continental Europe that have made such decisions, such as Grenoble and Lille in France. The expert system lies at the core of a model that draws on other modeling techniques to show the implications of the various strategies produced. The model will then be applied to a variety of cities in Britain, particularly those where discussion on the possible solutions to the problems of congestion are being conducted, such as Leeds, where both light-rail and guided bus systems are being considered, and Bradford, where trolley buses may be reintroduced. The cities in Europe to be examined are places in France and Germany such as those discussed and other interesting cases such as Essen, with its guided busway, and Amsterdam, where a light-rail extension to the metro was opened in 1990.

A major task is the identification of the appropriate experts who have been involved in making these decisions. The method being used is to start from local contacts with knowledge of the topic and to ask them who else to talk to. In this way a network of experts can be built up. A second method that may be used, especially for cities outside Britain, is to distribute a questionnaire by mail to cities in continental Europe, for example, to the general manager of the system, as identified from a source such as *Jane's Urban Public Transport Systems* (*7*); via contacts at the Union International des Transports Publics (UITP) in Brussels; or through direct contacts such as T. Ridley, mentioned earlier in the context of the Tyne and Wear Metro and the Docklands Light Railway and who is now at the University of London Centre for Transport Studies where the UTOPIA project is being undertaken, although he has no direct involvement in the project. [His presence will help to meet the sixth criterion on Ortolano and Perman's (*6*) list in the previous section.]

The questionnaires will be framed in such a way that they can be answered only by an expert. It will be essential to know who has actually responded to the questions. The questionnaire will include a request for a personal interview. This will be undertaken only if it is clear from the questionnaire and other soundings that the person concerned really is an expert. It could be possible for a person to fill in the questionnaire dishonestly, but this seems unlikely, and as knowledge is circulated it should be possible to eliminate any such cases.

Different experts will provide expertise on the basis of different experiences. This means that it will be possible to apply, say, the Essen experience or the Lille experience to a city like Leeds and come up with different proposals in the same way that one might if one took two experts to the same city. The expert system will explain how it comes to each solution. These can then be explored with the local planners in Leeds to see which one they prefer.

It is recognized that many decisions are essentially political. For example a particular type of technology may be produced locally, and supporting local manufacturing industry may be an objective. To some extent such factors, if they are known, can be incorporated into the expert system. It cannot replace the political process, but it can help to improve the process by making it more transparent. The ability of expert systems to explain their decisions is particularly useful in this context.

The methodology being used in the UTOPIA project is shown in Figure 1. The user will be the planner in the British city who

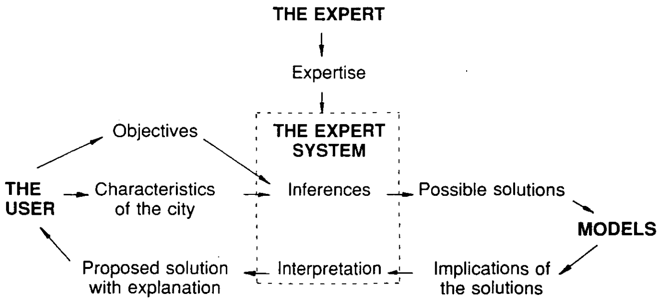


FIGURE 1 Methodology being adopted in UTOPIA study.

will define the objectives of the new system and provide information on the city. The objectives may be specified in terms such as capacity, speed, cost, and environmental effects. The expert system will incorporate various sets of expertise that have been encapsulated previously. Some possible solutions will be generated. Because an expert system is not ideal for handling complex mathematical functions, other models in, say, FORTRAN or C will be used to calculate the detailed implications of the system to be fed back through the expert system to provide an explanation to the user for why the chosen solution is appropriate. The user may then decide to revise the objectives, so the whole process is then repeated. Alternatively a different set of expertise can be used. The system is being designed to be interactive so that the planner can explore a range of options by using different criteria and consulting the knowledge of a range of experts. The system offers the opportunity to draw on a range of experts within a period of a few hours in a way that would probably be impractical if the experts had to be consulted in person.

PROGRESS ON UTOPIA PROJECT

As indicated the UTOPIA project started in January 1993. Initially the emphasis was placed on identifying appropriate public transport systems that should be studied, talking to various relevant people to help to identify experts and to build up knowledge, talking to British experts, and starting to develop the expert system.

As mentioned discussions have been held with T. Ridley, who was actively involved in the discussions about the Tyne and Wear Metro and the Docklands Light Railway, and further discussions will be held with him. More recently discussions about the decisions concerning the building of the Manchester Metrolink have taken place with experts. This is a light-rail system that opened in spring 1992. It uses two former suburban rail lines with street-running to link the former termini. The interview will be described here briefly to illustrate the nature of the process. The responses are based on notes taken by the author. The interview was tape-recorded and will be more systematically analyzed later for use in the expert system.

At the request of B. Tyson, one of the interviewees, a letter was sent in advance indicating the questions to be answered. These formed the basis of the discussion. They are provided below, with summaries of the main points of the responses.

INTERVIEW ABOUT MANCHESTER METROLINK

Place of interview	Offices of the Greater Manchester Passenger Transport Executive (GMPTE)
Date of interview	Wednesday, November 3, 1993
System being discussed	Manchester Metrolink
Interviewees	B. Tyson (Director of Planning and Promotion, GMPTE), T. Young (Operations Planning Manager, GMPTE)
Interviewers	R. Mackett, N. Tyler, M. Edwards (all CTS at UCL)

Question 1: What Alternatives Were Considered?

The following options were considered:

1. Closure of the two British Rail lines to Bury and Altrincham;
2. Continuation of the two lines, but with some investment;
3. A light-rail system, running on the two British Rail lines with street-running between the two city center termini;
4. As for Option 3, but with tunneling under the city center;
5. As for Option 4, but heavy rail, that is, a metro;
6. As for Option 3, but a busway; and
7. As for Option 6, but using guided buses.

This large number of options was considered because there was desire locally to look at a wide range and because the Department of Transport (that is, the central government department responsible for transport) said that it wanted a wide range to be considered.

Question 2: How Explicit was the Process of Deciding Between the Alternative Options?

It was an explicit process in which consultants were used to evaluate the alternatives. The patronage estimates for all the proposed systems were similar, so the decision was mainly based on costs. Tunneling was eliminated early on because of the high cost of access into and out of it and the lack of visibility of the system. This left Options 1, 2, 3, 6, and 7. Busways were then eliminated at the evaluation stage because of the high costs of removing the rail tracks. This left three options: closure, continued heavy rail with no rail connection between the two lines, and street-running light rail.

Question 3: If Alternative Technology was Considered, Would the Design of the System have been Different, for example, Alternative Routes, Stopping Points, or Interaction with other Traffic?

With a busway there would not have been so much segregation of the system from other traffic, and it would not have been necessary to move so many other services (for example, gas and electricity) from the affected streets. The former point means that congestion from cars, including misparking, would have had a greater adverse effect. The latter point occurs because light rail cannot be diverted when roadwork occurs, whereas a bus can.

Question 4: What factors were taken into account when deciding on the type of technology (for example, capacity, speed, and influence on demand)?

1. Capacity, to carry flows in the range of 1,000 to 5,000 passengers per hour, with a maximum of about 10,000 passengers per hour over the central sections;
2. Maximum speed of not less than 80 km/hr, with high acceleration and deceleration rates;
3. Ability to operate over the existing rail lines without extensive additional engineering costs;
4. High levels of reliability;
5. Acceptable environmental features;
6. Capability of expansion beyond the initial network;
7. Ability to run on the street (in the case of nontunneling options only);
8. Use of proven technology; and
9. Capability to carry large amounts of crosstown passenger movement.

Question 5: Have compromises been made because the vehicles run both off and on the streets? Was tunneling under the city center considered?

The system could never be driverless if street running was used. However the use of automatic vehicles was not seriously considered because of the desire for proven technology, the problems of keeping the line secure, and possible political problems of driverless vehicles in an area of high unemployment. Tunneling was considered, but it was rejected fairly early on in the decision process.

Question 6: To what extent have the level and method of funding influenced the design of the system?

The total level of funding affected decisions. With more funding the final system would have been of a higher quality, for example, refurbished suburban stations and better-quality seats in the vehicles.

The whole scheme has been implemented by using a DBOM (design, build, operate, and maintain) contract that will last for 15 years after the system opens. (The central government required the system to be built and operated by the private sector under contract to GMPTE.) There was a tendering process. The initial stage was to invite expressions of interest, and as a result of this 12 consortia were short listed. Of these, eight were selected for the first-stage tender. Five of these dropped out, leaving three that tendered. The differences between the three final designs included the vehicles, the overhead system, and the station design.

Question 7: What would you do differently if you were starting now?

GMPTE would have carried out more of the design and left less of it to the contractors. More thought should have been given to the design specification at the interfaces with third parties, such as British Rail and the city planning department. There should have been a more detailed reference specification. It might have been better to have had several small contracts instead of one large one. With a single large contract a contractor can hide delays but, on the positive side, must take into account the long-term maintenance implications of decisions at the design and building stages.

Question 8: Who actually decided on the type of system: politicians, managers, or technical staff?

Politicians actually made the decisions, with technical advice from the managers. Consultants were used to carry out much of the background work.

Question 9: What effects do you expect the systems to have on Manchester in terms of, for example, employment patterns and car use?

After 1 year patronage has already reached the level predicted for after 2 years. It appears to be attracting people out of their cars. There is anecdotal evidence that some people served by Metrolink are selling their second cars and even their first cars. One aim of building the system was to help the local labor market, which it has done. One of the four major aims of the Manchester Structure Plan is to retain the urban core, and it appears that Metrolink is likely to aid in that aim. It also helps to give an air of confidence to the city; for example, it featured prominently in Manchester's bid for the Olympic Games for 2000. Independent studies of the effects of the system are being carried out by the University of Salford and the consultants Oscar Faber TPA.

It can be seen that much useful information has been obtained and that much of it can be converted into statements of the form "if (premises) then (consequences)" for use in the control mechanisms in the expert system. Several volumes of reports produced at various stages in the decision-making process have been received and will be used to supplement the oral information summarized here.

On the technical side effort has been put into the design of the expert system. Much of the work has concentrated on the design of the Intelligent Cities Data Base. This will form part of the knowledge base of the system. It will also be used during the knowledge acquisition process, allowing experts and users to enter data on their cities in a systematic way, responding to questions from the computer. It will also provide the most appropriate value for a particular city in a particular year if none is available.

CONCLUSIONS

This paper has argued the need for better urban public transport systems. It has also suggested that cities in continental Europe tend to have a more positive approach to public transport than British cities, so there is scope for British cities to learn from experiences elsewhere. It is clear that there is a variety of public transport technologies available, and it is important to understand the implications of each. Choosing the appropriate type of system for a city requires considerable expertise. One way to apply the expertise from cities in continental Europe to British cities is to use expert systems. That is being done in the UTOPIA project. Although the work is still at an early stage it is showing great promise and is generating great interest.

ACKNOWLEDGMENTS

This work is being carried out as part of the project entitled An Intelligent Design Model for Multimodal Urban Public Transport Systems funded by the U.K. Science and Engineering Research Council (Grant GR/H 78481).

REFERENCES

1. *Lex Report on Motoring "The Consumer View,"* published by Lex Services Plc, with research by MORI, 1993.
2. *London Congestion Charging: A Review of Technology for Road Use Pricing in London.* Department of Transport. Her Majesty's Stationery Office, London, 1993.
3. Heywood, P. Heavy or Light Rail? In *Planning Passenger Railways* (N.G. Harris and E.W. Godward, eds.), Transport Publishing Co. Ltd., Glossop, Derbyshire, England, 1992.
4. Fujiwara, S. Shidami Human Science Town and Guideway Bus. Paper presented at the International Conference on Urban Development Policies and Projects, Nagoya, Japan, Oct. 1992.
5. Wachs, M. Learning from Los Angeles: Transport, Urban Form and Air Quality. *Transportation,* Vol. 20 No. 4, 1993, pp. 329–354.
6. Ortolano, L., and C. D. Perman. Applications to Urban Planning: An Overview. In *Expert Systems: Applications to Urban Planning* (T. J. Kim, L. I. Wiggins, and J. R. Wright, eds.), Springer–Verlag, New York, 1990.
7. *Jane's Urban Transport Systems 1992–93* (Bushell, C. ed.). Jane's Information Group, Coulsdon, Surrey, England, 1992.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Civil and Utilities Design Guidelines for Rail Transit Project

HARVEY L. BERLINER

The seemingly simple preparation of the civil and utilities drawings to be included in a set of transit facility construction contract documents requires designers who are familiar with transit design and reviewers who can look at a set of drawings and generate the proper comments. The information that needs to be contained within the civil and utilities drawings during both the preliminary and the detailed design stages is outlined in general terms. A guide for reviewers during a submission is provided, and their responsibilities are outlined. The designers and reviewers have performed their functions if the final drawings are properly coordinated and allow for efficient and correct construction.

It would take a book to recount all the problems that can arise during design and construction of a transit system. In that book many of the worst horror stories would begin apparently innocently: some minor drawing error or inconsistency was overlooked, only to emerge later as a major construction difficulty. Civil and utilities drawings in particular can become dangerously isolated from other types of drawings and can contain inconsistencies that hamper proper coordination. This paper is not the book of all that might go wrong but is a brief list of things that should be checked when reviewing civil and utilities drawings.

To visualize the problem more concretely consider the following scenario, which is not completely fictional. During the first review submission of drawings for a certain transit project, the reviewer found the civil and utilities drawings and the architectural and structural drawings to be oriented in opposite directions. When the reviewer requested that all drawings to be oriented in the same direction within the contract set, the designer replied that it was not unusual for a contract to have drawings with different orientations.

Later, when construction of the project was over budget and behind schedule, investigators traced the overruns and delays back to contract drawings poorly coordinated during design. Numerous change orders to correct the design deficiencies had delayed the contractor while the designer corrected the drawings. The "not unusual" inconsistency in drawing orientation had contributed to confusion and delay.

Such experiences show the hidden cost of poorly coordinated drawings: they can lead to confusion and misunderstanding, even when they are technically correct, and they may harbor actual undetected design errors. The goal of proper plan preparation is to minimize design errors before the final contract documents are advertised for bids. This is accomplished by proper coordination during design and indication of sufficient details on the drawings so that the contractor can understand and correctly build the project. Plan preparation depends on individual experience and background but must follow project guidelines to present the work consistently.

This paper outlines the typical civil and utilities drawings included in a rail transit contract, following the usual division of design into two phases: preliminary design and detailed design. The preliminary or 30 percent level study plans are prepared by a general consultant (GC) design team or by the transit authority and are then given to the detailed-design consultant (DDC) for completion. This paper describes what to include or verify during each phase for each group of drawings and what questions a reviewer must ask.

PRELIMINARY DESIGN

The preliminary plans are the first time that the project is laid out at a scale where the problems and conflicts can be identified and possible solutions analyzed. They also provide the basis for

- A more precise identification of right-of-way (ROW) requirements.
- A good engineer's cost estimate.
- The short list of proposed DDCs who will be asked to prepare their proposals.
- The selected DDC who will develop the design.

The following types of civil plans are included with the preliminary design drawings:

1. Cover sheet and index of drawings.
2. Plan and profile drawings; these drawings include track plan and profiles giving the following information
   - –Horizontal track alignment data with stationing (chainage).
   - –Rail profiles.
   - –Existing ground level.
   - –Contract limits.
   - –Station limits and platform centerline.
   - –Significant design features.
3. General arrangement drawings; these drawings include a plan and section of important features along the route. The contents of these drawings vary according to whether the design is for a bored tunnel, cut-and-cover tunnel, aerial section, or at-grade section, but the drawings generally include:
   - –Type of transit facility.
   - –Basic layout of structures.
   - –Type of drainage.
   - –Type of foundation.
   - –General information on existing structures.
   - –Existing utility information and proposed utility diversions.
   - –Typical sections.

Parsons Brinckerhoff Quade & Douglas, Inc., Two Gateway Center, 4th Floor, Newark, N.J. 07102-5098.

Before finalizing these drawings they should be carefully reviewed to ensure that the following basic information is included:

1. Are the track horizontal curves, vertical curves, and superelevations designed on the basis of the particular transit criteria? Review whether they satisfy the criteria.
2. Are streets and railroads designed on the basis of criteria for the different agencies involved with each?
3. If railroad tracks are involved, review the horizontal and vertical clearances between the transit facilities and the railroad tracks. Check the clearance requirements for that particular railroad. Check vertical clearances if the alignment crosses major streets.
4. Superimpose utility plans on the general arrangement drawings and determine whether any major utility relocation is required. Utility relocation should be kept to a minimum to avoid expensive and time-consuming relocation.
5. Review whether track horizontal and vertical curves are designed for the required design speed.
6. Review whether adequate sight distance is provided at street intersections.
7. Review whether typical sections, plans, profiles, and cross sections agree with each other.

During preliminary engineering accurate base mapping and existing utility mapping should be developed by the GC or transit authority. The base mapping provides the basis for many of the civil and utilities drawings throughout the design. Accurate existing utility maps are developed by

- Researching the as-built files of the affected municipalities and utility agencies.
- On-site investigation and survey.
- Digging exploratory bore holes in areas where it is necessary to know the exact location of underground utilities.

DETAILED DESIGN DRAWINGS

The preliminary design drawings establish the guidelines that the DDC must follow to develop the detailed design plans. However it is also the DDC's responsibility to thoroughly check all the information on the preliminary drawings to ensure that they meet the established guidelines and criteria. Any inconsistencies should be referred back to the preliminary designer for clarification.

The first step in developing the detailed civil and utility plans is laying out the drawing sheets to the same scale and sheet layout. The architectural, structural, electrical, and mechanical disciplines use their own distinct scales and sheet layouts. However, the civil and utilities drawings and the architectural, structural, electrical, and mechanical drawings must be oriented in the same direction. Common orientation helps ensure proper review and coordination throughout the design and gives the contractor and the engineer's field representatives a better understanding of the contract.

The detailed design drawings are usually reviewed by the GC or the transit authority during three submissions of the design period. These submissions are

1. In-progress submission or about 60 percent design level.
2. Prefinal submission or about 90 percent design level.
3. Final submission or 100 percent design level.

Some 15 types of civil and utilities drawings contained within a set of contract documents are outlined below, along with the types of information generally included on each. The drawings are grouped in two sections, first civil (10 types of drawings) and then utilities (15 types of drawings).

Civil Drawings

The civil drawings are the first plans that appear in the project set. These drawings lay out the project in total by defining

- Horizontal and vertical alignments.
- Grading, paving, and drainage requirements.
- ROW and easement needs.
- Traffic maintenance and traffic sign and marking requirements.

The civil drawings and a listing of what to include in each type of drawing are as follows:

1. Cover sheet, including contract title, contract number, contract description, and names of the DDC, GC, and transit authority.
2. Index of drawings. The drawing number and title in the index should be exactly the same as those in the title block of each drawing. Consecutively number each drawing and include the page number in the index.
   - –Identify all major streets.
   - –Show existing ground line on profiles.
   - –Provide coordinate grid and indicate at least two sets of north and east coordinates on each drawing.
   - –Provide the dimensions of typical sections of the transit alignment properly and describe their limits by stationing.
3. Alignment plans, plan and profile, and typical sections.
   - –Show track horizontal alignment data and vertical curve elevations. Identify all horizontal curves by code numbers. Include data sheets listing coordinates and stationing for all control points.
   - –Show profiles for all tracks. Indicate high and low points for all vertical curves.
   - –Indicate the horizontal and vertical clearances between the trackway and critical structures.
   - –Screen existing topography on the plan portion of plan and profile drawings.
   - –Outline tunnel, aerial structure, station structure, entrances, vent shafts, cross passages, and major utility crossings.
   - –Identify all major streets.
   - –Show existing ground line on profiles.
   - –Provide coordinate grid and indicate at least two sets of north and east coordinates on each drawing.
   - –Dimension typical sections of the transit alignment properly and describe their limits by stationing.
4. Grading and paving (restoration) plans, profiles, details, and sections.
   - –Show structure outlines, street lines, ROW lines, walls, sidewalks, ramps for handicapped individuals, curbs, medians, islands, alleys, drainage structures, fences, guardrails, and other surface features to be constructed or to be affected by transit construction.
   - –Dimension properly the driveways, access roadways, parking lots, and bus bays. Provide sufficient layout information

and alignment data on the drawings for their constructibility.

- Indicate all dimensions and alignment data needed to define and locate features not identified elsewhere.
- When the information is available, show (or reference other plans) type of pavement, curbs, and other details for areas to be paved, repaved, or restored.
- Define clearly the areas to be constructed or restored; include a reasonable area outside the limits of excavation and include any areas affected by utilities restoration.
- Show typical sections of highways, streets, parking lots, and bus bays. Thicknesses of pavement sections must be in accordance with local agency requirements.

5. Surface drainage plans, profiles, details, and sections. This information can be shown on the grading and paving plans depending on the complexity of the drainage system and whether the drawing scale will allow all information to be shown clearly on one drawing.
   - Properly indicate structure outlines and street lines.
   - Show layout of new and relocated drainage facilities.
   - Indicate on plan drawings the sizes of pipes and culverts and direction of flow, types of channels and gutters, and types of structures.
   - Develop profiles for all drainage lines and coordinate them properly with the plan drawings. Show all other utility crossings and any pertinent structural features on the drainage profiles.
   - Develop sections and details at the locations necessary to properly define the surface drainage and sewer systems.
   - Check to ensure that the surface drainage design is coordinated with the facility roof and floor drainage.
   - Check to ensure that all aspects of the transit facilities are adequately protected against flooding.
6. Cross sections.
   - Make certain that any cross sections required of the transit facilities, streets, driveways, access roadways, parking lots, and bus bays agree with the plans, typical sections, and profiles.
   - Indicate top-of-rail elevation for each transit section and pavement centerline (baseline) elevation for each roadway section.
   - Facilities on each drawing must be identified only once. This is done by the use of "Typ." to indicate typical facilities throughout a drawing.
7. ROW drawings.
   - Indicate the existing ROW and the permanent and temporary easement lines, including all the coordinates that the contractor needs.
   - Coordinate any necessary changes in the easement lines with the GC or transit authority.
   - Identify ample contractor work sites on the drawings. If a potential work site lies outside the easement lines, indicate it as a potential work site to be arranged by the contractor.
8. Traffic maintenance plan, construction access, temporary parking lots, and detours.
   - Develop plans indicating both existing traffic circulation and proposed circulation during construction.
   - If construction is to be performed in stages, outline in detail the traffic circulation for each stage. The traffic maintenance plan must be approved by the relevant agencies.
   - Detail any special construction access, temporary parking lots, or detours required for the contract.
9. Traffic signs and markings. Prepare plans of traffic signs, signaling, and pavement markings for each construction phase and for the final roadway layout.
10. Standard drawings. Prepare typical standard drawings in accordance with the details provided by the local agencies and the transit authority.

Utilities Drawings

To save money and time, some engineers include the utility design work as part of the civil drawings. Because of the importance and complexity of the utility work involved with any rail transit project, the utility design must be shown on its own set of drawings and coordinated closely with the civil design. These drawings define the utility work in plan, detail, and, where appropriate, profile. The utility drawings and a list of what to include in each type of drawing are as follows:

1. Composite plan of existing utilities.
   - Include existing utilities drawings for the entire project.
   - Indicate the outline of the transit structure and the centerline of the track alignment with stationing.
2. Composite plan of utilities rearrangement.
   - Indicate the transit structure details and track centerline.
   - Indicate relocation schemes for all affected utilities. Whenever possible relocate utilities permanently.
   - Indicate the utilities to be abandoned in place and those to be properly supported in place. Show any new utility construction to be performed by either the contractor or agencies.
   - Check that there is sufficient room above the cut-and-cover construction to support the utilities during construction, including the related manholes and handholes.
   - Check that there is sufficient ROW for any permanent or temporary utilities relocated outside the construction area.
   - Indicate properly the interfaces between the building services and outside utilities. Coordinate the utility drawings with the electrical and mechanical drawings.
3. Composite cross sections. Indicate on cross sections the transit structure and the elevation of the top slab, the treatment method for each affected existing utility, and the locations of the proposed utilities. The cross sections must agree with the rearrangement drawings. If this information can be shown clearly and at proper locations on the civil cross sections, it is not necessary to have separate utility composite cross sections.
4. Utility profiles. Provide profiles for all gravity sewers and all utilities that cannot be constructed at a uniform depth below ground surface. Profiles should indicate the following:
   - The relevant portion of the transit structure, the top slab elevation, and the existing and finish grade lines.
   - The slope, elevation, and connection method of the proposed utility.
   - Sizes, materials, and other details necessary for construction.
   - Crossing utilities.
5. Utility details.
   - Develop details specific to the contract. Ensure that they

agree with the rearrangement plans and the profile drawings.
–Include standard drawings in accordance with the details provided by the utility agencies.

Principles for Preparing All Civil and Utilities Drawings

Although each type of drawing has its own specific criteria, there are general principles applicable to all drawings. First, all drawings should be prepared in accordance with the drafting manual and computer-aided drafting and design (CADD) standards manual provided by either the GC or the transit authority. These manuals define the drafting standards to be used for all drawings. The other general principles are as follows:

- Indicate location in the titles of drawings whenever possible. Provide a key plan when applicable.
- Make all line work and lettering of sufficient size, weight, and clarity so that the half-size drawings can be read and scaled properly.
- Do not duplicate information in a set of documents. Use cross-referencing as required instead of making multiple copies in various locations. This will avoid confusion when a drawing must be changed.
- Review structural, architectural, electrical, and mechanical drawings regularly to ensure their agreement with the civil and utilities drawings. Such a type of review helps ensure proper coordination between disciplines.

REVIEW OF SUBMITTALS

Reviewers on the staff of either the GC or the transit authority review each submittal from the DDC for compliance with the project criteria and to verify that the design is progressing in an orderly manner. The first submittal is very important because it is the first time the entire project is laid out and presented in a contract set. This is the time to steer the DDC in the right direction for the final bid documents.

The reviewer must answer the following questions positively when reviewing a set of documents:

- Are the plans easy to read and understand?
- Do the plans follow the criteria and guidelines specified in the drafting manual and CADD standards manual?
- Are the layouts of the drawings well planned on the basis of the space available on the sheets?
- Have the technical criteria of the transit project been followed, and does the design allow for safe and economical constructibility?
- Have previously agreed-upon review comments been incorporated into the next submittal of the contract documents?
- Can everything shown on the plans be accurately laid out by the contractor in the field?
- Have the proper standard details been included in the contract set?
- Have all the utility rearrangements been coordinated with the utility agencies, and has the constructibility of utilities been checked for both the temporary and the permanent facilities?

CONCLUSIONS: GENERAL PRINCIPLES FOR CIVIL AND UTILITIES DRAWINGS

The care required to keep civil and utilities drawings accurate, up to date, and well coordinated with other contract drawings pays off in efficient construction. This payoff underlies all the specific rules listed in the paper. Thus the design and review process may be summed up in a few general principles.

