20

Roundabout intersections

At low/medium levels of traffic flow the control of traffic movements at intersections may be achieved by priority control. As has been discussed previously the form of priority control in the UK is that minor road vehicles give way to major road vehicles. On the continent of Europe nearside priority is sometimes used where vehicles give way to traffic approaching from the right while in Australia off-side priority is used.

As traffic flows increase delays with priority control become excessive and at high levels of flow grade-separated junctions are necessary. A junction of this form is however extremely expensive: in addition, land requirements are great and in urban and suburban areas interference with pedestrian flow can be considerable. For these reasons at-grade intersections either of the roundabout or signal control type are extremely important in urban areas. The characteristics of these junction types have been described by Millard[1] and a consideration of these characteristics will usually determine which type of junction is appropriate. They include:

(a) In urban situations land requirements are usually the deciding factor. If this is so, it will be found that the land required for a large island roundabout is greater than that needed for traffic signal control. This is especially true if flows on one pair of arms are low. On the other hand, if land purchase is necessary, it is often easier to acquire corner sites necessary for a roundabout than the long narrow strips needed when parallel widening of traffic signal approaches is carried out.

(b) Both conventional and mini-roundabouts have difficulty in dealing with unbalanced flows, especially during peak hours when the traffic entering on an arm is considerably greater than the traffic leaving by it. In such situations there is frequently a shortage of gaps in the circulating stream and, under off-side priority rule, delays may become excessive.

(c) Right-turning vehicles (left-hand rule of the road) cause difficulties with signal control when their numbers are large. Either late start or early cut-off facilities or a special phase must be provided causing reduced overall capacity at the junction. In such circumstances roundabouts offer advantages.

(d) Traffic signal control has difficulty in dealing with three-way junctions, especially where the flows are balanced. To a lesser extent this is true of junctions with five or more approaches.

Roundabouts have been commonly used in central city areas where traditionally they were used to resolve traffic and pedestrian conflicts in the large open squares which existed in the early part of the twentieth century. The central island frequently covered a large area of the square and was utilised for ornamental flower beds while traffic circulated around the surrounding carriageway. Increasing traffic demand and the pressure to allow pedestrians to cross the carriageway at surface level has resulted in many of these roundabouts being converted to signal control so that more positive control over traffic movements on an area wide basis may be exercised.

In suburban areas, roundabouts are frequently found at the intersection of radial and ring type roads where they are subjected to peak hour traffic demands due to commuter flows.

Roundabouts are also used on rural roads where traffic flows or the road type do not justify the provision of a grade-separated intersection. In these situations, speeds are high on the approaches to roundabouts and safety is an important consideration.

Roundabouts deal efficiently with traffic movements when there are three or four arms. With three arms, and well balanced flows a roundabout can be more efficient than signal control. When the number of arms exceeds four then direction signing and driver comprehension become difficult. With many approach arms the diameter of the roundabout increases, leading to possible higher circulating speeds and consequent safety problems.

In addition to the resolution of traffic conflicts, roundabouts are employed where there is a significant change in road type, a change from rural to urban conditions or when a significant change in road direction is required.

In current United Kingdom practice there are three basic types of roundabout: normal roundabouts, mini roundabouts and double roundabouts. There are variations of these types to form ring junctions, grade-separated roundabouts and signalised roundabouts.

The Department of Transport[2] defines a normal roundabout as one which has a one-way circulatory carriageway around a kerbed central island 4 m or more in diameter and usually with flared approaches to allow multiple vehicle entry.

Mini roundabouts are defined as a roundabout having a one-way circulatory carriageway around a flush or a slightly raised circular marking less than 4 m in diameter and with or without flared approaches. They have been widely used in urban areas where the speed limit does not exceed 30 mph. Physical deflection of vehicle paths to the left, an important factor in roundabout safety, may be difficult in urban junctions with fixed kerb lines and in these circumstances road markings should be used to induce some vehicle deflection. The circular marking varies in diameter from 1 to 4 m diameter and is domed to a maximum height of 125 mm. If space within the junction is very limited then the central island will be frequently run over by larger vehicles and in these cases the island is normally flush with the road surface. The Department of Transport advises that pedal cyclists may experience difficulty and if there are a substantial number of cyclists passing through the junction then signal control may be preferable.

A double roundabout is defined as an individual junction with two normal or mini roundabouts either contiguous or connected by a central link road or kerbed island. It is considered[2] that this form of junction will have advantages in the following circumstances: improving an existing staggered junction where it avoids the need to realign one of the approach roads, unusual junctions such as scissors junctions, joining two parallel routes separated by a river, railway or motorway, existing cross-roads where opposing right-turning movements can be separated over loaded single roundabouts, and junctions with more than four entries.

A solution to the problem of large roundabouts where entries are approaching capacity may sometimes be found by conversion to a ring junction where the usual clockwise circulation of vehicles around a large island is replaced by two-way circulation with three-arm mini roundabouts at the junction of each approach arm with the circulatory carriageway. Because of the two-way flow on the circulatory carriageway and the need for circulating vehicles to give way, adequate signposting is essential for efficient operation.

Roundabout control is also utilised in grade-separated intersections in the form of two-bridge roundabouts and dumb-bell roundabouts; these are discussed subsequently in Chapter 21.

When roundabouts are being designed it is usual to add at least one additional lane at the entry of the approach roads to the circulating area with a maximum addition of two lanes and a maximum entry width of four lanes. When only low flows are predicted in the future then widening may not be considered necessary but a minimum of two lanes in an entry is desirable The angle at which vehicles enter the circulating area is of considerable importance in the operation of a roundabout. If vehicles enter at an angle approaching 90 degrees to the circulating flow then vehicles are liable to stop quickly on entry and cause rear end collisions. Should they however enter at a small angle then drivers must look over their shoulders in an attempt to merge.

Safety at roundabouts is enhanced by limiting circulating speeds and this is achieved in geometric design by entry path curvature, which is a measure of the vehicle path deflection to the left imposed on vehicles entering the circulating area. The path of vehicles travelling straight through the roundabout is drawn with a flexible curve so that it is the centre line of the most realistic path that a vehicle would take in its complete passage through the junction on a smooth alignment without sharp transitions. The tightest radius of the entry path curvature is recommended by the Department of Transport[2] to be not greater than 100 m.

The size of a roundabout is largely determined by the inscribed circle diameter; the inscribed circle is that which approximately touches the outer edges of the circulating area. For a normal roundabout the smallest recommended inscribed circle diameter for United Kingdom construction is 28 m; if this is not possible a mini roundabout should be used. To meet deflection requirements the inscribed circle diameter should not be less than 40 m. For a mini roundabout the inscribed circle diameter should not be greater than 28 m.

Circulating carriageway widths should be constant and not exceed 15 m; in general the width should not be less than, or 20 per cent greater than the maximum entry width.

It is recommended[2] that the entry radius into the circulating area should lie between 6 m and 100 m with a normal value of approximately 20 m. Exiting radius should not be less than 20 m and preferably have a value of 40 m.

Accidents at roundabouts

Roundabouts are considered to be one of the safest forms of junctions. The Department of Transport has stated[3] that the average accident cost at a roundabout is approximately 30 per cent less than that at all other junctions, and about 60 per cent less than that on the approach carriageways.

In a similar manner to the estimation of personal injury accidents at priority intersections the Department of Transport COBA program estimates personal injury accidents at roundabouts by means of either an inflow (I) or a cross-product (C) model. Both models are of the form

$$\text{Annual number of accidents} = x\,(f)^y$$

In the inflow model the value of f is the value of the total inflow into the roundabout from all the approach road links expressed as thousands of vehicles AADT. In the cross-product model f is the value obtained by multiplying the combined inflow from two major opposing links by the sum of the inflows from the other links.

The COBA program categorises junctions in a very broad manner and for roundabouts the division is by the number of arms and by the generalised descriptions, standard, small and mini. For all three and four-arm roundabouts the formula which is employed is the cross-flow type and the values of the coefficient x and y are given in table 20.1.

TABLE 20.1
Coefficients for use in accident estimation models

Roundabout type	*No. of arms*	*Highest link standard*	x		y	
			rural	*urban*	*rural*	*urban*
Standard	3	single	0.033	0.760	0.033	0.760
	3	dual	0.033	0.760	0.033	0.760
	4	single	0.024	0.890	0.048	0.740
	4	dual	0.063	0.690	0.022	0.850
Small	3	single	0.033	0.760	0.033	0.760
	3	dual	0.033	0.760	0.033	0.760
	4	single	0.101	0.660	0.263	0.540
	4	dual	0.101	0.660	0.263	0.540
Mini	3	single	0.012	1.040	0.012	1.040
	3	dual	0.012	1.040	0.012	1.040
	4	single	0.070	0.640	0.070	0.640
	4	dual	0.070	0.640	0.070	0.640

A more detailed analysis of accidents at four-arm roundabouts has been carried out by Maycock and Hall[3] who analysed personal injury accident records from a sample of 84 four-arm roundabouts on main roads in the United Kingdom. The roundabouts studied comprised both conventional and small roundabouts in speed limit zones of 30 to 40 and 50 to 70 mph. A cross-product model of the form used in the COBA program was evolved and the value of x in the COBA equation was 0.095 for small roundabouts and 0.062 for conventional and dual-carriageway roundabouts; the value of y was found to be 0.68 for all roundabout types. Additionally arm-specific accident relationships for differing vehicle movements were developed and the reader should consult the reference quoted for detailed information.

The traffic capacity of roundabouts

Prior to 1966 traffic entering and passing through a roundabout resolved the vehicle-to-vehicle interactions by requiring the traffic streams to weave one with another as they passed around the central island. In this situation the traffic design of roundabouts consisted of designing the individual weaving sections of the carriageway around the central island, an individual weaving section lying between adjacent entry arms.

These weaving sections were designed using a formula developed by Wardrop[4] which incorporated parameters that described the geometric shape of the weaving section and also the proportion of weaving vehicles in the section.

Without any positive control of traffic, roundabouts tended to 'lock' under heavy traffic conditions because vehicles already on the roundabout were prevented from leaving by vehicles attempting to enter. For these traffic conditions a solution was the use of long weaving sections and large roundabouts.

To overcome the problem of locking, the 'give way to traffic approaching from the right' rule was introduced. Subsequently a number of investigations showed that because traffic was no longer weaving around the central island the original formula was no longer valid and the correct approach was to consider the capacity of the roundabout as the capacity of the individual entries of the roundabout.

A particular difficulty in deriving a formula for the capacity of roundabouts is the variety of geometric designs found in practice. Designs range from conventional roundabouts with approximately parallel-sided rectangular-shaped weaving sections, to irregularly shaped central islands and entries on to the roundabout with widths rather similar to that of the approach carriageway, and to the offside priority roundabout with circular central islands and flared approaches.

Recommendations for determining the capacity of a roundabout entry are given by the Department of Transport[5] based on the research carried out by the Transport and Road Research Laboratory[6].

The predictive equation for entry capacity (Q_e) is given by $k\ (F - F_cQ_c)$ when F_cQ_c is less than or equal to F (otherwise the entry capacity is negative)

where Q_e = the entry flow into the circulatory area in pcu/hour where a heavy goods vehicle equals two passenger cars,

Q_c = the flow in the circulating area in conflict with the entry in pcu/hour,

$$k = 1 - 0.00347\ (\phi - 30) - 0.978\ [(1/r) - 0.05]$$
$$F = 303X_2$$
$$F_c = 0.210t_D(1 + 0.2X_2)$$
$$t_D = 1 + 0.5/(1 + M)$$
$$M = \exp\ [(D - 60)/10]$$
$$X_2 = v + (e - v)/(1 + 2S)$$
$$S = 1.6(e - v)/l$$

A description of these symbols is given in table 20.1 and a precise definition is given in figure 20.1.

The equation for Q_e is applicable to all roundabout types except those that are incorporated into grade-separated intersections. In the latter case the term F is replaced by $1.11F$ and the term F_c is replaced by $1.4F_c$. The formula was derived from observations of traffic flow at roundabout intersections and the ranges of the geometric parameters at the observed roundabouts are given in table 20.2, together with the ranges of these parameters suggested for new design. It is also recommended that the circulatory carriageway width around the roundabout should be constant at 1 to 1.2 times the greatest entry width, subject to a maximum of 15 metres.

Whilst the use of the formula for Q_e will allow a check to be made that entry flows are below the entry capacity, there are difficulties when queueing commences. In these circumstances the entry and circulating flows are interdependent.

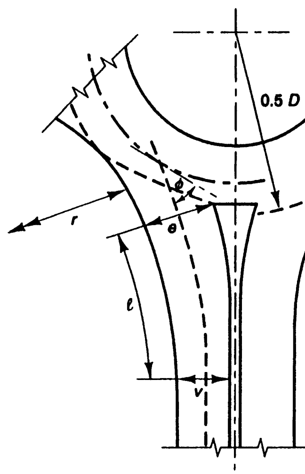

Figure 20.1

TABLE 20.2

Symbol	Description	Observed range	Recommended range for design
e	Entry width	3.6–16.5 m	4.0–15.0 m
v	Approach half-width	1.9–12.5 m	2.0–7.3 m
ℓ	Average effective flare length	1–∞ m	1.0–100.0 m
S	Sharpness of flare	0–2.9 m	
D	Inscribed circular diameter	13.5–171.6 m	15–100 m
ϕ	Entry angle	0–77°	10–60°
r	Entry radius	3.4–∞ m	6.0–100.0 m

ARCADY3

To enable delays to be predicted in this interactive situation the Transport and Road Research Laboratory has developed the computer program[7] ARCADY3 to model queues and delays at roundabouts. The model is also used as an aid to roundabout design.

