Structure of Competitive Transit Networks

Carlos F. Daganzo
WORKING PAPER
UCB-ITS-VWP-2009-6

UC Berkeley Center for Future Urban Transport
A **VOLVO** Center of Excellence

August 2009

STRUCTURE OF COMPETITIVE TRANSIT NETWORKS

Carlos F. Daganzo

Institute of Transportation Studies

University of California, Berkeley, CA 94720

(August 6, 2009)

ABSTRACT

This paper describes the network shapes and operating characteristics that allow a transit system to deliver a level of service competitive with that of the automobile. To provide exhaustive results for service regions of different sizes and demographics, the paper idealizes these regions as squares, and their possible networks with a broad and realistic family that combines the grid and the hub-and-spoke concepts. The paper also shows how to use these results to generate master plans for transit systems of real cities.

The analysis reveals which network structure and technology (Bus, BRT or Metro) delivers the desired performance with the least cost. It is found that the more expensive the system's infrastructure the more it should tilt toward the hub-and-spoke concept. Both, Bus and BRT systems outperform Metro, even for large dense cities. And BRT competes effectively with the automobile unless a city is big and its demand low. Agency costs are always small compared with user costs; and both decline with the demand density. In all cases, increasing the spatial concentration of stops beyond a critical level increases both, the user and agency costs. Too much spatial coverage is counterproductive.

1. INTRODUCTION

This paper examines the structure of urban transit systems that can deliver a level of service comparable to that of the automobile and the character of the cities in which this can be done at reasonable cost. These transit networks should provide good service between every pair of points in the city throughout the day, and be easily understood by the public. Only if they do this will they encourage auto users to leave their cars home when their daily plans include complex trip chains with impromptu and non-routine links.

To achieve these standards, transit systems must uniformly cover the service region in space and time with well-spaced transit stops and frequent reliable service. Good spatial coverage limits the walking time to/from every point in the service region, and good time coverage the waiting and transfer times.

Coverage should be dense enough to ensure that the sum of these times for any trip is comparable to the time that auto travelers spend walking to or from their cars, looking for parking and inside garages; i.e., not more than about 10 minutes. The system should also have a structure that guarantees in-vehicle-travel times comparable to those of auto trips; and just as importantly be cost-competitive and reliable. If all these criteria are met, public transit can become a viable alternative to the automobile because most auto trip chains could then just as well be completed by transit.

Other authors have asked structural transit questions. For example, Holroyd (1965) examined grids, Newell (1979) a hub and spoke system where the hub was a large street, and Wirasinghe et al (1977) corridors. These works determined the stop spacing, service frequency and, where appropriate, the line spacing that best balances the generalized user cost with the cost of providing service. Unfortunately, the structural families analyzed in these and other papers are too narrow to answer the type of question posed here. A more general family is needed.

Section 2 below describes the new family as well as the basic notation and assumptions. Section 3 expresses the systems' measures of performance in terms of key system descriptors. Then, using these formulae Section 4 formulates the design problem as a constrained optimization problem that answers our questions quantitatively. Section 5 discusses the results qualitatively, suggests model refinements, and proposes a way to develop master plans for real cities.

2. THE HYBRID CONCEPT

The service region is a square of side D (km) that generates Λ passenger trips per hour (p/hr) during the rush hour and an average of λ p/hr during the day's hours of service. Its origins and destinations are uniformly and independently distributed over the service region. The uniformity assumption is appealing because it does not require any extra parameters. Although it penalizes transit, compared with more centripetal distributions, this is a good thing because it sets a high bar for success.

Service is provided by buses or trains which are characterized by: their design capacity, C (p); their cruising speed including stops due to traffic and pedestrian interferences v (km/hr); the time lost per stop due to the required door operation, deceleration and acceleration, τ (hr/stop); and the time added per boarding passenger τ' (hr/p). We assume that the buses or trains can be operated at the cruising speed with only moderate variability in the headways; see Daganzo (2009) for a discussion of control methods.

Transit routes must lie on streets, which form a fine square mesh parallel with the square's sides. To provide uniform spatial coverage the square is covered with a square lattice of stops. This lattice is also oriented with the sides of the square, and its spacing is s (km). To provide good temporal coverage the service headway is assumed to be H (hr) in the central part of the square where all the transfers will take place, but it is allowed to be higher along its fringe. (Values of our coverage variables comparable with s = 0.5 km and H = 3 min seem reasonable if our transit system is to be competitive with the automobile.)

A focus of this paper is the system's structure and layout. The key choice on layout is whether each stop is covered by one or two lines. Figure 1a shows a hub and spoke pattern with trunk and branch that provides single coverage everywhere. Only one fourth of the square is shown because the other three wedges are obtained by rotating the picture 90°, 180° or 270°. Note each stop is associated with s km of infrastructure if one ignores the few stops where lines branch. Figure 1b shows a grid pattern that provides double coverage. Here, every stop is covered by two lines and is associated with $2s$ km of infrastructure. Therefore, the grid system includes more kilometers of line and is more expensive to operate. However, it also provides a higher level of service since users can travel from any origin stop to any destination stop without detouring from their shortest paths. Thus, the grid system should be preferred if enough people can share the rides to pay for its higher cost; i.e., if λ is high enough.

The choice does not have to be black or white, however. Figure 1c displays an example of a hybrid system that provides double coverage in a *central square* of side $d \le D$ and single coverage in its *periphery*. We shall see in the next sections that by varying the value of the ratio $\alpha \equiv d/D$ from $\alpha = s/D$ (hub and spoke) to $\alpha = 1$ (grid) we can find structures that outperform both extremes.

The hybrid system is intuitive and easy to navigate. For maximum transparency, maps could use a coordinate system to number the lines. For example, N-S lines could be numbered from left to right with letters (A, B, C…) and offshoots from their trunks also from left to right, Aa, Ab, … Ba, Bb, …etc. Likewise, its E-W lines could be numbered from top to bottom (1, 2, 3…), and their offshoots 1a, 1b,… 2a, 2b, … etc.

3. ANALYSIS

Formulae for key performance metrics of the hybrid network are now presented.

3.1 Agency metrics

Important agency metrics are: the total vehicular distance traveled per hour of operation, V (veh-km/hr); the vehicle hours traveled during the rush hour, M (veh); the infrastructure length L (km); and the peak vehicle occupancy during the rush hour, O (p/veh). The parameters V and M correlate with the agency's operating cost (M is also the required fleet size in operation); L correlates with the fixed capital costs; and O with the required vehicle size. The following approximate formulas are derived in Appendix A:

$$L = [D^2/s][1+\alpha^2]; \quad (1)$$

$$V = [2D^2/sH][3\alpha - \alpha^2]; \quad (2)$$

$$M = [V/v_c]; \tag{3}$$

where v_c is the vehicles' commercial speed including stops during the rush, which is given by:

$$1/v_c = [1/v + \tau/s] + [\tfrac{1}{2}\tau' \Lambda s H/D^2]/[3\alpha - \alpha^2]; \text{ and} \tag{4}$$

$$O = [\Lambda s H/D]\,[\max\{\tfrac{1}{2}(1-\alpha^2)/\alpha\,;\ (3-\alpha^4)/8\alpha + (D/s)(1-\alpha^2)^2/32\}]. \tag{5}$$

Equation (1) is based on the fact that there are $(D/s)^2$ stops, each consuming s km of infrastructure in the periphery and $2s$ in the center; and (3) on the fact that a fleet of M vehicles traveling at an average of v_c km/hr must jointly cover a distance of $V = Mv_c$ km in an hour. Note how all the cost measures (1, 2 and 3) increase with α as we had anticipated. Appendix A also shows that the central grid does an excellent job of distributing bus loads within the center. The critical load point is either a link feeding the center from the periphery, represented by the first argument of the "max" function in (5), or a link on the edge of the center itself and represented by the second term.