- The preliminary drawings are developed by the GC or the transit authority, and these drawings are used by the DDC as the guidelines to develop the detailed design drawings.
- The detailed design drawings are developed to the criteria and guidelines specified in the project's drafting manual and CADD standards manual. This will ensure consistency among all the contracts being prepared for the project.
- It is the DDC's responsibility to coordinate the drawings from the different disciplines within the contract set, despite each discipline's tendency to worry about its own drawings and to forget to coordinate changes with other disciplines. It is usually the responsibility of the civil and utilities designers within the DDC and of reviewers within the GC or transit authority to ensure the cross-discipline coordination.
- The review process is very important because it ensures that the DDC has followed the technical criteria of the project and that the design allows for the constructibility of the contract. It is also recommended that independent constructibility peer reviews be performed at the conclusion of the preliminary design phase and during the prefinal submission review.
- Resolving problems during design reduces the need for resolving problems during construction. This is more cost-effective. Therefore before the drawings are signed, the designers and reviewers must verify that the contract set indicates sufficient details so that the contractor can construct the contract efficiently and correctly.
- The designers and reviewers should never compromise their values for the sake of completing the contract documents on a particular schedule. This is sometimes extremely difficult within the tight design schedule imposed by the owners. However it is extremely important that the contract documents be correct and free of errors. An extra day taken during design to ensure correctness and constructibility can save weeks during construction to correct design errors.

Publication of this paper sponsored by Committee on Rail Transit System Design.

PART 2

Management, Marketing, and Fare Policy

Managerial Uses of Causal Models of Subway On-Time Performance

GARY HENDERSON AND VENGAL DARAPANENI

On-time performance (OTP) indicators reflect the performance of employees and the effectiveness of policies and organizational structure. However in New York City's subway system the operating environment includes factors beyond the control of operating employees, such as mechanical reliability of subway cars, route merges, scheduled headways, construction projects, crowding, passenger behavior, and so on. These factors differ from route to route and complicate comparisons of routes. Planning decisions affecting ridership levels, route design, and the operating environment can benefit from precise estimates of impacts on OTP, especially to help in evaluating the trade-offs involved. The results of a statistical causal model of on-time performance are presented. A set of hypothetically important variables was developed from New York City Transit Authority documents and train movement records. Data for over 54,000 morning rush hour trains traveling from terminals to central business district stations during 1988 and 1990 were used. The model quantifies the effect of variables on the probability that a train will be on time. How the results of such a model can be used to make predictions of performance that control for the operating environment are also shown, allowing performance comparisons between routes with different characteristics. How different OTP goals can be set for different routes is suggested. Finally by converting the results to odds ratios, it is shown how small improvements on routes with OTP of more than 90 percent can provide large benefits from the perspective of riders and how OTP measures obscure that fact.

On-time performance (OTP) indicators reflect employee performance and the effectiveness of managers, operating policies, and organizational structure. However performance levels also reflect the operating environment, for example, the mechanical reliability of subway cars, the frequency of route merges, the spacing between trains (scheduled headways), construction activity, crowding, and passenger behavior. These factors differ from route to route and complicate comparisons of one route with another. Is a 2 percentage point improvement in OTP for a long route with numerous merges equal to the same improvement for a short route with no merges? How can OTP goals be set for different routes?

The Office of the Inspector General (OIG) of the Metropolitan Transportation Authority (MTA) made a statistical causal model of New York City's rush hour subway service to account for differing infrastructures, route designs, and other variables. The model allows performance analysis to control for these factors. Are some routes with low OTPs really doing well given the operating environment? Should some routes with high OTPs be expected to do better? Quantifying the effects on subway performance also provides a planning tool that can be used to anticipate the impacts on timeliness caused by capital projects and the redesign of route structure. Although practitioners have long considered such factors in predicting service impacts, quantification of these effects is needed when change involves multiple variables. When some variables have opposite effects, an evaluation of the trade-offs requires some precision. For example MTA's proposal to extend the 63rd Street Tunnel requires evaluation of numerous route design options. The results of this model show that increased throughput achieved under some options would be offset by degraded reliability caused by the addition of merges.

FACTORS AFFECTING OTP

Two data bases covering morning rush hour subway performance in 1988 and 1990 were used (*1,2*). Each was produced from an analysis by OIG staff of New York City Transit Authority (NYCTA) train movement records. Together they include times for over 54,000 trains from terminals to central business district (CBD) stations, including all trains arriving between 6:00 and 10:00 a.m. From these data and other sources, we quantified hypothetically important variables. The variables found to have a significant impact on OTP were (a) number of route merges, (b) whether public schools are in session, (c) scheduled headway, (d) distance traveled, (e) stops, (f) crowding, that is, an index of ridership compared with scheduled capacity, (g) whether construction occurred the night before, and (h) mechanical reliability of subway cars, measured by mean distance between failure (MDBF). Other variables lacked useful data, were statistically insignificant, or were accidentally correlated with OTP.

Coefficients for the estimated effects of each variable on OTP (Table 1) were produced by logistic regression (*3–5*). OTP, the dependent variable, is dichotomous; each train is categorized as success or failure, as on time or late. Logistic regression estimates how variables affect the probability of being on time. It is a nonlinear model; the magnitude of the effect changes depending on the starting level of OTP, with the largest effects occurring when OTP is near 50 percent. This is necessary mathematically, because OTP cannot exceed 100 percent. The nonlinear behavior of OTP also provides an interesting perspective on how to measure performance, which will be discussed later. The coefficients produced by logistic regression are expressed in terms of logits—the natural logarithm of the ratio of successes to failures (the odds ratio). Table 1 gives three coefficients for each variable, using the 1988 and 1990 data both separately and combined. In Table 1 a negative sign, as for *merges*, indicates that an additional merge can be expected to hurt OTP, whereas a positive coefficient, as for *headway*, suggests that routes with longer headways are more likely to be on time.

A variable's effect on OTP depends on the starting level of OTP. It requires a conversion from logits back into OTP, and this is done in Table 2. The coefficient of each variable is given just

---

Office of the Inspector General, Metropolitan Transportation Authority, 100 Park Avenue, 14th Floor, New York, N.Y. 10017-5516.

TABLE 1 Estimates of Change for Causal Variables

VARIABLE	UNIT OF CHANGE	BOTH YEARS	1990	1988
MERGES	One	-0.292	-0.237	-0.310
CROWDING INDEX	1.0	-0.881	-1.077	-0.511
SCHOOL DAY	True	-0.132	-0.071	-0.156
MDBF Starting Level				
ANY	10,000 Miles	n.a.	0.029	0.098
10,000 MILES	10,000 Miles	0.305	n.a.	n.a.
30,000 MILES	10,000 Miles	0.113	n.a.	n.a.
60,000 MILES	10,000 Miles	0.066	n.a.	n.a.
TRIP LENGTH	1 Mile	-0.013	-0.010	-0.034
HEADWAY	1 Minute	0.107	0.041	0.172
NIGHTWORK	True	-0.127	-0.017*	-0.198
INTERCEPT		1.780	2.540	1.344

Notes:

The coefficients show the expected change in the natural logarithm of the odds of being on-time, i.e., the ratio of on-time probability to late probability. To see how this translates into OTP itself, refer to Table 2.

* The estimate is not statistically significant at the 95% confidence level. All coefficients except for Nightwork (1990) and Trip Length (1990) are significant at 99.9% confidence.

n.a. For the MDBF analysis of the combined 1988 and 1990 data, we used the natural log of MDBF to show how the strength of the effect differs for different starting levels of mechanical reliability. This non-linearity demonstrates diminishing returns (in terms of higher OTP) on investment. Such a logarithmic method applied to 1988 and 1990 data individually was not statistically significant.

Source: Analysis by OIG of NYCTA train movement records, General Orders, ridership counts, schedules, and other variables from 1988 and 1990. Analysis used logistic regression on the SAS system.

below the variable's name. In the column below that, the change in OTP expected for each additional merge, minute of headway, mile traveled, and so on, is provided for each level of OTP. For example an additional merge when OTP is only 5 percent would lower OTP to 3.8 percent. (To avoid confusion, we will use the term *percent* when we refer to OTP itself and the term *percentage point* when we refer to the change in OTP caused by the variable. For example, if OTP is 80 percent and a change in some variable causes a 10 percentage point decline, the resulting OTP would be 70 percent.) When OTP is 50 percent one more merge would lower OTP to 42.8 percent. Figures 1 and 2 illustrate the trends in Table 2.

Merges gives the number of times a given route converges on the same track with another route. For example the D route merges with the Q and B routes before West 4th Street (two merges). We did not include divergences: They should not delay trains or across-the-platform transfers, because rush hour trains are not supposed to be held for connections. Using data from both years, we estimated that each time a route merges OTP may drop as much as 7.2 percentage points (3 percentage points lower than when OTP is at 90 percent.) The effect of *merges* was greater in 1988 than in 1990, reflecting that OTP in general was lower in 1988 when mechanical reliability was so poor and schedule adherence so much worse than that in 1990 that the mistiming of trains at merge points was more acutely felt. Because the negative impact of merges results from the mistiming of train arrivals at the merge point, better schedule adherence can reduce the impact of this variable. However merges may always be a strong negative factor because of some inevitable lateness. For example schedule adjustments to manage service evenness are typically made by terminal dispatchers for one of the merging routes without conferring with the other terminal. A centralized, modernized control center planned by NYCTA may reduce the number of delays resulting from uncoordinated actions of decentralized, local decision makers, but we doubt that New York's system of merging routes can ever be completely rationalized.

Crowding index measures the ratio of the number of passengers to the capacity on that route (measured at the most crowded point) for each half-hour period. For example if the trains currently scheduled can carry 14,500 riders and 12,000 riders pass through

TABLE 2 Effects of Variables on Morning Rush Hour Subway OTP

							10,000 MILE IMPROVEMENT IN MDBF WITH STARTING MDBF OF		
ON-TIME PERCENT (OTP)	ONE MORE MERGE	SCHOOL IN SESSION	SCHEDULED HEADWAY (+1 MIN)	ONE MORE MILE	CROWDING INDEX (+0.1)	NIGHT WORK	10,000	30,000	60,000
	-0.292	-0.132	0.107	-0.013	-0.881	-0.127	0.305	0.113	0.066
1	-0.3	-0.1	0.1	0.0	-0.1	-0.1	0.4	0.1	0.1
5	-1.2	-0.6	0.5	-0.1	-0.4	-0.6	1.7	0.6	0.3
10	-2.3	-1.1	1.0	-0.1	-0.8	-1.1	3.1	1.1	0.6
20	-4.3	-2.0	1.8	-0.2	-1.4	-2.0	5.3	1.9	1.1
30	-5.8	-2.7	2.3	-0.3	-1.8	-2.6	6.8	2.4	1.4
40	-6.8	-3.1	2.6	-0.3	-2.1	-3.0	7.5	2.7	1.6
50	-7.2	-3.3	2.7	-0.3	-2.2	-3.2	7.6	2.8	1.6
60	-7.2	-3.2	2.5	-0.3	-2.1	-3.1	7.1	2.7	1.6
70	-6.5	-2.8	2.2	-0.3	-1.9	-2.7	6.0	2.3	1.4
80	-5.1	-2.2	1.7	-0.2	-1.4	-2.1	4.4	1.7	1.0
90	-3.0	-1.3	0.9	-0.1	-0.8	-1.2	2.4	1.0	0.6
95	-1.6	-0.7	0.5	-0.1	-0.4	-0.6	1.3	0.5	0.3
99	-0.3	-0.1	0.1	0.0	-0.1	-0.1	0.3	0.1	0.1

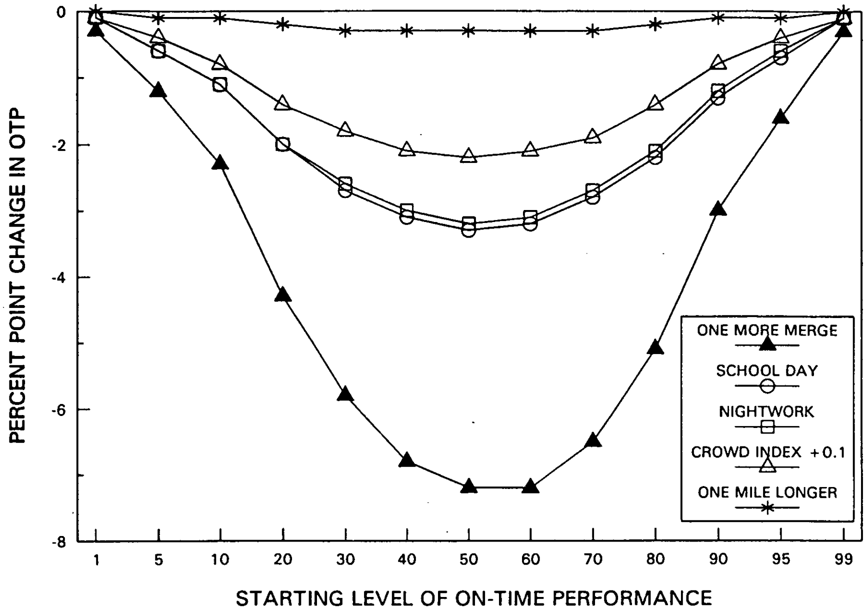
the maximum load station, then *crowding index* would be 0.83. The number of passengers came from NYCTA's annual counts of riders entering the CBD. Capacity is based on the service that NYCTA actually provided on the days of these counts. The index was calculated for each route for every half hour. In the 1990 data it ranged from the sparsely used downtown Q-route service (0.009 at West 4th Street between 6:00 and 6:30 a.m.) and the heavily used downtown E and F routes (1.193 at Fifth Avenue between 8:30 and 9:00 a.m.). The coefficient for *crowding index* (Table 1) gives the change in the logit for a change in the index from 0 to 1.0 (i.e., from no passengers to total capacity), but in the discussion of the measure—and in Table 2 and Figures 1 and 2—the expected change for an increase in the index of 0.1, for example, an increase from 0.5 to 0.6 was shown. It was found that an increase in the index of 0.1 would lower OTP by 2.2 percentage points when OTP is 50 percent and 1.4 percentage points when OTP is 80 percent.

In measuring crowding characteristics of passenger behavior was inevitably included. To the extent that the social norms of New York subway riders differ from those of riders in other cities and of other cultures, the estimates may not be applicable elsewhere. The policy implications may differ as well. As trains grow more crowded OTP tends to be lower because dwell times can increase dramatically when trains are crowded. In New York passengers on the platform appear more inclined to crowd aggressively around doors when trains are very crowded, perhaps because it often happens that the meek sometimes cannot board the train at all. This increases dwell time by slowing down deboarding passengers. Moreover crowded platforms require train operators to drive more slowly into stations to maintain safety.

The effects of all variables in the model were stronger in 1988 than in 1990 with the exception of *crowding index*. The effect of crowding was almost three times stronger in 1990 than in 1988. For example in 1988 the addition of 0.1 on the index, when OTP was 80 percent, would decrease OTP to 79.2 percent, but would decrease OTP to 78.2 percent in 1990. NYCTA reported that systemwide rush hour OTP improved from 89.6 percent in 1990 to 91.7 percent in the first 10 months of 1992, when ridership was lower. It is estimated that ridership loss accounts for one-fourth of the total improvement.

School indicates that New York City public schools were in session. Its values are the same for all routes. When school is in session the likelihood of a morning rush hour trip being on time declines by a maximum of 3.3 percentage points, probably because of the higher rate of pulled emergency cords, held train doors, and so on.

Nightwork indicates construction activity on the previous night. All work should have ended before the rush hour began, although

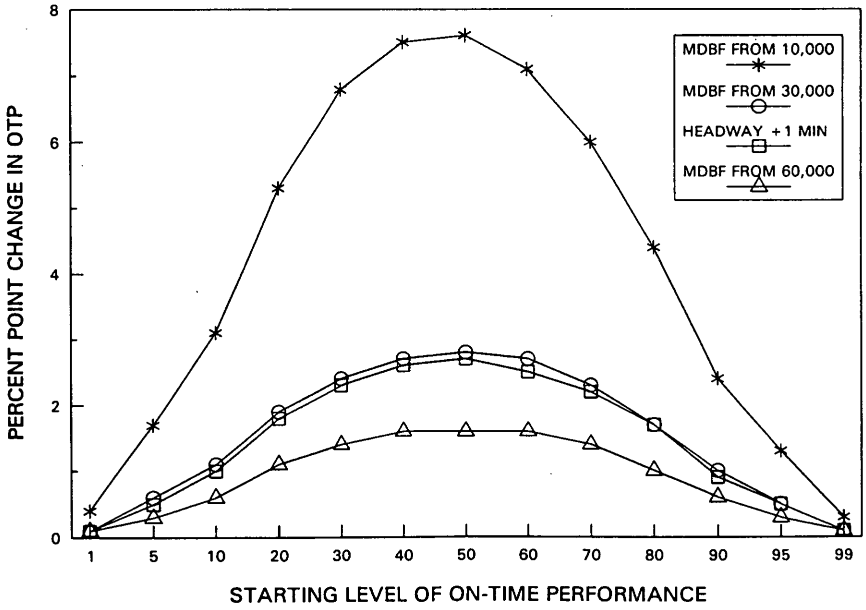


Source: Analysis of NYCTA morning rush hour subway data, 1988-90 (See Table 2)

FIGURE 1 Variables that decrease OTP.

track conditions are sometimes not perfect. If nightwork was scheduled the value of *nightwork* for every train on that route until 6:30 a.m. would be "true". All others are counted as "untrue." For passengers arriving in the CBD before 6:30 a.m. nightwork decreased the chance of being on time by a maximum of 3.2 percentage points. (In 1988 the effect could be as great as 4.9 percentage points.) This is caused when work cannot be finished on time or by slow work trains returning to the yard after construction duties. The coefficient for nightwork in 1990 was much weaker than that in 1988 and was not statistically significant. This suggests that NYCTA has been successful in scheduling and planning capital construction.

MDBF is the average monthly number of miles that trains travel before a mechanical failure causes a cancellation or a delay of



Source: Analysis of NYCTA morning rush hour subway data, 1988-90 (See Table 2)

FIGURE 2 Variables that improve OTP.

more than 5 min. The car class of every train was known and the monthly MDBF for that class was assigned to each record. MDBF is a crucial factor and perhaps the one most amenable to managerial intervention. The significantly improved MDBF of the subway fleet, primarily as a result of new car purchases and car overhauls, helped produce the improvement in New York City's subway reliability that occurred between 1988 and 1990. The coefficients in Table 1 give the effect of raising MDBF by 10,000 mi.

MDBF was treated to test whether improvement would have a greater effect when MDBF is low than when it is high. This approach was statistically significant when considering both years together. Therefore when combining data from both years, a different coefficient is given for cars with MDBFs of 10,000 mi than for cars with MDBFs of 30,000 or 60,000 mi. Table 2 gives three MDBFs to show that the effect depends on the magnitude of MDBF itself and on the starting level of OTP. However when 1988 and 1990 were looked at separately, this "nonlinear" approach was not statistically significant, so only one coefficient is listed for the individual years. When MDBF is 10,000 mi before a service disruption, improving it to 20,000 mi could improve OTP by as much as 7.6 percentage points. A 10,000-mi improvement is modest in light of the accomplishments of the new car and car overhaul programs. Now no cars have MDBFs as low as 10,000 mi, but in 1988 an MDBF of less than 6,000 mi was not unusual for certain car classes.

For most routes further improvement in mechanical reliability will not improve OTP by many percentage points. With an MDBF of 30,000 mi and OTP of, say, 80 percent, the expected improvement for raising MDBF to 40,000 mi is 1.7 percentage points. However with MDBF at 60,000 mi and OTP at 90 percent an increase in MDBF to 70,000 would raise OTP only to 90.6 percent. This conclusion was borne out by recent data. The 12-month rolling average MDBF for the subway fleet after August 1993 was 50,048 mi, a healthy improvement over the average in August 1992 of 41,452 mi. However this 8,500-mi improvement (more than 20 percent) had no measurable effect on OTP, which actually went down slightly during the same period (*6*). Similarly Table 1 shows that the effect for 1988 (0.098 for every mile) was three times greater than the effect for 1990 (0.029 for every mile). MDBF may have played itself out as a means of improving OTP, but not reliability generally, as will be seen.

Trip length gives the effect of adding a mile to the train's run. This variable was used in place of scheduled travel time because the latter varies during the rush hour; for example, the scheduled travel time of the Lexington Avenue Express between 125th Street and Grand Central is 7 min longer at the peak than early in the rush hour. The longer the trip, the more time there is for something to go wrong. Moreover the OTP standard is 5 min, not a percentage of the running time.

Headway is the scheduled time between trains at the most congested point. For example an A train from Lefferts Boulevard merges with the A train from Far Rockaway and the C train before Canal Street. The combined headway at Canal Street was used; for example, the time between an A train and a following C train is the C train's headway. The chance of being on time improves as headway grows larger. When more trains are scheduled and headway consequently decreases, OTP will decline by some amount. Adding 1 min to the headway of a route with 80 percent OTP raises OTP to 81.7 percent.

LESSONS FOR OTP MEASUREMENT THEORY

Logistic regression is the appropriate method for modeling a probability measure like subway OTP, and this provides an important insight into the uses and drawbacks of an OTP measure. Figures 1 and 2 illustrate that a given variable affects OTP most when OTP is about 50 percent—an extremely poor level of service—and least when it is closer to zero or 100 percent. As service improves it becomes more difficult to increase the OTP statistic; more resources are needed to produce the same percentage point change in OTP. However passengers may detect significant improvement in reliability even though the OTP statistic changes little (*7*).

Figure 3 shows how the odds ratio—the ratio of on-time to late trips—responded to changes in four variables. The change from 95 to 96 percent OTP may appear small, but in terms of the odds it is substantial; instead of experiencing a delay once for every 19 on-time trips, riders have one delay for every 24 on-time trips. By contrast the change in OTP from 75 to 76 percent is imperceptible: with OTP at 75 percent, riders are late once for every 3 on-time trips; with OTP at 76 percent, riders are late once for every 3.2 on-time trips.

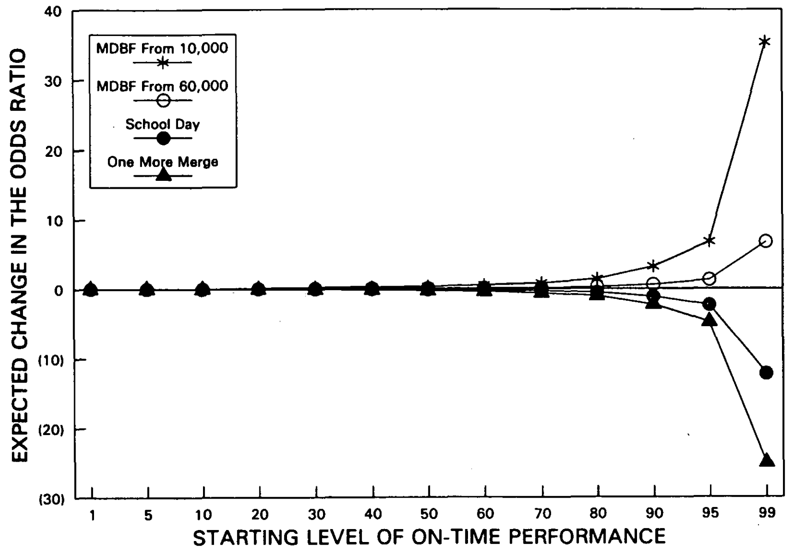
USING CAUSAL MODEL TO EVALUATE ROUTE PERFORMANCE

The results in Table 1 provide an equation for predicting OTP. These predictions, based on 1990 data and coefficients for that year, are given in Table 3. The intercept of 2.54 (a logit) translates into a base OTP of 92.7 percent. Each variable adds or subtracts from this intercept. For example the downtown no. 3 route has one merge before the CBD, which changes the logit to 2.303 (OTP 90.9 percent). Actual and predicted OTPs can be compared to see whether a route performed to expectations and to compare routes in terms of the variance between actual and predicted OTPs. Twenty-one of 33 routes or directions were predicted within 5 percentage points. The model had the hardest time predicting routes with high OTPs. This suggests that there are important quantitative factors—especially positive factors—that are not represented in the model.

Table 3 ranks each route or direction according to success against expectations. The J/Z route performed best. The downtown Q route was predicted to have the highest performance given its short length (to West 4th Street) and sparse ridership, but it finished 11th to 12th in terms of actual performance. Relative to its expected performance the downtown Q route was rated 31st out of 33. The downtown no. 5 and the uptown B routes were the most hopeful surprises. In terms of actual performance they were among the lowest 20th percentile, but given the merges and other obstacles experienced on those routes, they did better than expected.

CONCLUSIONS

Mechanical reliability has an obvious, positive impact on timeliness, and NYCTA's success in upgrading the subway car fleet had a significant effect on service throughout the period from 1988 to 1990. However further increases in MDBF will have diminishing



Source: Analysis of NYCTA morning rush hour subway service, 1988-90

FIGURE 3 Change in odds ratio.

returns in terms of OTP. This was predicted by the model, and recent NYCTA data support this conclusion.

A trade-off is involved in considering merges, because the merging structure provides beneficial routing alternatives when service disruptions occur. However the route merges that are so common to the New York City system hinder the delivery of timely service and constitute a major cause of delays according to the model. Each time a route merges, OTP may be as much as 7.2 percentage points lower (3 percentage points lower when OTP is at 90 percent). Because the negative impact of merges results from the mistiming of train arrivals at the merge point, better adherence to schedule can reduce the impact of this variable. The strong negative effect of merges on timeliness underscores the importance of careful scrutiny of route design plans that rely on additional merges. Even if significant increases in throughput and decreases in crowding can be achieved, the addition of merges can erode the benefits by impairing reliability.

Higher levels of ridership (holding the number of trains and all else constant) cause a decline in performance. As timeliness improved from 1988 to 1990, the negative effect of crowding on OTP became stronger. Although the effects of the other variables diminished, crowding emerged as a more serious problem. NYCTA measured a 2.1 percentage point improvement in rush hour OTP from 1990 to 1992, and the analysis suggests that one-fourth of this is because of lower levels of ridership.

Because the goal of NYCTA and transit advocates generally is to increase use of the subway, the adverse effect of increased crowding threatens to place constraints on service quality. Further improvement in subway service may be difficult as ridership levels increase from the recent slump, as they appear to be doing. Scheduling more trains can keep crowding levels constant on most routes. However on the most crowded routes more trains cannot be scheduled with current operating and safety rules and signalization. To make the matter more intractable, scheduling more trains makes headways smaller, and smaller headways are associated with lower OTPs. This analysis also suggests that reducing the number of trains (as is occasionally proposed in the interests of efficiency) may have unforeseen performance impacts, because the increase in headways from a service cut may be offset by an increase in crowding.

New subway cars being tested by NYCTA will carry more passengers and have design features that should allow passengers to board more quickly. If this helps reduce the times that trains spend at subway platforms, it could reduce crowding. Another way to increase throughput is to install new signal systems that permit more trains to operate in the peak interval.

A final possibility for increasing throughput to alleviate crowding is to change passenger behavior that produces unnecessarily long dwell times. Part of the strength of the crowding variable tested in the model results from passenger behavior. In a real sense passengers are members of the organization. Riders outside trains gather directly in front of doors, blocking exiting passengers, and riders inside cluster around doors, blocking entering passengers. Riders exiting from the middle of a car have tremendous difficulty making their way to the doors. Passengers hold doors for others. The clustering around doors by riders inside the train is being addressed by the design of the new cars; time will tell if this can help. Posters urge riders not to hold doors, but these focus on safety. NYCTA may need to consider public information campaigns that will inform riders how their actions delay trains.