Initially, there is assumed to be no circulating flow past the first entry, so that the entry flow will be either the entry demand or the entry capacity whichever is the least. This entry flow, after removal of vehicles that take the next exit, becomes the circulatory flow past the next entry. The entry flow at this next entry can then be determined by the circulating flow/entry flow relationship. This means that the circulatory flow and hence the entry flow at successive entries can be calculated in a clockwise manner. The process is iterated and the entry flows converge to their final values. Entry capacities are calculated as described above.

In the program successive short time intervals are considered during which the circulating flow, entry capacity and demand on the entering arm are known; hence the increase or decrease in queue length and the total queue length and delay can be calculated. This is achieved using equations 18.6 and 18.7 in Chapter 18.

Input data for the program is of three types: reference information giving details of the computer run, date and names of roads at the intersection being studied; geometric characteristics of the junctions; and traffic flow information.

Output for the program includes tables of demand flows, capacities, queue lengths and delays for each time segment, together with a diagrammatic representation of the growth and decay of the queues on each arm.

Facilities are also available in ARCADY3 to model pedestrian (zebra) crossings and to predict geometric delays (see Chapter 24) and likely accident frequencies.

Example

A roundabout at a grade-separated intersection has an inscribed circle diameter of 65 m, an entry width of 8.5 m and an approach half-width of 7.3 m. The effective length over which flare is developed is 30 m, the entry radius is 40 m and the entry angle is 60 degrees.

The design reference flows to be used for a preliminary design expressed as entry capacity (pcu per hour) are given in table 20.3. Calculate the reserve capacity of the intersection in these traffic conditions.

TABLE 20.3

From	To			
	N	S	E	W
N		850	200	100
S	700		450	250
E	150	350		700
W	350	450	350	

The flows given in table 20.2 are assigned to the roundabout as shown in figure 20.2. From this diagram the circulating flows around the central island are abstracted and these are shown in figure 20.3.

The capacity of an entry of a grade-separated roundabout may be calculated using the relationship:

$$Q_e = k\,(F - F_c Q_c)$$

$$\begin{aligned}
\text{where } k &= 1 - 0.00347\,(\phi - 30) - 0.978\,((1/r) - 0.05) \\
&= 1 - 0.00347\,(60 - 30) - 0.978\,((1/40) - 0.05) \\
&= 1 - 0.1041 + 0.02445 \\
&= 0.92035 \\
S &= 1.6\,(e - v)/l
\end{aligned}$$

$= 1.6(8.5 - 7.3)/30$
$= 0.064$
$X_2 = v + (e - v)/(1 + 2S)$
$= 7.3 + (8.5 - 7.3)/(1 + 2 \times 0.064)$
$= 8.3638$
$F = 1.11\ (303X_2)$
$= 1.11\ (303 \times 8.3628)$
$= 2813$
$t_D = 1 + 0.5/[1 + \exp((D - 60)/10)]$
$= 1+0.5/[1 + \exp((65 - 60)/10)]$
$= 1.1888$
$F_c = 1.4(0.210 t_D (1 + 0.2X_2))$
$= 1.4\ (0.210 \times 1.1888\ (1 + 0.2 \times 8.3638))$
$= 0.5846$
$Q_e = 0.92035(2813 - 0.5846Q_c)$
$= 2589 - 0.538Q_c$

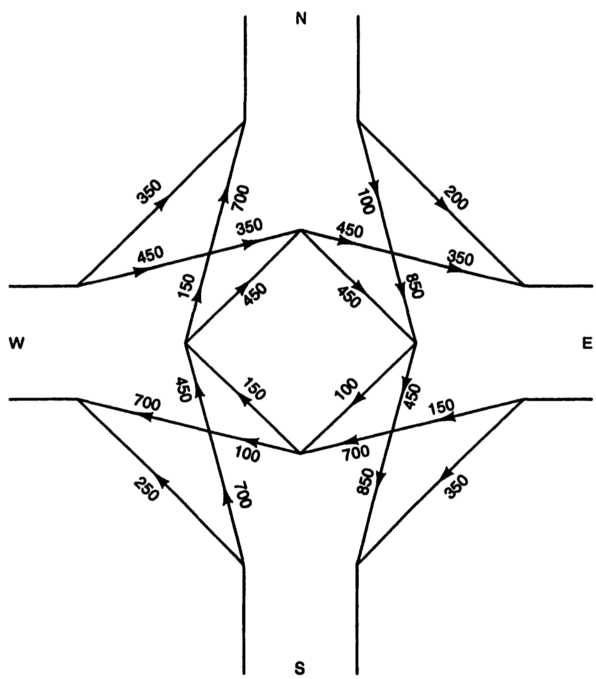

Figure 20.2

Using this relationship between entry capacity and circulating flow the entry capacity for the N, S, E and W arms can be calculated using the values of circulating flow obtained from figure 20.3; calculated values are entered in table 20.4.

Comparison of the calculated entry capacities and the entry flows given in table 20.3 shows that the entry flow is lower than the calculated capacity for each of the entries. The reserve capacity is calculated from capacity minus flow divided by entry flow, expressed as a percentage.

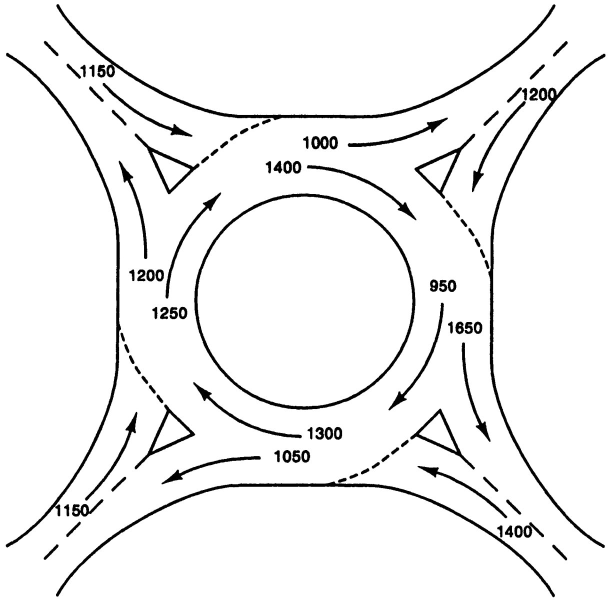

Figure 20.3

TABLE 20.4

Entry	Circulating flow (pcu/h)	Entry capacity (pcu/h)	Entry flow (pcu/h)	Reserve capacity (per cent)
N	1250	1917	1150	66.7
S	950	2078	1400	48.4
E	1400	1836	1200	53.0
W	1300	1890	1150	64.3

References

1. R. S. Millard, Roundabouts and signals, *Traffic Eng. Control*, **13** (1971), 13–15
2. Department of Transport, The geometric design of roundabouts, *Departmental Standard TD* 16/93, *Design Manual for Roads and Bridges*, Vol. 6, Section 2 (1993)
3. G. Maycock and R. D. Hall, Accidents at 4-arm roundabouts, *Transport and Road Research Laboratory Report* 1120, Crowthorne (1984)
4. J. G. Wardrop, The traffic capacity of weaving sections of roundabouts, *Proceedings of the First International Conference on Operational Research, Oxford, 1957*, English Universities Press, London (1957), pp. 266–80
5. Department of Transport, Junctions and accesses, determination of size of roundabouts and major/minor junctions, *Departmental Advice Note* TA 23/81, London (1981)
6. R. M. Kimber, The traffic capacity of roundabouts, *Transport and Road Research Laboratory Report* LR 942, Crowthorne (1980)
7. E. M. Hollis, M. C. Semments and S. L. Dennis, ARCADY: a computer program to model capacities, queues and delays at roundabouts, *Transport and Road Research Laboratory Report* 940, Crowthorne (1980)

21

Grade-separated junctions and interchanges

Grade-separated junctions have been defined[1] as ones which require use of an at-grade junction at the commencement or termination of the slip roads. This at-grade junction, in the form of either a major–minor priority junction or a roundabout together with the slip roads, can produce a diamond junction, a half-cloverleaf junction or a roundabout junction.

In the United Kingdom, decisions as to junction type are no longer made by the use of design standards which relate choice to traffic forecasts. Decisions are made on the basis of relative economic and environmental advantages of competing junction types. The choice of junction type is particularly difficult because of the sensitivity of delay to the future traffic forecast. Below a flow in the region of 85 per cent capacity, average vehicle delay is relatively uniform whilst above 85 per cent of capacity, average delays increase rapidly. As a result small differences in estimates of future demand flows or junction capacity can have a considerable effect on delay estimates. With grade-separated junctions in urban areas, environmental factors may have considerable influence on the choice of junction type.

Three-way junctions

At a three-way junction the usual layout adopted is the trumpet, illustrated in figure 21.1 It allows a full range of turning movements but suffers from the disadvantage that there is a speed limitation on the minor road right-turning flow due to radius size. Where topographical conditions make it necessary the junction can be designed to the opposite hand, as shown in figure 21.2, but this layout has the disadvantage that vehicles leaving the major road have to turn through a small radius.

Where a limited range of turning movements is required then one of the junctions of the form shown in figure 21.3 may be used. The layout shown in (c) may be used in a situation in which the two carriageways are a considerable distance apart and where a right turn exit on the right-hand lane would not be appropriate.

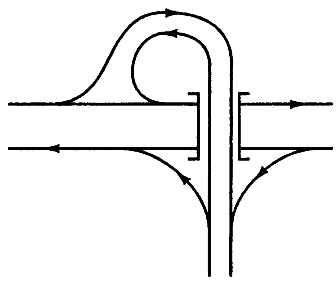

Figure 21.1 *The usual form of trumpet intersection*

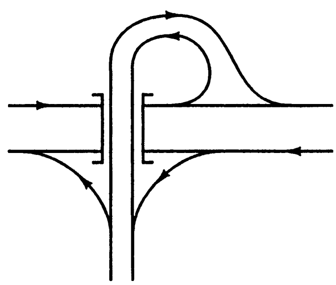

Figure 21.2 *A form of trumpet intersection to avoid existing development*

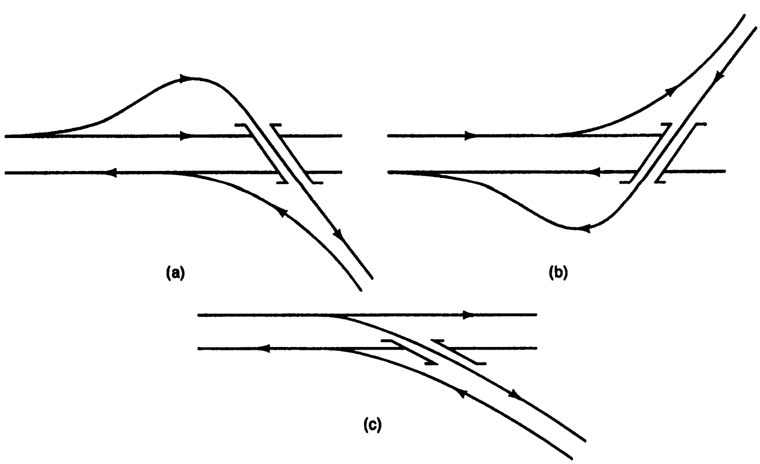

Figure 21.3 *Alternative forms of three-way junctions*

It may sometimes in the future become necessary to convert a three-way junction to a four-way junction, and the design should allow for this future conversion. A suitable form of junction is the partial bridged rotary shown in figure 21.4. Alternatively the trumpet intersection shown in figure 21.1 may be converted to a cloverleaf intersection.

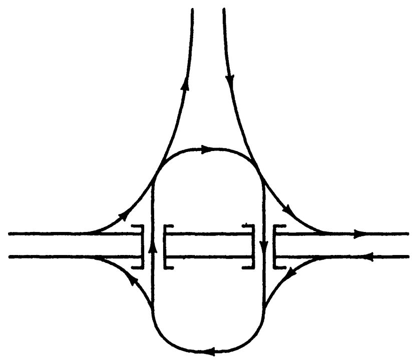

Figure 21.4 *A partial bridged rotary intersection*

Four-way junctions

Frequently four-way junctions are formed by the intersection of a major and a minor road. In such instances it is often possible to allow traffic conflicts to take place on the minor road. The simplest type of intersection where this occurs is the diamond, shown in figure 21.5. This is a suitable form of layout in which there are relatively few turning movements from the major road on to the minor road since the capacity of the two exit slip roads from the major road is limited by the capacity of the priority intersection with the minor road. Care needs to be taken with this and other designs by the use of channelisation, so that wrong-way movements on the slip roads are prevented.