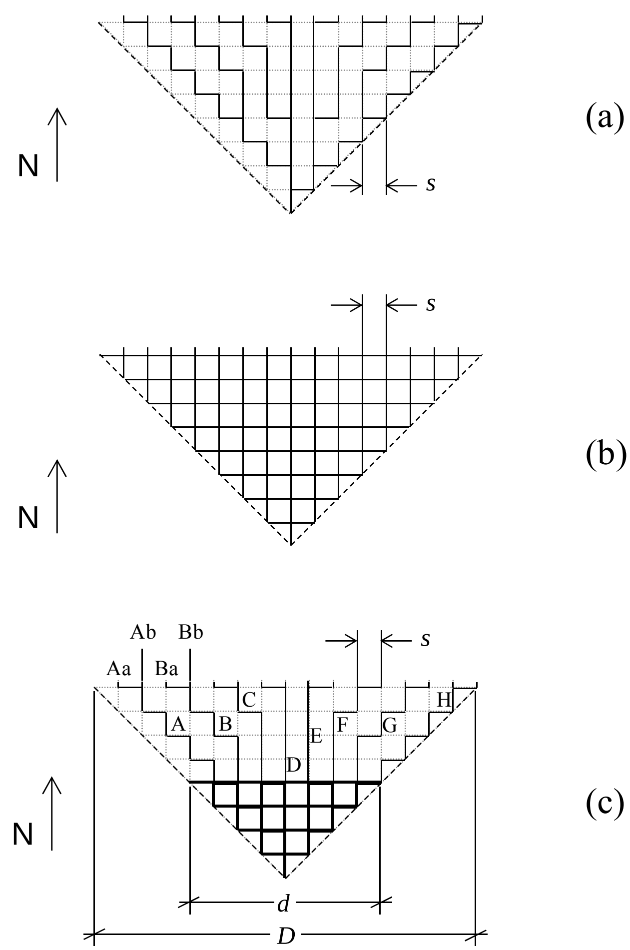

Figure 1. Possible system layouts: (a) hub and spoke; (b) grid; (c) hybrid.

3.2 User metrics.

Important user metrics (averages per passenger) are: the waiting time, W (hr); the walking access time, A (hr); the in-vehicle-travel-distance (km) E, and time, T (hr); and the expected number of transfers, e_T. We are also interested in the total passenger distance traveled per hour, P (p-km/hr) because the latter is proportional to the fares that can be extracted.

These user metrics depend on how transit users choose routes. Therefore it is assumed that people: (i) use the closest stops to their origins and destinations; (ii) travel between these stops with the least possible number of transfers and as directly as possible; (iii) transfer at the first opportunity; and (iv) randomly choose the initial direction of travel when both their origin and destination are in the center. These heuristics are simple, realistic and yield unique paths with near minimum travel times. Optimum rules are too complex for people to use and do not significantly affect the macroscopic properties of the final assignment. The following approximate formulas are derived in Appendix A:

$$A = s/w, \quad \text{where } w \text{ is the "design" walking speed.} \tag{6}$$

$$W = [H]\,[(2+\alpha^3)/(3\alpha) + (1-\alpha^2)^2/4] \qquad \text{for } \alpha > s/D, \tag{7}$$

$$E = [D]\,[2/3 + (1-\alpha)^3(4+5\alpha+3\alpha^2)/12], \tag{8}$$

$$T = E/v_c \approx [D(1/v + \tau/s)]\,[2/3 + (1-\alpha)^3(4+5\alpha+3\alpha^2)/12], \tag{9}$$

where the approximation holds if the third term in the expression for $1/v_c$ can be neglected,

$$e_T = 1 + ½(1-\alpha^2)^2, \text{ and} \tag{10}$$

$$P = \lambda E = [\lambda D]\,[2/3 + (1-\alpha)^3(4+5\alpha+3\alpha^2)/12]. \tag{11}$$

Note, for $\alpha = 1$ the equations reduce to what is expected for the case of grids.

4. TRANSIT NETWORKS FOR DIFFERENT CITIES

4.1 Formulation

Here we look for the values of s, H and α that minimize the sum of the agency costs and the generalized user costs when these costs are weighted appropriately. Without loss of generality, we express this sum, z, in terms of passenger riding time (hr). We reduce walking time to riding time by using a low walking speed, w = 2 km/hr, instead of the walking speeds observed in reality, which are somewhere between 5 and 6 km/hr. This low value recognizes both the discomfort associated with walking and the natural delays that pedestrian encounter when crossing streets. On the other hand, we do not value differently waiting and riding time because the waits of our system should be low and predictable, and therefore not annoying. We convert agency money into passenger riding time with a parameter, μ (\$/hr), that cannot be precisely defined because it balances two different interest groups. Despite its political nature, μ should not be too different from the prevailing wage. Therefore, we use μ = 20 (\$/hr) in our examples. We assume that infrastructure costs and operating costs per hour are of the form $\$_L L$ and $\$_V V+\$_M M$, respectively, where the coefficients $\$_L$ (\$/km-hr), $\$_V$ (\$/veh-km) and $\$_M$ (\$/veh-hr) are known unit costs.

Cities are characterized by their dimension D and their average trip generation rate λ (p/hr). The peak factor Λ/λ is assumed to be $\Lambda/\lambda = 2.5$ in all our examples. Three modes will be considered: Bus, BRT, and Metro. The bus system is operated in traffic; the BRT system is segregated from traffic and benefits from signal priority; the Metro system has a totally separate right of way and higher peak speed but is operated underground. We assume that the bus and BRT systems use advanced fare collection methods so that the passenger-dependent part of the average stop time is small; we use $\tau' = 1$. The systems' key characteristics are summarized in Table I.