Performance is likely to be slightly worse during official school days, another indication of the importance of passenger behavior. Nighttime construction lowered OTP in 1988 and 1990, but the 1990 result was not statistically significant, suggesting that improvements in operating procedures by NYCTA worked.

The research results show that relying on OTP to measure service quality may obscure a significant improvement once OTP has reached a high level. Odds ratios better reflect the improvement that passengers experience. Odds ratio show that the most significant improvement can be achieved, for a given amount of change in OTP, only after OTP has surpassed 90 percent. This is especially important in the context of state and local government fund-

TABLE 3 OTP Predictions and Comparisons for 1990 Morning Rush (7:00 to 9:00 a.m.)

RANK	ROUTE	DIRECTION	PREDICTION	ACTUAL	VARIANCE	ACTUAL RANK
1	J/Z	Downtown	87.3	98	10.7	1
2	N	Downtown	83.4	93	9.6	3-4
3	No.3	Downtown	83.5	92	8.5	5-7
4	No.7(loc)	Downtown	82.1	90	7.9	9
5	M	Downtown	88.2	96	7.8	2
6	No.3	Uptown	85.7	93	7.3	3-4
7	No.7(exp)	Downtown	82.0	89	7.0	10
8	No.5	Uptown	78.6	85	6.4	16-18
9	R	Uptown	85.5	91	5.5	8
10	C	Uptown	78.9	84	5.1	19-23
11	No.5	Downtown	77.1	82	4.9	26-28
12	B	Uptown	77.7	82	4.3	26-28
13	No.2	Uptown	84.2	88	3.8	11-12
14	Q	Uptown	82.3	86	3.7	13-15
15	B	Downtown	82.4	86	3.6	13-15
16	No.2	Downtown	76.4	80	3.6	29-30
17	No.1	Downtown	89.0	92	3.0	5-7
18	D	Downtown	73.2	76	2.8	32
19	L	Uptown	89.3	92	2.7	5-7
20	N	Uptown	83.4	86	2.6	13-15
21	No.4	Downtown	81.9	84	2.1	19-23
22	D	Uptown	82.9	85	2.1	16-18
23	R	Downtown	82.8	84	1.2	19-23
24	No.4	Uptown	82.8	84	1.2	19-23
25	A	Uptown	82.4	83	0.6	24-25
26	E	Downtown	76.6	77	0.4	31
27	A	Downtown	85.8	84	-1.8	19-23
28	F	Uptown	87.4	85	-2.4	16-18
29	No.6	Downtown	84.4	82	-2.4	26-28
30	M	Uptown	85.4	83	-2.4	24-25
31	Q	Downtown	90.9	88	-2.9	11-12
32	C	Downtown	85.7	80	-5.7	29-30
33	F	Downtown	76.0	70	-6.0	33

RANK shows the ranking of routes from best to worst, i.e., in the order given by "VARIANCE" (the difference between actual and predicted performance). ACTUAL RANK gives the ranking of routes according to their actual 1990 performance, as measured by the OIG (column called 'ACTUAL').

Source: ACTUAL OTP was calculated by the OIG; it is not an official NYCTA statistic. Predictions were made using the logistic regression coefficients listed in Table 1 for 1990.

ing decisions. Legislators may believe that a system with an OTP of 90 percent has already achieved an excellent level of service and that additional investment is not needed because it will not improve OTP by much.

OTP for each route was predicted by using the coefficients estimated by the model and compared the predictions with actual OTP. Assuming that a model that more accurately predicts OTP can be made, managers can use its predictions in several ways. (a) Routes ranked at the bottom—which perform worse than expected—can be targeted for managerial initiatives to improve performance. (b) Line superintendents can be given realistic performance improvement targets on the basis of the variances calculated in this way. NYCTA asked line superintendents to raise OTP by 1 percentage point in 1993. That may be unrealistic for some routes, and others may be expected to do better. (c) Employees on routes that have severe handicaps—for example, the oldest equipment, many merges, and crowding—may be assigned pay differentials in proportion to difficulty to attract the most experienced and capable workers and managers.

ACKNOWLEDGMENT

This research was funded in large part by the Federal Transit Administration (Section 8 Technical Studies grant NY-08-0185).

REFERENCES

1. *The Inspector General's Evaluation of Morning Rush Hour Subway Performance, 1987–1988.* New York State Inspector General for the Metropolitan Transportation Authority, Aug. 14, 1990.

2. *Commuting to Work by Subway: The Inspector General's Review of Morning Rush Hour Subway Performance.* New York State Inspector General for the Metropolitan Transportation Authority, Dec. 10, 1991.
3. Hosmer, D. W., and S. Lemeshow. *Applied Logistic Regression.* John Wiley & Sons, Inc., New York, 1989.
4. Hanushek, E. A. and J. E. Jackson. *Statistical Methods for Social Scientists.* Academic Press, Inc., New York, 1977.
5. *SAS/STAT User's Guide, Version 6*, 4th ed., Vol. 1. SAS Institute, Inc., 1989.
6. *Transit Authority Committee Agenda.* New York City Transit Authority. Oct. 1993.
7. Henderson, G., H. Adkins, and P. Kwong. Subway Reliability and the Odds of Getting There On-Time. Public Transit Research: Management and Planning 1991. In *Transportation Research Record 1297*, TRB, National Research Council, Washington, D.C., 1991.

The findings reflect the views of the authors and not necessarily the official views or policies of FTA or the New York Metropolitan Transportation Council (which administered the grant).

Publication of this paper sponsored by Committee on Rail Transit Systems.

Survey of Fare Policies at Large Transit Systems

DENNIS HINEBAUGH AND DANIEL K. BOYLE

The Center for Urban Transportation Research at the University of South Florida designed and conducted a national survey of fare policies for the Metro-Dade Transit Agency (MDTA). The purpose of the survey was to determine fare policies at other large transit agencies nationwide, including fare levels by mode, transfer charges, treatment of intermodal fares, discounts for multitrip purchases, pricing of monthly passes, time of day differentials, and distance-based fares. Seventeen of the 20 systems (including MDTA) responded to the survey. The range of fare policies is summarized by category for the 17 transit agencies that responded. The implications of the results are also discussed.

This paper presents the results of a national survey of fare policies at large transit agencies and has been prepared as part of a larger project undertaken by the Center for Urban Transportation Research (CUTR) to develop a long-term fare policy for the Metro-Dade Transit Agency (MDTA) in Dade County, Florida. The purpose of this survey was to determine fare policies at other large transit agencies nationwide, including transfer charges, treatment of intermodal fares, discounts for multitrip purchases, pricing of monthly passes, time of day differentials, and distance-based fares.

Before the survey CUTR reviewed the American Public Transit Association (APTA) 1991 fare summary (*1*), which by the time of the survey was outdated and did not contain all of the necessary information. Shortly after the survey was conducted APTA published its new and more detailed 1993 fare summary (*2*). The survey results provide a greater level of specificity in certain areas, particularly in the comments offered by respondents.

FARE POLICY SURVEY

In February 1993 CUTR contacted 20 of the largest transit systems in the United States to request information on fare policies. Systems were then mailed a Fare Policy Survey form. Seventeen of the 20 systems (including MDTA) responded to the survey. The following sections summarize the range of fare policies by category from the 17 responding transit agencies. A copy of the fare survey is contained in the full report (*3*), which also includes summaries of each transit system surveyed, pertinent fiscal year 1991 Section 15 information, and additional fare information.

Local Bus Fares

Local bus full fares range from a low of $0.40 in San Antonio to a high of $1.50 in Philadelphia, as presented in Table 1 and Figure 1. The average fare is $1.06. Discounted fares for elderly individuals range from free (off-peak) in Philadelphia and Pittsburgh to $0.60 in Atlanta, Chicago, MDTA, and New York, for an average of $0.35. The average fare for disabled individuals is $0.44, with a low of $0.15 in Boston and a high of $0.75 in Philadelphia. Fares for students range from $0.20 in San Antonio to $1.50 in Philadelphia, with an average of $0.73.

Express Bus Fares

Express bus service is operated in 15 of the 17 systems in the group of survey respondents. Table 2 and Figure 2 present the express bus fares at these 15 systems. Full fares range from a low of $0.75 in San Antonio to a high of $4.00 in New York, with a group average of $1.54. As shown in Figure 3 the ratio of express to local bus full fares averages 1.45, with a range of between 1.0 in four cities and 3.2 in New York. Fares for elderly individuals range from free (off-peak) in Pittsburgh to $2.00 in New York, with an average of $0.66. Fares for disabled individuals range from $0.25 in San Francisco and Dallas to $2.00 in New York, with an average of $0.71. Fares for students show an average of $1.14 and a range from $0.25 in San Francisco and Dallas to $4.00 in New York.

Heavy Rail Fares

Metrorail service in Dade County is categorized as heavy rail. Heavy rail service is operated in 10 of the 17 systems in the group of survey respondents. Heavy rail full fares are included in Table 3 and Figure 4 and range from a low of $0.85 in Boston to a high of $1.50 in Chicago, Cleveland, and Philadelphia, with a group average of $1.25. Rail full fares in Boston, Chicago, and Cleveland are higher than local bus fares. Fares for elderly individuals range from free in Philadelphia to $0.75 in Chicago, with an average of $0.47. Fares for disabled individuals range from $0.20 in Boston to $0.75 in Chicago and Philadelphia, with an average of $0.54. Fares for students range from $0.40 in Boston to $1.50 in Philadelphia, with an average of $0.97.

Light Rail Fares

Light rail service is operated in 10 of the 17 systems. Dade County does not have a light-rail system. Table 4 and Figure 5 present the light rail fare information. Light rail full fares range from a low of $0.85 in Boston to a high of $1.50 in Cleveland and Philadelphia, for a group average of $1.14. Fares for elderly individ-

---

Center for Urban Transportation Research, College of Engineering, University of South Florida, 4202 East Fowler Avenue, ENG 118, Tampa, Fla. 33620.

TABLE 1 Local Bus Fares

City/System	Full Fare	Elderly	Student	Disabled
Atlanta	$1.25	$0.60	$1.25	$0.60
Baltimore	$1.25	$0.45	$0.85	$0.45
Boston	$0.60	$0.15	$0.30	$0.15
Chicago	$1.25	$0.60	$0.60	$0.60
Cleveland	$1.25	$0.50	$1.00	$0.50
Dade County	$1.25	$0.60	$0.60	$0.60
Dallas	$0.75	$0.15	$0.25	$0.25
Los Angeles	$1.10	$0.45	$1.10	$0.45
New Jersey	$1.00	$0.45	$0.45	$0.45
New York	$1.25	$0.60	$0.60	$0.60
Philadelphia	$1.50	*$0.00	$1.50	*$0.75
Pittsburgh	$1.25	*$0.00	$1.25	*$0.60
Portland	$0.95	$0.45	$0.70	$0.45
San Antonio	$0.40	$0.20	$0.20	$0.20
San Francisco (MUNI)	$1.00	$0.25	$0.25	$0.25
San Jose	$1.00	$0.25	$0.50	$0.25
Washington D.C.	$1.00	$0.30	$1.00	$0.30
Average of 17 Systems	**$1.06**	**$0.35**	**$0.73**	**$0.44**

* Off-peak only

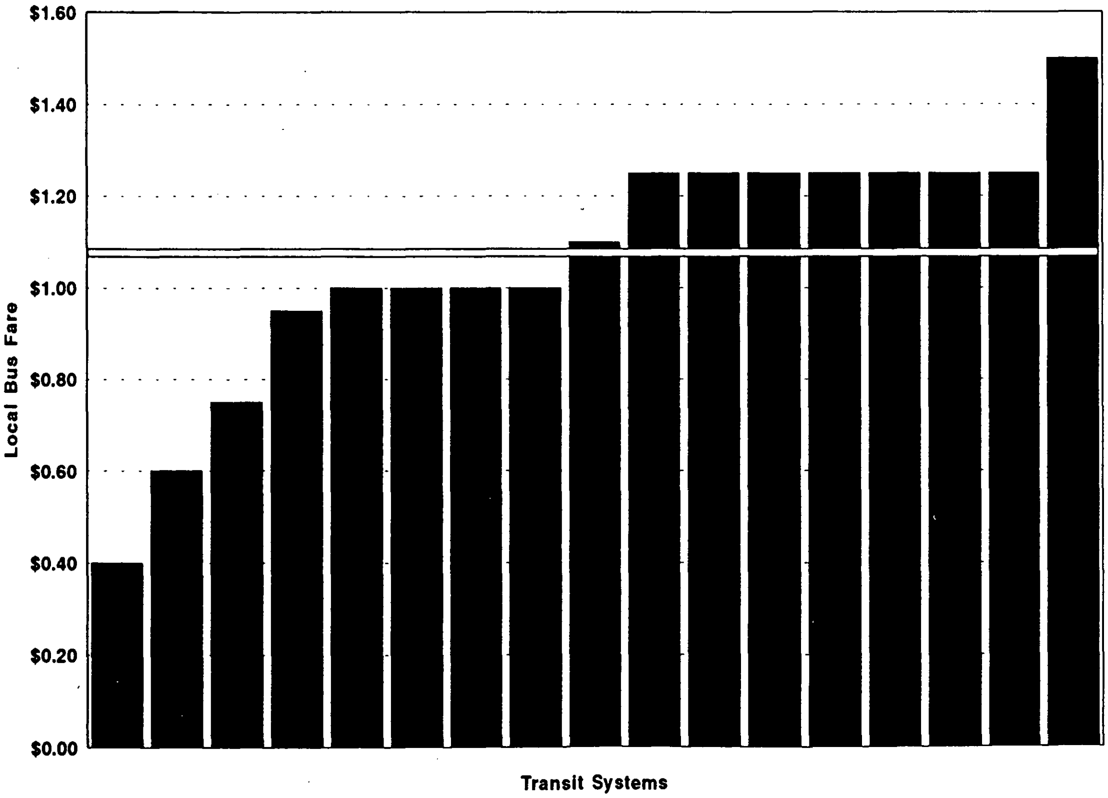


FIGURE 1 Local bus full fares.

TABLE 2 Express Bus Fares

City/System	Full Fare	Elderly	Student	Disabled	Express to Local Full Fare Ratio
Atlanta	$1.25	$0.60	$1.25	$0.60	1.0
Baltimore	$1.55	$0.75	$1.55	$0.75	1.24
Boston	$1.50	$0.75	$0.75	$0.75	2.5
Chicago	$1.50	$0.85	$0.85	$0.85	1.2
Cleveland	$1.50	$0.50	$1.00	$0.50	1.2
Dade County	$1.50	$0.75	$0.75	$0.75	1.2
Dallas	$1.75	$0.15	$0.25	$0.25	2.3
Los Angeles	$1.10	$0.45	$1.10	$0.45	1.0
New Jersey	$1.25	$0.55	$0.55	$0.55	1.25
New York	$4.00	*$2.00	$4.00	*$2.00	3.2
Pittsburgh	$1.25	*$0.00	$1.25	*$0.60	1.0
San Antonio	$0.75	$0.35	$0.35	$0.35	1.88
San Francisco (MUNI)	$1.00	$0.25	$0.25	$0.25	1.0
San Jose	$1.50	$1.50	$1.50	$1.50	1.5
Washington D.C.	$1.65	$0.50	$1.65	$0.50	1.65
Average of 15 Systems	**$1.54**	**$0.66**	**$1.14**	**$0.71**	**1.45**

* Off-peak only

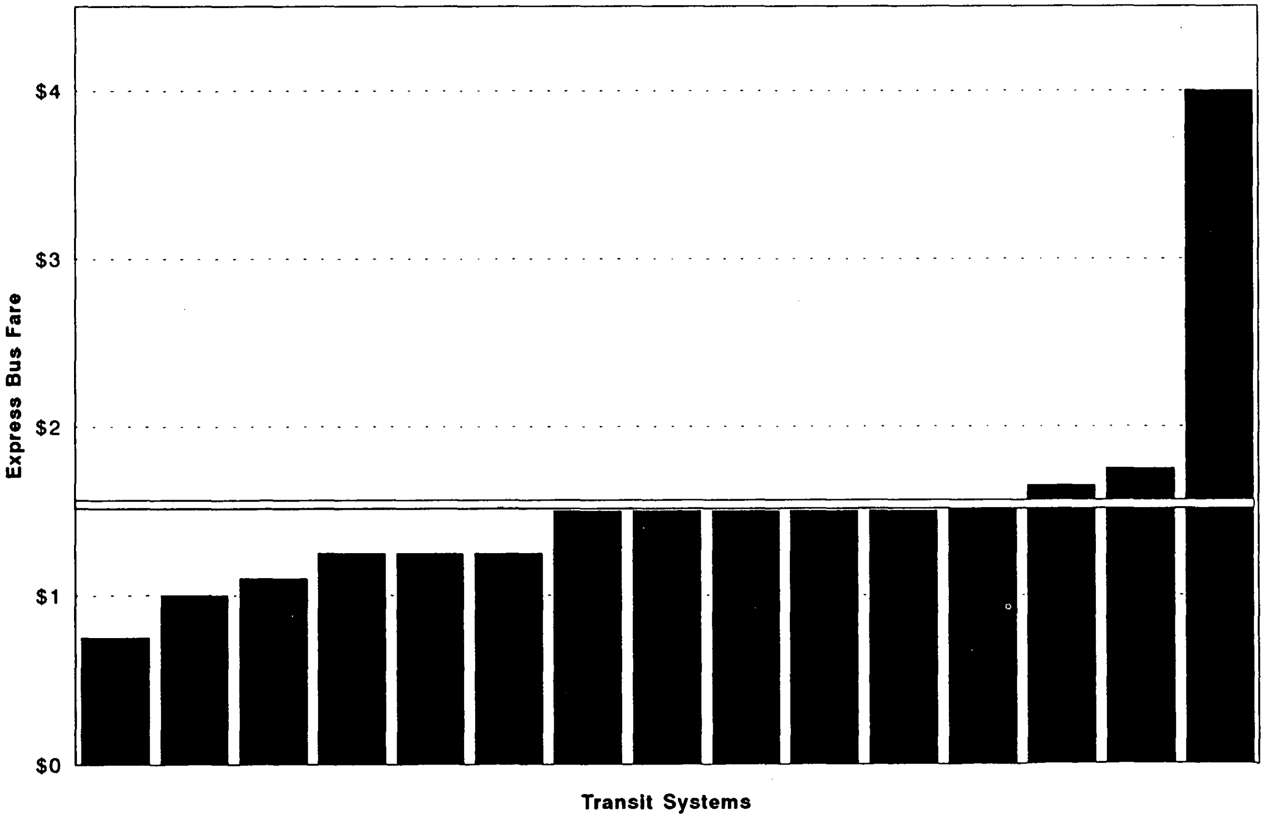


FIGURE 2 Express bus full fares.

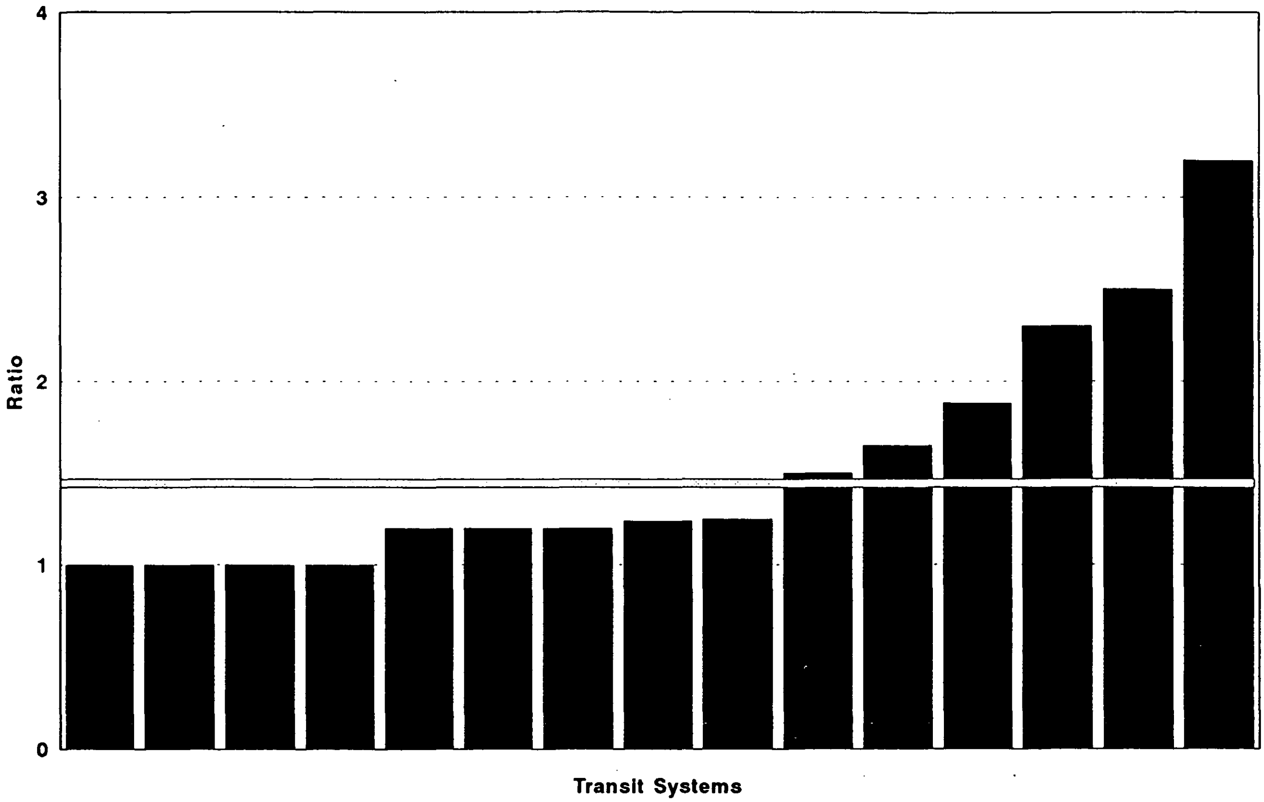


FIGURE 3 Express to local full fare ratio.

TABLE 3 Heavy Rail Fares

City/System	Full Fare	Elderly	Student	Disabled
Atlanta	$1.25	$0.60	$1.25	$0.60
Baltimore	$1.25	$0.45	$0.85	$0.45
Boston	$0.85	$0.20	$0.40	$0.20
Chicago	$1.50	$0.75	$0.75	$0.75
Cleveland	$1.50	$0.50	$1.00	$0.50
Dade County	$1.25	$0.60	$0.60	$0.60
Los Angeles	$1.10	$0.45	$1.10	$0.45
New York	$1.25	+$0.63	+$0.63	+$0.63
Philadelphia	$1.50	*$0.00	$1.50	*$0.75
Washington D.C.	$1.00	$0.50	$1.00	$0.50
Average of 10 Systems	**$1.25**	**$0.47**	**$0.97**	**$0.54**

* Off-peak only
\+ $1.25 fare includes return trip

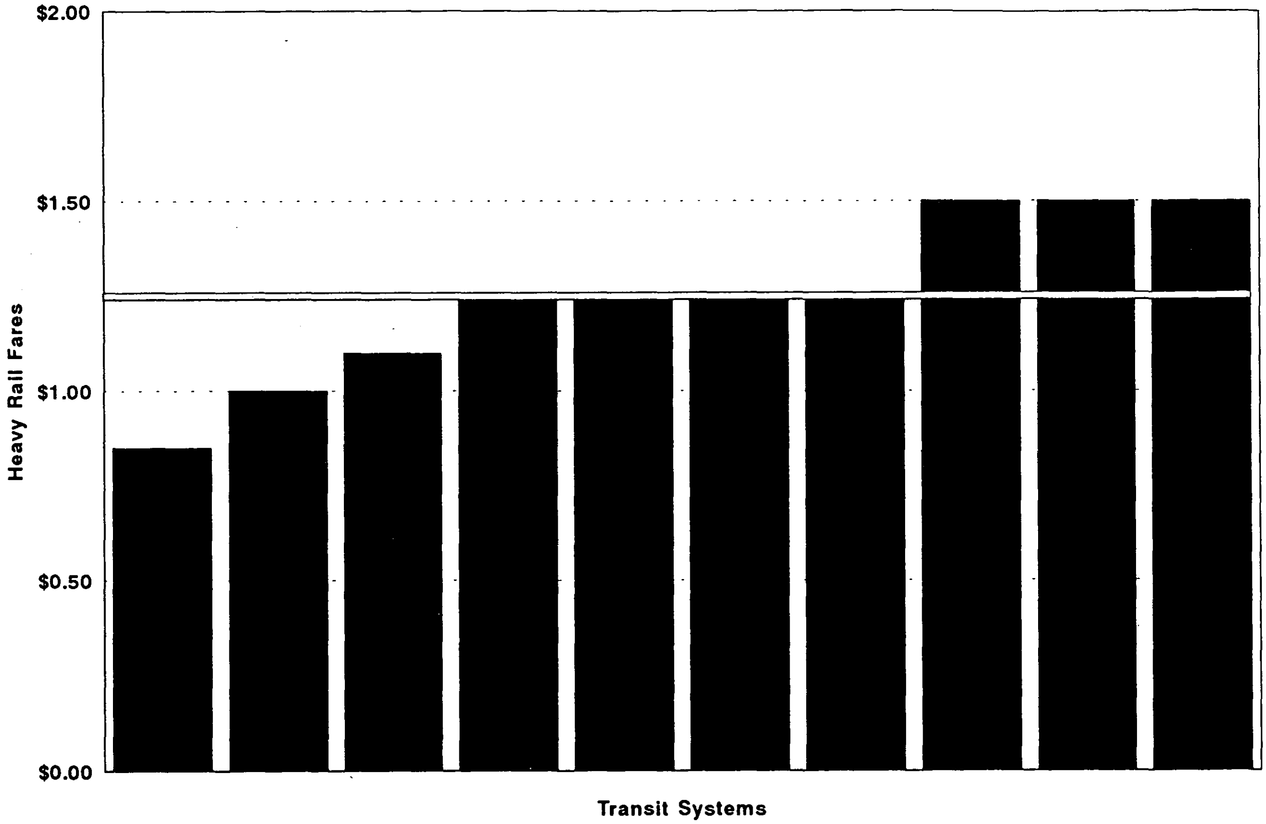


FIGURE 4 Heavy rail full fares.