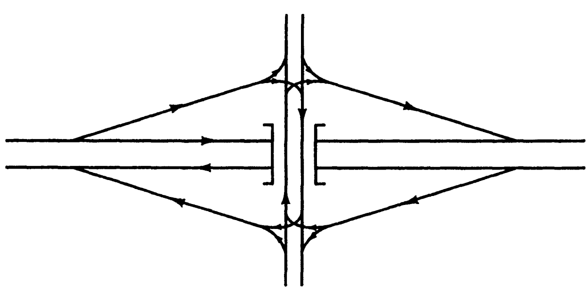

Figure 21.5 *A diamond intersection*

In some circumstances site conditions may prevent the construction of slip roads in all of the four quadrants as required by the diamond. A solution in this case would be the construction of a half cloverleaf as shown in figure 21.6(a). Where site conditions make it necessary or where right-turning movements from the major road are heavy then the opposite hand arrangement as shown in figure 21.6(b) may be used, with the disadvantage however of a more sudden speed reduction for traffic leaving the major road. As the traffic importance of the minor road and the magnitude of the turning movements increase additional slip roads may be inserted into the partial cloverleaf design until the junction layout approaches the full cloverleaf design.

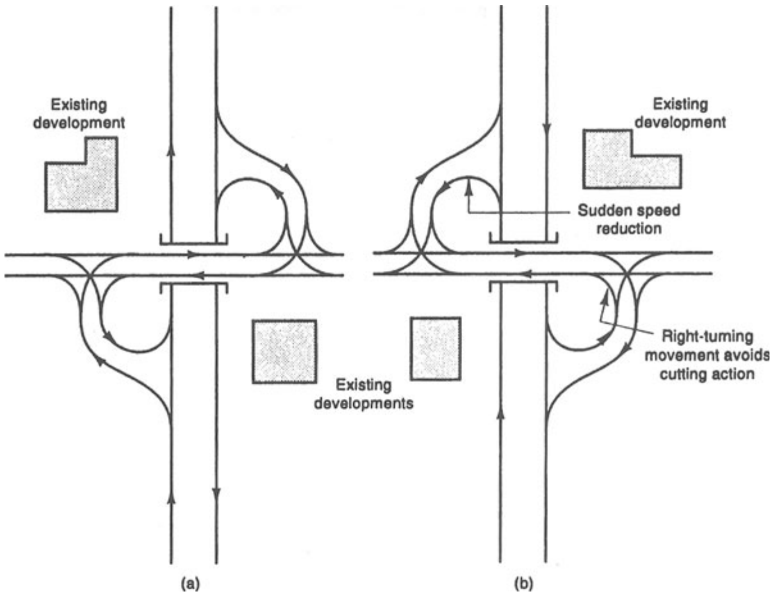

Figure 21.6 *Partial cloverleaf designs*

An advantage of both the diamond and the half-cloverleaf junctions is that only one bridge is required, but it is recommended[1] that in its design provision should be made for the future likely construction of a ghost island on the minor road.

Where major traffic routes intersect, it is no longer possible for the traffic conflicts to be resolved on the road of lesser traffic importance and the simpler junction types previously described are then likely to be unsatisfactory, both from the point of view of capacity and accident potential. The design of intersections of this type is always carried out after consideration of the individual directional traffic flows, having regard for the topographical features of the area. Nevertheless it is possible to discuss the general forms of these intersections and the traffic flows that are likely to warrant their construction. When turning movements from the route of greater traffic importance cannot be handled by the diamond intersection, then the usual solution in Great Britain has been the grade-separated roundabout as shown in figure 21.7.

The advantages of this type of junction are that it occupies a relatively small area of land and has less carriageway area than other junctions of this type. It also allows easy U-turns to be made, a factor of some importance in rural areas where intersections may be widely spaced or in urban areas where a number of slip roads may join the major route between interchanges.

As the traffic importance of the minor road increases then a disadvantage of this form of junction is the necessity for all vehicles on the road of lesser traffic importance to weave with the turning traffic from the other route. For this reason the capacity of the weaving sections limits the capacity of the intersection as a whole.

A junction intermediate between the diamond and the two-bridge roundabout is the dumb-bell roundabout where priority junction slip road terminations of the diamond junction are replaced by two compact roundabouts. This form of junction requires only one bridge and has a reduced landtake compared to the diamond or two-level roundabout.

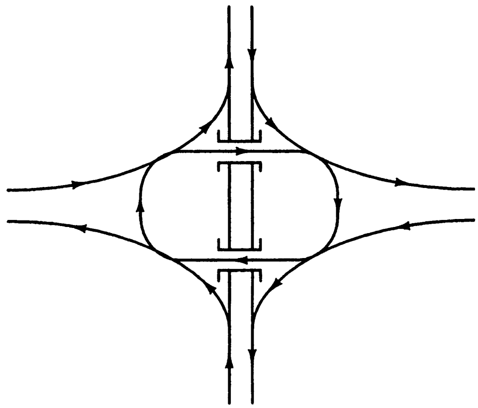

Figure 21.7 *Grade-separated roundabout intersection*

As the traffic flow on the minor crossing road increases then a disadvantage of the two-level roundabout junction is the necessity for all vehicles on the crossing road to pass through the circulating area of the roundabout.

An alternative form of intersection which, while popular in the United States, has found limited application in Great Britain is the cloverleaf, shown in figure 21.8. With this form the straight ahead traffic on both routes is unimpeded and in addition left-turning movements may be made directly from one route to the other. The cost of structural works is also less than with the grade-separated roundabout

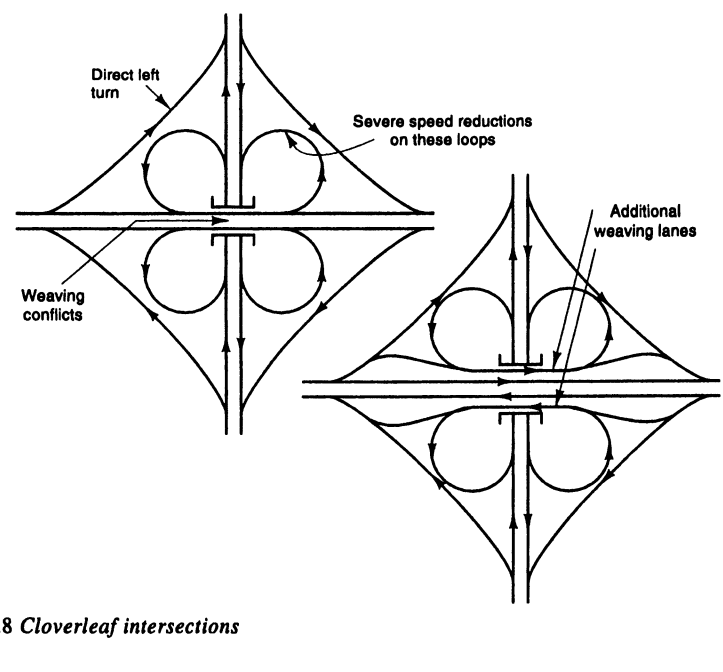

Figure 21.8 *Cloverleaf intersections*

because only one bridge is required although the bridge will normally be wider. On the other hand the carriageway area is greater for the cloverleaf.

Operationally the cloverleaf intersection has many disadvantages. Vehicles leaving both routes to make a right turn have to reduce speed considerably because of radius restrictions and there is also a weaving conflict between vehicles entering and leaving each route, which peak at the same time. This conflict can however be reduced by the provision of a separate weaving lane on either side of each carriageway but this will increase carriageway and bridge costs. There are also four traffic connections on to each route and U-turns may present some difficulty to drivers unfamiliar with the junction. Sign posting has also been stated to present some difficulties.

Where heavy right-turning movements in one direction are anticipated at a clover leaf type intersection this movement has frequently been given a direct connection by a single link as shown in figure 21.9, requiring extra bridgeworks but eliminating one loop with its small radius and restricted speed.

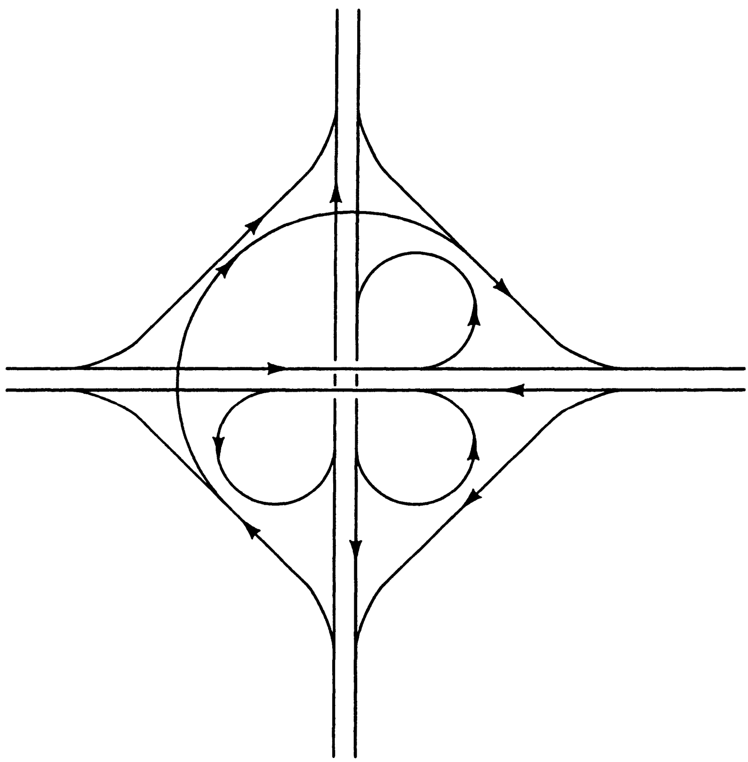

Figure 21.9 *A cloverleaf intersection with a direct link for a heavy right-turning movement*

Where two major routes of equal importance intersect, a solution which has been frequently adopted is the use of a roundabout to deal with turning movements while straight flows on both major routes are unimpeded. This layout is shown in figure 21.10.

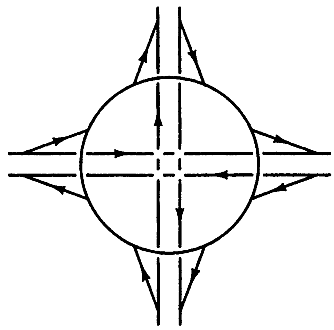

Figure 21.10 *A three-level grade-separated intersection*

The three-level roundabout has advantages compared with larger and more complex interchanges because of its lower carriageway area and reduced landtake. It has however high structure costs, and if traffic growth is greater than predicted then operational problems may arise with queueing vehicles on the roundabout entries. For this reason the three-level roundabout is particularly suitable when right-turning flows and the proportion of heavy vehicles in the turning flows are relatively low.

Interchanges

In contrast to grade-separated junctions, interchanges provide uninterrupted movement for vehicles travelling from one mainline carriageway to another mainline carriageway, via link roads using a succession of diverging and merging manoeuvres.

A form of interchange which has been used when high turning flows are predicted is the four-level layout shown in figure 21.11. This layout has a low land-

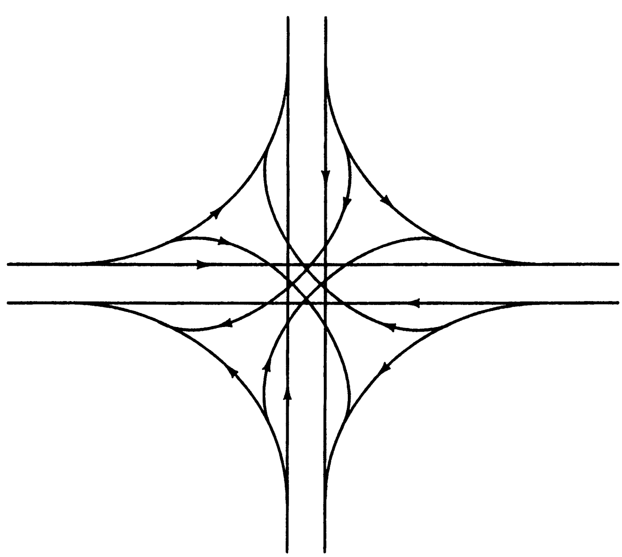

Figure 21.11 *Direct connection between two primary distributors or motorways*

take, relatively little structural content and does not contain loops. It is a form of interchange however which is considered visually intrusive and in addition has a large number of conflict points. An alternative layout to the four-level interchange is the three-level interchange incorporating two loops shown in figure 21.12.

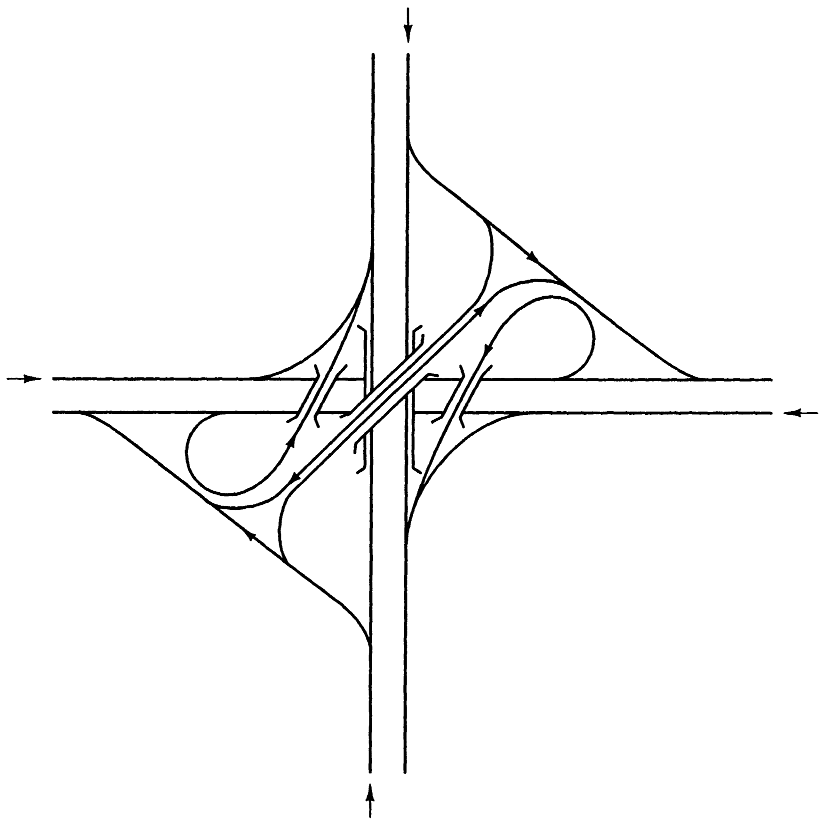

Figure 21.12 *Three-level interchange incorporating two loops*

Whilst this interchange has fewer conflict points it is more expensive structurally than the four-level and is still considered visually intrusive. What is referred to as the cyclic or two-level interchange is shown in figure 21.13; it uses reverse curve links to achieve a low number of conflict points. Its structural content is high and it requires extensive landtake but it produces little visual intrusion and is favoured where land is available.