Table I assumes in agreement with observation that Metro stops take significantly longer, so τ is greater for Metro. This happens not just because of Metro cruises faster, but also because it uses rail which requires more gradual changes in speed. The cruising speed v for the bus system assumes that the city is congested. The parameter δ is a fixed penalty for transfers expressed in terms of an equivalent distance walked. Whereas bus transfers involve minimal walking given the structures that emerge from analysis, Metro transfers do not. So we use a larger value of δ for Metro. A fixed access and egress penalty should also be applied to the Metro system to reflect the inconvenience of traveling underground but this is not done. So we are slightly favoring Metro in our analysis. The last two columns show rough values for the amortized infrastructure cost and the vehicle operating cost.[1]

To formulate the optimization problem, convert our proxies for operating cost, V (veh-km/hr) and M (veh-hr/hr), into hours of riding per passenger. Do this by means of the weights: $\pi_V = \$_V/(\lambda\mu)$ and $\pi_M = \$_M/(\lambda\mu)$. Likewise, to convert our proxy for infrastructure cost, L (km), into the same metric, multiply it by $\pi_L = \$_L/(\lambda\mu)$. And finally, to convert the expected number of transfers e_T into riding time, use the weight (δ/v_W). Therefore, the transit optimization problem is:

---

[1] For the former we use a straight-line amortization over 12 years of 365 days, assuming 18 operating hours per day. The capital cost of Metro is taken to be 72×10^6 \$/km, assuming extensive tunneling, and the cost of BRT 1/10th of the cost of Metro--although it could be less for forms of BRT that do not require separate stations as we propose here. As a point of reference, the Chengdu and Beijing Metros cost 450×10^6 RMB/km and the Beijing BRT 40×10^6 RMB/km (Nie and Sun, 2009). The Bus infrastructure is assumed to cost 1/10th of BRT's since its requirements are minimal. Unit operating costs include a component expressed per unit time (including driver wages and vehicle depreciation—mostly the former) and a component expressed per unit distance (including vehicle operation and maintenance). I was unable to locate reliable sources for Metros, but it seems reasonable to assume that the cost per km of running and maintaining a train is more than twice the cost of running a bus.

$$\min\{z = [\pi_V V + \pi_M M + \pi_L L] + [A + W + T + (\delta/w)e_T] : \ s \geq 0, H \geq 0, 0 \leq \alpha \leq 1, O \leq C\} \qquad (12)$$

Since this problem only has three decision variables it can be solved exactly for each of the three modes, for any given values of λ and D. Each solution includes the ideal network structure and operating frequency for each mode as well as the agency and user cost. As such, the solutions reveal how the three modes stack up against each other.

Table I: Modal characteristics

Modes	**C** (p)	**τ** (s)	**τ'** (s/pax)	**v** (km-hr)	**δ** (km)	**$\$_L$** (\$/km-hr)	**$\$_V$** (\$/veh-km)	**$\$_M$** (\$/veh-hr)
Bus	120	30	1	25	0.03	9	1	30
BRT	150	30	1	40	0.03	90	1	30
Metro	1000	45	0	60	0.2	900	3	40

4.2 Test

To test the quality of approximations (1) to (11), they were applied to data from Barcelona's bus network to see how their predictions matched some observed performance metrics. The demand side is characterized by Λ = 50,000 pax/hr, λ = 20,000 pax/hr and D = 10 km. The first row of Table I describes well Barcelona's existing supply side, except for the bus cruising speed which is v = 21.4 km/hr (Barcelona is quite congested) and $\$_L = 0$. The following structural variables best describe the existing network: s = 0.2 km, α = 0.88 and H = 12 min. The predicted (vs. observed) values for the three agency metrics for which data were available are: v_c = 11.1 (vs. 11.7) km/hr; L = 887 (vs. 895) km; and M = 839 (vs. 891). The corresponding operating cost is 1.8 \$/pax. The predicted average travel times are: A = 6 min (walk); W = 12.3 min (wait); T = 36 min (riding); and 54 min (total door-to-door). The travel time numbers assume reliable operations. So in reality they should be slightly larger. The numbers are consistent with typical travel times by bus in Barcelona. Barcelona's network has not been optimized, however. Therefore, let us now do so with (12).

4.3 Optimization results: Barcelona and other generic cities

Table II shows the result. Its last two columns are the components of the total cost corresponding to the agency and to the users' level of service. The table assumes that v = 25 km/hr for the Bus alternative since this is a more typical value than 21.4 km/hr. However, the agency costs do not change perceptibly if one uses the latter; and although the user cost increases by about 3 min, to 46min, this is still 8 min better than current performance in Barcelona. The BRT alternative reduces travel time even more, by 16 min to only 38 min.

Table II: Base scenario (Barcelona): λ = 20,000 pax/hr; D = 10 km.

Modes	α	s (km)	H (min)	O (pax)	M (veh)	v_c (km/hr)	z (min)	z_A (min)	z_U (min)
Bus	.89	.45	4.5	61	665	16.7	48	5	43
BRT	.77	.51	3.5	79	491	23.5	46	8	38
Metro	.45	.93	2.5	200	178	33.2	75	21	54

The table's first column is the network structure parameter. Its square $(\alpha^*)^2$ is the fraction of the city's surface area that should receive double coverage. This fraction is 80% for bus, 58% for BRT and 18% for Metro. Not surprisingly, it declines with the cost of infrastructure provision. The recommended spacing for bus service is also wider than typical. The reason is that increased spacing allows buses to travel faster and the reduced riding time compensates users for the longer access times. Also worth noting is that the frequency of service is quite high for all three modes, but not high enough to fill a dedicated lane with buses for the BRT alternative. Since room is available in this lane, other vehicles could be allowed to use it when not interfering with buses; see Eichler and Daganzo (2006) for a way to do this. Finally, note that both bus alternatives are significantly better than Metro.

BRT is also quite competitive with Auto. The door-to-door travel time by BRT is 38 min, vs. an estimated 26 min by automobile—or 28 min for Barcelona with its lower cruising speed.[2] The BRT agency cost (8 min or \$2.6) is also comparable with the mileage cost portion of an auto trip, which is \$2.2, using "zip car" rates. These numbers do not include parking hassles and fees, garaging costs, infrastructure use charges or congestion tolls. If to compare "apples" with "apples" infrastructure charges are excluded from the costs of both modes, and parking and garaging costs are included with both, then the auto mode can easily turn out to be 10 min (\$3) more expensive than BRT. So BRT is competitive.

These results are affected somewhat by our choices of technological constants, but not enough to change the qualitative conclusions. For the BRT alternative for example, the elasticities of z_A and z_U are: –0.14 and –0.24 with respect to v; +0.10 and +0.18 with respect to τ; and much smaller with respect to τ'.

[2] The auto trip optimistically includes only 10 min for finding and paying for parking plus walking at 2 km/hr at the two trip ends.

The values are qualitatively similar for the Bus alternative. The low impact of passenger boarding times is explained because with the short proposed headways so few passengers board a bus every km—only 7 for the BRT system and 9 for Bus—that their contribution to bus delays is minuscule: well under 4% of the travel time.

Tables III, IV and V illustrate the effects of demand intensity, sprawl and city size by displaying the same results for three alternative scenarios with: (a) 4 times the demand with no spatial growth, which could represent Barcelona five years from now if a system of the proposed type is deployed; (b) 4 times the area but the same demand, which could represent a more sprawled mid-size city in the US; and (c) 4 times the area and 4 times the demand, which could represent a large dense city such as Paris (France). Note, in all the tables user costs are several times greater than agency costs. This happens because agency costs can be reduced as much as desired by choosing arbitrarily large H and s, but user costs cannot be reduced below the time needed to overcome distance ($2D/3v$) no matter what one does.