TABLE 4 Light Rail Fares

City/System	Full Fare	Elderly	Student	Disabled
Baltimore	$1.25	$0.45	$0.85	$0.45
Boston	$0.85	$0.20	$0.40	$0.20
Cleveland	$1.50	$0.50	$1.00	$0.50
Los Angeles	$1.10	$0.55	$1.10	$0.55
New Jersey	$1.00	$0.45	$0.45	$0.45
Philadelphia	$1.50	*$0.00	$1.50	*$0.75
Pittsburgh	$1.25	*$0.00	$1.25	*$0.60
Portland	$0.95	$0.45	$0.70	$0.45
San Francisco (MUNI)	+$1.00	+$0.25	+$0.25	+$0.25
San Jose	$1.00	$0.25	$0.50	$0.25
Average of 10 Systems	**$1.14**	**$0.31**	**$0.80**	**$0.45**

* Off-peak only

\+ Cable Car full fare is $3.00; elderly/disabled/student fare is $1.00; monthly pass is valid for cable car full fare

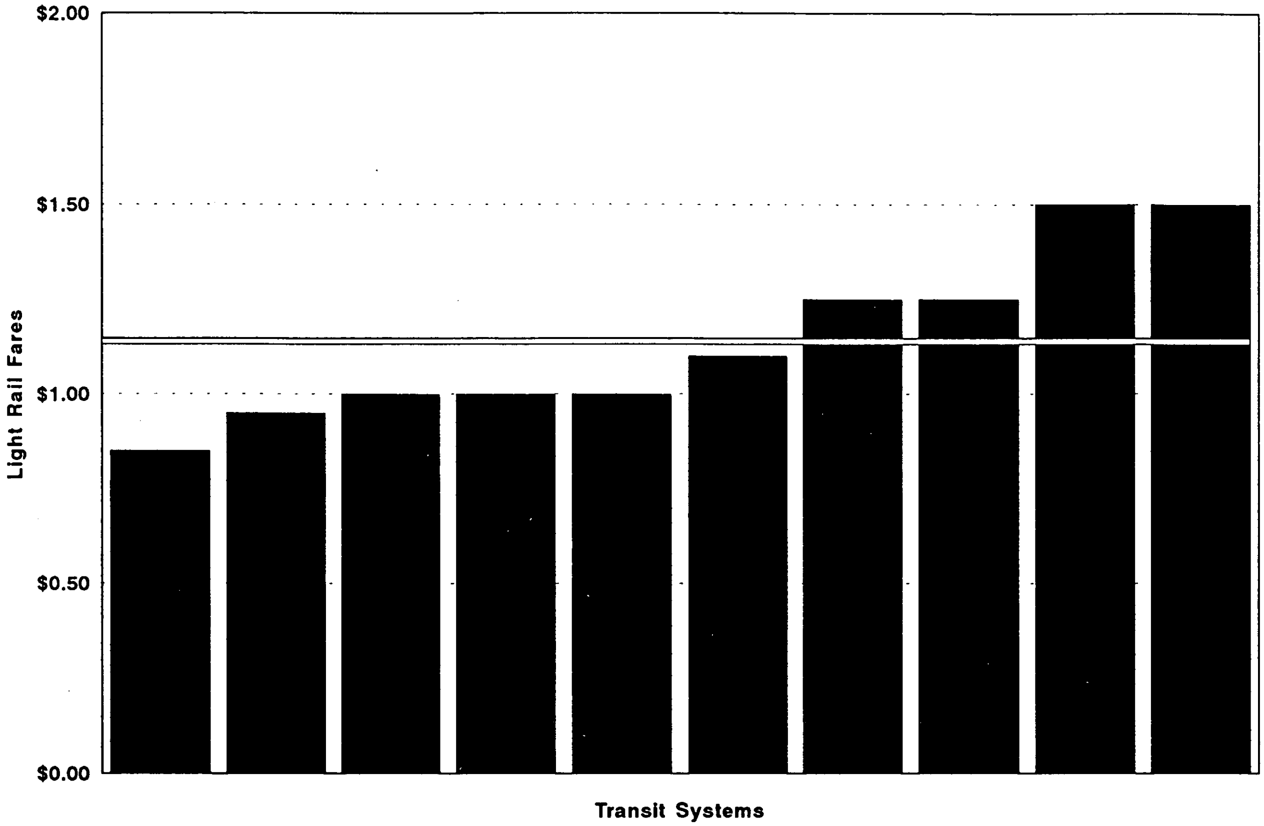


FIGURE 5 Light rail full fares.

uals range from free in Philadelphia and Pittsburgh to $0.55 in Los Angeles County, for an average of $0.31. Fares for disabled individuals range from $0.20 in Boston to $0.75 in Philadelphia, for an average of $0.45. Fares for students are lowest ($0.25) in San Francisco and highest ($1.50) in Philadelphia, with an average of $0.80. Light rail fares match either local bus fares in systems without heavy rail or the heavy rail fare.

Peak Differential

Five of the 17 transit systems currently have a peak surcharge. Chicago has surcharges of $0.25 (full fare) and 0.15 (other fare categories) only on its bus system in the peak, resulting in a $1.50 peak fare on both bus and rail. Philadelphia has a surcharge on full fares in the peak on only its light rail system ($0.25 for zone 1; $0.40 for zone 2). Elderly and disabled passengers must pay full fare in the peak in both Philadelphia and Pittsburgh. San Jose provides a midday discount for full fare bus and light rail passengers, with a midday fare of $0.50. In Washington, D.C., the number of zones traveled for the base fare is restricted to three in the peak period. Peak surcharges range from a low of $0.15 for elderly and disabled individuals and students in Chicago to a high of $1.50 for elderly patrons in Philadelphia.

Transfers

Transfers are currently free on 9 of the 17 systems, as shown in Table 5 and Figure 6. At those systems that charge a transfer the cost ranges from $0.25 at three systems (including MDTA) to the full fare rate (no transfers given) in Boston.

Monthly Passes

Of the survey respondents, only New York and Washington, D.C., do not currently offer monthly passes, although Washington, D.C., is in the process of introducing a 28-day rail-only pass priced at $100. Table 6 and Figure 7 present the monthly pass costs. Full fare monthly passes range from a low of $16.00 for local bus in San Antonio to a high of $78.00 in Chicago. The average cost for a full fare local bus pass (not including zone charges) is $40.07 and $13.88 for a discounted pass (not shown in Table 6). The average cost for a full fare light rail pass is lower ($36.90), whereas express bus and heavy rail pass costs are higher on average.

Table 6 and Figure 8 provide information on the breakeven point for monthly pass purchasers. This is the number of trips at which the cost of the monthly pass equals the sum of the cost of single fares. The average breakeven point is 37 trips per month, with a low of 30 trips in San Jose and a high of 52 trips in Chicago.

Other Passes

Most systems offer passes other than their monthly pass. These passes include weekly passes (seven systems), two-week passes

TABLE 5 Transfer Charges

City/System	Full Fare	Discount
Atlanta	Free	Free
Baltimore	$0.10	Free
Boston	Additional Fare	Additional Fare
Chicago	$0.30	$0.15
Cleveland	Free	Free
Dade County	$0.25	$0.10
Dallas	Free	Free
Los Angeles	$0.25	$0.10
New Jersey	$0.45	
New York	*Free	*Free
Philadelphia	$0.40	$0.40
Pittsburgh	$0.25	$0.10
Portland	Free	Free
San Antonio	Free	Free
San Francisco (MUNI)	Free	Free
San Jose	Free	Free
Washington D.C.	+Free	+Free
Average of 17 Systems	**$0.15**	**$0.05**

* Free transfers between same modes and between bus and subway at a limited number of stations; otherwise additional full fare is required

\+ Within District of Columbia

(three systems), punch passes (one system), daily-only passes (five systems), weekday-only passes (three systems), weekend-only passes (two systems), student or college student passes (four systems), and annual passes (two systems). Weekly pass prices (with no additional zone charges) range from $11.00 in Atlanta, Baltimore, and Pittsburgh to $20.50 in Chicago. The range of costs for daily passes is from $2.00 (San Antonio and San Jose) to $5.00 (Boston).

Tokens

Eleven of the systems sell tokens for fare payment, as illustrated in Table 7. In New York and Boston tokens are not discounted. Discounts on tokens range from approximately 5 percent in Atlanta and Baltimore to a high of 30 percent in Philadelphia. Only Chicago offers a token for elderly and disabled individuals and students.

Tickets

Seven of the systems sell tickets as a method of fare payment. These are also shown in Table 7. Dallas does not offer a discount on their tickets. Ticket discounts range from 5 to 10.5 percent.

Distance-Based Fares

Ten of the 17 systems (Baltimore, Boston, Cleveland, Los Angeles, New Jersey, Philadelphia, Pittsburgh, Portland, San Antonio, and Washington, D.C.) currently have distance-based surcharges. These surcharges vary by mode of service and by the number of zones traveled. Some zonal boundaries match political jurisdictions (i.e., county lines), although the size and number of zones also vary.

Magnetic Fare Cards

Six of the systems (Atlanta, Boston, Chicago, Miami, Philadelphia, and Washington, D.C.) currently use magnetic fare cards. Cleveland planned to start using magnetic fare cards later in 1993. All of the magnetic fare card systems are multimodal, but it is unclear from the survey responses whether the fare cards can be used on all modes.

Credit Card Purchases

Eight of the systems (Baltimore, Boston, Dallas, Los Angeles, New Jersey, Philadelphia, Pittsburgh, and San Jose) accept credit cards and checks for the purchase of fare instruments (passes, tokens, tickets).

Average Fare Per Unlinked Trip

The average fare per unlinked trip is derived from Section 15 statistics for fiscal year 1991 (*4*). The average fare is calculated by dividing passenger fare revenue by unlinked trips on all modes of service. Data for New Jersey and Philadelphia are not included here because their Section 15 passenger fare data incorporate commuter rail, which inflates the average fare. Table 8 and Figure 9 present information on average fares. The average for 15 systems is $0.52, with a high of $0.75 in New York and a low of $0.28 in San Antonio.

Fare Box Recovery Ratio

The 1991 Section 15 report is also the source of the fare box recovery ratio (*4*). The ratio is obtained by taking the passenger fare revenue as a percentage of operating expenditures on all modes of service. Data for New Jersey and Philadelphia are excluded because of the inability to separate commuter rail data in the Section 15 report. As shown in Table 8 and Figure 10, fare box recovery ratio for the 15 systems ranges from 11 percent in San Jose to 48 percent in Washington, D.C., with an average of 32 percent.

SURVEY OBSERVATIONS AND IMPLICATIONS

The information obtained in the survey has been helpful to MDTA in evaluating its system with respect to similar transit systems nationwide. The findings highlighted below are also of interest beyond Dade County.

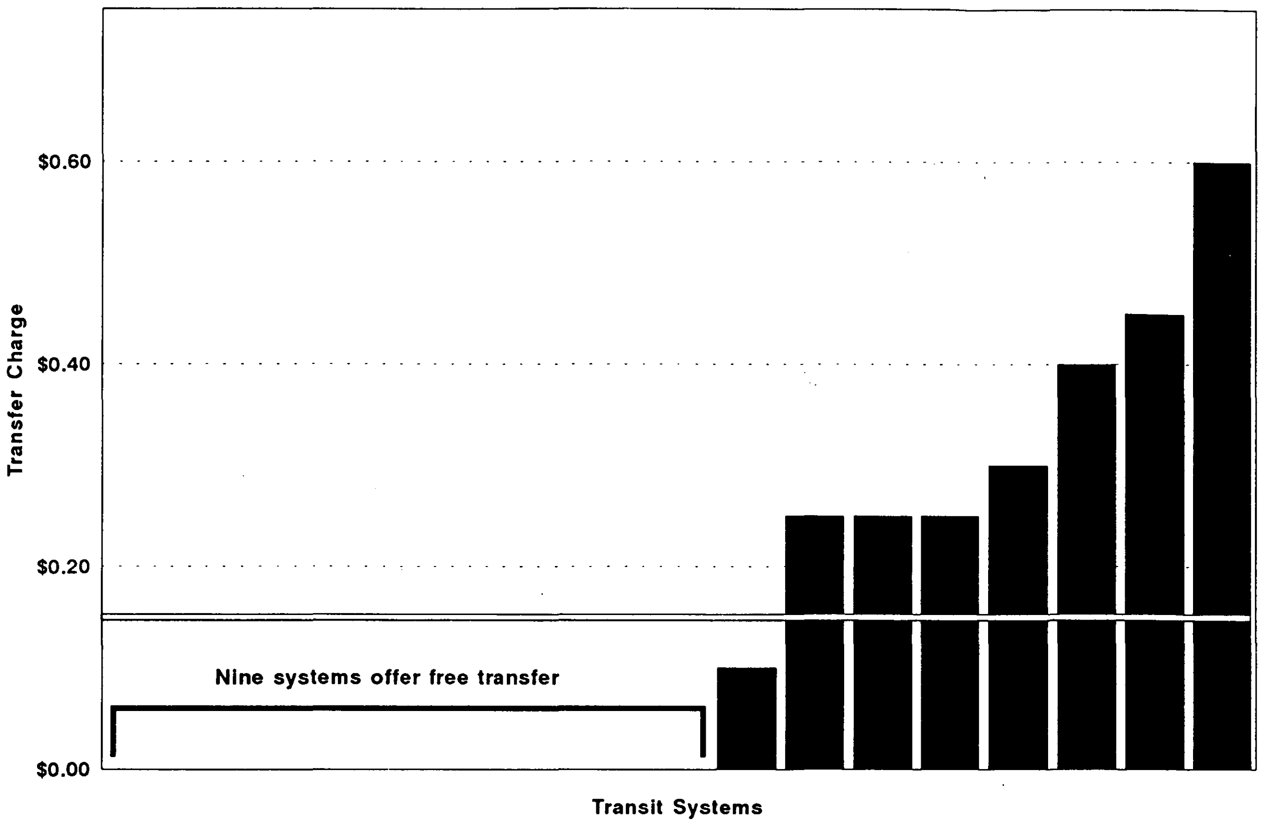


FIGURE 6 Full fare transfer charges.

TABLE 6 Monthly Pass Costs

City/System	Local Bus	Express Bus	Heavy Rail	Light Rail	Break-even Number of Trips (Local)
Atlanta	$43.00	$43.00	$43.00	--	34
Baltimore	$42.00*	$52.00*	$42.00*	$42.00*	34
Boston	$20.00	$64.00	$27.00	$27.00	32
Chicago	$78.00	$78.00+	$78.00	--	52
Cleveland	$45.00	$54.00	$54.00	$54.00	36
Dade County	$60.00	$60.00	$60.00	--	48
Dallas	$23.00	$54.00	--	--	31
Los Angeles	$42.00	$42.00	$42.00	$42.00	38
New Jersey	$41.00*	$59.00	--	$41.00*	41
Philadelphia	$58.00	--	$58.00	$58.00	39
Pittsburgh	$40.00	$40.00	--	$40.00	32
Portland	$31.00	--	--	$31.00	33
San Antonio	$16.00	$30.00	--	--	40
San Francisco (MUNI)	$32.00	$32.00	--	$32.00#	32
San Jose	$30.00	$45.00	--	$30.00	30
Average of all Systems	**$40.07**	**$50.23**	**$50.50**	**$36.90**	**37**

*	Base zone only; zone charges apply to monthly pass price
+	Plus $0.25 surcharge per boarding
#	Monthly pass valid on cable car

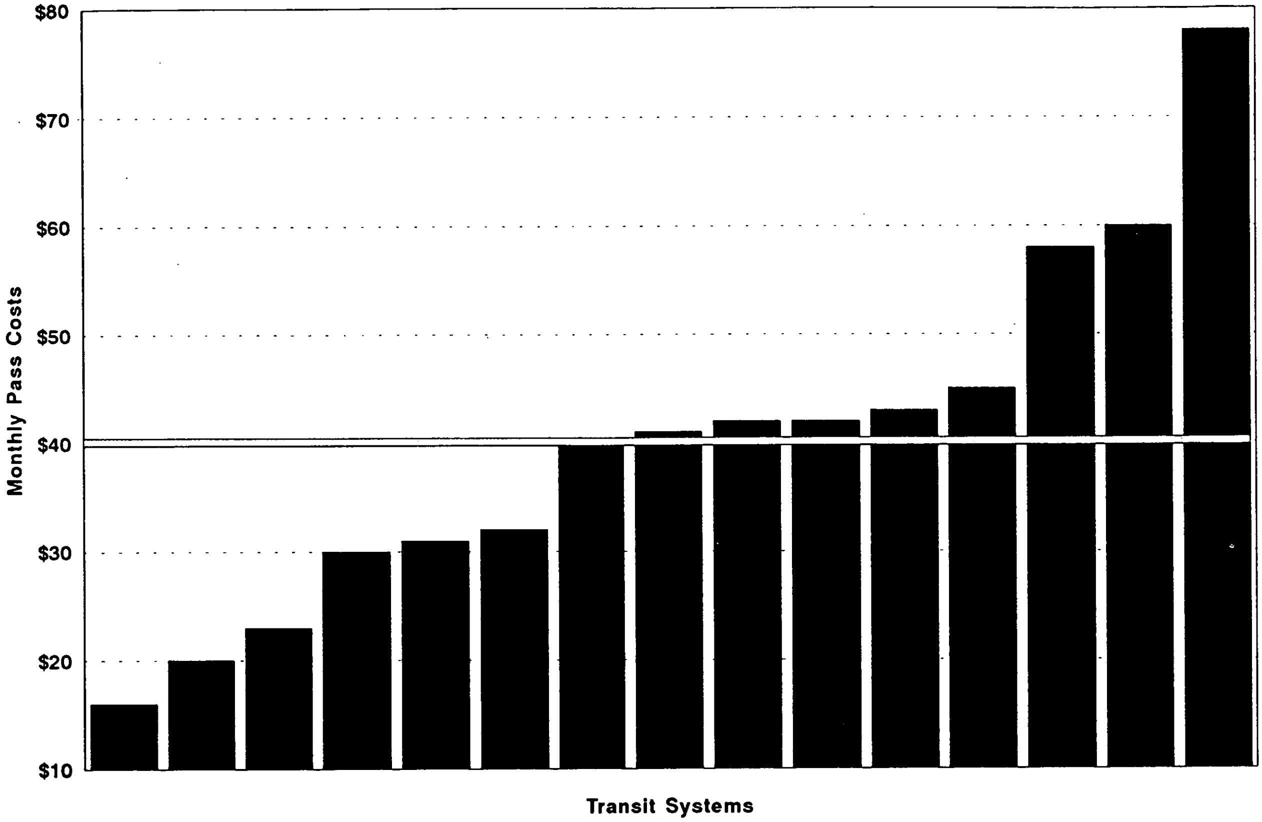


FIGURE 7 Local bus monthly pass costs.

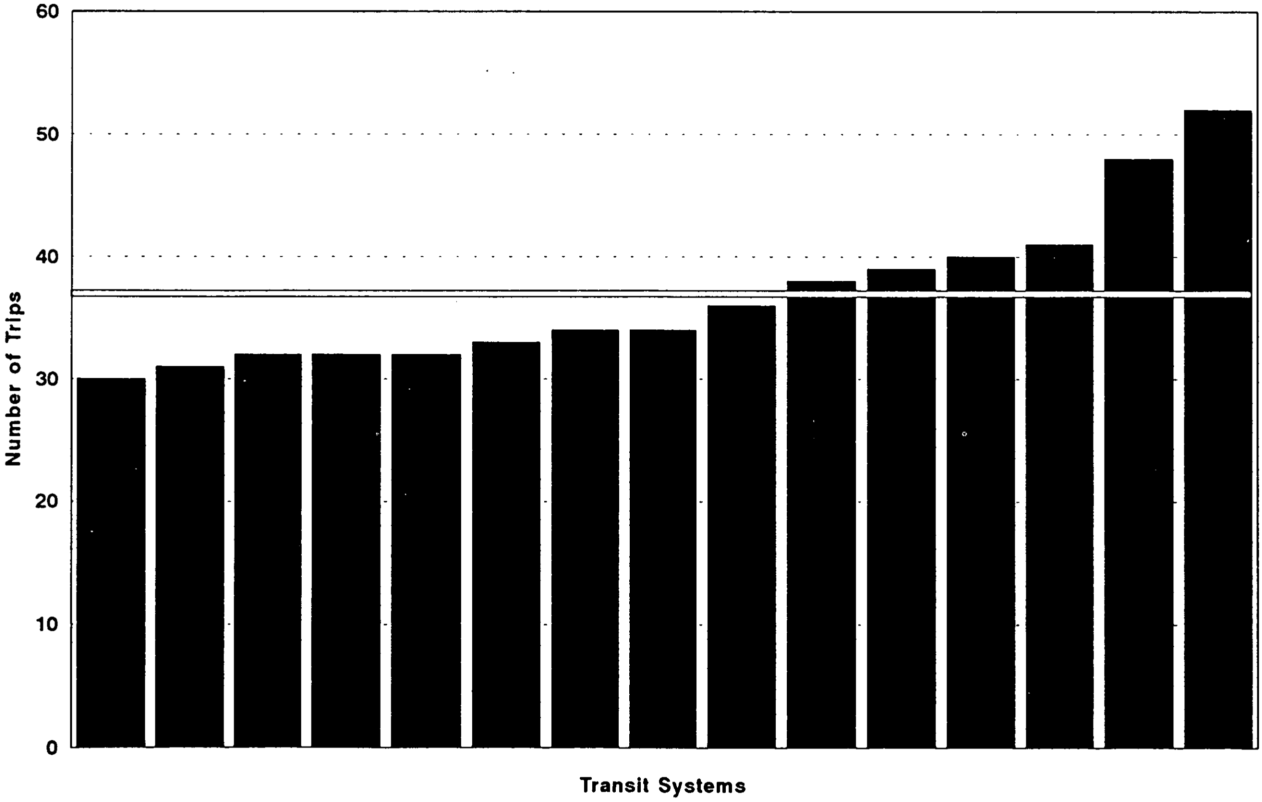


FIGURE 8 Breakeven number of trips (local bus service).

TABLE 7 Token and Ticket Costs

City/System	Token	% Discount	Ticket	% Discount
Atlanta	$1.20	4.0%	--	--
Baltimore	$1.20	4.0%	$1.20	4.0%
Boston	$0.85	0.0%	$0.71	16.4%
Chicago	$1.25	16.7%	--	--
Cleveland	--	--	*$1.19	4.8%
Dade County	$1.00	20.0%	--	--
Dallas	--	--	$0.75	0.0%
Los Angeles	$0.90	18.2%	--	--
New Jersey	--	--	Varies	Varies
New York	$1.25	0.0%	--	--
Philadelphia	+$1.05	+30.0%	--	--
Pittsburgh	--	--	$1.15	8.0%
Portland	--	--	$0.85	10.5%
San Francisco (MUNI)	$0.90	10.0%	--	--
Washington D.C.	$0.90	10.0%	--	--
Average of all Systems	**$1.05**	**11.3%**	**$0.98**	**7.3%**

Cost is average cost per trip

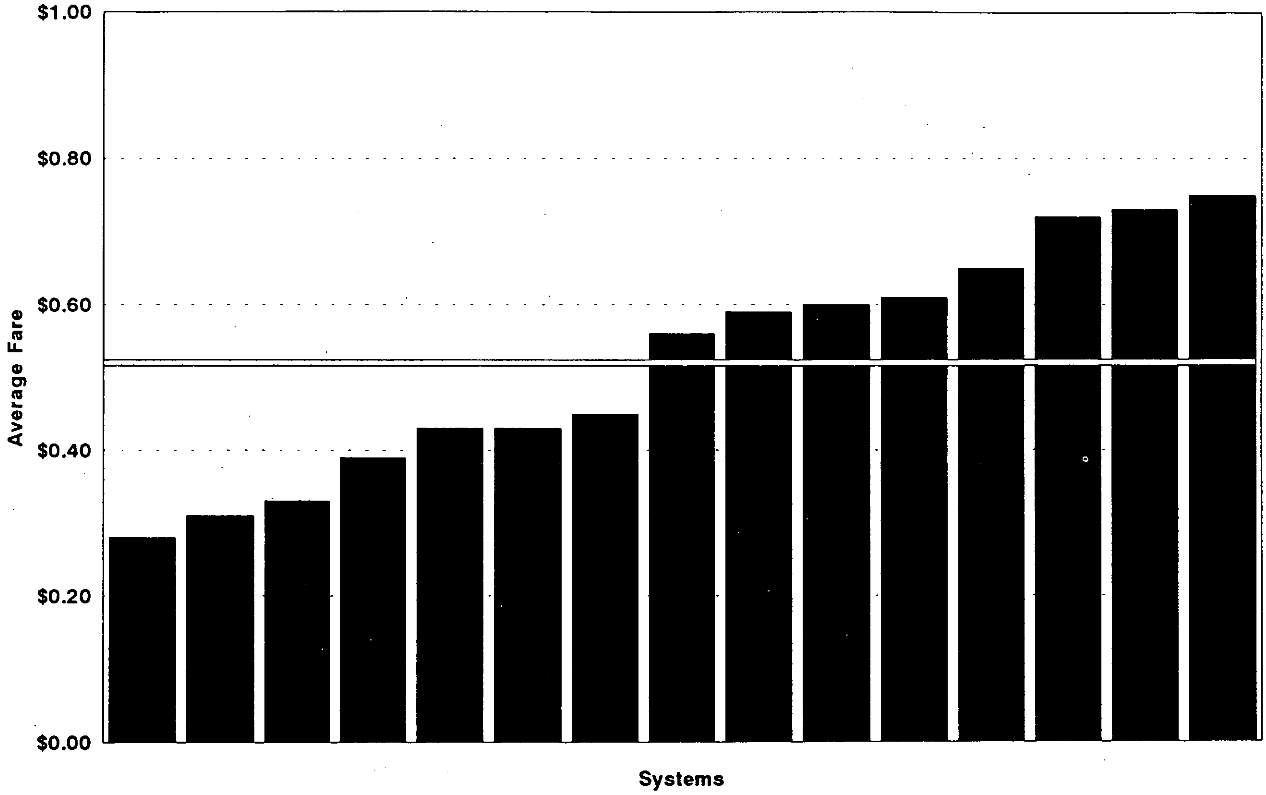
* $1.43 for rail (4.7% discount)

\+ Student only

TABLE 8 Average Fares Per Unlinked Trip and Fare Box Recovery Ratios

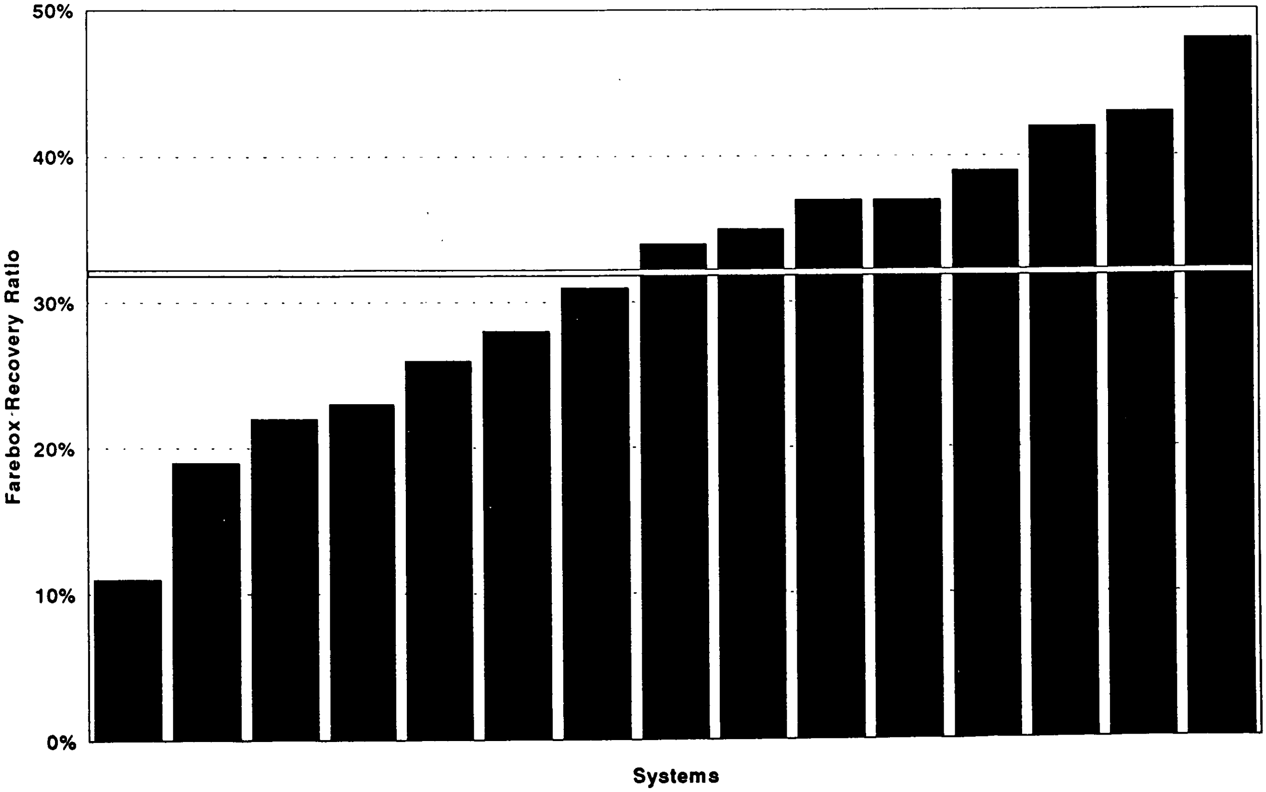
City/System	Average Fare per Unlinked Trip	Farebox Recovery Ratio
Atlanta	$0.43	37%
Baltimore	$0.61	37%
Boston	$0.39	22%
Chicago	$0.59	42%
Cleveland	$0.60	26%
Dade County	$0.72	34%
Dallas	$0.43	19%
Los Angeles	$0.56	39%
New York	$0.75	43%
Pittsburgh	$0.65	35%
Portland	$0.45	28%
San Antonio	$0.28	23%
San Francisco (MUNI)	$0.33	31%
San Jose	$0.31	11%
Washington D.C.	$0.73	48%
Average of 15 Systems	**$0.52**	**32%**

Source: Derived from 1991 Section 15 Statistics



Source: Derived from 1991 Section 15 Statistics

FIGURE 9 Average fare per unlinked trip.