Reference

1. Department of Transport, Layout of grade separated junctions, TD 22/92 *Departmental Standard Design Manual for Roads and Bridges,* Volume 6, Section 2, Part 1 (1992)

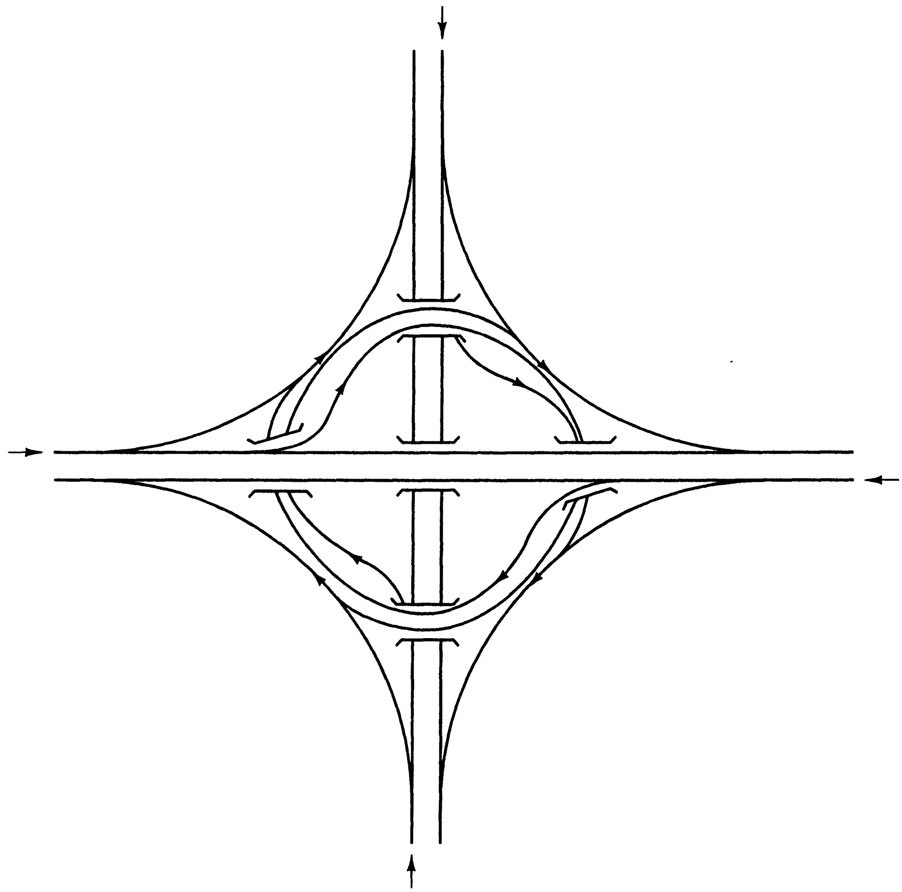

Figure 21.13 *Two-level cyclic interchange*

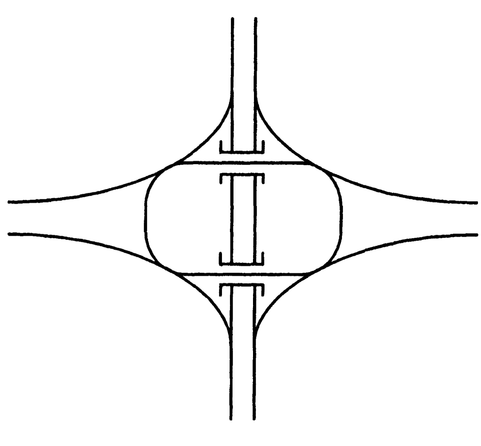

Figure 21.14(a)

Problems

From the following junction types (a)–(i) select a layout that is likely to be satisfactory for the site and traffic conditions described in (1)–(6).

(a) A grade-separated rotary intersection with all the minor road flow passing around the rotary section as shown in figure 21.14(a).

(b) A grade-separated rotary intersection with only turning movements passing around the rotary section as shown in figure 21.14(b).

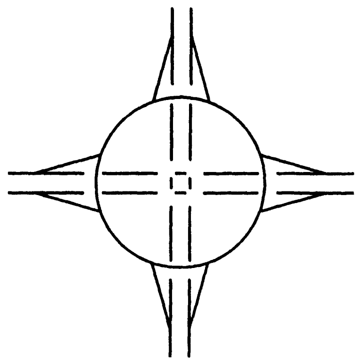

Figure 21.14(b)

(c) A single-level four-way signal-controlled intersection as shown in figure 21.14(c).

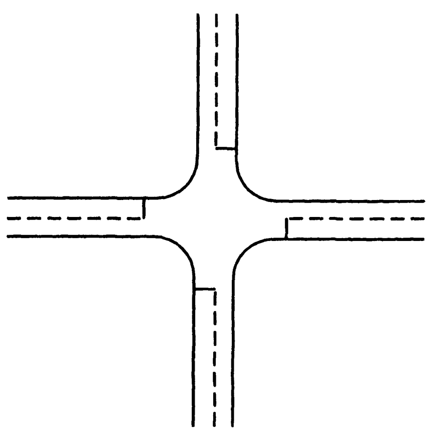

Figure 21.14(c)

(d) A single-level four-way roundabout intersection as shown in figure 21.14(d).

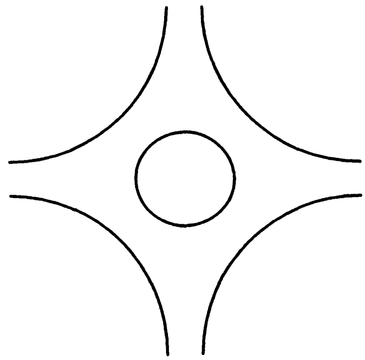

Figure 21.14(d)

(e) A semi-cloverleaf intersection as shown in figure 21.14(e).

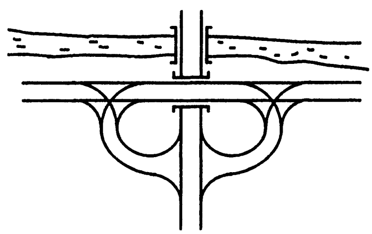

Figure 21.14(e)

(f) A diamond-type intersection as shown in figure 21.14(f).

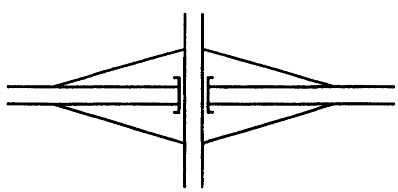

Figure 21.14(f)

(g) A trumpet intersection as shown in figure 21.14(g).

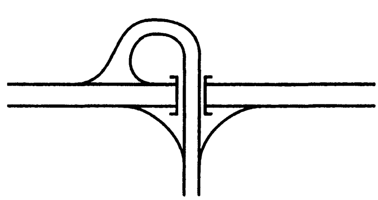

Figure 21.1 4(g)

(h) A semi-grade-separated roundabout as shown in figure 21.14(h).

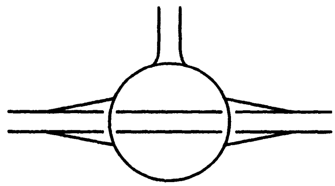

Figure 21.14(h)

(i) A grade-separated intersection with direct connections for all movements as shown in figure 21.14(i)

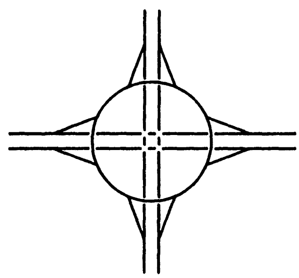

Figure 21.14(i)

(1) Two major rural highways intersect at a right-angled T-junction with heavy right-turning movements from both highways. The traffic flows and accident potential are so great that grade separation is desirable.
(2) In an urban area a heavily trafficked principal traffic route crosses a less heavily trafficked highway and there are a considerable number of turning movements between the highways.
(3) Two motorways intersect in a rural area and it is desired to allow for a full range of turning movements between the two routes.
(4) An all-purpose highway is crossed by a motorway but it is not anticipated that there will be many turning movements between the two highways. There are no topographical limitations on the design of the junction.
(5) The junction is as described in (4) but a river runs parallel with the all-purpose highway preventing any entry or exit highways on one side of the all-purpose highway.
(6) A right-angled T-junction is to be constructed between a motorway and an all-purpose highway and it is anticipated that in the future the all-purpose highway will be extended so that a four-way connection will be required.

Solutions

The junction layouts that would be chosen for the given situation are:
(1) At this intersection grade separation is desirable and as future extension of the three-way layout is not considered, then the trumpet intersection (g) would appear suitable.
(2) As this intersection is heavily trafficked and there is a considerable number of turning movements then grade separation is desirable. It is an urban

situation where land is likely to be limited and so a grade-separated roundabout (a) with all the traffic from the less heavily trafficked route passing around the roundabout would be used.

(3) At the intersection of two motorways where provision for all turning movements has to be made then it is usual in Great Britain to use a grade-separated rotary intersection (b) where a relatively compact layout is required or an intersection with direct connections (i) where more space is available or where speed reductions for the turning movements are undesirable.

(4) Where an all-purpose highway crosses a motorway and there are few turning movements and no topographical limitations then a diamond intersection (f) is usual.

(5) Where a diamond intersection cannot be provided because of physical obstructions than a semi-cloverleaf intersection (e) allows all the slip roads to be provided on one side of the minor road.

(6) At a T-junction between a motorway and an all-purpose highway where it is required in the future to convert the T-junction into a four-way junction then a semi-grade separated roundabout (h) may be preferred. This layout allows the easy extension of the minor road into a four-way junction.

22

Merging, diverging and weaving at grade-separated junctions and interchanges

Merging, diverging and weaving occur at and between many types of junctions in both urban and rural situations. Most frequently these traffic actions are associated with junctions on routes with a high standard of geometric design such as is found on the British motorway and trunk road system.

In British practice a grade-separated junction is one where the commencement or termination of slip roads connects with an at-grade roundabout or major–minor priority junction. An interchange gives uninterrupted movement from one mainline to another mainline carriageway by link roads using merging and diverging movements.

Weaving between traffic streams can occur in both urban and rural situations and is often inherent in urban traffic management systems.

Design standards for merging, diverging and weaving design in the United Kingdom are given by the Department of Transport[1]. A collective term used for link and slip roads is 'connector roads' and their design speeds are given in table 2.1. In many types of grade-separated junctions and interchanges, vehicles are required to change their direction of motion by travelling around loops; the minimum radius of these loops is related to speed. For loops on to or off a motorway the minimum radius

TABLE 22.1
Connector road design speed (km/h)

Mainline design speed	Urban: (i) 100 or (ii) 85 km/h			Rural: (i) 120 or (ii) 100*A* km/h		
Connector road	Interchange link	Slip road	Link road	Interchange link	Slip road	Link road
Design speed	(i) and (ii) 70	(i) and (ii) 60	(i) 100 or 85 (ii) 85 or 70	(i) and (ii) 85	(i) and (ii) 70	(i)120 or 100*A* (ii) 100*A* or 85

is 75 m; for loops on to all-purpose roads the value is 30 m, and off all-purpose roads 50 m.

Maximum flow levels used in the design of merging, diverging and weaving areas are based on flows observed in the United Kingdom and differ according to road class as shown in table 22.2.

TABLE 22.2
Maximum values of hourly flow (km/h)

Road class	*Mainline, two-lane links and slip roads >6 m*	*Single-lane links and slip roads < 5 m*
All-purpose	1600	1200
Motorway	1800	1350

The proportion of heavy goods vehicles in the flow and uphill gradients affect design and it is necessary to correct design hour flows for these effects using the factors given in table 22.3.

TABLE 22.3
Percentage correction factors for gradient

Percentage heavy goods vehicles	*Mainline gradient*		*Merge connector gradient*		
	< 2%	*>2%*	*< 2%*	*>2%<4%*	*>4%*
5		+10		+15	+30
10		+15		+20	+35
15		+20	+5	+25	+40
20	+5	+25	+10	+30	+45

The design flow used in merging, diverging and weaving calculations depends upon the road type. For main urban roads as defined by the *Traffic Appraisal Manual*[2] the 30th highest annual hourly flow is used; for Inter-Urban and Recreational road types the 50th and 200th highest hourly flows respectively are used. These flows should be predicted for 15 years after the opening of the highway.