Table III reveals economies of density. More demand in the same service region reduces both the fares required to pay for the system and the door-to-door travel times. The Metro system becomes more competitive, but BRT still "wins". The BRT system now gains a considerable edge over the automobile. Especially since a denser city would imply lower automobile speeds and higher parking charges, but lower agency costs as the table shows. Note too that the parameter that changes most from Table II is the service frequency, and that the geometric structural parameters change little. This is good because the latter are more difficult to adapt to changing demand. It means that a transit system should perform well over time as its demand changes if one just adjusts its service frequencies on the different lines. Thus, forecasting demand is not so important for design purposes.

Tables IV and V illustrate the effect of spatial growth, both with constant demand and with constant demand density. The geometric structural parameters now change considerably, which should not be surprising since trip lengths are proportional to D. For a given population, increasing city size lowers α whereas for a fixed population density, city size has the opposite effect. In both cases the stop spacings are larger than in Tables II and III since the city is bigger. Note too, that Metro is the least competitive technology in all scenarios and BRT the best. For the Big City, BRT produces a door-to-door travel time which is 12 min greater than the 42 min required with Auto. However, the BRT agency cost is also 7.5 min lower (\$5.2 − \$2.7 = \$2.5 = 7.5 min), and this is before subtracting from it the BRT infrastructure cost or adding to both alternatives the parking and garaging costs. Thus, BRT is still quite competitive in the Big City, albeit not as much as in the Small Dense City. Tables IV and V, however, also reveal that BRT suffers a 15 min (\$5) penalty if our Big City morphs into a Sprawled City; as a result, transit is no longer broadly competitive.

Table III: Increased demand scenario: λ = 80,000 pax/hr; D = 10 km.

Modes	α	s (km)	H (min)	O (pax)	M (veh)	v_c (km/hr)	z (min)	z_A (min)	z_U (min)
Bus	.92	.39	2	86	1862	15.8	43	3	40
BRT	.87	.40	2	105	1330	20.9	38	4	34
Metro	.72	.63	2	247	571	27.4	54	11	43

Table IV: Sprawl scenario (US city): λ = 20,000 pax/hr; D = 20 km.

Modes	α	s (km)	H (min)	O (pax)	M (veh)	v_c (km/hr)	z (min)	z_A (min)	z_U (min)
Bus	.84	.70	5	66	1305	19.1	80	11	69
BRT	.69	.82	4.5	109	743	27.9	77	16	61
Metro	.31	1.65	2.5	260	235	41.3	130	41	89

Table V: Big city scenario (París): λ = 80,000 pax/hr; D = 20 km.

Modes	α	s (km)	H (min)	O (pax)	M (veh)	v_c (km/hr)	z (min)	z_A (min)	z_U (min)
Bus	.92	.56	3.5	110	2627	17.8	71	5	66
BRT	.83	.61	3	150*	1882	25.1	62	8	54
Metro	.57	1.05	2.5	396	724	35.0	88	21	67

4.4 Comments: BRT effect on cars; BRT bus crossings; spatial coverage.

BRT's beneficial effect on cars: In all our examples BRT buses carry somewhere between 1300 and 3000 p/hr on the critical link, and not much less on other central links. This is a considerably greater flow of travelers than a city lane can carry with single-occupant cars. Thus, if a sufficiently high portion of these

BRT travelers had formerly used car, the BRT system would have created extra space for cars. This benefit can be magnified by opening the BRT lane to cars, provided they do not interfere with BRT buses.

BRT Bus crossings: We have assumed that BRT buses can cruise at 40 km/hr. This is reasonable in the periphery. Engineering calculations show that a speed v above 41 km/hr can be maintained by signal preemption with intersections as closely spaced as 100m if bus delay arises only at the signals and the bus speed limit is 45 km/hr. These calculations, however, ignore that BRT lines cross in the center and that bus-to-bus conflicts may create additional delays. Fortunately, these crossing delays turn out to be minor because if priority is alternated between the N-S and the E-W lines at the crossing points then BRT buses only have to yield to higher priority buses every $2s$ km. For the data in our examples such delays can be shown to reduce v by less than 1 km/hr.

Spatial coverage: Refer now to (1)-(11) and note: although agency costs decline with s, user costs are not monotonic increasing. In fact, consideration of (6) and (9) (i.e., the only user cost equations in which s appears) reveals that the user cost reaches a minimum when: $s^c \approx (w\tau D/[2/3 + (1-\alpha)^3(4+5\alpha+3\alpha^2)/12])^{½}$. (This, of course, should not be surprising, since there is a well-known trade-off between walking and riding time as one changes s; see Vuchic and Newell, 1968). Since user costs decline when $s < s^c$, there is no justification whatsoever for increasing spatial coverage so much that s dips below s^c. Note this critical value is independent of H but increases with the square root of τ and D. For the BRT option of Table II, s^c is about 0.33 km; i.e., about 10 stops per km^2. For slow boarding times and bigger cities, s^c is even larger.

Spatial coverage influences performance more than temporal coverage. For example, for the BRT case of Table II, the elasticities of z_A and z_U are: –0.97 and +0.36 with respect to s, but only –0.25 and +0.07 with respect to H. This should not be surprising if one remembers that wait is a much smaller fraction of the total time than riding time and that infrastructure costs (affected by s but not by H) are a significant part of the agency cost in the case of BRT. Thus spatial coverage deserves more attention than temporal coverage when designing a system.

5. DISCUSSION

Implementation issues: Although the formulas of this paper are general, the examples have assumed that buses are operated efficiently; i.e., that: (i) their headways are not allowed to fluctuate by more than a minute or so; (ii) they can travel at the speed of traffic when mixed with it, or substantially faster when segregated from it; (iii) they are not delayed by other buses (or by traffic) when accessing or emerging from their stops; and (iv) passenger board expeditiously. This kind of efficiency is easily within grasp. Adaptive control schemes that can maintain bus speeds while keeping headway fluctuations to within one minute already exist; underutilized bus lanes can be opened intermittently to private vehicles when the buses are not there; private vehicles can be banned from passing buses at stops—and buses can be endowed with cameras linked with the police to enforce the rule; and fare collection machines can be positioned deep in the bus, perhaps with surveillance cameras too, to prevent boarding queues that spill back onto the street.

Model predictions and the real world: Despite its simplicity, the proposed model does a good job of reproducing qualitatively some patterns in the real world. For example the results suggest that bus systems in cities like those of our tables should have a grid structure, and some real-world bus systems (e.g., Chicago) have this trait. The model also suggests that LRT/Metro systems in medium-size cities should only offer double coverage in a small part of a city, and again one finds that some real systems (e.g. Stockholm) have this trait. In fact, the hybrid structures proposed in this paper, when appropriately deformed to account for geography, do seem to describe a good number of real transit systems.