Source: Derived from 1991 Section 15 Statistics

FIGURE 10 Fare box recovery ratio.

One interesting observation is that the concept of pricing transit service in accordance with its cost or its level of service has not been implemented on a large scale. Peak period surcharges are not generally assessed, despite the higher cost of operating peak service. Distance-based fares are more common but are far from universal. The fare differential between premium service (e.g., express bus) and regular service is less than might be expected.

Recent concepts that have developed among transit professionals in the past decade have also not been generally implemented. The extent to which significant discounts are offered for prepaid single-ride fare media (such as tokens or tickets) is not great, indicating that the deep-discount concept has not achieved widespread acceptance. At the same time, monthly passes continue to be priced at relatively low breakeven points, despite resulting revenue losses.

Technological advances and the spread of existing automated fare collection technology have the potential to affect fare policies. Improvements in fare collection methods certainly can affect the feasibility of certain fare options and the ease of administering complex fare structures. There is no evidence that fare collection technology determines fare policy. Essential policy questions have been and will continue to be decided independently of technological considerations.

What drives fare policy decisions? Although the fare survey has produced a considerable amount of technical information, it did not address the political realm. Political feasibility is perhaps the major determining factor in decisions on fare levels and overall fare policy. Low fares, simple fare structures, and a definable sense of equity (often translated to a flat fare structure) among socioeconomic groups and neighborhoods are likely to be higher political priorities than cost recovery issues. These political priorities are reflected in systems for which the existence of a dedicated local funding source allows a transit system to keep its fares low.

The mixture of fare structure issues, technological developments, and political priorities is not complete without a consideration of the extent of customer orientation on the part of a transit agency. Total quality management concepts are migrating from the private to the public sector, and many transit agencies are scrambling to become more customer oriented. Although pricing policies based on market segmentation with regard to fare sensitivity are undoubtedly efficient, these do not necessarily keep the customer as pleased as do low fares and a good deal on a monthly pass.

The fare survey results cannot address all these implications regarding fare policy. What the survey has provided is a clear indication of baseline conditions related to fare levels, policies, and innovations for the larger transit systems in the United States.

REFERENCES

1. *Transit Fare Summary*. American Public Transit Association, Washington, D.C., 1991.
2. *Transit Fare Summary*. American Public Transit Association, Washington, D.C., 1993.
3. *Metro-Dade Transit Agency Transit Fare Policy and Strategies, Technical Memorandum Number 1: Current and Historical MDTA Fare Policies/Fare Policies at Other Major Transit Systems*. Center for Urban Transportation Research, Tampa, Fla., July 1993.
4. *Transit Profiles: Agencies in Urbanized Areas Exceeding 200,000 Population*. FTA, Washington, D.C., Dec. 1992.

Publication of this paper sponsored by Committee on Public Transportation Marketing and Fare Policy.

Evaluation of Transit Telephone Information at SCRTD

RANDOLPH W. HALL, EMMANUEL LE COLLETTER, AND YOUNGBIN YIM

The way transit-related telephone information is provided and how it is perceived and used by callers to complete a planned trip are investigated. A test and survey are conducted among a sample of both transit and nontransit users in Los Angeles, using the Southern California Rapid Transit District (SCRTD) telephone information service. Participants were asked to request specific directions from either an operator or the synthesized voice system and to record the directions they received. A sample of the phone calls made to SCRTD was analyzed. The telephone information service appeared to be well received by the general public, especially by frequent transit riders. Telephone operators at SCRTD obtained high ratings in terms of friendliness, clarity of directions, speed of speech, and callers' confidence in the accuracy of directions. The synthesized voice system (ARTI) did not obtain the same high ratings as SCRTD's human operators. Clarity of directions and clarity of speech of the automated voice were rated low. Because the ability to interact was valuable to callers, the noninteractive synthesized voice system did not appear as effective as human operators in transmitting information to callers. Participants recorded less information and, on average, made more mistakes when interacting with ARTI than with the operators. The potential of a users telephone information service to generate more transit trips when the information is provided by a human operator appeared higher among users who are dependent on transit than among those who use transit infrequently. In that sense, the research established a link between the quality of the information provided and the propensity to use transit.

Southern California Rapid Transit District (SCRTD) is one of the largest urban transit operators in the United States. Created in 1964, it provides bus and light rail service in the greater Los Angeles region. Its 1,442-mi^2 service area covers Los Angeles County and large parts of Ventura, Orange, and San Bernardino counties. SCRTD carries 1.3 million riders in a weekday and 409 million annually (1991). The network includes more than 200 bus routes and one light rail line stretching 22 mi between downtown Los Angeles and Long Beach. SCRTD operates a fleet of 2,500 buses and 54 light rail trains. Since the completion of this study, subway service (red line) began operation in January 1993. Although SCRTD was merged into LACMTA in April 1993, the authors refer to it as SCRTD because the study was completed before the merger.

To help potential transit users find their way on this extended network, SCRTD provides a telephone information service offering itinerary, schedule, and fare information. The system operates from a computerized data base that includes routes of 24 transit operators in the Los Angeles area. Currently, 105 operators handle 12,000 calls daily, for a total of 4 million calls annually.

The objective of this study is to investigate the effects of transit telephone information on trip behavior. It explores the way transit information is recorded and used by callers and looks at the potential impact on their choice of transportation. It also describes the public's perception of telephone information services, those delivered by operators and those provided by SCRTD's synthesized voice system. Finally, the study examines a sample of telephone calls made to the service to gain an understanding of how callers interact with operators to obtain travel information.

LITERATURE REVIEW

This part of the paper presents an overview of the studies previously conducted on the topics of transit information, the way information services are perceived and used by transit users. The literature review includes a summary of the research done on the trip behavior of users who were provided with transit information.

Effectiveness of Transit Telephone Information Systems

Since the 1950s, the transit industry has faced fierce competition from the automobile, which severely eroded public transportation's demand base. Facing constant declines in ridership and ever-increasing operating costs, transit operators have tried marketing public transportation to reverse the downward trend.

Marketing efforts in the transit industry have included radio and television ads, maps and timetables, newspaper announcements, and telephone information services. The latter has distinct advantages over other media: it is easily accessible to most of the population; it enables information to be updated quickly to reflect changes in service; and more important it permits information to be tailored to the specific needs of individual customers.

Telephone information service is now considered an essential public service, and it is provided by almost every transit agency in North America. Traditionally, travel information was provided by agents who referred to a hard-copy data base consisting of street and route maps, schedule information, and daily bulletins of service modifications. This mode of information retrieval, although still widely in use, is extremely labor-intensive and therefore costly to operate. Starting in the 1970s, various efforts have been undertaken to computerize part of the operation to reduce cost and increase call-handling capacity.

In North America, the effort has taken two forms: one that focuses on computerized information storage and retrieval within

---

R. W. Hall, University of Southern California, School of Business Administration, Information and Operations Management Department, 400C Bridge Hall, Los Angeles, Calif. 90089. E. Le Colletter, University of California at Berkeley, Institute of Transportation Studies, Berkeley, Calif. 94720. Y. Yim, University of California at Berkeley, Institute of Transportation Studies, Richmond Field Station, 1301 S. 46th Street, Building 452, Richmond, Calif. 94804.

the traditional telephone information system (TIS) environment; the other way has been to develop fully automated schedule information to provide callers with stop-specific arrival times without involving agents. The former trend aimed at improving agent productivity and information consistency; the latter aimed at increasing ridership by reducing uncertainty associated with waiting at bus stops.

Computerized Data Retrieval Within the Existing TIS Environment

SCRTD and the Washington Metropolitan Area Transit Authority (WMATA) were among the first transit agencies to develop an automated data storage and retrieval system for telephone information purposes (*1*).

Los Angeles Experience The population of Los Angeles has been described as transit ignorant because of low transit use. Indeed, in spite of its large size, SCRTD accounts for only 4 percent of all trips in the Los Angeles area (*1*). Trip patterns in Los Angeles do not focus on the traditional CBD but are diffuse and complex. People have to travel very long distances and typically lack geographical knowledge of their destination point. In that context, the provision of telephone information services is especially important.

Since 1977, all requests for information at SCRTD have been handled through the Computerized Customer Information System (CCIS). The system allows agents to retrieve quickly any transit-related information. Given any origin and destination in the service area, CCIS produces a set of possible itineraries, from which the operator can pick one that is most adapted to a caller's needs. CCIS was first implemented to increase agents' productivity, but it has had other positive impacts, such as improved accuracy and consistency of agents' responses, elimination of a hard-copy data base, reduction in the training period for new agents, and improved job satisfaction for employees.

CCIS was upgraded in 1992 to include the automatic delivery of itineraries through the use of a synthesized voice. The new information network ARTI, the name for the regional transportation information network, allows callers to input a code for their origins and destinations using a touch-tone phone, without talking to an operator. Origin and destination codes for all major cross streets are published in a catalog that is readily available. Employer members of the Corporate Transit Partnership also have a seven-digit code that can be processed by ARTI.

Washington Experience The case of Washington, D.C., is interesting because it provides an evaluation of the cost-effectiveness of a transit telephone information system comparable to SCRTD. In 1983, WMATA conducted a phone survey and reported that 82 percent of 602 respondents took the trip about which they had called for assistance (*1*). Of these people, 67 percent stated that they would not have taken the trip by transit without the information they had obtained from the telephone service. Using these results and the average fare, WMATA estimated that the telephone information service brought $520,000 in additional revenues to the authority, beyond the telephone information service costs (assuming conservatively that each effective call generated only one ride).

Although 80 percent of the callers did not have a car available at the moment they called, 56 percent stated that they had an automobile available at certain times. These callers were not entirely transit dependent, which confirmed the essential marketing role of telephone information systems to attract choice riders. The survey also revealed that 56 percent of the trips made by the callers took place during the off-peak period, compared with a 37 percent average systemwide. It suggested that telephone information services are most useful to the off-peak rider who may or may not also be a regular transit user.

Automated Schedule Information Systems

The other form of automated transit telephone information systems provides callers with fully automated schedule information (stop-specific arrival times), without involving a telephone operator. These systems are aimed at increasing ridership by reducing uncertainty associated with waiting time at bus stops, especially at night or during poor weather.

Such a system was first implemented in Canada. It is marketed under the name "Telerider." The first version was tested in Mississauga, Ontario, and included both bus position monitoring and arrival time calculation capabilities to provide callers with real-time information based on the actual position of the bus rather than its scheduled location. The version of Telerider that is now sold throughout North America, however, does not include the complex real-time capabilities; it gives only scheduled arrival times. The information is transmitted without the need of an agent because bus stop locations are coded in the phone numbers themselves.

The impact of such service on ridership has been investigated, and conflicting results have been found. Ross and Soberman (*2*) examined route by route transit riderships in four Canadian cities before and after the implementation of Telerider. In Winnipeg, Mississauga, Kitchener, and Guelph, no clear correlation could be established between the deployment of Telerider and ridership. However, Diewald et al. (*3*) showed that Telerider in Ottawa had a clearly positive impact on ridership in each time period of the day (peak, midday, and evening) and also contributed to improved public perception of the transit authority.

Telerider has also been implemented in the United States. Transit authorities of Salt Lake City, Pittsburgh, Columbus, and San Diego all have tested the system (*1,2*). Telerider was found to have no significant impact on ridership in Salt Lake City or San Diego, whereas in Columbus, Telerider bus routes performed 7 percent better than the control routes. In light of these results, the Central Ohio Transit Authority implemented Telerider systemwide in 1983 (*2*).

Effect of Transit Information on Transit Trip Behavior

Another way to look at the effectiveness of transit information is to observe the trip behavior of users who were provided with directions. The authors summarize the results of a few studies looking at that aspect. Geehan and Deslauriers (*4*) found that few people correctly interpret transit timetables. Only 9 percent of 580 participants were able to determine from a schedule when the next bus would pass their home. The researchers also reported that between 23 percent and 69 percent of the respondents were unable to use a transit route guide to plan an actual transit trip. Bronzaft et al. (*5*) examined whether the inability to interpret transit information correctly would adversely affect transit trip time. Twenty students were provided maps of the New York City subway system and asked to visit five stations on a set itinerary. All subjects took longer than necessary because they chose indirect or inappropriate routes or had to backtrack.

Hall (*6*) also studied the influence of trip information on route choice. A set of 50 students was observed as they traveled to a specified location accessible by several bus lines 1 ½ mi away. Twenty students received no information, 15 received maps only, and another 15 received both maps and schedules. The subjects who had maps reached their destination significantly faster than those without information. However, subjects provided with both maps and schedules took longer than those using maps alone. This result indicates that riders are unable to adjust their route to take advantage of a specific schedule. A surprisingly wide variety of itineraries was obtained among the students, and very few subjects took the shortest route to reach the destination. The experiment showed that route choice was a dynamic process because subjects constantly reevaluate their decisions in light of the last information they receive.

All previous studies agree that human ability to plan a trip from map and schedule information is somewhat limited, which may explain the demand for transit telephone information services that provide itineraries on a personalized basis. However, no study to date has looked at the way such travel information is interpreted and used. That is the objective of this paper.

EXPERIMENT

The experiment was designed to investigate the effects of transit telephone information on trip behavior. The objective of the experiment was to determine how efficiently transit information is provided by operators and how closely it is recorded and used by callers. The callers' response was analyzed using data collected from a test and survey; at the same time, directions provided by operators were analyzed using a transcript of sample phone calls.

Test and Survey of Callers

A test and survey among transit and nontransit users were conducted at Broadway Plaza, an indoor shopping mall in the heart of downtown Los Angeles, during two weeks in August 1992. The survey was conducted between 11:00 a.m. and 2:00 p.m. by University of California students. Participants were offered a $5 gift certificate. The sample size was 120. The test was composed of four parts. The first portion consisted of a survey of participants' travel behavior in general, looking at the mode, purpose, frequency, and length of their most frequent trips. Special emphasis was placed on reasons for use or non-use of public transportation, including a detailed assessment of participants' perceptions of transit in the Los Angeles area.

The second part of the test was aimed at providing a precise portrait of the effectiveness of the information provided by telephone information services. Participants were asked to call the information service and to write down the directions they obtained. For this part, participants were divided into two experimental groups. During the first week, participants asked the SCRTD operator for directions to their most frequent destination. During the second week of the experiment, all participants were instructed to request directions for a single standardized trip, which was to be taken at a specific time, and to include the use of two buses and a transfer. The experiment was expanded during the second week to compare the human operator with ARTI, the system that permits SCRTD to deliver directions through a synthesized voice; about half of the participants started with the operator and the other half began with ARTI. This design was used to control for any bias in the second call based on information obtained through the first call. During each phone call, surveyors collected information on the quality of service provided; waiting, holding, and answering times were recorded.

In the third part, participants were asked about the quality of the service provided in the phone call in terms of clarity and usefulness of directions as well as performance of the operator or the synthesized voice system. Participants' perceptions of transit in the area were assessed once again to determine whether the phone call had had any impact.

Finally, to evaluate whether the sample was representative of the general population, the last part of the survey consisted of questions characterizing socioeconomic profile. Overall, this revealed that the sample was comparable to the daytime population in downtown Los Angeles; that is, 60 percent are male, 53 percent drive alone, 37 percent use transit, and 7 percent carpool.

Analysis of Recorded Information

The recorded information was analyzed quantitatively only for the standardized trip group because the directions provided were the same for everyone. This part of the experiment led to interesting conclusions regarding the efficiency of the human operator as compared with the synthesized voice.

In this test, participants were asked to request information for a trip from Knott's Berry Farm in Buena Vista to Temple and Grand Streets in downtown Los Angeles. Two buses had to be taken to complete the trip on the SCRTD network. The complete directions provided by the information service were the bus number, direction, bus line, departure point, and the arrival point for the first bus and the second bus. In some cases, the operator may have omitted some of them. Participants were also asked to write down the given directions on the questionnaire as clearly as possible. They were allowed to draw maps or whatever figures they might find helpful. However, not one participant drew a map, which seems to indicate that information provided orally is not perceived or remembered in visual form. Instead, all participants wrote down the directions in a sequential fashion, noting the bus lines and the boarding and arrival points with their corresponding departure and arrival times. Although there were wide variations in the clarity and completeness of the recorded directions, all participants used the same basic format.

The total number of correct elements recorded was used to measure the caller's ability to obtain and comprehend essential information. As mentioned earlier, every caller got information from both a human operator and ARTI, the synthesized voice, with half the callers reaching the operator first and the other half starting with ARTI. Participants recorded more information when the directions were provided by a human operator. They were able to record, on average, 12.5 elements out of 17 from the operator as compared with 10.5 from ARTI. Table 1 shows the distribution of the number of elements recorded by the participants for the two modes. Directions provided by the synthesized voice were incorrectly interpreted more often than were directions provided by the operator.

Although many elements were not recorded, the most important directions were written down by most participants. Bus numbers, the directions of the buses, scheduled departure and arrival times, and the location of the transfer point were all recorded by at least 80 percent of the participants talking to an operator and by at least 65 percent of those dealing with the voice. Less crucial information, such as the fare and the exact starting point (Knott's Bus Terminal), was often overlooked by those using a human operator. Those hearing the synthesized voice were more apt to record this information. Human operators would often skip these details (as was seen in an analysis of transcripts), whereas ARTI is programmed to deliver all pertinent information.

It was also determined that regular transit users tend to record less information than infrequent users. That was observed among transit users speaking with a human operator, who recorded only 11.6 elements on average compared with nonusers, who recorded 13.0. Of those hearing the synthesized voice, 9.3 elements were recorded by transit users compared with 11.3 recorded by car users. The proportion of information recorded by transit users is consistently less for all the elements. The difference between the two groups is more marked for two elements: the fare, as would be expected, and the bus line, which indicates the street on which the bus is running. Indeed, transit riders seem satisfied with just a bus line number, whereas car users typically recorded the street as well.

Post-Call Questionnaire

After the phone call, participants were asked to assess the quality and usefulness of the service in terms of clarity of directions, clarity and speed of speech, friendliness of the operator or voice, completeness of directions, and their confidence in the accuracy of the information provided. The author's summary of these responses is mostly concerned with those of participants who dealt with both the human operator and the synthesized voice.

Again, waiting, holding, and answering times were recorded by surveyors during the phone calls. Waiting time was the time needed to first get an answer to the call and reach the service. Holding time was the delay between reaching the service and the moment that callers could request information. Answer time was the time taken by the operator or ARTI to transmit the directions. Values obtained for these three service times ranged from 0 to 10 min. Waiting time was similar for the operator and ARTI (2.5 min on average), holding time was less for the operator (1.3 min compared with 2.3 min), and answer time was slightly less for the operator (1.1 min versus 1.4 min). However, current version of ARTI significantly reduces waiting time to less than 1 min.

TABLE 1 Number of Correct Elements and Errors in the Recording of Participants

Number of Correct Elements	Human operator (44 cases)	Synthesized voice (40 cases)
Less than 5	2%	15%
Between 5 and 9	7%	17%
Between 10 and 12	25%	28%
Between 13 and 16	66%	40%
All (17)	0%	9%
AVERAGE:	**12.5**	**10.5**
Number of Errors	**Human operator**	**Synthesized voice**
None	71%	60%
One	29%	35%
Two	0%	5%

The human operator performed better in every category, according to the survey. Waiting time is similar for both modes because participants who listened to ARTI reached an operator first and then specifically asked for the synthesized voice to deliver the directions. Although a phone number was available to reach ARTI directly, surveyors discovered that it involved excessive and irregular waiting times.

One explanation for the longer answer time for ARTI was that it repeated directions to the callers 1.83 times on average, while the operator repeated them only 1.18 times. These are average values for the first phone call. On the second call, once the participants were familiar with the directions from their previous call, the average number of repetitions went down to 1.30 for ARTI but remained the same for the operator.

Results further showed that instructions given by a human operator were more intelligible and clearer than the ones provided by the synthesized voice. Indeed, 87 percent of the participants rated the instructions given by the operator "easy to follow" compared with only 24 percent for ARTI. On the other hand, only 2 percent stated that the operator was "difficult to follow," whereas 24 percent said so for ARTI. Similarly, when prompted about the clarity of the speech, 91 percent of the participants said the operator pronounced words very clearly, whereas only 24 percent said the same for ARTI.

Three quarters (76 percent) of the participants were satisfied with the speed of the operator when giving directions; the rest judged that it was somewhat too fast. However, only a third (33 percent) of the participants were satisfied with ARTI's speech speed. More than 22 percent thought it was too slow, and 45 percent thought it was too fast.

The operators were perceived to be friendly by 82 percent of the participants. On the other hand, the synthesized voice was rated as neutral by 75 percent of callers. Most participants (89 percent) also were very satisfied with the operator's promptness in answering their questions. Because ARTI does not enable people to interact directly, this question was not asked of ARTI users.

Participants further expressed confidence in the accuracy of information provided by the service. Almost everyone (96 percent) trusted the directions given by the operator. ARTI's credibility was almost as high at 89 percent. However, just two thirds (67 percent) of the participants judged that the information was complete enough to get them to their destination.

Transit users gave higher ratings to the telephone information service than did nonusers. They were consistently more satisfied with the clarity of information, the speed of the speech, the friendliness of the operator, and the promptness of the answers to their questions. They were also more confident with the accuracy and completeness of the information provided than were nontransit users. This result might be explained by transit users' ability to handle transit-related information better and their more positive attitude toward SCRTD in general.

One of the objectives of the experiment was to observe how the availability of precise and personalized transit information would affect travelers' propensity to use public transit. Therefore, participants were asked how the directions they received would affect their chances of using transit. Nearly one half of the participants indicated that transit information would influence their choice of transportation mode, if the information appears to be reliable, accurate, and helpful (*7*). The experiment also indicated that frequent users are more receptive to information services than those who use transit infrequently. Fifty-five percent of transit users stated that the information they received encouraged them to take transit as compared with 35 percent of the nontransit users surveyed. In addition, the manner in which the information was provided influenced their propensity to use transit. Fifty-six percent of the participants said the information provided by the human operator encouraged them to ride transit, whereas only 36 percent of the same people said so when the same information had been transmitted by ARTI.

When asked directly how they preferred to receive transit information by telephone, participants overwhelmingly (95 percent) selected the human operator. The main reasons given were the possibility of interaction (mentioned by 91 percent of participants), clarity of directions (83 percent), clearer pronunciation of names (79 percent), and the better speech speed (76 percent).

Participants were also asked about the most appropriate medium to transmit transit information. Telephone is the preferred medium, as it's supported by 75 percent of participants. Written information was second choice; other sources, such as one's employer, personal computer, or FAX were only marginally of interest.

Finally, participants' perceptions of the transit system in Los Angeles were assessed once again after the information call on the same 1-to-5 scale. No major differences in satisfaction level were found on most points; however, a decrease in satisfaction was observed concerning bus travel time, once participants were given an example of a typically long (1 hr 16 min) bus commute. Satisfaction with SCRTD employees was higher after the call, probably because participants appreciated the telephone operators.

Analysis of Phone Calls Made to SCRTD

In addition to the survey, a sample of phone calls made to SCRTD was recorded, and the recorded transcript was analyzed to gain a better understanding of the flow of information between callers and operators. Three hours of conversation were recorded while the survey was underway in August 1992. Three operators collaborated in the task. The sample contained a total of 92 phone calls.

As shown in Figure 1, a typical phone call starts with a request for a specific type of information from the caller. Next, the operator usually helps the caller to clarify the request. The operator then inputs the details of the request into the computer and waits for possible solutions to be generated, while the caller stays on hold. The operator then chooses the most appropriate solution out of the ones proposed by the computer and details it the caller. Generally, the caller accepts the given solution and completes the call after confirming the information. In some cases, callers tell the operator they are not satisfied with an aspect of the proposed solution, such as the route or the schedule. When this happens, the process starts again and the operator tries to find a new solution that better fits the needs of the caller.

Eighty percent of the 92 calls are requests for an itinerary, as indicated in Table 2. Itinerary calls are those for which callers provide the operator with an origin and destination and ask for the most convenient way to complete the trip using the SCRTD network. Most itinerary calls concern a single trip with a specific time constraint specified (Table 2). Schedule information accounts for one-tenth of the calls; there are relatively few requests for fare information. The high proportion of itinerary calls in Los Angeles compared with other American cities reflects the large size and complexity of the SCRTD network (in Washington, D.C., for example, itinerary calls account for only two-thirds of the total) (*1*).

To examine the duration of itinerary phone calls, the total number of interactions was used as an approximation of the exact length in minutes. An interaction represents any statement or question made by one of the interlocutors (either caller or operator). Interactions are not of equal length, but, on average, their number is a reasonable approximation for the call length. The average phone in the sample includes 21.5 interactions, although some phone calls contain as few as five and one contained 73. Figure 2 shows the distribution of phone call lengths. One of the recorded calls was part of the call-in experiment. This call was completed with only seven interactions, much less than the average phone call. This appears to indicate that the experiment did not portray entirely the complex interactions that usually occur between the caller and the operator. That the trip was not intended to be taken also explains the haste with which the sample call was handled by both the caller and the operator. When the caller is really planning to undertake the trip, he or she usually asks for more details and for a repetition of the directions, which makes phone calls significantly longer.