Design of merging lanes and diverging lanes

In the design of merging and diverging lanes at grade-separated junctions and intersections, the following factors are considered:

(a) The number of lanes and the design hour flows upstream and downstream of the merging and diverging areas on the main carriageway.
(b) The number of lanes and the design hour flows on the merging or diverging link or connector.
(c) The traffic composition and the gradient of the mainline carriageway and the merging or diverging links; correction factors are given in table 22.3.

(d) Maximum hourly flows per lane based on United Kingdom experience; these differ for all-purpose roads and for motorways, and details are given in table 22.2.

(e) The relationship between mainline and entry flow for merging lanes and the requirement that merging and diverging flows should not exceed the upstream and downstream flow for merging and diverging lanes respectively.

Using these limiting factors the Department of Transport has developed flow region diagrams for merging and diverging lane design, as shown in figures 22.1 and 22.2.

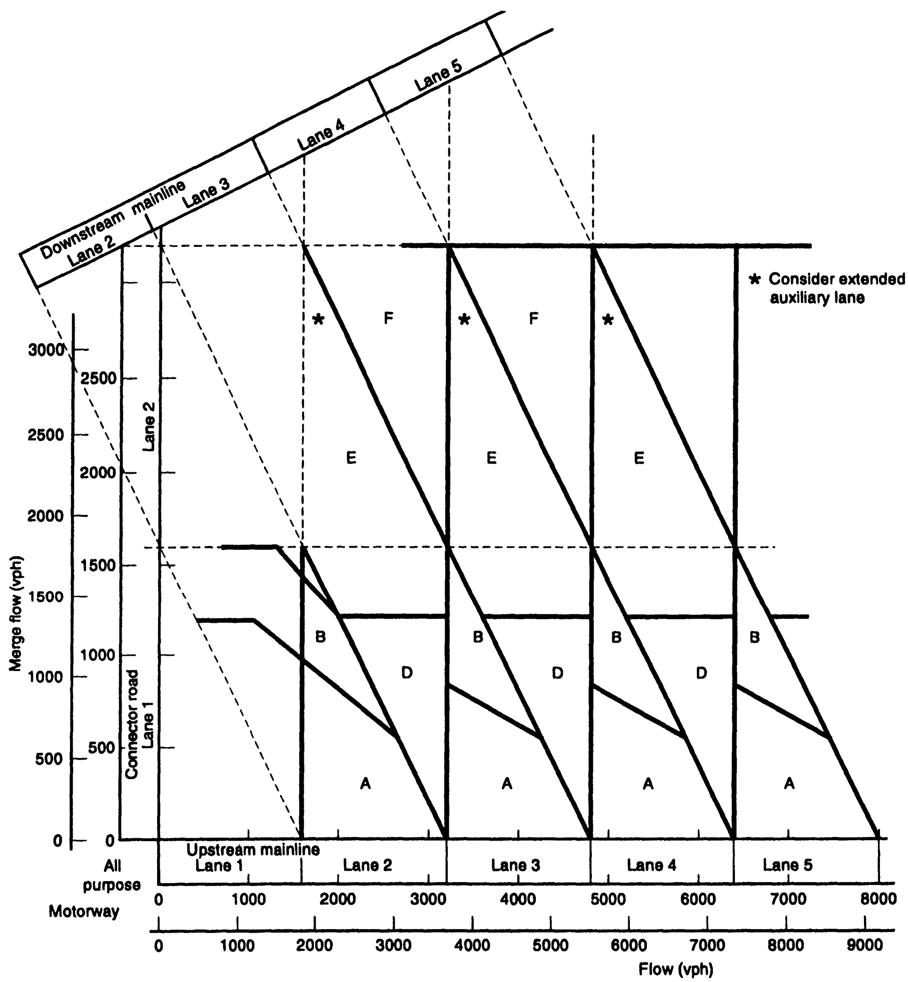

Figure 22.1 *Merging diagram (ref. 1)*

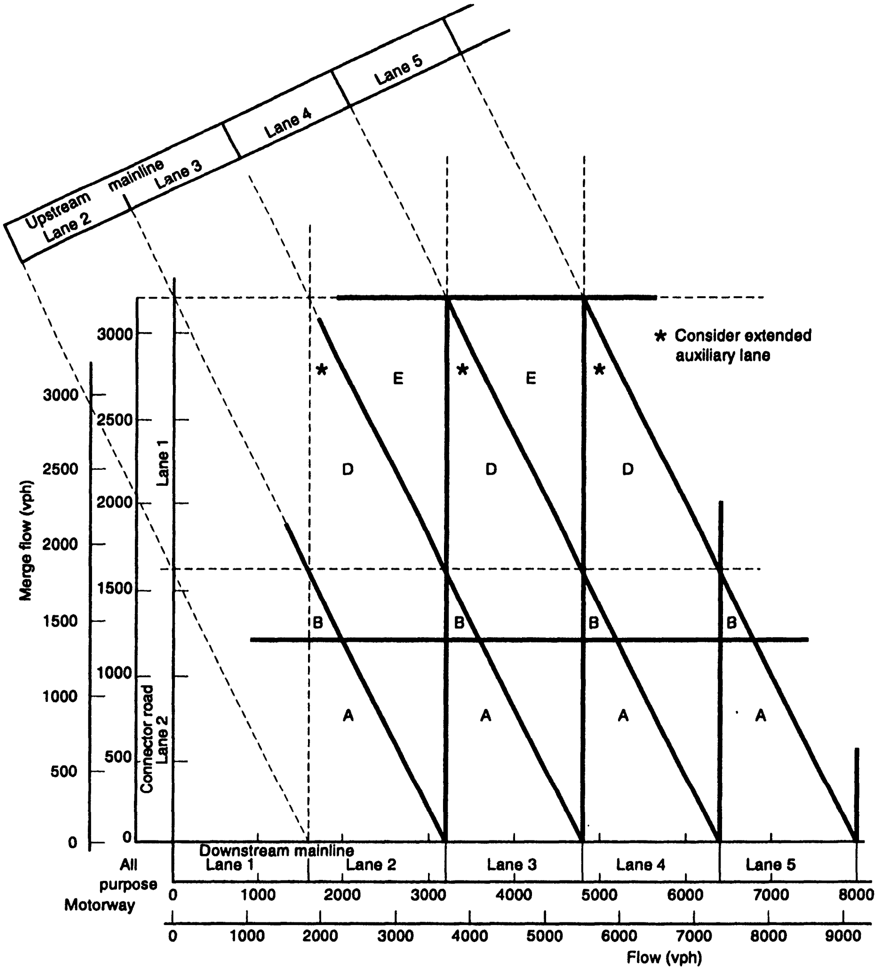

Figure 22.2 *Diverging diagram (ref. 1)*

The following merge/diverge layouts correspond to the letters in figures 22.1 and 22.2:

A: Taper (or normal) merge/diverge
B: Parallel merge/diverge
C: (merge) Ghost island
C: (diverge) Mainline drop at taper diverge
D: (merge) Mainline gain
D: (diverge) Mainline lane drop at taper diverge
E: (merge) Mainline gain at ghost island
E: (diverge) As layout D, but with two lanes off
F: (merge) As layout E, both lanes added

These layouts are illustrated in ref. 1.

Design of weaving sections

Weaving takes place in many situations on highways where traffic streams merge, diverge and cross whilst travelling in the same general direction. This form of weaving is to be found on main carriageways between intersections and is caused by conflicts in the paths of entering, leaving and straight through vehicles. It is also to be found on link roads within free-flow intersections and in the area of some junctions where entering and leaving vehicles conflict.

To design the weaving section it is necessary to know the design speed of the mainline carriageway upstream of the weaving area; the mainline is normally taken as the one carrying the major flow. The maximum allowable hourly flows per lane used in the design of weaving sections are given in table 22.2 and vary with carriageway width and road type.

The design of a weaving section makes use of two graphs which are given in reference 1. The length of the weaving section is given by the larger graph of figure 22.3 which relates the minimum length of the weaving section to the total weaving flow for differing ratios of the maximum allowable hourly flow per lane to the upstream design speed. The smaller graph of figure 22.3 relates the length of the weaving section to the design speed, the greater of the two weaving lengths is then used for design. On rural motorways the desirable minimum weaving length is recommended as 2 kilometres but in extreme cases when the predicted traffic flows are at the lower end of the range given for the carriageway width being considered then an absolute minimum length of 1 kilometre may be considered.

To calculate the required width of the weaving section the minor weaving flow is multiplied by a weighting factor which takes into account the reduction in traffic flow caused by weaving. This factor depends upon the ratio of the minimum length to the actual length of the weaving section.

The number of lanes required is given by

$$N = \frac{Q_{nw} + Q_{w1}}{D} + \left(\frac{2 \times L_{min}}{L_{act}} + 1\right)\frac{Q_{w2}}{D}$$

where N = required number of traffic lanes,
Q_{nw} = total non-weaving flow (veh/h),
Q_{w1} = major weaving flow (veh/h),
Q_{w2} = minor weaving flow (veh/h),
D = maximum allowable mainline flow (veh/h/lane),
L_{min} = minimum weaving length from figure 22.3 (m),
L_{act} = actual weaving length available (m).

It can be seen that the maximum value of the weighting factor is 3 when the actual length of the weaving section is equal to the minimum length as given by figure 22.3. The Department of Transport gives the following advice when the value of N is not an integer[3]. Obviously if the junction can be moved then the actual weaving length will change and the value of N can approximate to a whole number of lanes. If this is not possible then if the size of the fractional part is small and the weaving flow is low then rounding down is possible, conversely a high fractional part and a high weaving flow will favour rounding up to an additional lane. Consideration should be given to the uncertain nature of future traffic predictions, the difficulty of obtaining land in urban areas and the difference between urban commuter and recreational flows.

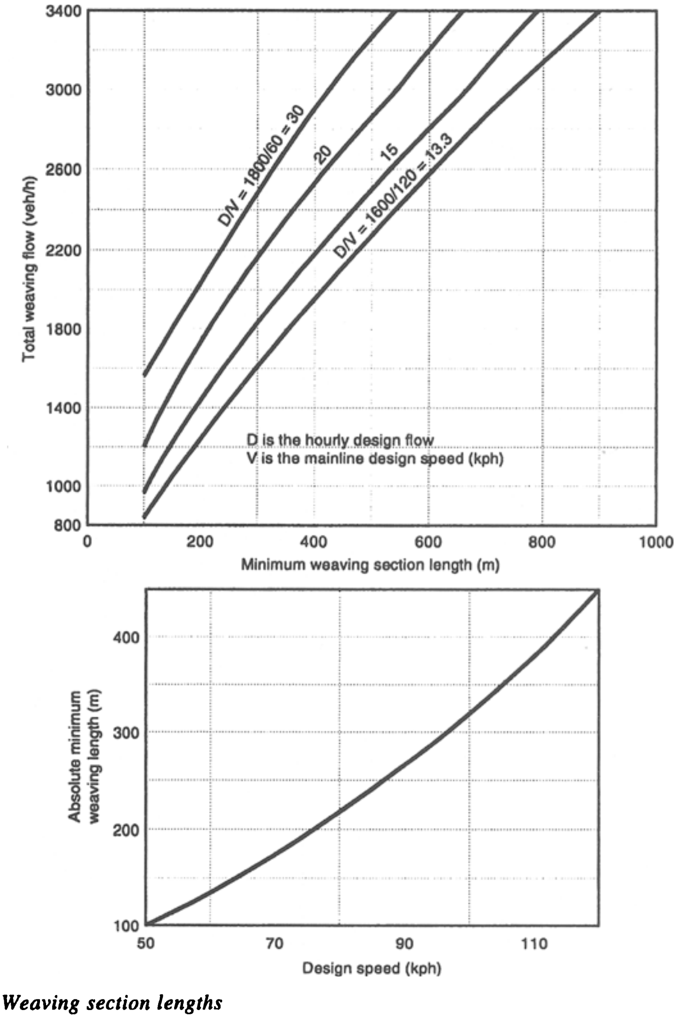

Figure 22.3 *Weaving section lengths*

References

1. Department of Transport, Layout of grade separated junctions, *Departmental Standard* TD 22/92, *Design Manual for Roads and Bridges*, Volume 6, Section 2, Part 1 (1992)
2. Department of Transport, *Traffic Appraisal Manual*, London (1982)
3. Department of Transport, Layout of grade separated junctions, *Departmental Advice Note* TA 48/86, London (1986)

Problems

(1) The design hour traffic flows at a merge at a rural motorway to motorway interchange are: upstream mainline 2500 veh/h (20 per cent heavy goods vehicles), entry link 700 veh/h (15 per cent heavy goods vehicles). The two design hour flows coincide in time, the mainline gradient is 3 per cent uphill and the link gradient is 3 per cent downhill. Select a suitable design configuration.

(2) The design hour traffic flows through a weaving section on a rural all-purpose link road are: total non-weaving flow 3000 veh/h, major weaving flow 1500 veh/h, minor weaving flow 1000 veh/h. The percentage of heavy goods vehicles in the flow is 10 per cent and the gradient through the weaving section is 1 per cent. The length of the weaving section is 700 m. Determine the required width.