The one thing the model does not explain is why Metros are so prevalent in reality—recall that Metros were predicted to be the inferior alternative in all cases studied. Model error could be the cause of this anomaly but this is unlikely because models of the type we have used typically produce results with 1% to 2% accuracy; see Ouyang and Daganzo (2006). So what is a likely cause? In some instances poor technical analyses, perhaps driven by political agendas, could be a cause. Accidents of history could be another. But this does not explain all the metros one sees. A more likely cause is that a city may not have the willingness (e.g. the NIMBY effect due to pollution, noise and accidents) or the ability (enough suitable streets) to dedicate sufficient street space to buses. The latter is a real problem in Chengdu. These barriers to implementation can be reflected in the model by including in the infrastructure and operating cost of bus service its externalities on the urban space, together with a constraint limiting the number of infrastructure corridors. Metro could then emerge in many big-city cases as the winning alternative.

Modifications to the model: The systems' performance can be slightly enhanced if one allows for additional decision variables in the model. For example, one could replace the central square grid of lines by a rectangular grid with extra stops between transfer points. This more flexible design distributes benefits less evenly but can reduce total cost by a small percentage. Analysis shows that the marginal benefit of an extra stop declines quickly with the number of extra stops already provided; so we propose that at most one stop should be provided between transfer points. Two modifications to our grid structure could be considered: a (0-1) grid, in which the lines of the N family would be spaced s km apart and those of the S family $2s$ km apart—or the reverse—and a (1-1) grid in which both families would have spacing $2s$. In both cases buses would stop at all the lattice points along their lines. Grid (1-1) does not cover all the points on the lattice but does not increase average walking times unreasonably. Barcelona could benefit slightly from a (0-1) grid.

Other generalizations to our basic structure are possible. For example a city could try to run a system where transfers are not required. However, a transferless system that would provide a door-to-door distance comparable with that of a hub-and-spoke system requires that each point be visited by D/s lines instead of 1. This would be prohibitively expensive. Alternatively, the city could introduce slightly higher coverage levels by using lines that run diagonally to the grid, but this too turns out to be unreasonably costly. Another possibility is to design the central grid as a superposition of two perpendicular copies of the arterial system proposed in Newell (1979). This system would allow all trips to be completed with only 1 transfer, but unfortunately at the cost of less direct travel within the grid for some passengers, a less understandable network and longer bus routes.

More beneficial for big cities is the provision of multiple transit systems (with different s and different v) each optimized to serve long trips, short trips and the severely handicapped, or combinations of these categories. The model in this paper can be used to design a good set of systems. Also beneficial could be districting schemes for lines in the periphery that would allow some buses to skip stops before reaching the center, e.g., as proposed in Clarens and Hurdle (1975). This, however, turns out to be significantly advantageous only when the peripheral routes include many stops; i.e., when D/s is large and α is small—in other words, if the city is large and its demand low.

Creating a master plan: The proposed model can be used to develop master plans with a multi-step process similar to that proposed in Daganzo (1991) for logistics systems. Three steps are recommended: (i) broad design; (ii) detailed layout; (iii) adaptation.

Step (i) is taken by solving (12), and obtaining target ranges for s, H and α, which define the broad design. An idealized drawing can then be made to aid with visualization. Refinements of (12) are possible. If desired, for example, one can model the city as a rectangle instead of a square because rectangles can be more easily adapted to urban forms. Such a model would use one or two additional decision variables and would require minor revisions to (1)-(11), and (12). If appropriate, these formulas could also be revised to recognize that there may be a more concentrated rate of trip generation and attraction in a central area. Constraints can also be added if the problem demands it; e.g., to ensure that the separation between corridors is consistent with the number of available streets. Finally, although the design step should focus on the periods of peak demand, attention should also be paid to the choice of service variables (s and H) during the off-peak, weekends and nights.

Step (ii) (detailed layout) is taken by constructing a route map that conforms as closely as possible to broad design targets of step (i), including a headway management plan. In this step one tries to place stops and lines at key nodes and demand generators, such as connecting points with the intercity transportation network. Because real demand is lumpy, one should also allow for flexibility in the selection of headways so they can be shorter on lines and times where the demand is high. One can also adjust the line and stop spacings slightly, reducing them in areas of high demand, e.g. as in Wirasinghe and Ghoneim (1981). The detailed map should be constructed by deforming the idealized drawing--keeping the number of stops, the total length of the lines and the size of the central area roughly fixed. It is as if one were given a plate of spaghetti strands (the bus lines on the idealized drawing) and were asked to rearrange them on the plate (the map) as evenly as possible ensuring that they: (i) pass through the key locations, (ii) duplicate as little as possible pre-existing lines and (iii) stay within the set of available streets.[3] Although the final design is unlikely to look much like the idealized drawing of step (i), the predictions of step (i) should remain roughly valid because the agency's service metrics depend mostly on the weight of the spaghetti, which should not change much, and the user metrics depend mostly and rather sluggishly on the evenness of the spaghetti distribution, which we try to preserve.

Step (iii) is the implementation step. If implementation is to be gradual e.g., by deploying one line at a time, one should start with those projects that best complement the existing system and generate the most demand. Note however that a selection of an initial line using these criteria alone, without the constraint of a master plan, could conflict with the plan and have adverse consequences later on. Of course, although the master plan is crucial, it should be adjusted based on field observations as the project is rolled out.

Future work: The ideas in this paper can be a building block toward systematically answering big-picture questions about the use of urban space by multiple transportation modes; and for other purposes as well. Questions to be answered can be operational (e.g., how to apply congestion pricing recognizing its effects on transit), tactical (how to serve the handicapped) and strategic (how to guide land use development), but all questions can be better answered with a simple model of transit performance such as the one proposed in this paper. Of some interest too is the development of computer-aided design tools to perform step (ii). This has been already been done for simple logistics problems with considerable success (see Robusté et al, 1990, and Ouyang and Daganzo, 2006); and there is no reason why a similar approach cannot be developed for transit systems.

ACKNOWLEDGEMENT

This research was supported by the U.C. Berkeley Center of Excellence on Future Urban Transport. The study is the scientific basis under Barcelona's tentative plan to develop a system of high-performance "Metro-buses". The comments of A. Lopez (Director of Mobility Services for Barcelona), I. Armengol (Director of Bus services for TMB, Barcelona) and my colleagues F. Robuste, M. Estrada, M. Roca and D. Cuscurita (CENIT, Barcelona) are greatly appreciated.

[3] The set of available streets should not be treated as totally fixed: if some streets are not available because they run in the opposite direction, consideration should be given to reversing their direction.

REFERENCES

Clarens, G., Hurdle, V.O. (1975), "An operating strategy for a commuter bus system." *Transportation Science* 9(1), 1-20.

Daganzo, C.F. (1991), *Logistics Systems Analysis*, Springer-Verlag, Heidelberg, Germany.