The calls requiring a longer treatment were classified under several categories of problems, as presented in Table 3. (Note that the percentages do not add up to 100 percent because some calls fit into more than one category). As the table indicates, only 62 percent of the calls were a direct answer from the operator. These calls are significantly shorter than the others; they are completed in 15 interactions on average. For 11 percent of the calls, the operator had to force the caller to formulate the request in a precise way, such that the operator could input it into the computer. This percentage does not include the numerous calls for which the operator asked the caller about small details, such as the departure or arrival time of the trip. Notice also that in one call out of 10 the caller is not satisfied with the first solution proposed by the operator. These calls are on average twice as long as other ones. Cases in which the caller is not familiar with the locations described by the operator also make up 8 percent of the total calls. Such calls are twice as long as more direct ones, because the operator has to provide detailed access and egress information. In these cases, the operator's knowledge of the exact bus stop location proved to be important.

To conclude the analysis of the phone transcripts, all the specific elements of information that were transmitted by the operator during the itinerary calls in the sample were extracted. Table 4 summarizes the results and compares them to what was recorded by the survey participants (those inquiring about the standardized trip from a human operator). The last column presents the percentage of participants who correctly recorded each element of information.

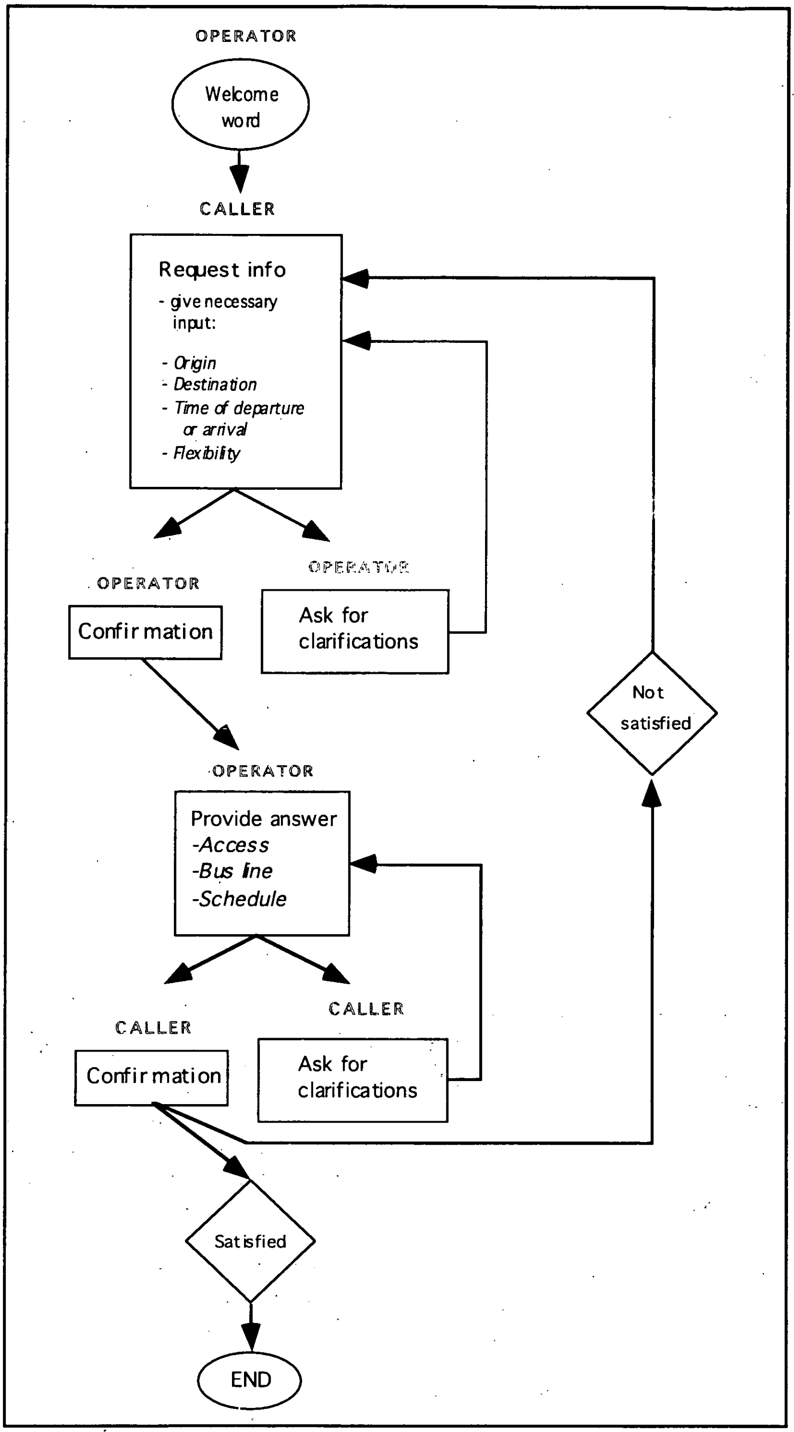


FIGURE 1 Structure of a phone call for transit information.

Detailed access information was provided in 24 percent of the phone calls sampled. Two cross streets first were named and their names repeated at the caller's request in another 23 percent of the calls; no single access information was given in the remaining half of the calls. Similarly, precise egress information was provided in only 31 percent of the calls, including all those in which destination points were referred to by an address.

Only one scheduled departure time was provided in most of the cases (73 percent); two departure times were provided for 11 percent of the calls, three or more for only 3 percent of the calls. Thirteen percent of the itineraries did not include any schedule information.

The comparison presented in Table 4 indicates that there is good consistency between the information provided and what is re-

TABLE 2 Type of Information Requests Made at SCRTD Telephone Information Service

Itinerary		80%
single trip, no time specified	4%	
single trip, with time specified	62%	
round-trip, with time specified	14%	
Schedule information		10%
Fare and other transportation related information		3%
Transfer to other services		7%

TABLE 3 Problem Calls and Their Average Length

Type of Problem	Proportion	Average Length
Calls without problems	62%	15
Operator asks for precision about request	11%	25
Caller does not accept the solution proposed by the operator	9%	32
Caller unfamiliar with locations	8%	32
Caller has wrong or incomplete information about his/her request	6%	21
Caller repeats given information	6%	52
Operator has to repeat information	5%	40
Caller does not understand	3%	43
Caller keeps changing request	1%	59

corded by the callers. Note, however, that the numbers are not directly comparable because they come from two different samples. The first column refers to the 92 phone calls sampled, whereas the second column refers to the 120 participants in the call-in experiment.

CONCLUSION

The authors investigated how transit telephone information is provided and how it is interpreted and used by callers planning a trip. Findings can be summarized as follows:

- Telephone information services appear to be well received by the general population, especially by frequent transit riders. People view the telephone as the most effective means to transmit transit information. Telephone operators obtain high ratings in terms of friendliness, clarity of directions provided, speed of speech, and confidence in the accuracy of directions provided.
- A synthesized voice system did not obtain the same high ratings as the human operators. Public satisfaction with the automated voice's clarity of directions and speech was low.
- Human operators are preferred overwhelmingly to the synthesized voice, mainly because of users' ability to interact with operators when getting directions. Analysis of phone transcripts reveals that much information is exchanged through multiple interactions between the caller and the operator instead of merely being provided by the operator.
- A synthesized voice system did not appear as effective as human operators in transmitting information to callers. Participants recorded less information and made more mistakes on average when interacting with ARTI than with the operators.
- Telephone information services' potential to generate more transit trips is higher if people requesting information are already frequent transit riders and the information is provided them by a human operator. Therefore, there is a link between the quality of transit information provided and the propensity to use public transportation.

ACKNOWLEDGMENTS

This study was performed as part of the PATH program of the University of California, in cooperation with the state of California, Business and Transportation Agency, Department of Transportation, and FHWA. The authors acknowledge those students of the University of California who conducted the experiment, especially Brian Pfeifle, who supervised the students. Special thanks to SCRTD and Broadway Plaza for their information and support.

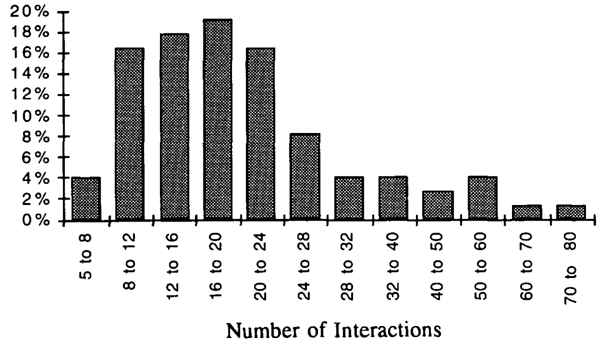


FIGURE 2 Number of interactions between caller and operator during a phone call.

TABLE 4 Directions Provided by Operators and Recorded by Callers

Element of information	Percentage of calls where the information is....	
	provided by the operator	recorded by the caller
Access	24%	11%
Departure point	91%	80%
Bus line	100%	98%
Direction	88%	91%
Departure time	87%	90%
Arrival point	96%	83%
Arrival time	80%	86%
Fare	8%	21%
Egress	31%	_
Other	21%	_

REFERENCES

1. Cutler, M. R., and R. F. Potter. The Effectiveness of Telephone Information in Transit. UMTA, Washington, D.C., Feb. 1984.
2. Ross, D. R., and R. M. Soberman. Evaluation of Public Transit Automated Telephone Information Systems. Transportation Development Center, Montreal, Quebec, Canada, May 1987.
3. Diewald, W. J., W. H. Frost, and W. Bamberg. Assessment of Transit Passenger Information System. UMTA, Washington, D.C., June 1983.
4. Geehan, T., and P. Deslauriers. Transit Route Guide—An Investigation and Analysis. *Bus Ride*, Vol. 14, No. 4, 1978, pp. 56–58.
5. Bronzaft, A. L., S. B. Dobrow, and T. J. O'Hanlon. Spatial Orientation in a Subway System. *Environment and Behavior*, Vol. 8, 1976, pp. 575–594.
6. Hall, R. W. Traveler Performance and Information Availability: An Experiment in Route Choice. *Transportation Planning and Technology*, Vol. 8, 1983, pp. 177–189.
7. Yim, Y. B., R. W. Hall, and E. Le Colletter. Dynamic Traveler Information and Transit Ridership. *Proc. Pacific Rim TransTech Conference*, Vol. I, July 25–28, 1993.

Publication of this paper sponsored by Committee on Public Transportation Marketing and Fare Policy.

PART 3

Technology

Personal Rapid Transit Study in Gothenburg, Sweden

Bo Blide

The Gothenburg Traffic Authority found that new traffic policy goals identified during recent years are very difficult to achieve by conventional public transport techniques. Consequently new system concepts must be studied. Personal rapid transit (PRT) is one of them. The PRT project in Gothenburg, Sweden, was initiated in 1991. After prestudies an enlarged study, which is described, was started in summer 1992. The object of the study was to investigate whether PRT could take over as the only public transport system and replace the present system (light rail and buses). The work concentrated on four parallel activities: (a) design of a PRT track network covering a major part of the city and the central parts of two adjacent communities, (b) establishment of the travel demand in the area (trip matrices), (c) development of a control system suitable for a large PRT system, and (d), development of a simulation program for the analysis of PRT system functions, with special emphasis on operational strategies, travel standard, capacity, productivity, and resources needed. The approach to the problem and the techniques used and developed in the study are described. The result of the study is that it appears theoretically possible to operate very large PRT systems. The system studied in Gothenburg includes 700 track km (counted in single tracks), 650 stations, and 17,000 vehicles. One question still to be answered is whether it is possible to attain satisfactory reliability for all the components involved.

Gothenburg is Sweden's second largest city, with 433,000 inhabitants. Including the suburbs, the population is 730,000. The present public transport system is a mixed light rail-bus system with nine tram lines and 30 bus routes. The number of daily trips by public transport in the city is about 300,000, including trips by passengers who transfer from regional bus and train services. Up to 65 percent of public transport operations in the city are financed by local taxes.

During recent years new traffic policy goals that are practically impossible to realize with a conventional public transport system have been identified. Consequently new system concepts must be studied. Personal rapid transit (PRT) is one of them.

STUDY OBJECTIVES

The PRT project in Gothenburg was initiated in 1991. After prestudies an enlarged study, which is described in this paper, was started in summer 1992. The object of the study has been to investigate whether PRT could take over as the only public transport system and replace the present system (light rail and buses).

MODEL CONCEPT

According to a decision by the Traffic Committee of Gothenburg the model PRT concept should be a system of cars with rubber wheels on elevated tracks. In some sensitive parts of the city, mainly in the central area, it is anticipated that tunnel solutions must be discussed.

TRAVEL DEMAND

Trip Matrices

The calculations of travel demand were based on existing public transport trip matrices and existing statistics concerning car travel (area-area matrices).

The information was translated into station-station matrices for the alternative PRT networks studied. Matrices have been prepared for the time periods 0600 to 0900, 0900 to 1500, and 1500 to 1900 hr. The statistical material also made it possible to study shorter time periods, down to 30 min. For the capacity tests of the networks the half-hour of the morning and afternoon peak periods with maximum ridership levels were used.

Figure 1 shows the distribution of the present public transport ridership between 0500 and 2200 hr on a half-hour basis.

The share (in percent) of the maximum hour and half-hour in each time period is (a) time period—0600–0900, 0900–1500, 1500–1900; (b) maximum hour—50, 22, 34 percent; and (c) maximum half-hour—30, 11, 17 percent.

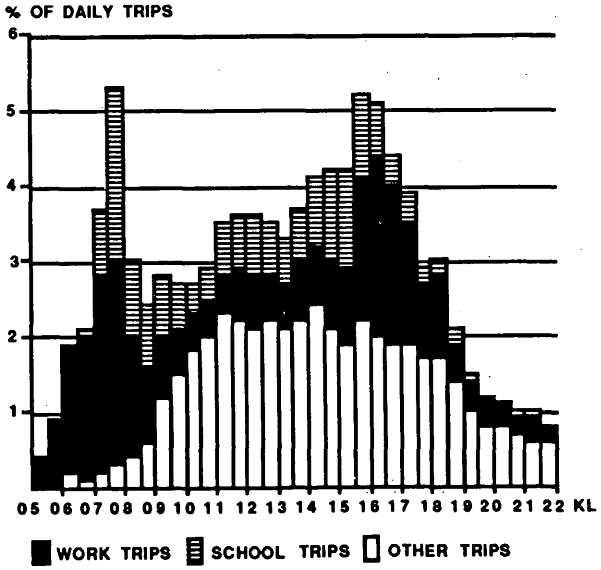


FIGURE 1 Distribution of present public transport ridership between 0500 and 2200 hr on half-hour basis.

Gothenburg Traffic Authority, Box 2403, S-40316 Gothenburg, Sweden.

Increased Public Transport Ridership

A citywide PRT system should be able to cope with a considerable increase in public transport ridership in comparison with the present ridership. There can be various reasons for such an increase.

- Improved travel standard makes today's public transport riders travel more.
- Better level of service attracts a portion of the present car riders, especially for origin-destination combinations for which present public transport services are poor.
- Car riders are "forced" to leave their cars by traffic policy measures, such as high parking fees and road pricing.
- A change of the "global" conditions for car use, such as supply and price of fuel.

In the study described here an expansion of public transport ridership is created by a transfer of car riders from the car trip matrices to the matrices containing the present public transport trips. The matrices are prepared on a station-station level for the PRT system.

The transfer of car trips has not been made evenly for all origin-destination combinations, because the present modal split is not the same for the central area and the rest of the city. In the city as a whole 25 percent of the trips made are by public transport, but for the central area the figure is considerably higher (55 percent). By testing the capacity (in percent) of the PRT system the transfer was accomplished in a stepwise manner as shown in the in-text table.

Transferred Car Riders	*Central Area (percent)*	*Other Trip Combinations (percent)*
Step 1	10	24
Step 2	20	48
Step 3	30	72

In Step 3 almost 60 percent of the car riders during the morning peak period was transferred, corresponding to an 80 percent increase in the number of public transport riders. The simulations indicated that the PRT system could cope with this situation. The figures on the performance of the system refer to Step 3, which is believed to cover the transfer potential of people attracted by the high standard of the new system and people forced to use the new system by traffic policy measures that are already available or being discussed.

TIME GAPS

The possible minimum time gap between two cars is a basic factor in the design of a PRT system and for the calculation of its capacity. The time gap is defined as the shortest possible time distance between two cars that can be allowed if Car 2 in Figure 2 should be able to stop without any car damage or personal injuries if Car 1 comes to an abrupt stop.

The time gap depends on the following factors:

- In what way Car 1 stops.
- Time it takes to inform Car 2 of the stop.
- Time it takes for Car 2 to evaluate the information and to give braking order.
- Time it takes for the brakes to work to full effect.
- Time for deceleration to a standstill.

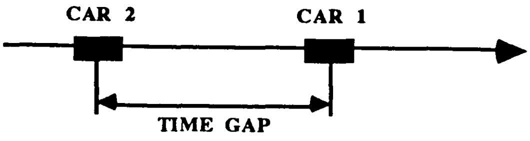


FIGURE 2 Time gap.

Of these time factors the last one is by far the longest.

Concerning the way in which Car 1 stops, two cases have been studied: (a) the car is brought to a brick-wall stop, (b) the car completely breaks down, slides along the track, and stops as fast as friction allows. Assuming an operating speed of 13 m/sec and a maximum deceleration of 6 m/sec^2 the minimum time gap is 1.6 sec in Case a and 0.8 sec in Case b. In simulations of the system an average speed of 10 m/sec (36 km/hr) has been used to compensate for lower speeds at curves and switches.

A special risk case arises if a car is brought to a stop according to Case b and finally stops at a merge point where two track links meet. In this case a special detecting system in the track is needed if the time gap of 0.8 sec is used.

The choice of time gap depends, however, not only on technical factors but also on psychological ones. To what extent are people prepared to ride in automatically guided vehicles at the actual speed and distance gaps? No such studies have been done within the framework of the present study. It should be observed, however, that the construction time for a system of the actual size is long, and is probably the time before the minimum time gaps must be used. Therefore there will be time for people to get used to the new situation.

In the evaluation of the capacity and travel standard of the PRT system the effects of both time gaps were studied.

OPERATIONAL STRATEGIES

A strict PRT operation means that the passenger rides directly from start station to the destination, alone or with the company he or she chooses. It was estimated that the average number of people per car in peak hours would be 1.25, the same as in private cars today. This implies a poor use of the car and track network capacity, especially considering that, on average, less than 50 percent of the cars in the system is running with passengers. The others are either on their way to a new mission or are waiting at stations or depots.

It was found to be important to increase the occupancy of the cars, at least in the morning and afternoon peak periods, to increase the ability of the system to cope with a growing number of passengers.

An increase in the occupancy requires an organized coordination of trips. There are two main ways of doing that: (a) route operation and (b) ride sharing.

Route operation was abandoned at an early stage. The important quality of direct trips would require a great number of routes even if a considerable portion of the trips would still be strictly PRT. Furthermore the route network would have to be changed several times during the day because of variations in travel patterns.

A more flexible way to coordinate trips would be through some sort of organized ride sharing. Ride sharing between prechosen pairs of stations was found to have little effect. The same arrangement for groups of stations had a better effect on car occupancy but had two disadvantages: (a) there would be several intermediate stops for many passengers and (b) the prechosen groups of stations would have to be changed a number of times each day because of the varying travel pattern. Instead it was decided to test a more dynamic type of ride sharing that is more flexible and better exploits the possibilities of a modern control system.

A station is used for ride sharing if the number of incoming passengers is at least two per minute. Passengers with destinations likely to result in ride sharing wait for a certain maximum time, in the present study 3 min, before their car arrives. Then the car leaves whether fellow passengers have arrived or not. The matching of passengers is based on the first passenger's destination order. Other passengers are accepted if they have the same destination or a shorter or longer trip in the same direction. In the present study "the same direction" means that no passenger in a ride-sharing group has a destination that results in more than a 30 percent detour for anyone. The simulations indicate that this detouring possibility will be used to a fairly low degree.

Ride sharing is arranged only from one and the same station, which means that the passenger "knows" the company he or she is going to ride with from the start and there are no unpleasant surprises en route. Moreover tests with a pick-up-en-route strategy showed little effect on average car occupancy.

With this type of ride sharing it was possible to increase the average car occupancy from 1.25 to 1.90 (52 percent) in peak hours, which leads to a 30 to 35 percent decrease in the size of the car fleet needed.

In Figure 3 the trip from Station 0 to Station 1 is initially booked. A new passenger to Station 2 is accepted if Station 2 is close to the route between Stations 0 and 1 (within the small oval in Figure 3). Alternatively a passenger to Station 3 can be accepted if Station 1 is close to the direct route from Stations 0 to 3 (within the large oval in Figure 3). The ovals represent a 30 percent increase in the riding time compared with that of the shortest route.

Assume that the trip to Station 2 is accepted. With two destinations locked (Stations 1 and 2), a third station can be accepted according to one of the following examples.

Points 0, 1, and 2 are now given. Station 4 can be accepted if it is close to the route between Stations 0 and 2. Station 5 can be accepted if it is close to the route between Stations 2 and 1. Station 3 can be accepted if Station 1 is close to the route between Stations 0 and 3 (Figure 4).

Matched passengers are gradually grouped until they fill up a car or until the first passenger has waited for 3 min.

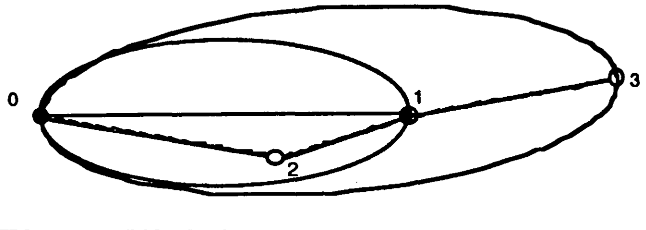


FIGURE 3 Ride sharing.

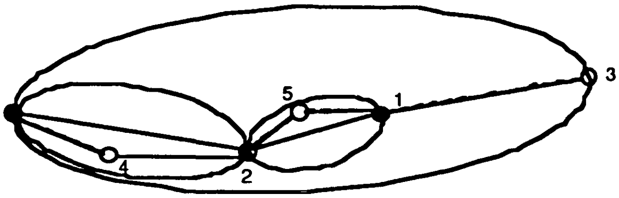


FIGURE 4 Extended ride sharing.

EMPTY CAR HANDLING: DEPOT SYSTEM

The handling of empty cars is one of the big problems in a large PRT system. In the initial computer simulations of the Gothenburg system it became clear that the stations would not be able to house all their needed empty cars without becoming unacceptably large. Therefore special depots for empty cars were introduced at strategic locations in the network to (a) secure the provision of empty cars to subareas in a way that gives short and "guaranteed" waiting times and (b) work as buffers for empty cars to minimize the sizes of the stations. For a more detailed description of vehicle distribution see the paper by Andréasson (this Record).

CONTROL SYSTEM

Four control system principles have been studied and compared: synchronous control, asynchronous control, quasisynchronous control, and point-synchronous control. The characteristics of the four control system principles are briefly described.

Synchronous Control

- A car does not start its transport mission until a time gap is available and booked the whole way (through all the switches) between the start station and the destination.
- The search for and booking of time gaps must be made in a central computer, which then updates booking tables for each switch. The supervision of time gaps and cars is made via the communication system between local systems and the cars.
- The cars are driven at a speed that must be coordinated and synchronized with the generated time gaps.
- In case of disturbances in the system all cars must stop and wait for a replanning of routes and time gaps.

Asynchronous Control

- Each car in the system is allowed to adjust its running according to events that occur en route. The cars can, within certain limits, accelerate and decelerate as traffic conditions demand. The cars behave like cars on a road system.
- A transport mission can start without the whole route to the destination being planned and clearance through all switches secured. Continuous route choices are made during the trip.
- During capacity disturbances in the track network either queues are allowed to be formed or the control system gives the cars alternative routes without disturbing the reliability of the system.

Quasisynchronous Control

- The generation of time gaps is made at a central level, but the assignment of time gaps is made locally (at switches).
- A transport mission starts as soon as a time gap is available at the start station.
- The cars operate at a constant speed, which must be the same as and synchronized with the time gaps. At merge conflicts or disturbances the cars are allowed to back a distance corresponding to a multiple of the time gap.
- At disturbances in the system queues may develop, but redistribution of routes can be made. Queues can be organized at time gap or shorter distances.

Disadvantages of Synchronous, Asynchronous, and Quasisynchronous Controls

Synchronous, asynchronous, and quasisynchronous controls were found to have disadvantages when applied to a large PRT system. The synchronous control is very rigid in its operation, centralized, and sensitive to disturbances (each disturbance has a significant consequence on the operation). The use of asynchronous control carries the risk of congestion, which is difficult to control. Quasisynchronous control corresponds best to the demands of a large system, but it demands the synchronism of time gaps, which complicates the control at switches.

Point-Synchronous Control

It was decided to try to combine the advantages of the three types of controls in a new one: the point-synchronous control. The following principles from the other types of controls were chosen:

- Synchronous control: the assignment of time gaps and an even speed at switches to provide high capacity.
- Asynchronous control: simplicity, in which each car controls its own speed; decentralization, in which all decisions are made in cooperation between the car and the following switch; robustness, in which disturbances are dealt with locally without central replanning; and flexibility, in which rerouting can be made continuously.
- Quasisynchronous control: time gap assignment for the next switch and methods for giving priority and for the continuous choice of routes.

The point-synchronous control principle can be said to be asynchronous control with adaptation to locally created time gaps at switches. The speed is adjusted by the car for arrival at the next switch at the right time and the right speed, just as a pedestrian advances toward a self-revolving door.

SIMULATION SYSTEM

The purpose of the simulation studies was to evaluate different track networks, control principles, and operation strategies. The evaluation included several aspects: (a) travel standard, (b) capacity, (c) productivity, and (d) resources needed.

For simulation results, see the later section System Analysis. For a description of the simulation system see the paper by Andréasson (this Record).

RELIABILITY

Extensive automation of different functions in society has been going on for a long time. Automation as such is consequently not an unknown concept. There have been a number of applications with automated vehicles in the industrial sector for more than a decade. In the public transport area there are well-tried automated systems for rail-bound vehicles as well as buses.

None of these systems, however, has anything like the complexity of a large PRT system in the number of possible origin-destination combinations, the number of switches, and the number of vehicles.

When reading the section System Analysis one should bear in mind that the results are based on simulations that presume 100 percent reliability for all components. Even if the control system can deal with disturbances in the operation, a PRT system of the actual dimensions will demand reliabilities of both vehicles and other installations that are greater than those of today's vehicles and systems.

NETWORK STUDIES

Network Design

A combination of two network models has been in the design of track networks, the spider net model and the grid network, both of which are in general use in traffic planning.

The spider net model is generally used to design plans for the main arterials for car traffic in a city. The model has the advantage of easing traffic pressure on the central area and of creating direct connections in tangential travel combinations.