Solutions

(1) From table 22.3 the corrected upstream mainline flow is

$$2500 \times 1.25 = 3125 \text{ veh/h}$$

The corrected entry link flow is 700 veh/h.
Figure 22.1 indicates these flows fall into flow region D.
A 'mainline gain' design is therefore appropriate, with two lanes upstream of the merge, on the mainline, and three lanes downstream.

(2) From table 22.3 a correction for heavy goods vehicles and for gradient is not required. From figure 22.3 for a total weaving flow of 2500 veh/h, a mainline speed of 120 km/h and a maximum value of hourly flow of 1600 veh/h (tables 22.1 and 22.2), the minimum length of the weaving section is 560 m.

The number of lanes required in the weaving section is given by

$$N = \frac{Q_{nw} + Q_{w1}}{D} + \left(\frac{2 \times L_{min}}{L_{act}} + 1\right)\frac{Q_{w2}}{D}$$

$$= \frac{3000 + 1500}{1600} + \left(\frac{2 \times 560}{700} + 1\right)\frac{1000}{1600}$$

$$= 2.81 + 1.63$$

$$= 4.44$$

A choice has to be made as to whether to round up or round down the required number of lanes. Many factors would be considered in the practical case: land requirements, the nature of the peak flows, that is, commuter or recreational, and uncertainty of future predictions.

23

Queueing processes in traffic flow

Queueing theory originally developed by A. K. Eriang in 1909 has found widespread application in the problems of highway traffic flow[1].

In any highway traffic situation it is necessary to know: the distribution of vehicle arrivals into the queueing system; whether the source of vehicle arrivals is finite or infinite; whether queue discipline is first-come first-served, priority or random selection; the number of service stations whereby the vehicle may exit from the system and the distribution of service times for each service station.

A typical system occurs at the entry or exit of a stream of vehicles into or from a parking garage. The vehicles arrive at the parking garage at random, form a queue and enter or leave the garage on a first-come first-served basis. The times required by vehicles to pass through a garage entrance or exit form an approximation to an exponential distribution.

While the input distribution is assumed as random and the service time distribution is exponential in the above example, the application of queueing theory to traffic engineering has been mainly developed around the regular, random and Erlang distributions.

When vehicles arrive at random the numbers of vehicles arriving in successive intervals of time can be represented by the Poisson distribution and an understanding of queueing theory can be obtained from this simple case of Poisson-distributed vehicle arrivals in a single lane, in which vehicles depart with an exponentially distributed service rate.

Consider a traffic queue where $P(n, t + \mathrm{d}t)$ is the probability that the queue contains n vehicles ($n > 0$) at time $t + \mathrm{d}t$. There are three ways in which the system could have reached this state if it is assumed that $\mathrm{d}t$ is so small that only one vehicle could have arrived or departed. These are:

(a) a vehicle did not arrive or depart in time t to $t + \mathrm{d}t$;
(b) the queue contained $n - 1$ vehicles at time t and one arrived in time $\mathrm{d}t$;
(c) the queue contained $n + 1$ vehicles at time t and one departed in time $\mathrm{d}t$.

Now with Poisson distributed arrivals

$$P(n) = \frac{(qt)^n}{n} \exp(-qt) \tag{23.1}$$

where $P(n)$ is the probability of n vehicles arriving in time t when the mean rate of vehicle arrival is q. From equation 23.1

$$P(0) = \exp(-q\,\mathrm{d}t) \tag{23.2}$$

where $P(0)$ is the probability of zero arrivals in time dt and

$$P(1) = (q\,\mathrm{d}t)\exp(-q\,\mathrm{d}t) \tag{23.3}$$

where $P(1)$ is the probability of one arrival in time dt.
Expanding equation 23.2

$$P(0) = (1 - q\,\mathrm{d}t + q^2\,\mathrm{d}t^2/2!\ldots)$$

Expanding equation 23.3

$$P(1) = q\,\mathrm{d}t(1 - q\,\mathrm{d}t + q^2\,\mathrm{d}t^2/2!\ldots)$$

Neglecting second and higher powers of dt

$$P(0) = (1 - q\,\mathrm{d}t)$$

and

$$P(1) = q\,\mathrm{d}t$$

Similarly the probability of 0 and 1 departures from the queue are

$$P(0) = (1 - Q\,\mathrm{d}t)$$

and

$$P(1) = Q\,\mathrm{d}t$$

where Q is the mean rate of departure from the queue. Where $n > 0$ the system can reach a state of n vehicles at time $t + \mathrm{d}t$ in the following manner

$$\begin{aligned} P(n, t+\mathrm{d}t) = {} & P(n, t) \times P\text{ (a vehicle does not arrive or depart)} \\ & + P(n-1, t) \times P\text{ (a vehicle arrives)} \\ & + P(n+1, t) \times P\text{ (a vehicle departs)} \\ = {} & P(n, t)\,[(1 - q\,\mathrm{d}t)(1 - Q\,\mathrm{d}t)] \\ & + P(n-1, t)\,(q\,\mathrm{d}t) + P(n+1, t)\,(Q\,\mathrm{d}t) \end{aligned}$$

Ignoring second and higher powers of dt

$$P(n, t + \mathrm{d}t) = P(n, t)(1 - q\,\mathrm{d}t - Q\,\mathrm{d}t) + P(n - 1, t)\,(q\,\mathrm{d}t) + P(n + 1, t)\,(Q\,\mathrm{d}t)$$

$$\frac{P(n, t + \mathrm{d}t) - P(n, t)}{\mathrm{d}t} = -P(n, t)\,(q + Q) + P(n - 1, t)\,(q) + P(n + 1, t)\,(Q) \tag{23.4}$$

In the limit for the steady-state solution the rate of change is zero. Hence

$$0 = -P(n)(q + Q) + P(n - 1)(q) + P(n + 1)\,(Q)$$

or

$$(1 + q/Q)Pn = (q/Q)P(n - 1) + P(n + 1) \tag{23.5}$$

Similarly when $n = 0$, there are two ways in which the queue can contain n vehicles at time t + dt, a change of type (a) or a change of type (c)

$$P(0, t + \mathrm{d}t) = P(0, t)(1 - q\,\mathrm{d}t) + P(0 + 1, t)(Q\,\mathrm{d}t)$$

$$\therefore \frac{P(0, t + \mathrm{d}t) - P(0, t)}{\mathrm{d}t} = P(0 + 1, t)(Qt) - P(0, t)(qt)$$

As before the steady state of the queue probability of n vehicles in the system is

$$P(1) = P(0)\,(q/Q)$$

From equation 23.5 when $n = 1$

$$P(2) = (q/Q)^2\,P(0) \tag{23.6}$$

Similarly when $n = 2$

$$P(3) = (q/Q)^3\,P(0)$$

By induction it can be shown that

$$P(n) = (q/Q)^n\,P(0) \tag{23.7}$$

When the queue size may be infinite

$$P(0) + P(1) + P(2) + P(3) + \ldots + P(\infty) = 1$$

From equation 23.7

$$P(0) + (q/Q)\,P(0) + (q/Q)^2\,P(0) + (q/Q)^3\,P(0) + \ldots + (q/Q)^\infty P(0) = 1$$

$$P(0)\,(1/(1 - q/Q)) = 1$$

$$P(0) = (1 - q/Q)$$

Also

$$P(n) = (q/Q)^n(1 - q/Q) \tag{23.8}$$

The expected number in the queue $E(n)$ is given by

$$\begin{aligned}
E(n) &= \sum_{n=0}^{\infty} nP(n)\\
&= 0 \times P(0) + 1 \times P(1) + 2 \times P(2) + \ldots + nP(n)\\
&= (q/Q)P(0) + 2(q/Q)^2 P(0) + \ldots + n(q/Q)^n P(0)\\
&= (q/Q)P(0)(1 + 2(q/Q) + \ldots + n(q/Q)^{n-1})\\
&= (q/Q)P(0)\,(1/(1 - q/Q)^2)\\
&= q/(Q - q)
\end{aligned} \tag{23.9}$$

Because there is a probability that the queue will be zero the mean queue length $E(m)$ will not exactly be one less than the mean number in the queue $E(n)$

$$\begin{aligned}
E(m) &= \sum_{n=1}^{\infty} (n - 1)P(n)\\
&= \sum_{n=0}^{\infty} P(n)n - P(n) + P(0)\\
&= E(n) - q/Q
\end{aligned} \tag{23.10}$$

As well as the expected number in the queue and the mean queue length, the waiting time w before being taken into service and the total time in the queue v are of considerable importance in the study of traffic phenomena.

The waiting time distribution may be considered in two parts.

Firstly there is the probability that the waiting time will be zero, which means that a queue does not exist, and

$$P(0) = 1 - q/Q, \qquad n = 0$$

Secondly there is the probability that the waiting time for a vehicle is between time w and time $w + \mathrm{d}w$

$$P(w < \text{wait} < w + \mathrm{d}w) = f(w)\,\mathrm{d}w \qquad n > 0 \tag{23.11}$$

Such a delay is possible as long as there is a vehicle in service, which may be expressed as

$$P(n \geqslant 1) = \sum_{n=1}^{\infty} Pn \tag{23.12}$$

For the waiting time for a vehicle to be exactly between w and $w + \mathrm{d}w$ all the vehicles in the queue ahead of the one being considered except the one immediately ahead, must depart in time w and the one immediately ahead must be served in time $\mathrm{d}w$. This is the product of the two probabilities:

$$P(n-1, w) = \frac{(Qw)^{n-1}}{(n-1)!} \exp(-Qw)$$

from equation 23.1 and

$$P(1, \mathrm{d}w) = Q\,\mathrm{d}w$$

Substituting these probabilities in equation 23.11 and summing over equation 23.12

$$f(w)\,\mathrm{d}w = \sum_{n=1}^{\infty} P(n)P(n-1, w)P(1, \mathrm{d}w)$$

$$= \sum_{n=1}^{\infty} (q/Q)^n(1 - q/Q) \times \frac{Q\,\mathrm{d}w}{(n-1)!}(Qw)^{n-1} \exp(-Qw)$$

$$= q(1 - q/Q)\,\mathrm{d}w \exp(-Qw) \sum_{n=1}^{\infty} \frac{(wq)^{n-1}}{(n-1)!}$$

$$f(w) = (q/Q)(Q-q) \exp(-w(Q-q)) \qquad w > 0 \tag{23.13}$$

The moment generating function for waiting times is given by

$$M_w(\theta) = \int_0^{\infty} \exp(\theta w) f(w)\,\mathrm{d}w$$

$$= \int_0^{\infty} \exp(\theta w)(q/Q)(Q-q) \exp(-w(Q-q))\,\mathrm{d}w$$

$$= (q/Q)(Q-q)\int_0^{\infty} exp(-w(Q-q-\theta))\,\mathrm{d}w$$

$$= (q/Q)(Q-q)/(Q-q-\theta)$$

$$M'w(0) = E(w) = q / [Q(Q-q)] \tag{23.14}$$

The average time an arrival spends in the queue is given by $E(w)$ plus the average service time $1/Q$

$$E(v) = 1/(Q - q) \tag{23.15}$$

For more generalised cases when the service time can no longer be described by a negative exponential distribution the expected number in the queue when arrivals are random is given by

$$E(n) = \frac{q}{Q} + \left(\frac{q}{Q}\right)^2 (1 + C^2) / 2\left[\left(1 - \frac{q}{Q}\right)\right] \tag{23.16}$$

where C is the coefficient of variation of the service time distribution, that is the ratio of the standard deviation to the mean.

If the service is exponential then $C^2 = 1$ and equation 23.16 reduces to equation 23.9.

If the service is regular $C^2 = 0$ and

$$E(n) = \frac{q}{Q}\left(1 - \frac{q}{2Q}\right) \Big/ \left(1 - \frac{q}{Q}\right) \tag{23.17}$$

In this case it has been shown that the average time a vehicle spends queuing is given by

$$E(w) = q/2Q(Q - q) \tag{23.18}$$

To illustrate the use of queueing theory in highway traffic flow it is necessary to find some situation in which vehicles are delayed and allowed to proceed in accordance with the simple situations previously considered.

Typical situations occur when vehicles have to stop to enter or leave parking facilities during a period of exceptional demand. Queues can frequently be observed at the entrance to or exit from car parks on public holidays, at weekends at the coast and at sporting events.

It is desirable that the rate of arrival q is reasonably constant and, to make this possible, observations may be divided into shorter periods of time, each with separate value of q.

In some cases the service rate may vary with demand but where queues are forming it is usually possible to assume that the service rate is approximately constant.

As an example of the application of queueing theory observations were made of the number of vehicles waiting to enter two parking areas. One was a multistorey parking garage equipped with automatic entry control equipment and the other was a surface car park where drivers paid the attendant as they entered. In both instances vehicles were delayed as they queued to enter the park and drivers on the approach had considered only one choice of entry gate.

Observations made of vehicle arrivals at the entrance to the surface car park are included to demonstrate the statistical technique whereby vehicle arrivals are shown to be random or non-random.

Where vehicles arrive at random then the numbers of vehicles arriving in successive time intervals may be represented by the Poisson distribution.

Then the probability of n vehicles arriving in a given interval of time t may be calculated from

$$P(n) = \frac{(qt)^n \exp(-qt)}{n!} \tag{23.19}$$

This distribution is often referred to as the counting distribution because it describes the number of vehicles arriving at a given point on the highway.