Daganzo, C.F. (2009), "A headway-based approach to eliminate bus bunching: systematic analysis and comparisons". Transportation Research Part B, (in press) doi:10.1016/j.trb.2009.04.002

Eichler, M., Daganzo, C.F. (2006) "Bus lanes with intermittent priority: strategy, formulae and an evaluation." *Transportation Research Part B* 40, 731-744,

Holroyd, E.M. (1965), "The optimum bus service: a theoretical model for a large uniform urban area", in *Vehicular Traffic Science* (L.C. Edie, R. Herman and R. Rothery editors), Proc. 3rd Int. Symposium on the Theory of Traffic Flow, Elsevier, New York, N.Y.

Newell, G.F. (1979), "Some issues relating to the optimal design of bus routes." *Transportation Science* 13, 20-35.

Nie, B., and Sun, Z. (2009), private communication: Bin Nie, Deputy Director, Chengdu Municipal Transportation Commission, and Zhuangzhi Sun, Deputy Chief Engineer, Beijing Transportation Research Center.

Ouyang, Y., Daganzo, C.F. (2006), "Discretization and validation of the continuum approximation scheme for terminal system design." *Transportation Science* 40, 89 - 98.

Robusté, F., Daganzo, C.F., Souleyrette, R., (1990) "Implementing vehicle routing models." *Transportation Research Part B* 24, 263286.

Vuchic, V.R., Newell, G.F. (1968), "Rapid transit interstation spacings for minimum travel time." *Transportation Science* 2(4), 303-339.

Wirasinghe, S.C. and Ghoneim, N.A. (1981), "Spacing of bus stops for many to many travel demand." *Transportation Science* 15, 210-221.

Wirasinghe, S.C., Hurdle, V.F., Newell, G.F. (1977), "Optimal parameters for a coordinated rail and bus transit system." *Transportation Science* 11(4), 359-374.

APPENDIX A: PROOFS

Recall Fig. 1c and refer to Figure A1, below. Let us place the origin of a system of (x, y)-Cartesian coordinates oriented with the sides at the center of the square, and define cordons in the shape of a square similar to the edge of the region but β times smaller. The perimeter of a cordon is $4D\beta$. The cordon $\beta = \alpha$ marks the edge of the center. The N, S, E or W side of a cordon will be called a barrier.

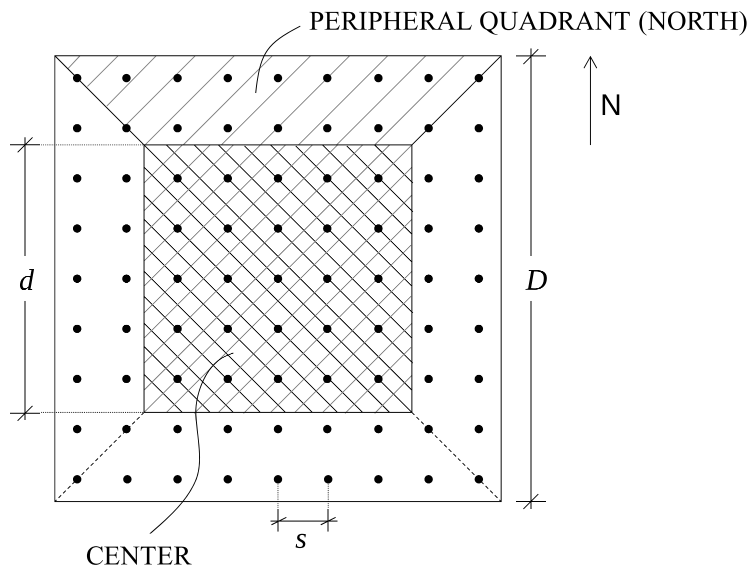

Figure A1. The city and its regions

Vertical and horizontal transit lines have the following properties: (a) they are spaced s distance units apart in the center, where they provide double coverage; (b) each line is contained in a single N-S or E-W hemisphere never crossing the boundary between peripheral quadrants; (c) lines branch as necessary to provide single coverage in the periphery; and (d) all lines reach the edge of the city, where buses turn around. It follows from (a)-(d) that the flow of buses across any barrier is conserved. Figure 1c also shows that the maximum number of transit lines crossing a barrier cannot exceed the length of the barrier divided by s: $\beta D/s$. Consideration of this and similar figures also reveals that this upper bound is met periodically as β increases. Hence, it is reasonably tight and we shall use it as an approximation. We are now ready to begin.

RESULT 1: *The total length of the two-way infrastructure system is given by* (1): $L \approx [D^2/s][1+\alpha^2]$.

Proof: Points that receive single coverage are associated with s km of two-way infrastructure. Points with double coverage are associated with an extra s km of infrastructure. If we ignore the few points on the periphery where the lines branch we see that the fraction of points receiving double coverage equals the ratio of the areas of the center and the region: α^2. Thus, the average length of infrastructure associated with each point is: $s[1+\alpha^2]$. Finally, since number of points in the region is $[D^2/s^2]$ the result follows. □

RESULT 2: *The total vehicle distance traveled per hour is given by* (2): $V \approx [2D^2/sH][3\alpha - \alpha^2]$.

Proof: We consider first the center and then the periphery. Since the headway is H everywhere in the center, the distance traveled is the ratio of the length of the routes to be covered and the headway. The length of these routes is twice the length of the infrastructure in the center. The latter is $2[D^2/s][\alpha^2]$ as per Result 1. Hence, the total distance traveled per hour in the center is: $4[D^2/sH][\alpha^2]$. The periphery has to be handled differently because the headways are not constant in it. So, let us derive first the distance traveled by the average vehicle in a peripheral quadrant between slices β and $\beta+d\beta$. Since the vehicle must traverse the slice, it must travel $\frac{1}{2}Dd\beta$ in the direction perpendicular to the barrier. The vehicle must also be displaced laterally. Note that a vehicle traveling along the center of the slice requires no displacement but a vehicle traveling along the edge is displaced $\frac{1}{2}Dd\beta$ units horizontally. Thus, the average displacement is on average: $Dd\beta/4$. Combining these two distances and multiplying the result, $[3Dd\beta/4]$, by the one-way flow of buses across the barrier we obtain the one-way distance traveled per hour in the slice. To obtain this flow recall that it is the same across all barriers. Therefore it can be expressed as the ratio of the number of lines emerging from the center ($\alpha D/s$) and the headway H: $[(\alpha D/s)/H]$. Obviously then, the one-way distance traveled per hour in the slice is: $[3Dd\beta/4][(\alpha D/s)/H]$. Now, integrate this from $\beta = \alpha$ to $\beta = 1$ to obtain the total one-way distance traveled per unit time in a peripheral quadrant. The result is $[3D(1-\alpha)/4][(\alpha D/s)/H]$. Since there are 4 quadrants and they are covered in 2 directions the total peripheral distance becomes: $[6D(1-\alpha)][(\alpha D/s)/H] = 6[D^2/sH][\alpha-\alpha^2]$. Finally, add to this value the central distance obtained earlier, $4[D^2/sH][\alpha^2]$, to obtain the total: $[2D^2/sH][3\alpha-\alpha^2]$. This is the sought result. □

Let us now examine the expected number of transfers and other passenger metrics. To do this it will be convenient to classify passenger trips among the following categories: (a) internal if both the origin and destination are in the center; (b) external if neither trip end is in the center; and (c) mixed otherwise. The mixed category is further subdivided into two equal-sized subcategories (c1) and (c2) depending on whether a trip's origin and the destination are in the same hemisphere or not. Note form Fig. A1: trips in category "c1" require two transfers. All other trips can be completed with at most one transfer.