The grid model is used mainly in older parts of cities that were designed in blocks and with little thought of the larger travel needs of the inhabitants. For both car and PRT traffic the grid model has the advantage that the corridors are easy to arrange for one-way traffic, which can provide higher capacity, simpler intersection design, less space demand, and lower costs.

To exploit the economic and space-saving advantages of the grid model, it was used in the central area of the city and in some residential and industrial areas according to the principle described below. The major configuration of the citywide track network follows the spider net model. The network for the final simulations is shown in Figure 5.

The network has the following dimensions:

- 728 track km (counted in single tracks and including station and depot tracks).
- 391 stations (654 station directions; see later for information concerning stations).

Capacity

The theoretical capacity of a track link at a speed of 10 m/sec is, if all time gaps have one car, 2.250 cars/hr at a time gap of 1.6 sec and 4.500 cars/hr at a time gap of 0.8 sec.

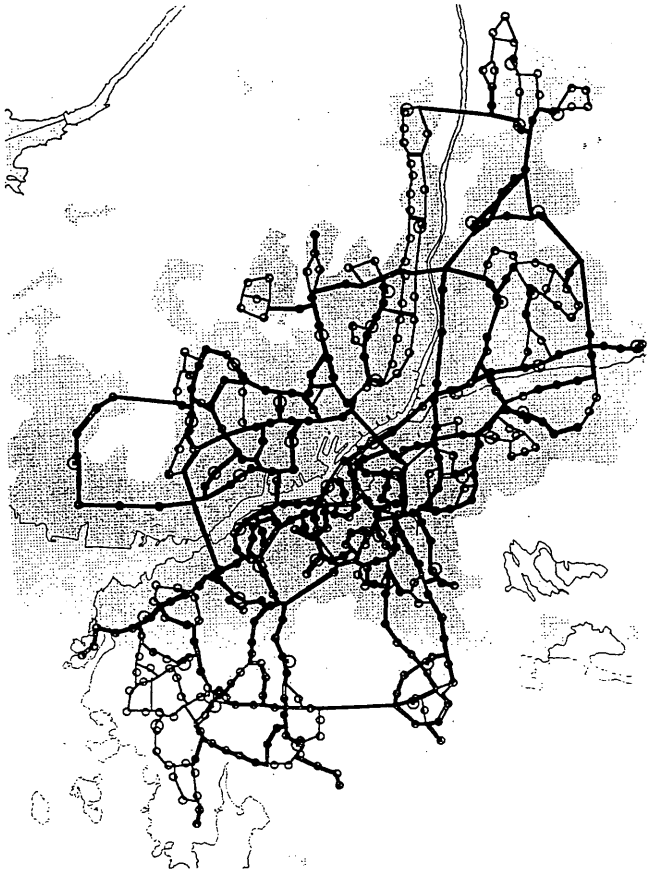


FIGURE 5 PRT network in Gothenburg study.

The simulation, however, used the possibility that empty cars could be run with a shorter safety distance between them. This resulted in link use of up to 2.800 cars/hr in the 1.6-sec time gap case. The capacity of a track link depends on the percentage of empty cars in the flow and in which order they arrive (the possibility of forming platoons). The greatest benefit of this type of operation is reached when both capacity demand and empty car percentage are highest in the central area.

STATIONS

Design

There are two main types of stations: on-line stations and off-line stations. On-line stations have the disadvantage that cars stopping at the station delay other cars. They can only be used on links with very low traffic flows or as cul-de-sac stations on links with no other stations. In the Gothenburg system practically all stations are off-line (Figure 6).

In two-way track sections the two station directions can be split up, which means that the station pattern is better at covering the area and implies shorter average walking distances. In contrast to route traffic a PRT system offers a direct trip from all station directions even if the riding time is a bit longer from the most unfavorable direction. It is the passenger's choice whether he or she wants to walk or ride longer (Figure 6).

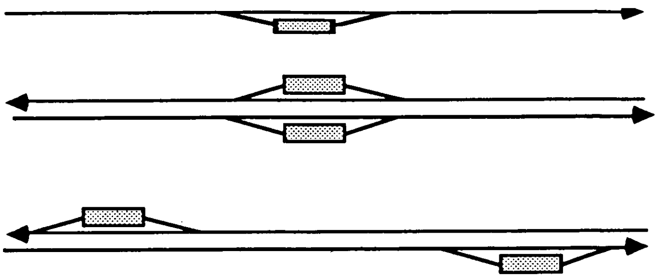


FIGURE 6 Off-line stations.

The principal function of an off-line station in the Gothenburg study is shown in Figure 7.

The standard station has one entrance from and one exit to the main track. The station has an arrival zone, an empty car buffer zone, and a departure zone. The car buffer zone is installed to provide as short a waiting time as possible. The number of empty cars in the zone should correspond to the car consumption during a time period equal to the running time from the nearest car depot. If the station is used for ride sharing it should be provided with an extra exit track as shown in Figure 7.

With ride sharing, arranged according to the principles described in the section Operational Strategies, cars can use the arrival zone of a station and continue to the main track without being delayed by the car buffer zone or disturbing the operation at the departure zone. This additional exit also has the function that arrived and emptied cars can easily be sent away if the buffer zone is full.

Capacity

The capacities of the stations and the main tracks should harmonize to take care of the forecasted number of passengers.

There are no empiric figures concerning the maximum capacities of PRT stations. In different papers it has been assumed that one car could depart every 10 sec. To be on safer ground preliminary capacity calculations were carried out. They indicate that a station where three cars can be boarded simultaneously could send away at least 500 passengers per hour.

SYSTEM ANALYSIS

Simulations and analyses of the PRT system of the four loading cases presented in the section Travel Demand were performed.

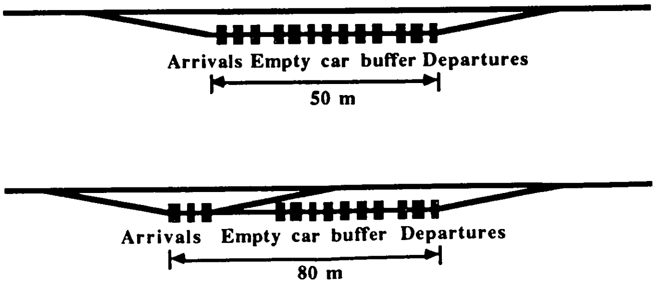


FIGURE 7 Stations without and with ride sharing.

The performance figures presented refer to the highest-load case. Simulations were made for both the 1.6- and the 0.8-sec time gaps.

TRAVEL STANDARD

The travel standard that a PRT system can offer is a crucial factor in the evaluation of the system. It should motivate the costs and other efforts connected with its realization.

The concept of travel standard has many components. Concerning most of the components PRT is superior to conventional public transport. A few important ones follow.

- *Travel time.* The travel time is shorter because of 100 percent direct trips, shorter trip lengths, higher average speeds, and the complete absence of disturbances from other traffic.
- *Headway.* The concept of headway does not exist in a PRT system. Instead there is a waiting time for an ordered car that is considerably shorter than the average waiting time in the present public transport system. The so-called hidden waiting time often discussed in connection with conventional public transport disappears completely.
- *Transfers.* The transfer frequency is 0 percent.
- *Punctuality.* The concept of punctuality does not exist because there are no timetables.
- *Comfort.* All passengers are seated.

The travel standard components mostly used in traffic planning are travel time and trip length, the first as a good measure of the "travel sacrifice" of the passengers and the second as a measure of the "traffic work" and a base for the calculation of energy consumption. Both components are easy to quantify and are thereby useful instruments in a comparison with other types of public transport systems.

Comparison with Present Public Transport System

Table 1 shows the average travel time (excluding walking time, which is estimated to be the same as that in the existing system) in the PRT system outlined here compared with that of the present public transport system (light rail and bus). For the present system the figures represent an average for the morning period from 0600 to 0900 hr, whereas the figures for the PRT are valid for the half-hour of the same period in which ridership is at a maximum. No transfer penalty has been included in the calculations for the present system.

TABLE 1 Average Travel Time in PRT System Compared with That in Present Public Transport System

System	Riding time	Waiting time	Transfer time	Total	%
Present, real time	17.3 min	5.1 min	1.9 min	24.3 min	100
PRT, real time					
0 car riders 2)	10.3 min	1.5 min	0 min	11.8 min	48
30/72 car riders	11.2 min	1.5 min	0 min	12.7 min	52
Present, weighted time 1)	17.3 min	10.2 min	3.8 min	31.3 min	100
PRT, weighted time 1)					
0 car riders 2)	10.3 min	3.0 min	0 min	13.3 min	42
30/72 car riders	11.2 min	3.0 min	0 min	14.2 min	45

1) Riding time weight 1, waiting and transfer times weight 2.
2) Concerning car riders see Travel Demand

The detailed presentation below refers to the half-hour of the morning peak period in which ridership is at a maximum. The same observations for the afternoon peak period are then provided.

Trip Length

The average trip length is 6.7 km at a time gap of 1.6 sec and 6.4 km at a time gap of 0.8 sec. This is because of the higher capacity use of the network at the longer time gap, which leads to a more extensive rerouting (bottlenecks). Some of these bottlenecks could be removed by a more carefully detailed design of the network. There was, however, no time or reason for such a detailed study.

In the present public transport system the average trip length is 7.2 km. The difference is an effect of the spider net model used for the PRT system (see the section Network Studies), whereas the present system is radially oriented.

At this point it should be observed that the present public transport riders are mainly riding in the travel combinations where public transport provides good service. A major part of the transferred car riders would have considerably longer trip lengths in the present public transport system.

Travel Time

As shown in Table 1 the travel times of the present public transport riders would be reduced by more than 50 percent in a PRT system. For the transferred car riders the travel times would not be as short as they are today by car, but PRT would be a fairly acceptable alternative.

Waiting Times

The waiting times at stations are approximately the same in both time gap alternatives (1.5 min). Part of the waiting time is needed for the short walk from the ticketing machine to the departure point. The average waiting time includes the somewhat longer waiting times for some of the ride-sharing passengers. Some 99 percent of the passengers have waiting times shorter than 4 min.

Delays

At high capacity use of the system queues can develop at certain merge points. This is observed by the control system, which reroutes the traffic around these points. In the simulation, however, this does not happen instantaneously and certain delays can arise before the queue has disappeared. In a full-scale control system the rerouting can be accomplished more quickly, which will reduce delays. The average delay because of speed adaptations at switches is small, less than 1 minute in both time gap alternatives.

The maximum delay, which happens to a very limited number of passengers, is also small in the time gap alternative of 0.8 sec and is probably even smaller outside peak hours. In the time gap alternative of 1.6 sec the maximum delay was found to be more than 5 min in the highest loading alternatives, but it involves a limited number of passengers. The maximum delay depends on the bottleneck effects discussed and it is possible to reduce this by a more detailed network design. It indicates, however, a rather large sensibility for disturbances, as is the case in a car traffic system with a high capacity use.

Capacity

The system can cope with the high travel demand without any inconveniences other than the maximum delays presented. The number of passengers at the stations at the end of the half-hour in which ridership is at a maximum corresponds to a normal inflow, and the passengers will have the same waiting time as the average. With the present daily variation of traveling the PRT system could provide more than 600,000 trips per day.

Productivity

Fully 40 percent of the passengers in the half-hour in which ridership is at a maximum are ride sharing, which results in an average car occupancy of 1.9 persons.

The number of passenger and car kilometers is practically the same in the two time gap alternatives, 1.6 and 0.8 sec. The number of empty car kilometers is about 45 percent of the total.

The number of transport missions per car hour in the period in which ridership is at a maximum is slightly more than two, which indicates that each car carries out 20 to 25 missions per day and serves 35 to 45 passengers per day.

Resources

The car fleet needed for the operation in the half-hour of the morning period in which ridership is at a maximum is 15,000 cars with the time gap alternative of 0.8 sec and 17,000 cars with the time gap alternative of 1.6 sec.

Depots

The PRT system has been provided with 45 depots for empty cars (see the section Empty Car Handling). The depots have the following functions:

- To secure the provision of empty cars to subareas in a way that gives short and guaranteed waiting times.
- To work as buffers for empty cars to minimize the sizes of the stations.

The sizes of the depots needed vary between 50 and 250 cars, depending on the size of the subarea (number of stations) it must serve. A more sophisticated control system in which not all empty cars must go via a depot should bring down the depot sizes.

Altogether the depots can house about 7,000 cars, which means a total track length of 25 km.

Stations

The 654 stations (one directional) are of various sizes, depending on the different needs for empty car buffers. Only 9 percent of the stations demand a car buffer of more than 10 cars. A detailed study of the depot function would probably reduce the size of the largest stations to this figure or less.

Night Storage of Cars

The outlined depots can house about 40 percent of the car fleet and the stations can house about 30 percent. The other 30 percent must be stored somewhere else at times when the PRT system is closed or is seldom used. The most economical way of doing this is to use the track sections in the network. Such sections can be one direction in two-way track links that are closed in a way that travel is not affected except for longer trips, because of rerouting around them. The track length needed for 5,000 cars is about 20 km, which can easily be found in the network.

AFTERNOON PEAK PERIOD

Travel times and trip lengths during the afternoon period are practically the same as those during the morning period.

The waiting times, both the average and the maximum, do not show any differences between the two time periods, nor do the delays because of high capacity use.

Ride sharing is somewhat higher in the afternoon period, probably because a higher percentage of the passengers will have their start stations in a concentrated central area. The average occupancy of the cars is, however, the same as that in the morning period because of the lower concentration of passengers at the rest of the stations.

Also in the afternoon the travel demand can be accommodated without any passenger queues at stations at the end of the peak hour.

The percentage of productive cars is the same as that in the morning period.

The number of transport missions per car hour in the afternoon peak period is slightly lower than that in the morning peak period.

The same total car fleet was used for the simulations in both time periods.

CONCLUSIONS

The following conclusions drawn from the studies presented in this paper and the discussions during the work can be made.

- A PRT system provides a travel standard that is clearly superior to a conventional public transport system. The travel time (excluding walking time) would be reduced by 50 percent or more. Walking times are estimated to be the same as those in the present system.

This superior travel standard is also created for origin-destination combinations in which the public transport service presently is poor, which makes PRT an acceptable alternative to car riding even if today's car travel times cannot be provided on all trips.

- A PRT system that covers most of the city can be given a capacity that allows an increase in public transport ridership by up to 80 percent, corresponding to a transfer of close to 60 percent of the present car riders in the city during peak hours.

This is estimated to cover the possible transfer of car riders because of the attractiveness of the PRT system and the number of people who can be encouraged to leave their cars by present or future traffic policy measures.

- A PRT system with the characteristics described above is large. It includes some 700 track km and more than 600 stations. As a comparison, the total length of the present tram and bus routes is about 600 km and the number of stops is somewhat greater than 600.
- The realization of such a system is expensive, because an automated system demands an infrastructure of its own that is completely separate from that for other traffic.
- A large PRT system will demand a high degree of reliability of all the components included, hardware as well as software. Estimation of the possibility of reaching the necessary reliability demands extensive studies and full-scale tests.
- The architectural aspects of the PRT system have not been studied in detail. Discussions, however, point to the difficulty of using elevated tracks in the central area of the city. Tunnels may have to be considered.
- From a land use point of view the effects of an increased and more evenly distributed accessibility are of a long-term nature.
- Present knowledge is not sufficient to implement a PRT system large enough to function as the only public transport in a city. An immediate decision is not necessary, however.

Publication of this paper sponsored by Committee on New Transportation Systems and Technology.

Vehicle Distribution in Large Personal Rapid Transit Systems

INGMAR ANDRÉASSON

Point-synchronous control for large personal rapid transit (PRT) systems that offers a possibility of increasing the link capacity for empty vehicles is described. Two different principles for the redistribution of empty vehicles are outlined. Depots for empty vehicles are introduced so that the passenger stations can be kept smaller. The methods were developed and tested on the basis of a simulation model for PRT systems. The model includes facilities for ride sharing and a number of control options. The described methods have been evaluated on tentative PRT networks for Gävle and Gothenburg, Sweden. The Gothenburg network has 654 stations, 4,200 links, and 12,800 vehicles. Simulation results indicate that point-synchronous control performs better than synchronous control, close spacing of empty vehicles can increase link capacity by up to 80 percent, depots can reduce station sizes to one-third and reduce call time to one-third, and waiting times are the same with or without depots.

A personal rapid transit (PRT) system provides on-demand transport for individual passengers or small parties. The service is direct between off-line stations in a network of tracks and switches.

The travel demand is generally unbalanced, particularly in the peak hours, with some stations serving mainly as origins and others mainly as destinations. Hence empty vehicles end up where they are not needed, with growing deficits at other stations. In a large network with long-distance commuting the task of balancing supply and demand of empty vehicles constitutes a major problem.

In a PRT system with high demand link capacity is a limiting factor and over half of the vehicle flow on many links may be empty vehicles. The time to call an empty vehicle may be very long. It is therefore necessary to call vehicles before they are needed on the basis of the predicted supply and demand at each station.

With random passenger arrivals and large call times vehicle buffers at stations would need to be quite large. Such large stations would be difficult to place in the city structure.

This paper describes methods for empty vehicle management first in systems in which stations are used as sole vehicle buffers and second in systems with separate depots for empty vehicles. So-called point-synchronous control is described. Point-synchronous control allows different speeds for and closer spacing of empty vehicles. A simulation model was used to develop and evaluate various control ideas. The conclusions are based on simulation results for a large tentative PRT network for the city of Gothenburg, Sweden.

POINT-SYNCHRONOUS CONTROL

Simulation models have been developed for synchronous control and what is called point-synchronous control. In *synchronous control* all vehicles move synchronously over the network and no vehicle is allowed to depart from a station before a free slot is available all the way to the destination.

With *point-synchronous control* a vehicle starts as soon as the exit is free and then modifies its path and speed as it goes along. A vehicle books its passage through the next node and then controls its speed on the basis of the remaining time and distance to its booked passage time (Figure 1). Booking points on two incoming links are located at equal distances to the merge point. The synchronization in each node ensures that conflicts are avoided and that the node capacity is used to its maximum.

Bookings and route choices are made successively and locally on the basis of the momentary situation. The control is decentralized to each vehicle and its next downstream node. Point-synchronous control allows for different speeds on different links, different speeds for different vehicles, and different slot frequencies for different nodes.

Point-synchronous control can also be viewed as a method of realizing collision avoidance in an asynchronous control system, ensuring synchronism at merge points.

PRT SIMULATION MODEL

Simulation models have been developed both for synchronous control and for point-synchronous control. The latter model is described here. A description of synchronous simulation has been provided elsewhere (*1*). The purpose of the models is to evaluate alternative PRT networks for given demand matrices and to study alternative control principles.

Passenger groups are generated randomly, with exponential intervals and group size being either fixed (typically 1.25 persons) or random, with probability p^n of group size n. Passengers interchanging from route services (buses or trams) can be generated in bursts at fixed intervals. Origins and destinations of trips are drawn from given matrices of probabilities. Passengers may be matched for ride sharing in various ways as discussed in the paper by Blide (this Record). Passengers are placed in queues at the departure station in the best direction of departure (stations on double-track sections).

Each vehicle is traced through the system as time is incremented in steps equal to the time slot between node passages (typically 1.6 sec). It gets successive slot allocations for passing each node and is given route directions on the basis of the quickest path in each diverge as described below.

Simulations start with vehicles distributed over stations according to expected demand. An initial period (typically 30 min) is simulated before statistics are collected.

LogistikCentrum, Taljegårdsgatan 11, S-431 53, Mölndal, Sweden.

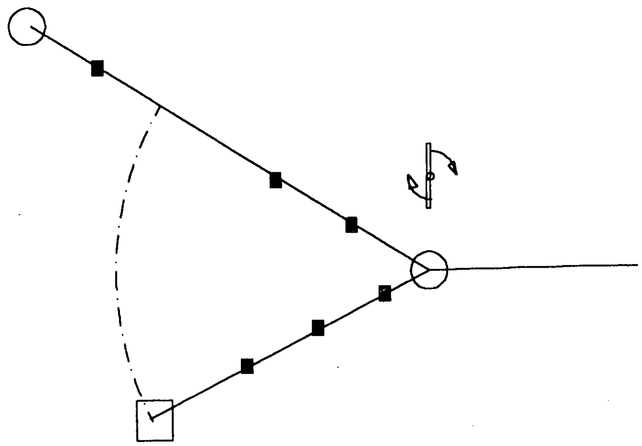


FIGURE 1 Point-synchronous control resembles pedestrians approaching self-revolving door.

The model has been programmed in Simula on a Macintosh computer. Graphics and animations are available.

POINT-SYNCHRONOUS VERSUS SYNCHRONOUS CONTROL IN GÄVLE

The same PRT network (Figure 2) for the city of Gävle (91 stations, 372 links, and 1,350 vehicles) was simulated both with synchronous control and with point-synchronous control.

The Gävle studies have been summarized previously (*1*). Further studies are being done for the system in Gävle to (a) recommend stages of implementation, (b) outline architectural design and integration into the street environment, and (c) analyze the socioeconomic costs and benefits of the possible implementation of a PRT system.

Simulation results reported previously (*2*) and summarized in Table 1 indicate that point-synchronous control performs better than synchronous control in all respects studied.

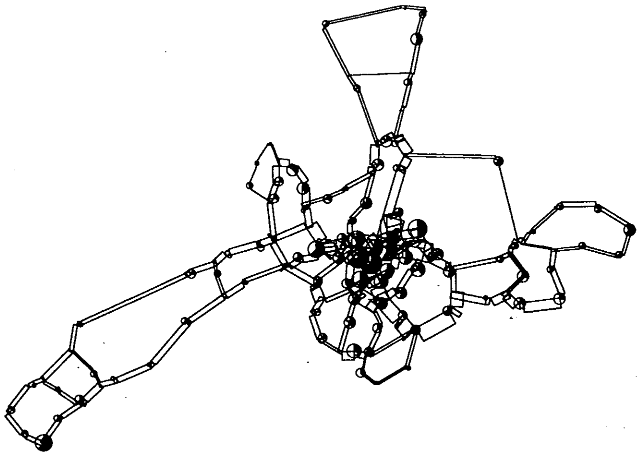


FIGURE 2 PRT network for Gävle showing passenger flow and use of stations (evening peak).

TABLE 1 Performance of Different Control Systems

	Synchronous control		Point-synchronous
Vehicle fleet needed	1346	vehicles	-4 %
Waiting-time	2.2	mins	-14 %
Riding-time	11.5	mins	-2 %
Maximum link flow	1700	veh/hr	+ 30 %

An important additional advantage of point-synchronous control is its robustness toward disturbances. A vehicle breakdown in a synchronous system would affect the overall plan, and each vehicle would have to be rescheduled. In a point-synchronous system all control is local and can immediately adapt to the new situation. A blocked link would cause the nearest upstream diverge to send all vehicles the other way. Diverges farther back would change their directions as soon as the path tables were recalculated.

CHOICE OF ROUTES

Each vehicle carries its destination and transmits that to the next downstream diverge controller. Each diverge controller keeps a table of the shortest running times to each destination. The diverge controller checks the latest observed running times to the following diverges downstream (left and right), adds the remaining running times (from tables) to the desired destination, picks the shortest alternative unless it is blocked, and transmits a left or right back to the vehicle.

The tables of running times are recomputed regularly and are based on statistics on link times including delays. A modified forward label correction method building the shortest path trees between nonstation diverges was used. Recalculation of all 2 million paths in Gothenburg by this method takes 160 sec on a Powerbook 170 (25 MHz).

PLATOONING OF EMPTY VEHICLES

Empty vehicles do not need a safety distance between them. They can be allowed to pick up on each other to form platoons so that several empty vehicles can pass through a node in the same time slot. This will increase the capacity of the system. Safety distances need to be ensured before and after loaded vehicles only. It is advantageous if empty vehicles group together, and the weaving algorithms are made to do that.

With 16-m slots (1.6 sec at 10 m/sec) and 3.5-m vehicles it is possible to pack four empty vehicles into the same slot. The theoretical link capacity increases from 2,250 to 9,000 vehicles/hr if all vehicles are empty. In simulations with a mix of loaded and empty vehicles link flows of more than 4,000 vehicles/hr (+80 percent) with a time slot of 1.6 sec have been observed.

With the link capacity increased by platooning the average simulated delays owing to congestion were reduced from 0.9 to 0.6 min per passenger trip (−30 percent).

WEAVING EMPTY VEHICLES IN MERGES

A weaving algorithm for merges has been designed to serve two purposes: (a) to maximize node capacity by weaving empty vehicles together and (b) to avoid queues spilling back past the upstream node, hindering flows in other directions.

In each merge and in each time step a vehicle is selected for passage by the following algorithm (selecting the first alternative whose condition is fulfilled):

1. If the outgoing link has no free space then the merge is blocked,
2. If one incoming link is empty then choose the other,
3. If one incoming vehicle has not arrived then choose the other link,
4. If the previous vehicle was loaded or both incoming vehicles are loaded then choose the link with the least free space behind the incoming queue (to avoid spillback of queues),
5. If the next vehicle on the link from where the previous (empty) vehicle came is also empty then choose that vehicle, or
6. Otherwise take the empty vehicle from the opposite incoming link (and add an extra space for weaving)

Weaving of empty vehicles and avoiding of queue spillback are important for the overall network capacity.

REDISTRIBUTION BETWEEN STATIONS

Various procedures for the redistribution of empty vehicles between passenger stations have been evaluated by simulations. They are based on predicted demand at each station and actual as well as predicted supply (including vehicles scheduled to arrive). Vehicles are called when the supply falls below a call level and are sent off when the supply exceeds a (higher) send level. These levels are individual for each station and are calculated from the travel demand matrix. Currently the following procedure is used:

1. For each station (and direction) the number of vehicles needed per minute is determined from the travel demand matrix. For stations with ride sharing (high demand) the vehicle need is reduced. The required fleet is calculated as the vehicles needed at stations to cover the demand during a given number of minutes (8 to 30 min depending on the average trip time).
2. Vehicles are initially placed at each station in proportion to its demand. The call level is specified as a fraction of the initial supply (typically about one-third).
3. As passengers are ready to depart from a station with a vehicle supply below the call level, an empty vehicle is called from the nearest station with a surplus (over its call level).
4. For each station a supply forecast is maintained:
    - Supply forecast = vehicles at the station + vehicles on their way in.
    - When a station gets overfilled it sends off a vehicle to the station with the lowest forecast supply in relation to its call level.

DEPOTS FOR EMPTY VEHICLES

Special parking areas for empty vehicles—so-called depots—have been introduced into the network so that the station buffers can be made smaller.

The distribution and redistribution of empty vehicles with depots have been made as follows:

1. The vehicle demand is calculated in the same way as it is without depots.
2. Vehicles are placed at each single station to cover the need only during the time it takes to bring an empty vehicle from the nearest depot. This initial supply is also the call level for that station. Remaining vehicle need for each station is placed at the nearest depot. The call level of a depot is specified as a fraction of its initial supply.
3. As passengers are ready to depart from a station with a vehicle supply below the call level an empty vehicle is called from the nearest depot.
4. For each depot a supply forecast is maintained:
    - Supply forecast = vehicles at the depot − unserved calls out + vehicles on their way in.
    - When a station gets overfilled it sends off a vehicle to the depot with the lowest forecast supply in relation to its call level. When a depot gets empty or the supply forecast falls below the call level then it calls a vehicle from the nearest other depot with a surplus (forecast exceeding the call level).