The numbers of vehicles arriving at the entrance to the surface car park or at the end of the queue in successive 60-second intervals were observed. At this situation marked changes in the arrival rate were not expected and so observations were continued for a period of 3000 seconds and the mean arrival rate taken as q.

To test the form of the arrival distribution and also to determine the mean rate of arrival the observations were tabulated as shown in table 23.1.

Two further sets of data were obtained at this site and the values of both q and Q derived from the observations are included in table 23.4.

It can be seen from tables of chi-squared that there is no significant difference at the 95 per cent level between the observed and the theoretical distributions.

The mean time taken for a vehicle to enter the car park, the service time, was observed as the time interval between successive vehicles in the queue moving away from the attendant or passing beneath the raised barrier in the case of the parking garage.

TABLE 23.1

Vehicle arrival distribution at the entrance to a surface car park

No. of vehicles arriving in a 60 s interval	*Frequency of observed intervals*		*Theoretical frequency ($P(n)\ \Sigma f_0$)*	*Chi-squared*
(n)	(f_0)	($f_0 n$)	(f_t)	(χ^2)
0	1	0	0	
1	2	2	2.40	0.01
2	4	8	4.86	
3	9	27	7.71	0.22
4	10	40	9.18	0.07
5	7	35	8.74	0.35
6	6	36	6.94	0.13
7	4	28	4.72	
8	3	24	3.83	
9	2	18	1.49	0.01
10	1	10	0.71	
11	1	11	0	
	Σ50	Σ239		Σ0.79

$$\text{mean headway} = \frac{60 \times 50}{239} = 12.6 \text{ s}$$

$$q = 1/12.6 = 0.08 \text{ veh/s}$$

$$\text{arrival volume} = 3600/12.6 = 286 \text{ veh/h}$$

Table 23.2 gives the observed distribution of service times at the entrance to the surface car park. The general exponential nature of the service time can be seen. Over 50 per cent of drivers are able to pay the attendant and receive a receipt in less than 8 seconds. A smaller number of drivers require a longer period to tender the fee and receive change. No driver takes longer than 28 seconds to enter the car park after arriving at the entrance.

TABLE 23.2

Vehicle service distribution at entrance to surface car park

Service time class interval (seconds)	*Observed frequency* (f_0)	$f_0 t$	*Theoretical frequency* f_t	χ^2
0–3.9	80	160	87	0.56
4–7.9	53	318	55	0.07
8–11.9	41	410	35	1.03
12–15.9	28	392	22	1.64
16–19.9	18	324	14	0.29
20–23.9	12	264	9	1.00
24–27.9	8	208	6	4.00
28	0	0	10	
	Σ240	Σ2076		Σ8.59

$\bar{t}_s = 2076/240 = 8.7$ seconds

Where the service time distribution is exponential the probability of drivers requiring service times between the class limits $t - \Delta t/2$ and $t + \Delta t/2$ may be calculated from

$$\frac{\Delta t}{\bar{t}_s} \exp - \frac{t}{\bar{t}_s}$$

where Δt is the class interval,

t is the class mark,

$\bar{t}_s$ is the mean service time.

Observed and theoretical values are compared in table 23.2 and it can be seen from tables of χ^2 that there is no significant difference at the 5 per cent level.

Observations of the service time at the entrance to the multistorey parking garage show a considerably different form of distribution. In this case the drivers had only to drive into the entrance bay, take a ticket and drive into the garage when the barrier had been automatically raised. Table 23.3 gives details of these observed service times for one of the 3000-second periods of observation made at this garage.

At the same time as observations were made of the arrival and service time distributions a note was made of the queue length at 100-second intervals.

The average delay to vehicles entering the car park was calculated from

$$\frac{\Sigma \text{ sum of queue lengths} \times 100}{\text{no. of vehicles arriving}}$$

The observed delay to queueing vehicles for each 3000-second period of observation together with the mean arrival and service rates are given in table 23.4.

Theoretical delays were calculated assuming an exponential distribution of service times (equation 23.14) and a regular distribution of service times (equation 23.18).

TABLE 23.3
Vehicle service distribution at entrance to multistorey car park

Service time class interval (seconds)	*Observed frequency* f_0	$f_0 t$
0–3.9	27	54
4–7.9	120	720
8–11.9	31	310
12–15.9	5	70
	Σ183	Σ1154

$\bar{t}_s = 1154/183 = 6.3$ seconds

It can be seen from table 23.4 that observed delays at the entrance to the surface car park can be approximately represented by equation 23.14. On the other hand delays at the entrance to the multistorey car park, where service is approximately regular because it is only necessary to take a ticket and move through a barrier, exhibit characteristics midway between those given by equations 23.14 and 23.18.

Using the equations derived from queueing theory it would be possible to form an estimate of delays of higher arrival volumes. This allows a balance to be obtained between delays to queueing vehicles and the cost of opening additional entrances.

TABLE 23.4

Arrival volume (veh/h)	*q (veh/s)*	*Mean service time*	*Q (veh/s)*	*Observed average delay*	*Theoretical delay Eq. 23.14*	*Theoretical delay Eq. 23.18*
Surface car park						
288	0.0800	8.7	0.1149	25	20	10
293	0.0814	8.7	0.1149	23	21	11
327	0.0908	8.7	0.1149	35	32	16
Multstorey car park						
506	0.1406	6.3	0.1587	41	55	24
497	0.1381	6.3	0.1587	33	49	21
533	0.1481	6.3	0.1587	61	94	44

Reference.

1. P.M. Morse, *Queues, Inventories and Maintenance*, Wiley, New York (1963)

Problems

A census point is set up on a highway where vehicle arrivals may be assumed to be random and the one-way traffic volume is 720 veh/h. All vehicles are required to stop at the census point while a tag is attached, the operation taking a uniform time interval of 4 s.

(a) Is the expected number of vehicles waiting at the census point 1.4, 2.4 or 3.0?
(b) Is the average waiting time of a vehicle at the census point 6.5 s, 8.0 s or 9.4 s?

(c) The enumerators are replaced by untrained staff and the time taken to attach a tag now has the same mean value but the standard deviation of the time taken is noted to be 2 s. Will the number of vehicles waiting at the checkpoint be increased by more than 50 per cent?

Solutions

(a) At a queueing situation in which arrivals are randomly distributed and where the service times are uniform then the expected number of vehicles waiting at the census point is given by equation 23.17

$$E(n) = \frac{q}{Q}\left(1 - \frac{q}{2Q}\right) \Big/ \left(1 - \frac{q}{Q}\right)$$

where q is the mean rate of arrival of vehicles,
Q is the mean rate of departure of vehicles.
In this example

$$\begin{aligned} q &= 720/3600 \text{ veh/s} \\ &= 0.2 \text{ veh/s} \\ Q &= 1/4 \text{ veh/s} \\ &= 0.25 \text{ veh/s} \end{aligned}$$

$$\begin{aligned} E(n) &= \frac{0.2}{0.25}\left(1 - \frac{0.2}{2 \times 0.25}\right) \Big/ \left(1 - \frac{0.2}{0.25}\right) \\ &= 0.8(1 - 0.40)/(1 - 0.8) \\ &= 2.4 \end{aligned}$$

24

Geometric delay at non-signalised intersections

Delay to vehicles at intersections is an important factor in the choice of intersection type and frequently attention is focused on queueing delay which occurs at peak flow times. There is however a further source of delay which is found regardless of traffic conflicts, caused by vehicles slowing down and subsequently accelerating as they negotiate the junction. As this form of delay is determined by the size and shape of the junction, it is referred to as 'geometric delay'. Because geometric delay takes place throughout the whole of the day it is often a substantial portion of the whole of the delay.

In the United Kingdom methods for estimating geometric delay have been largely based on observations made by the Department of Civil Engineering at the University of Southampton. A series of reports on geometric delay at priority junctions, roundabouts and large intersections has been drawn together in *Transport and Road Research Laboratory Supplementary Report* 810[1]. In this report two basic methods of analysis have been used. Firstly, category or regression analysis was used to develop predictive equations for the delay for each vehicle manoeuvre in terms of junction geometry and approach link speeds. Secondly, a model was developed for the delays for each element of the manoeuvre and these elements were then summed to give the total journey time for the whole manoeuvre. The delay is then obtained by subtracting the theoretical journey time assuming that the links meet at the centre of the junction.

The category analysis indicated that for light vehicles delays at priority junctions, grade-separated roundabouts and diamond intersections, geometric delay could be estimated from table 24.1.

Regression analysis, for at-grade roundabouts, grade-separated roundabouts, trumpet intersections and motorway links gave the following expressions for geometric delay (seconds):

At grade roundabouts: $0.11ED + 0.72(Y - V) + 3.06$
Grade-separated roundabouts: $0.08ED + 1.27(Y - V) + 2.38$
Trumpet intersections: $0.17R + 34$

TABLE 24.1
Geometric delay (seconds) determined by category analysis

	Left turn		*Right turn*		*Straight ahead*	
	side	*main*	*side*	*main*	*side*	*main*
Priority junction[a]	7.8	5.7	10.6	6.5	12.2	0
Grade-separated roundabout	10[b]		28[b]		11	0
Diamond intersection	15		19		0	0

[a]For all movements except main ahead, add 2 s if mean link speeds > 65 km/h, and 1.4 s if visibility is substandard.
[b]Add 3 s for a flyover when travelling from the minor road to the motorway.

Motorway links: $0.045ED + 0.4$
where ED is the extra distance due to negotiating the intersection (m)
Y is the average of approach and exit link speeds (m/s)
V is the average speed within the intersection (m/s)
R is the loop radius (m).

The synthetic model is based on a set of empirical approximations where the actual speed profiles are represented in simplified forms. For example, for an at-grade roundabout the link approach speed V is assumed constant up to the point at which constant deceleration starts until the vehicle speed reaches the speed V_b at which it enters the circulating section of the roundabout. During its passage around the circulating section a vehicle has an average speed $V_{b\,c}$ and an exit speed from the circulating section of V_c, whilst in large roundabouts the speed within the circulating section may be higher than at entry and exit. This refinement did not increase the overall accuracy of the model and V_b has been taken as equal to V_c. The vehicle then accelerates at a constant rate until it reaches the link exit speed V_d.

Similar models were adopted for use at trumpet intersections where it was noted that speeds on the loop section were reasonably constant, at priority intersections and at diamond intersections where the points of entry and exit were considered to be coincident. For gyratory systems, motorway links and other types of intersection with complex internal sections, speeds within the intersection were related to geometric features. These features were kerb entry and exit radii, sight distances, inscribed circle diameter, turning radii, entry and exit angles, and lengths of slip roads.

In the observations made to develop the geometric delay models, vehicles were divided into light and heavy vehicles. Light vehicles were defined as those with four tyres or less, and heavy vehicles were other vehicles. For this reason heavy vehicles covered a wide range of vehicle types from six wheel vans to multi-axled articulated and other commercial vehicles including buses and coaches. As a result geometric delay for heavy vehicles varied considerably according to intersection type. In *Transport and Road Research Supplementary Report* 810[1] it is recommended that for free-flowing motorway links the delay to a heavy vehicle should be the same as for a light vehicle; for priority junctions and diamond intersections light vehicle delay should be increased by 25 per cent and at all other intersections light vehicle delay should be increased by 15 per cent.

Reference

1. M. McDonald, N. B. Hounsell and R. M. Kimber, Geometric delay at non-signalised intersections, *Transport and Road Research Laboratory Supplementary Report* 810, Crowthorne (1984)

Problem

Calculate the geometric delay at a priority junction for vehicles turning left from the major road to the minor where the following speed and geometric conditions exist:

Speed on major road entry link 20 m/s (V_a)
Speed on minor road exit link 18 m/s (V_d)
Entry kerb radius 10 m (*ER*)

Solution

Using category analysis (table 24.1), geometric delay = 5.7 + 2.2 = 7.9 seconds. The solution for the synthetic model is as follows:

Overall geometric delay = $JT - [D_{ab}/V_a + D_{cd}/V_d]$. JT is the overall journey time and is equal to $t_{ab} + t_{cd}$ where a is the point on the major road where vehicles commence to decelerate, b and c are coincident for a priority intersection and represent the junction of the major and minor roads, and d is the point on the minor road where vehicles complete their acceleration and attain the minor road link speed. t_{ab} is the deceleration time and t_{cd} is the acceleration time. D_{ab} is the deceleration distance and D_{cd} is the acceleration distance, and

$$t_{ab} = (V_a - V_b) / [1.06(V_a - V_b) / V_a + 0.23]$$

$$t_{cd} = (V_d - V_c) / [1.11(V_d - V_c) / V_d + 0.02]$$

$$D_{ab} = (V_a^2 - V_b^2) / 2[1.06(V_a - V_b) / V_a + 0.23]$$

$$D_{cd} = (V_d^2 - V_c^2) / 2[1.11(V_d - V_c) / V_d + 0.02]$$

$$V_b = V_c = 1.67\sqrt{(ER)}$$

$$V_b = V_c = 1.67\sqrt{(10)} = 5.28 \text{ m s}^{-1}$$

$$t_{ab} = (20 - 5.28) / [1.06(20 - 5.28) / 20 + 0.23]$$
$$= 14.57 \text{ s}$$

$$t_{cd} = (18 - 5.28) / [1.11(18 - 5.28) / 18 + 0.02]$$
$$= 15.81 \text{ s}$$

$$D_{ab} = (20^2 - 5.28^2) / 2[1.06(20 - 5.28) / 20 + 0.23]$$

$= 184.19$ m

$D_{cd} = (18^2 - 5.28^2)/2[1.11(18 - 5.28)/18 + 0.02]$

$= 184.06$ m

$JT = t_{ab} + t_{cd}$

$= 14.57 + 15.81$

$= 30.38$ s

Overall geometric delay $= JT - [D_{ab}/V_a + D_{cd}/V_d]$

$= 30.38 - [184.19/20 + 184.06/18]$

$= 10.94$ s

25

The environmental effects of highway traffic noise

Noise: measurement of sound levels

In our industrial society the number of sources of sound are steadily increasing and when these sounds become unwanted they may be classed as noises. Sound is propagated as a pressure wave and so an obvious measure of sound levels is the pressure fluctuation imposed above the ambient pressure.