So let us now examine the number of transfers and the access travel time. We conservatively assume that trips in category c1 always require 2 transfers and the rest 1. Then, we have:

RESULT 3: *The expected number of transfers per trip is given by* (10): $e_T = 1 + \frac{1}{2}(1-\alpha^2)^2$.

Proof: Since the probability that a trip end falls in the periphery is $(1-\alpha^2)$, the fraction of trips in category c must be $(1-\alpha^2)^2$. The result follows since categories c1 and c2 contain the same number of trips. □

RESULT 4: *The expected walking time at the origin and destination is given by* (6): $A = [s/w]$.

Proof: Remember that travelers are assumed to access and egress from the system at the closest stops to their real destinations. Thus, the maximum distance traveled at each end is s and the average ½s. Thus, the average combined distance is s and the average combined time is $A = s/v_W$. □

RESULT 5: *The expected waiting time per user including the origin and all transfer stops is given by* (7): $W = [H]\,[(2+\alpha^3)/(3\alpha) + (1-\alpha^2)^2/4]$ for $\alpha > ½s/D$.

Proof: The expected wait has three components: (i) at the origin stop, W_O; (ii) at the last transfer point, W_D; and (iii) at the intermediate transfer point, W_T, only for those trips requiring such transfer. Recall that the flow of buses across a barrier of a peripheral quadrant is $[(\alpha D/s)/H]$, and that the number of lines crossing the barrier is approximately $\beta D/s$. Therefore, since the approximate headway is the reciprocal of the flow per line $[\beta D/s]/[(\alpha D/s)/H] = (\beta/\alpha)H$, the average wait for passengers accessing the system on the barrier is approximately one half of that quantity: ½$(\beta/\alpha)H$ if $\beta > \alpha$. (Note: this expression is meaningful only if αD exceeds the value for the pure radial system, which is ½s since we are working with a finite grid; therefore the results about to be derived only hold if $\alpha > ½s/D$.) Of course, if the passenger starts the trip from the center, $\beta \leq \alpha$, the wait is ½H. Therefore, a combined expression for the expected wait at the origin conditional on β is: $[W_O|\beta] \approx ½H \max\{1, \beta/\alpha\}$. Now, since β is triangularly distributed in [0, 1] with p.d.f., 2β, the unconditional expectation W_O is approximately given by integrating $\beta H \max\{1, \beta/\alpha\}$ in [0, 1]. The result is: $W_O \approx H(2+\alpha^3)/(6\alpha)$. Waits for buses going to a destination are identically distributed as those from an origin in the same ring. Therefore, $W_O + W_D \approx H(2+\alpha^3)/(3\alpha)$. Finally, since the middle transfer always happens on the edge of the center and users can take any line serving this edge, their wait is on average ½H. And, since only $(e_T - 1) = ½(1-\alpha^2)^2$ of the users experience this wait, a random user's total expected wait is $W = W_O + W_D + e_T H/2 \approx H(2+\alpha^3)/(3\alpha) + (1-\alpha^2)^2.H/4$, as claimed.. □

RESULT 6: *The expected in-vehicle travel distance per trip is given by* (8): $E = [D]\,[2/3+\Delta]$, where $\Delta = (1-\alpha)^3(4+5\alpha+3\alpha^2)/12$.

Proof: We assume for economy of notation and without loss of generality that $D = 1$. It is well known that the expected distance between two random points in a L_1 square is 2/3. Therefore, we just need to show that the circuity added by the transit network is: $(1-\alpha)^3(4+5\alpha+3\alpha^2) / 12$. Note that only external trips (category b) experience circuity and that the probability of such a trip is $(1-\alpha^2)^2$. Thus, the desired result is $\Delta = \Delta_b(1-\alpha^2)^2$, where Δ_b is the expected circuity of an external trip. We express Δ_b as the difference between: the expected distance on the network E' and the expected L_1 distance E''. Thus, $\Delta =$

$(E' - E'')(1-\alpha^2)^2$. Expressions for these two distances are now derived. Because the steps are numerous but simple, only the logic is described.

Derivation of E': A trip on the network is composed of (i) an inbound radial segment connecting the origin with a point on the boundary of the central area; (ii) an outbound radial segment connecting a second boundary point with the destination; and (iii) a middle segment connecting the two boundary points. Consideration shows that the average L_1 distance between a random point on the β-cordon and the center is $3\beta/4$. Therefore, the distance of a radial connector segment is $E(3\beta/4) - 3\alpha/4$, since only the part of a radius external to the center should be considered. (Note, the expectation of the first term should be evaluated over the (trapezoidal) distribution of β in the interval $[\alpha, 1]$.) The result of this calculation is $\frac{1}{2}[(1+\alpha)^{-1} - \alpha/2]$. So the distance for the inbound and outbound radii combined is: $(1+\alpha)^{-1} - \alpha/2$. Note now that the middle segment never includes any circuity; therefore its average length is the average L_1 distance between two random points on the α-cordon. This distance is $11\alpha/12$. [This is true because if the two points lie on perpendicular sides the average distance equals one side of the cordon; and if they lie on parallel sides it equals 5/6 of a side: 1/3 in the direction of the sides, and ½ in the perpendicular direction.] Thus, combining all three segments we find: $E' = (1+\alpha)^{-1} + 5\alpha/12$.

Derivation of E'': The L_1 distance between two points in the periphery has two equal components: one in each direction. So, let us focus on the x-direction to find $E''/2$. Divide the periphery in two zones: (A) the two $(1\times\frac{1}{2}(1-\alpha))$-rectangles on the north and south sides of the center; and (B) the two $(\frac{1}{2}(1-\alpha)\times\alpha)$-rectangles on the east and west sides. The probability that a point falls in zone A given that it fell on the periphery is $p_A = (1-\alpha)/(1-\alpha^2) = 1/(1+\alpha)$; the probability that it falls in zone B is $p_B = \alpha/(1+\alpha)$. Now note, if both points fall in zone A the expected x-distance is 1/3 since the distribution of both x-coordinates is uniform. However, if one or both points fall in part B we must recognize that the distribution of the x-coordinate of the point in part B is uniform with a gap in the center. Using this fact, standard probability calculations reveal that the average x-distance is: $(4+2\alpha)/12$ if both points are zone B; and $(4+\alpha+\alpha^2)/12$ if one point is in zone A, and the other in zone B. So the expected x-distance is $E''/2 = (p_A)^2[1/3] + (p_B)^2[(4+2\alpha)/12] + (2p_A p_B)[(4+\alpha+\alpha^2)/12]$. A few algebraic manipulations reduce this expression to: $E'' = 2/3 + [\alpha^2 + 2\alpha^3]/[3(1+\alpha)^2]$.