With the introduction of depots in the Gothenburg network the average call time for empty vehicles was reduced from 14 to 5

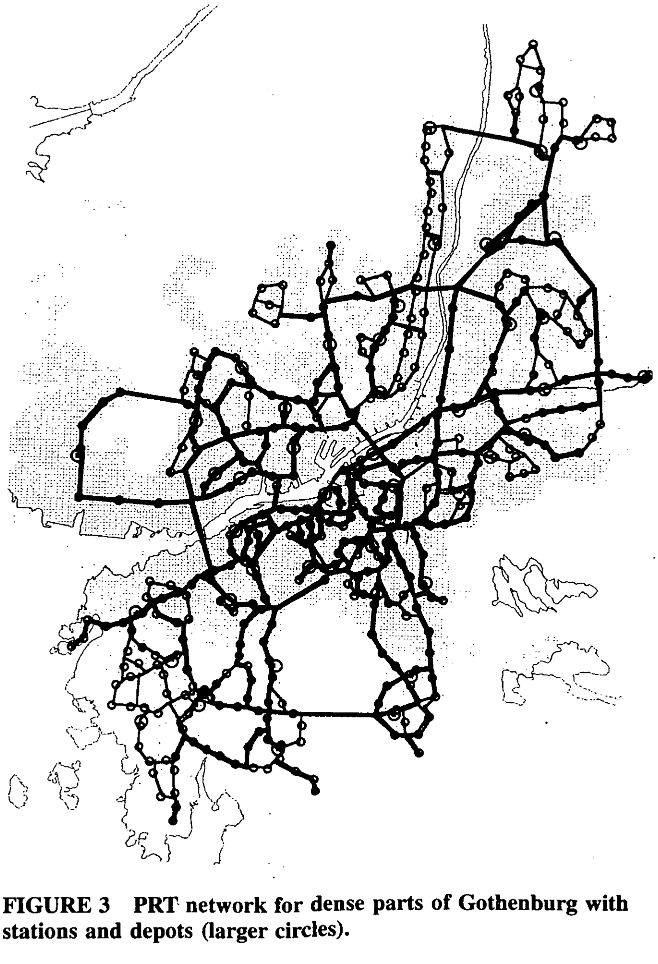


FIGURE 3 PRT network for dense parts of Gothenburg with stations and depots (larger circles).

min. Waiting times for passengers remained the same (average 1.3 min). Empty running mileage increased by 23 percent, so that the fraction of the fleet running empty at any time went up from 27 to 34 percent. Station sizes could be significantly reduced. Without depots the 90th percentile of station sizes was 21 vehicle births. With depots the same percentile went down to six vehicle births.

The average depot has room for 250 vehicles, which is needed to park 88 percent of the fleet at night. The remaining 12 percent of the fleet is parked at stations. During the peak half-hour 6 percent of the fleet is at stations and 22 percent is at depots, totaling 28 percent standing vehicles.

APPLICATION FOR GOTHENBURG

A tentative PRT network (654 single stations and 4,200 links) has been designed to cover the needs for the entire city of Gothenburg (see the paper by Blide, this Record). Into that network were introduced 45 depots for empty vehicles. The number of depots and their locations were not optimized. They were placed where space was thought to be available, with the objective of developing and testing the principle of depot management (Figure 3).

The results discussed in this paper are from a case with about 12,800 vehicles providing about 60,000 passenger trips in the peak hour (600,000 trips per day). Ride sharing is assumed between trips from the same origin going in the same direction (up to 3 min of waiting, two intermediate stops, and 30 percent detour for any passenger). Table 2 is a summary from one of the runs.

The PRT study for Gothenburg is now focusing on a smaller network for the central city. That network has about 30 km of track and 42 stations and is planned to supplement existing bus and tram routes.

CONCLUSIONS

- Point-synchronous control performs better than synchronous control.
- Closer spacing of empty vehicles can increase link capacity by up to 80 percent.
- Depots can reduce station size to one-third and reduce call time to one-third.
- Waiting times are the same with or without depots.

ACKNOWLEDGMENTS

The simulation model for synchronous control was developed with funding from the Swedish Transport and Communication Research Board, which also funded the applications for Gävle. The model for point-synchronous control and the applications for Gothenburg were funded by the Gothenburg Traffic Office. The

TABLE 2 Excerpt from Output of Gothenburg Simulation

```
SUMMARY GOTHENBURG PRT NETWORK version 4.1, 7.30-8.00 am.
   716 track kms
   654 single stations
  1396 diverges and equally many merges
 12795 cabs
   1.6 secs slot interval
    10 m/sec normal speed
    20 secs for loading + acceleration
     2 pass/min at station for sharing,    max  3.0 mins wait
    30 % acceptable detour for sharing,    max    2 intermediate stops
    30 initial minutes without statistics
    30 minutes in study period
 27191 passengers departed in study period
    23 % of stations have ride-sharing     with 65 % of passengers
    57 % matched departing share stations,    = 37 % of all
   1.3 minutes waiting for cab,            max 18.6  99 % < 3.0
  11.0 minutes riding,                     max 51.4
   0.6 mins congestion & stopping delay,   max  6.3
   6.3 kilometers average trip,            max 30.2
    34 kilometers/h average speed
   6.6 minutes per empty trip,             max 34.8
   4.2 kilometers per empty trip,          max 22.0
  12.8 minutes call time
  1.76 passengers per loaded cab
   4.3 passengers departed per cab hour
   2.1 loaded cabs departed per cab hour
    39 % of cabs running with passengers
    27 % of cabs running empty
    34 % of cabs waiting at stations or depots
   507 cabs/hour on average link,          max 2849
   467 passengers/hour on average link,    max 3556
 64131 vehicle kilometers empty
 85240 vehicle kilometers with passengers
150320 passenger kilometers
```

PRT networks for Gothenburg were designed under the leadership of B. Blide, project manager for the Gothenburg PRT study.

REFERENCES

1. Andréasson, I. *Simulation Studies of PRT Systems*. Report DNR 90-291-732. Swedish Transport and Communications Research Board.
2. Andréasson, I. Simulation of Large PRT Systems for Swedish Cities. *Proc., 3rd ASCE Conference for APMs*, 1993.

Publication of this paper sponsored by Committee on New Transportation Systems and Technology.

Overview of Automated People Mover System in Taipei

L. DAVID SHEN AND SHI-SHENG LAI

The $920 million automated people mover (APM) system in Taipei, Taiwan, Republic of China is part of an ambitious $17.4 billion Taipei Rapid Transit System. The 11.6-km, 12-station APM system was scheduled to open in early 1994. The fully automated, driverless, rubber tire on concrete track system is based on the French VAL system design with a maximum capacity of 30,000 passengers per hour per direction. The minimum headway is 60 sec, and the top speed is 80 km/hr (50 mph). It is estimated that 27,340 passengers will use the Taipei APM system during the peak hour in the year 2001. The peak-hour traffic is expected to more than double to 55,900 passengers in 2021. Approximately two-thirds of the passengers on the Taipei APM system are expected to walk to and from the APM stations. This is the first time that a medium-capacity trunk-line APM system has been designed to accommodate significant traveling needs in a major city of more than 2 million people. If the operation turns out to be successful, similar applications of APM as a medium- to heavy-capacity trunk-line service could be expected in other cities.

Construction of the $920 million automated people mover (APM) system in Taipei, Taiwan, Republic of China (ROC), is almost finished. The system was scheduled to open in early 1994. It is part of the ambitious $17.4 billion Taipei Rapid Transit System (TRTS) (*1*). The 11.6-km 12-station APM system is the only medium-capacity transit system of TRTS. The fully automated, driverless, rubber tire on concrete track system is based on the French VAL system design with a capacity of 5,000 to 30,000 passengers per hour per direction (*2*).

Taipei, the provisional capital and largest city of the ROC on Taiwan, has undergone dramatic economic growth in the last decade. With the highest foreign exchange reserve ($86.5 billion) in the world and one of the highest standards of living in Asia, Taiwan has slowly transformed itself from a predominantly agricultural society into a major industrial power. As a result Taipei's population has increased a minimum of 2 percent per year since 1978, when it already had a population of slightly more than 2 million (*3*).

The increase in automobile ownership during the same period of time, however, has been phenomenal. In 1980 Taipei had 182,000 cars. The figure multiplied threefold to more than 570,000 cars in 1992 and continues to grow at an alarming rate of 5,000 cars per month. In addition it has more than 860,000 motorcycles, and the number keeps on growing at a rate of 7,000 new motorcycles per month.

The concentration of people and vehicles in this densely populated city of only 274 km^2 has created a nightmare for its 2.7 million residents. Traffic jams are so common that city streets look more like parking lots than roads during rush hours. Taipei is rated as one of the busiest and most congested cities in Asia, and probably the world.

To alleviate traffic congestion, pollution, and the strain on inadequate transportation facilities in this bustling city, the government of the ROC has started an ambitious mass rapid transit system project with four high-capacity rapid rail transit lines and one medium-capacity APM line to improve traffic flow in the Taipei metropolitan area (population, 6 million; area, 1,824 km^2). The Taipei APM system is going to be the first medium-capacity automated trunk-line APM service ever been built in a major city of more than 2 million people. The purpose of this paper is to present an overview of this challenging project.

BACKGROUND

With a population density of 9,854 people per km^2 (25,226 people per mi^2), Taipei has one of the highest population densities in the world. Uneven development and overconcentration of population are major causes of excessive growth in the Taipei metropolitan area. In the last 2 years the value of urban properties has inflated an average of three times, and land value has quadrupled in Taipei. For example land in central Taipei is now worth $2,000/$\mathrm{ft}^2$, and today the cost of an apartment averages about $900/$\mathrm{ft}^2$ (*4*).

The kind of traffic found in Taipei would make commuters in Los Angeles decide to stay home for a day of television rather than brave the chaos. Sorting out the day-to-day traffic of a city of 3.5 million commuters, over half a million cars, 4,200 buses (operating on 300 routes), 860,000 motorcycles, and 36,000 taxis is an impossible task for any transportation engineer (*5*;*6*,p.111). A severe lack of parking spaces, crowded narrow roads, ubiquitous traffic accidents, endless traffic jams, and the smog created by the uncontrolled growth of motor vehicles are just a few items in the endless list of urban headaches in Taipei. When citizens are asked questions about political, economic, or diplomatic issues, chances are there will be as many opinions as there are people. Yet change the topic to traffic and there will be a universal agreement: the place is a mess.

In an effort to address the above growth-related problems the government of the ROC is implementing the large-scale Six-Year National Development Plan totaling $300 billion to improve the infrastructure systems on the island for economic development into the 21st century. Nearly one-third of the expenditures will be spent on transportation projects. The $920 million Taipei APM system is part of this ambitious plan and is scheduled to be the first rapid mass transit system to open. The planned route of the Taipei APM system is shown in Figure 1. The abbreviation for the APM system is BR, which stands for brown line. Because of

---

L. D. Shen, Department of Civil and Environmental Engineering, Lehman Center for Transportation Research, Florida International University, The State University of Florida at Miami, Miami, Fla. 33199. S.-S. Lai, Department of Rapid Transit Systems, 9th Floor, 16, Nan-King East Road, Section 4, Taipei, Taiwan, 105 Republic of China.

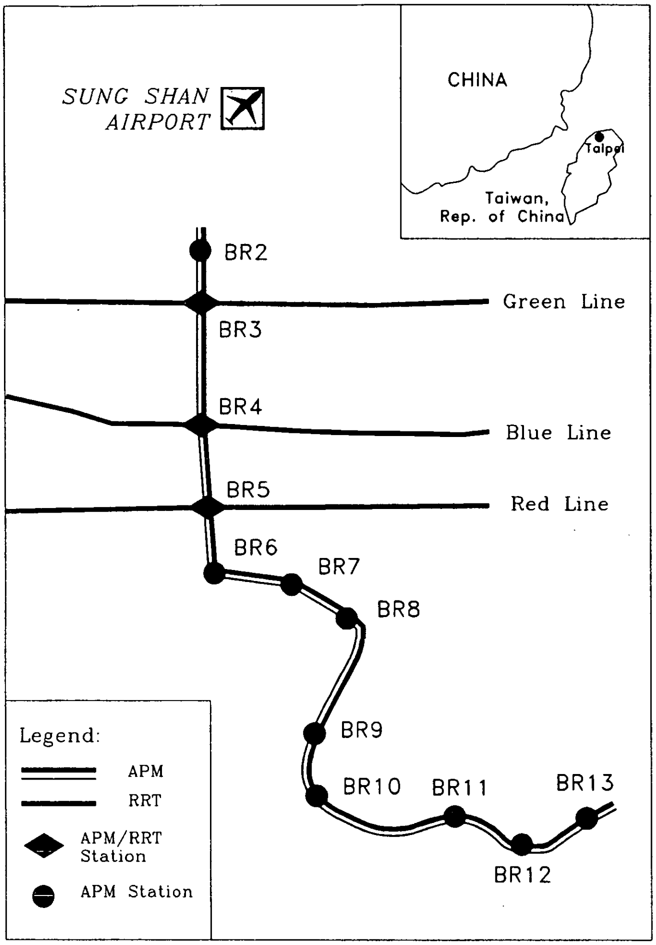


FIGURE 1 Planned route of Taipei APM system.

construction and testing problems the original opening date had been pushed from 1992 to 1994.

This is the first time that a medium-capacity trunk-line APM system has been designed to accommodate significant traveling needs in a major city: up to 30,000 passengers per hour per direction. If Taipei's APM system turns out to be successful a new chapter in mass transit is certainly going to be recorded.

SYSTEM

Except for a 0.8-km tunnel section the remaining 10.8 km of the Taipei APM system is elevated. The vertical clearance of the guideway structure is between 5 and 7 m. The span between the cast-in-place supporting columns is 25 m long. Because of the high cost of land, nearly the entire Taipei APM system is elevated over the medians of existing major arterials. For route design the minimum horizontal radius is 35 m and the maximum vertical grade is 6 percent (*7*).

The air-conditioned APM vehicle is 2.56 m wide and 3.51 m high (interior height is 2.045 m). The length of the two-car train is 27.56 m. The vehicle is powered by two 120-kW direct current motors on 750 V. The Taipei APM system is designed to operate in four-car consists. The normal capacity (4 people per m^2) for each car is 24 seated passengers and 60 standing passengers. The crash capacity (6 people per m^2) for each car is 24 seated passengers and 101 standing passengers (*2*).

The driverless vehicle is fully automated. The minimum headway is 60 sec. The maximum capacity is 30,000 passengers per hour per direction. The average operating speed is 34 km/hr. The cruising speed is 60 km/hr. The top speed is 80 km/hr (*2*).

All station platforms are of lateral design. The $263 million vehicles and system contract awarded to MATRA of France calls for 102 air-conditioned, wide-body VAL 256 vehicles to be operated initially in four-car consists; to accommodate future traffic all stations are designed for six-car consists (*8*). Elevators and escalators are provided at all stations to facilitate the use of the stations for disabled people. There are two types of station design (*9*).

1. *Type I: Elevated station over the street median.* The entrance and exit are located at the building adjacent to the roadway, and the building is connected to the station by an enclosed pedestrian bridge. For passengers who want to change the direction of their travel, they must take the elevators or escalators to the third floor, which is above the track, to complete their changes. Station BR 6 at the Science and Technology Building is an example of Type I design (Figure 2). Except for Stations BR 9 and BR 12, all stations are of Type I design. The cross-section diagram for Station BR 12 (Mucha Station) is shown in Figure 3. The profile diagram for Mucha Station is shown in Figure 4.

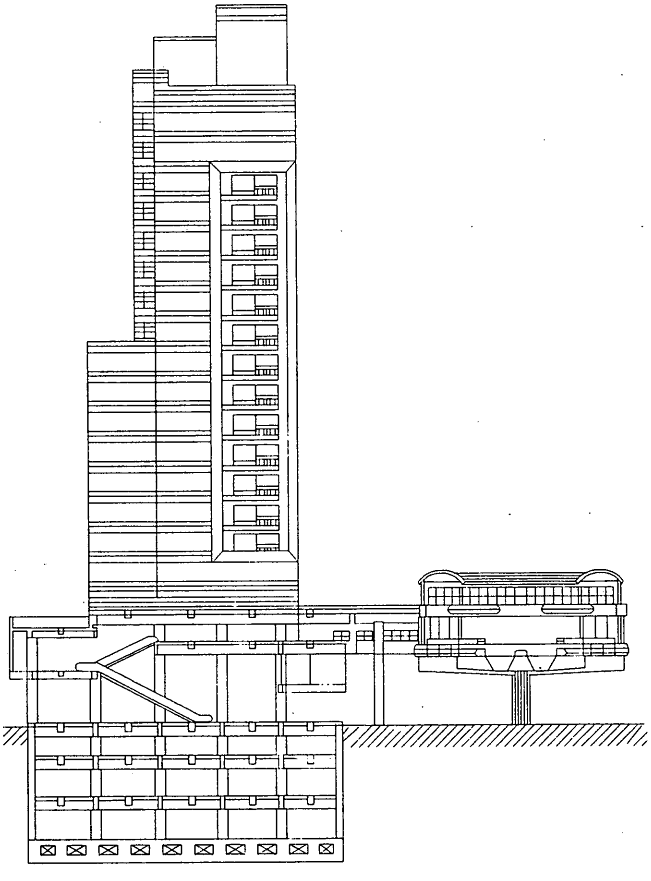


FIGURE 2 Cross-section diagram of Science and Technology Building Station.

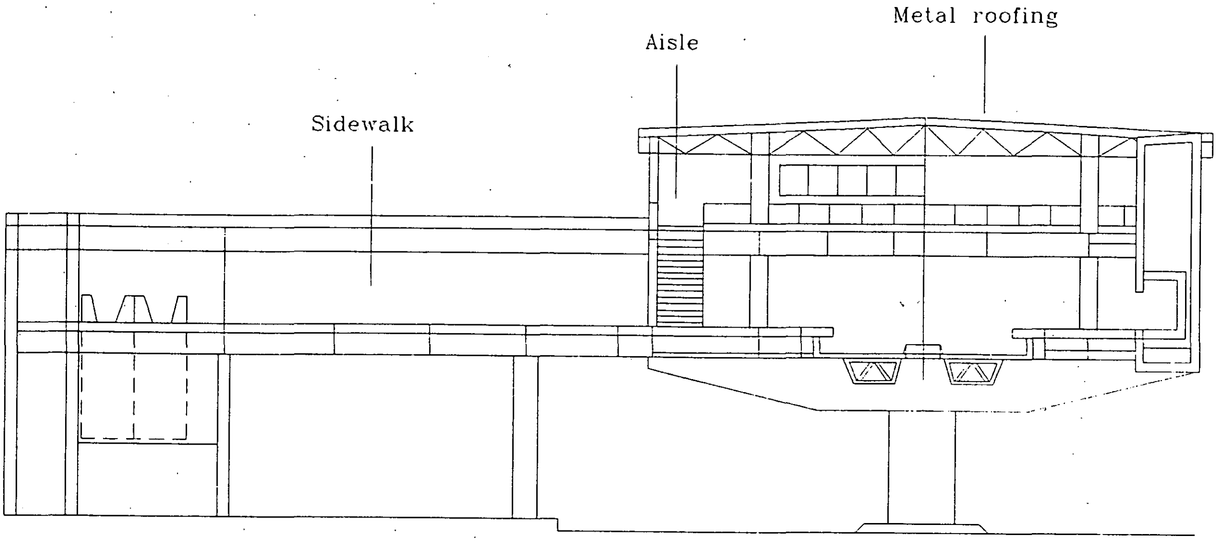


FIGURE 3 Cross-section diagram of Mucha Station.

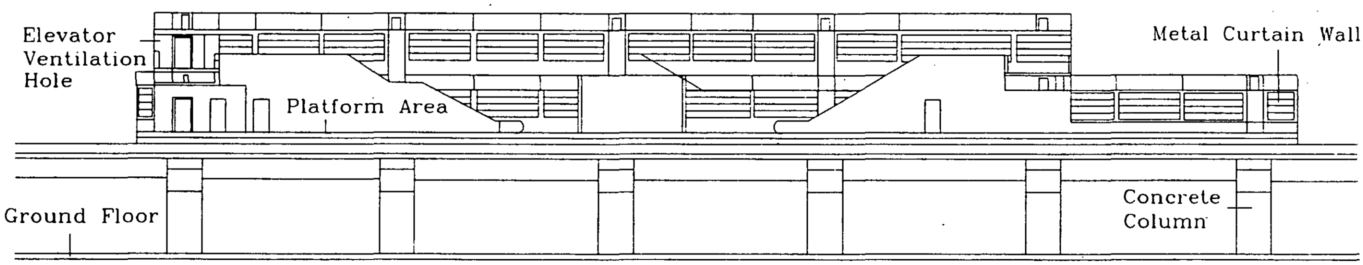


FIGURE 4 Profile diagram of Mucha Station.

2. *Type II: Integrated station.* The entrance and exit are located at the ground floor of the same station building. Stations BR 9 and BR 12 are of such design.

It is estimated that 27,340 passengers will use the Taipei APM system during the peak hour in 2001. The peak-hour traffic is expected to more than double to 55,900 passengers in 2021 (*7*). This kind of heavy peak-hour traffic is unheard of for any existing APM system. However because of the high population density and extremely crowded streets in Taipei, these figures have a high possibility of becoming reality. The peak-hour passenger traffic forecast for the 12 stations in Taipei's APM system is shown in Table 1.

When it opens in 1994, approximately two-thirds of the APM system's passengers are expected to walk to and from the APM stations. The remaining one-third are expected to use bus, taxi, park-and-ride, and kiss-and-ride as feeders for this medium-capacity transit system (*7*). The proposed operating schedule for the Taipei APM system is shown in Table 2 (*10*).

SUMMARY

The $920 million Taipei APM project is only part of a bigger government plan—the $300 billion Six-Year National Development Plan—that is looking ahead to Taiwan's infrastructure requirements for economic development into the 21st century. This is the first time that a medium-capacity trunk-line APM system has been designed to accommodate the significant travel needs of a major city of more than 2 million people. Because of construction and testing problems the original opening date had been pushed from 1992 to 1994. However if the operation is successful similar applications of APM as medium- to heavy-capacity trunk-line service could evolve in other cities.

TABLE 1 Peak-Hour Passenger Traffic Forecast for Taipei APM System

Station Name	Station No.	Year	Entering Passengers	Exiting Passengers
Sun Yat-Sen Jr. High	BR 2	2001	680	1,290
		2021	6,600	17,800
Nanking E. Rd.	BR 3*	2001	4,130	8,460
		2021	5,900	18,600
Chunghsiao E. Rd.	BR 4*	2001	1,440	2,870
		2021	4,900	14,200
Ta-An	BR 5*	2001	2,210	2,580
		2021	6,900	10,600
Science-Tech Bldg.	BR 6	2001	3,170	2,970
		2021	5,200	5,100
Liuchangli	BR 7	2001	480	440
		2021	4,400	2,200
Linkuang	BR 8	2001	3,040	960
		2021	4,800	2,800
Hsinhai	BR 9	2001	850	240
		2021	1,500	500
Wanfan Hospital	BR 10	2001	6,150	2,550
		2021	9,300	2,900
Wanfan Community	BR 11*	2001	1,200	100
		2021	1,800	400
Mucha	BR 12	2001	2,840	1,300
		2021	2,400	600
Taipei City Zoo	BR 13	2001	1,150	510
		2021	2,200	1,400
	Total	**2001**	**27,340**	**24,270**
		2021	**55,900**	**77,100**

* Transfer station to Rapid Transit (RRT) lines.

TABLE 2 Proposed Operating Schedule for Taipei APM System

Station to Station	Distance (meters)	Travel Time (seconds)	Dwell Time (seconds)
Northbound Train			
BR 13 - BR 12	680.12	60.20	17.00 (BR 13)
BR 12 - BR 11	514.02	49.20	17.00 (BR 12)
BR 11 - BR 10	1137.64	98.10	17.00 (BR 11)
BR 10 - BR 9	758.19	101.30	17.00 (BR 10)
BR 9 - BR 8	1598.09	118.70	17.00 (BR 9)
BR 8 - BR 7	821.42	65.50	17.00 (BR 8)
BR 7 - BR 6	1132.43	113.70	17.00 (BR 7)
BR 6 - BR 5	745.00	59.50	17.00 (BR 6)
BR 5 - BR 4	891.99	67.40	17.00 (BR 5)
BR 4 - BR 3	1268.01	86.80	17.00 (BR 4)
BR 3 - BR 2	931.00	69.50	17.00 (BR 3)

Station to Station	Distance (meters)	Travel Time (seconds)	Dwell Time (seconds)
Southbound Train			
BR 2 - BR 3	931.00	69.40	17.00 (BR 2)
BR 3 - BR 4	1267.99	86.70	17.00 (BR 3)
BR 4 - BR 5	892.01	67.60	17.00 (BR 4)
BR 5 - BR 6	745.00	59.70	17.00 (BR 5)
BR 6 - BR 7	1135.57	113.40	17.00 (BR 6)
BR 7 - BR 8	821.59	65.70	17.00 (BR 7)
BR 8 - BR 9	1593.06	119.60	17.00 (BR 8)
BR 9 - BR 10	763.71	103.70	17.00 (BR 9)
BR 10 - BR 11	1140.32	100.90	17.00 (BR 10)
BR 11 - BR 12	513.01	48.10	17.00 (BR 11)
BR 12 - BR 13	683.28	61.50	17.00 (BR 12)

REFERENCES

1. Shen, L. D. An Ambitious Six-Year Development Plan for the Future in Taiwan. *ITE Journal*, May 1993, pp. 18–22.
2. *Medium Capacity Rapid Transit System—An Introduction.* Department of Rapid Transit Systems, Taipei Municipal Government, Taipei, Republic of China, 1991.
3. Shen, L. D. Taipei: Rapid Transit on the Rim. *Mass Transit,* Sept./Oct. 1992, pp. 68–70.
4. Yeung, I. This Land is My Land. *Free China Review,* Vol. 41, No. 8, Aug. 1991, pp. 4–11.
5. Hwang, J. Sliding Toward Gridlock. *Free China Review,* Vol. 41, No. 3, March 1991, pp. 4–13.
6. Taipei Rapid Transit Systems and Traffic Congestion. *The Journalist,* Dec. 6–12, 1993.
7. *Medium Capacity Rapid Transit System.* Department of Rapid Transit Systems, Taipei Municipal Government, Taipei, Republic of China, 1991.
8. *Taipei APM System.* MATRA Transport, Montrouge Cedex, France, 1991.
9. *Taipei Transit—Brown Line Medium Capacity Rapid Transit System.* Department of Rapid Transit Systems, Taipei Municipal Government, Taipei, Republic of China, 1990.
10. *Taipei Transit—Brown Line Proposed Schedule.* Department of Rapid Transit Systems, Taipei Municipal Government, Taipei, Republic of China, 1991.

Publication of this paper sponsored by Committee on New Transportation Systems and Technology.