To conclude the proof replace E' and E'' in the expression $\Delta = (E' - E'')(1-\alpha^2)^2$ by the formulas we just found to obtain: $\Delta = \{(1+\alpha)^{-1} + 5\alpha/12 - 2/3 - [\alpha^2 + 2\alpha^3]/[3(1+\alpha)^2]\}(1-\alpha^2)^2$. This result can be simplified to become: $\Delta = (1-\alpha)^3(4+5\alpha+3\alpha^2)/12$. □

COROLLARY 1: The total distance that passengers travel in the system per unit time is given by (11): $P = \lambda E = [\lambda D]\,[2/3 + (1-\alpha)^3(4+5\alpha+3\alpha^2)/12]$. Their average in-vehicle travel time obeys the first equality of (9): $T = E/v_c$.

Proof: The first part of the corollary follows from Little's formula. The second is self-evident. □

RESULT 7: *The expected commercial speed during the rush hour is given by* (4):

$$1/v_c \approx [1/v + \tau/s] + (\tfrac{1}{2}\tau'\Lambda sH/D^2)/(3\alpha-\alpha^2).$$

Proof: Let γ be the average number of passengers a vehicle collects per unit distance traveled. Then, the time a vehicle requires to travel a unit distance during the rush hour ($1/v_c$) should include the time consumed while: (i) overcoming distance ($1/v$); (ii) stopping (τ/s); and (iii) collecting passengers ($\tau'\gamma$). Thus, we need to show that $\gamma \approx (\tfrac{1}{2}\Lambda sH/D^2)/(3\alpha-\alpha^2)$. An approximation for γ is the ratio of the trips generated per hour during the peak (Λ) and the vehicle-km traveled per hour (given by Result 2 as $V \approx [2D^2/sH][3\alpha-\alpha^2]$.) Thus, $\gamma \approx (\tfrac{1}{2}\Lambda sH/D^2)/(3\alpha-\alpha^2)$. □

COROLLARY 2: The second equality of (9) holds. The number of vehicles in operation during the rush hour is given by (3): $M = V/v_c$. □

Proof: The first result is self-evident on combining Results 6 and 7. The second result holds because V is by definition the number of km covered by M vehicles in one hr; i.e., Mv_c . □

RESULT 8: *The expected vehicle occupancy on the critical load point during the rush hour is approximately given by* (5): $O \approx [\Lambda sH/D]\,[\max\{\tfrac{1}{2}(1-\alpha^2)/\alpha\,;\,(3-\alpha^4)/8\alpha + (D/s)(1-\alpha^2)^2/32\}]$.

Proof: The expected vehicle occupancy on the critical load point is the product of the critical link flow and H. The critical load point can be either in the periphery or in the center. Since the critical load will be the maximum of these two, it suffices to show that the critical flows are $[\Lambda s/D][\tfrac{1}{2}(1-\alpha^2)/\alpha]$ in the periphery and $[\Lambda s/D][(3-\alpha^4)/8\alpha + (D/s)(1-\alpha^2)^2/32\}] = [\Lambda s/D](3-\alpha^4)/8\alpha + \Lambda(1-\alpha^2)^2/32$ in the center.

Critical flow in the periphery: Since the peripheral routes serve either as passenger collectors or distributors their critical links are those closest to the center. For the inbound routes, the average flow across all these links is the ratio of the trip generation rate in the periphery $\Lambda(1-\alpha^2)$, and the number of inbound links into the center $4\alpha D/s$; i.e., $[\Lambda s/D][(1-\alpha^2)/4\alpha]$. However, not all inbound links carry the same flow. Consideration shows that the number of stops on an inbound route and therefore the flow on its last link before entering the center can vary across routes by as much as a factor of 2; see Fig. 1c. Therefore, we multiply the average by a factor of 2 to obtain the desired result.

Critical flow in the center: Again, we calculate first an average and then correct for the distribution. Consider all the links along the equator of our city (the line = 0) that can carry flow from north to south. There are $\alpha D/s$ links of this type, and all are in the center. Every trip with an origin in the N-side of the equator and destination in the S-side uses once one and only one of these links. Trips of this type are generated at a rate of $\Lambda/4$. So the average flow across all the equatorial links is $\Lambda s/4\alpha D$. (The total flow crossing other parallels is less; so we focus on the equator.) Now focus on the distribution. Dive all trips crossing the equator from N to S into two classes containing: (i) trips with one transfer on each side of the equator; and (ii) the rest. These classes are relevant because trips in class (i) flow exclusively on the extreme two links of the equator, whereas those of class (ii) are spread across all links. A trip is in class (i) if and only if it has both its origin and destination in the E or W peripheral quadrants—with the origin on the north side of its quadrant and the destination on its south side. Therefore, these trips are generated at a rate, $\Lambda[(1-\alpha^2)^2]/16$. And since this rate is divided evenly among two links, the class-(i) flow on each of these links is: $\Lambda[(1-\alpha^2)^2]/32$.

The trip generation rate for class (ii) is obviously $\Lambda/4 - \Lambda[(1-\alpha^2)^2]/16$. This flow will be divided across the equatorial links in proportion to the size of their catchment areas. The catchment areas from the center are the same for all links, but those on the periphery can be twice as big for the links on the edge as they are for the links on the center. The easiest way to make this correction is by pretending that the trip generation rate in all relevant catchment areas is twice as large as in reality, and then dividing evenly across all links the artificially elevated trip generation rate. The relevant catchment area includes all trips with: (a) origin in the N peripheral quadrant and destination in the S peripheral quadrant (and therefore with generation rate $\Lambda[((1-\alpha^2)/4)^2]$); (b) origin in the N peripheral quadrant and destination not in the S peripheral quadrant but on the south side of the equator (and generation rate $\Lambda[(1-\alpha^2)][½- (1-\alpha^2)/4]$); and (c) origin not in the N peripheral quadrant but in the north side of the equator and destination in the S peripheral quadrant (and generation rate $\Lambda[(1-\alpha^2)][½- (1-\alpha^2)/4]$). Therefore the trip generation rate in the relevant catchment area is the sum of these three rates: $\Lambda[(1-\alpha^2)/4][1 - (1-\alpha^2)/4]$. To double this rate, simply add it to the total rate, $\Lambda/4 - \Lambda[(1-\alpha^2)^2]/16$. The result is: $\Lambda/4 - \Lambda[(1-\alpha^2)^2]/8 + \Lambda[(1-\alpha^2)/4 = (\Lambda/8)(3 - \alpha^4)$. Now divide by the number of links $\alpha D/s$ to obtain the class-(ii) flow: $(\Lambda s/\alpha D)(3-\alpha^4)/8$. Thus, the final flow on the critical links in the center is the sum of class-(i) and class-(ii) flows; i.e.: $\Lambda[(1-\alpha^2)^2]/32 + (\Lambda s/D)(3-\alpha^4)/8\alpha$. □