If the graph of pressure against time for a single frequency is examined it is found to have a maximum amplitude (P_m) and in sound-pressure measurements it is the root-mean-square pressure that is recorded. Some sounds are a combination of many frequencies while others are composed of a continuous distribution of frequencies. When this occurs the root-mean-square pressure values of all the individual frequencies are added together.

Using pressure units to describe sound levels requires a considerable range of numbers. It is frequently stated that the quietest sound that most people can hear has a sound pressure level of approximately 20 micropascals ($\mu N\ m^{-2}$) while at 100 m away from a Saturn rocket on take-off the sound pressure level is approximately 200 kPa. Rather than use a measurement system with this range, the ratio of a sound pressure to a reference pressure is used so that the sound pressure level is given in decibels by the ratio

$$20 \log_{10} \frac{\text{pressure measured}}{\text{reference pressure}} \text{ decibel (db)}$$

The reference pressure is taken as 20 μPa.

Some idea of the range of sound-pressure levels measured in decibels can be obtained from the values given in table 25.1.

When sound-pressure levels are measured adjacent to a highway, a meter measuring in dB might indicate the same value when a fast moving motor cycle with a high pitched or high-frequency engine note passes and when a slow moving goods vehicle passes with a lower frequency note. The reason why the high pitched note is usually found more annoying than the lower one from the goods vehicle is that the human ear is more sensitive to sounds with higher frequencies than it is to sounds with lower frequencies.

TABLE 25.1
Some typical sound-pressure levels expressed in dB

Sound	*Approximate sound pressure level (dB)*
pneumatic drill	120
busy street	90
normal conversation	60
quiet office	50
library	40
quiet conversation	30
quiet church	20

If a sound level scale is going to be useful for measuring annoyance to human beings it must take this effect into account. Such a scale is measured in dB(A), the sound level measurements then being obtained by an instrument that weights the differing frequency components according to the curve given in figure 25.1. This

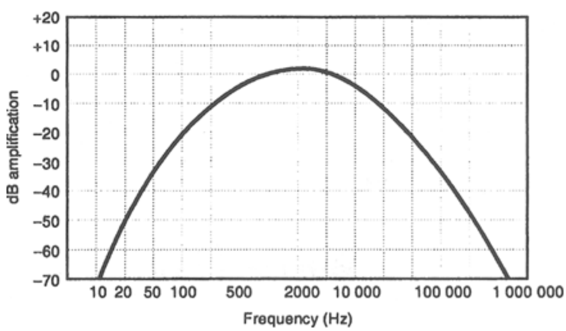

Figure 25.1 *The 'A' weighting curve for sound-level meters*

results in those frequencies that are relatively high or low receiving less weighting than those in the range 1 to 4 kHz. With a sound level meter reading in dB(A) it would therefore be found that the higher frequency note of the fast motor cycle would give a higher reading in dB(A) than the goods vehicle, although both produce the same sound pressure level measured in dB.

Sometimes it is necessary to know not only the sound level in dB or dB(A) but also the contribution that differing frequencies make to the overall sound. To obtain this information the sound is analysed by an instrument that passes it through a system of filters and allows the relative proportions to be determined.

Road traffic noise differs from most other sources of noise in that the level of noise varies both considerably and rapidly. If the variations of sound-pressure level with time are recorded then a record of the type shown in figure 25.2 is obtained. At low sound-pressure levels the noise emitted from vehicles does not cause a great deal of

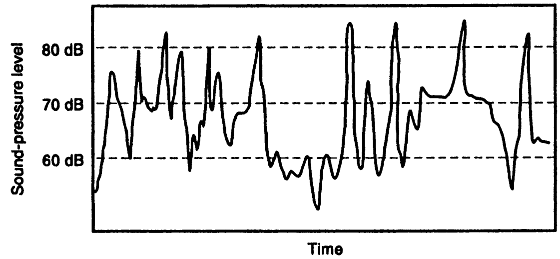

Figure 25.2 *The variation of sound-pressure level with time for highway traffic*

annoyance but at higher levels the annoyance is considerable. For this reason many measures of noise nuisance specify a sound-pressure level that is exceeded for 10, 20, 30 per cent etc. of the time.

When considering a scale that can be used to express a level of noise that should not be exceeded, it should be noted that this scale should be capable of expressing the relative effect on people of the noise being measured. Scholes and Sargent[1] have stated that the unit selected should meet at least the following requirements. Firstly the unit should correlate reasonably well with the criterion of dissatisfaction chosen so that noise levels measured using the unit will represent subjective reactions. Secondly a reasonably accurate set of design rules should be available, covering the estimation of noise exposure from traffic data and the estimation of the performance of noise control techniques, in terms of the chosen unit.

The London Noise Survey[2] measured noise levels at 540 sites and the subjective reactions of 1300 residents. During this survey particular values noted were the L_{10} level (the sound pressure level in dB(A) exceeded for 10 per cent of the time) and the L_{90} level (the sound pressure level in dB(A) exceeded for 90 per cent of the time). These levels represent the extremes of the range that were recorded for 80 per cent of the time. These values are referred to as the 'noise climate' and, together with the L_{10} value, are often quoted in the Wilson Committee Report[3].

Using data from interviews with 1200 residents at 14 sites in the London area where the roads were all straight, level and carrying free-flowing traffic, Langdon and Scholes[4] developed the 'Traffic Noise Index' (TNI). They found this index correlated well with dissatisfaction with noise conditions when the TNI was given by

$$4(L_{10} - L_{90}) + L_{90} - 30$$

A study carried out in Sweden[5] showed a good correlation between noise disturbance and three measures of noise, L_{10}, L_{50} and a noise exposure index based on the energy mean of the noise level, and given by the expression

$$L_{eq} = K \log \frac{1}{100} \Sigma\, 10^{L_i/K f_i}$$

where K is an empirically determined constant,
L_i is the median sound level for the 5 dB(A) interval i,
f_i is the percentage time that a sound level is in the ith interval.

This unit has not however been found to correlate well with experience in the United Kingdom. The difference in performance between the two countries is considered to be caused by the lower noise levels and the greater variability experienced in Swedish traffic conditions.

Another unit has been proposed[6] to cover a range of noise sources, whether highway traffic noise, aircraft noise or laboratory noise. It is referred to as the 'Noise Pollution Level' (LNP) and is given by the following expression

$$\mathrm{LNP} = L_{eq} + 2.56\sigma$$

where L_{eq} is the energy mean noise level of a specified period,
σ is the standard deviation of the instantaneous sound level considered as a statistical time series over the same specified period.

It has been found that the 'Noise Pollution Level' can express annoyance with traffic noise as well as the 'Traffic Noise Index' but both of these units suffer from the fact that their prediction under a wide range of circumstances is still uncertain. For this reason the Building Research Station[1] has proposed that, as an interim measure until further investigations are completed, the average L_{10} taken over the period 6 a.m. to 12 midnight on a weekday would provide a suitable standard for measuring traffic noise nuisance in dwellings. In a few years time however it is expected that sufficient experience will have been gained to use a unit incorporating the variability of traffic noise.

This unit is used to describe the noise exposure of a dwelling and is measured at 1 m from the mid-point of the facade of the building. It is the arithmetic average of the hourly levels in dB(A) just exceeded for 10 per cent of the time. These hourly values are obtained from sampling. The duration of each sample should include the passage of at least 50 vehicles and preferably 100 vehicles.

A maximum acceptable level of noise nuisance

The setting of an acceptable or a maximum level of noise presents considerable difficulties in compromising between what is desirable and what is physically and economically possible. The Wilson Committee[3] made a number of recommendations in 1963 in terms of L_{10} levels averaged over a 24-hour period. The maximum L_{10} levels inside buildings recommended by this Committee are given in table 25.2.

TABLE 25.2
Wilson Committee recommendation for maximum L_{10} levels indoors

Situation	*Day*	*Night*
country areas	40 dBA	30 dBA
suburban areas	45 dBA	35 dBA
busy urban areas	50 dBA	35 dBA

Scholes and Sargent[1] have suggested that as an interim standard a value of L_{10} (6 a.m.–midnight) of 70 dB(A) at residential facades should not be exceeded. The Noise Advisory Council has also recommended that as an act of conscious public policy existing residential development should in no circumstances be subjected to a noise level of more than 70 dB(A) on the L_{10} index.

The prediction of noise levels

The major factors which influence the generation of road traffic noise are:

(a) the traffic flow;
(b) the traffic speed;
(c) the proportion of heavy vehicles;
(d) the gradient of the road;
(e) the nature of the road surface.

In addition the following factors influence the noise level at a reception point distant from the highway:

(f) attenuation of the sound waves due to distance between source and receiver and also due to ground absorption;
(g) obstruction of the sound waves by buildings or noise barriers;
(h) obstruction of the sound waves due to a restricted angle of view of the source line from the reception point;
(i) reflection effects.

When predicting traffic noise levels by the procedure given in the Department of Transport Memorandum, *Calculation of Road Traffic Noise*[7], the factors (a)–(e) are used to predict the basic noise level (in terms of the hourly L_{10} or the 18-hour L_{10}) and the factors (f)–(i) are used to modify the basic noise level to obtain the prevailing noise level at a reception point.

Initially the basic noise level in terms of the 18-hour L_{10} or hourly L_{10} noise level is determined by the 18-hour or hourly traffic flow for a normalised source to receiver distance of 10 m when the mean traffic stream speed is 75 km/h, there are no heavy vehicles in the flow and the roadway is level. The traffic flow (Q) is the two-way flow from 06.00 h to 24.00 h or the hourly flow (q) except when the two carriageways are separated by more than 5 m or where the heights of the outer edges of the two carriageways differ by more than 1 m. In these cases the noise level of each carriageway is evaluated separately.

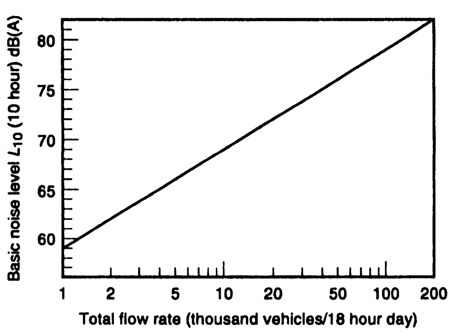

Figure 25.3 *Basic noise level L_{10} (18 h)*

The relationship between L_{10} (18 h) basic noise level and traffic volume is shown graphically in figure 25.3 and expressed mathematically as

$$L_{10}\ (18\ \text{h}) = 29.1 + 10 \log Q\ \text{dB(A)}$$

also $$L_{10}\ (\text{hourly}) = 42.2 + 10 \log q\ \text{dB(A)}$$

A correction has to be made for the mean traffic stream speed and the percentage of heavy vehicles where these differ from 75 km/h and zero per cent respectively.

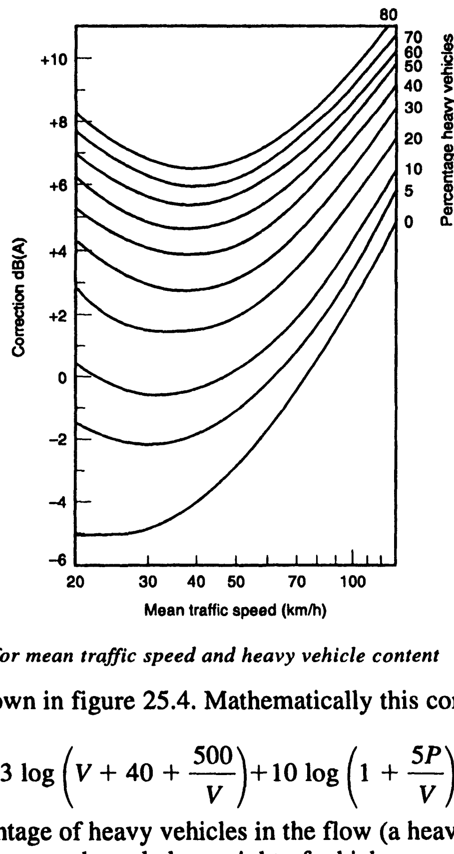

Figure 25.4 *Correction for mean traffic speed and heavy vehicle content*

The correction is shown in figure 25.4. Mathematically this correction is given by:

$$\text{Correction} = 33 \log \left(V + 40 + \frac{500}{V}\right) + 10 \log \left(1 + \frac{5P}{V}\right) - 68.8\ \text{dB(A)}$$

where P is the percentage of heavy vehicles in the flow (a heavy vehicle is any vehicle, other than a motor car, the unladen weight of which exceeds 1525 kg).

The traffic speed (V) may be obtained in either of two ways. It may be the prescribed highest mean speed in any one year within a 15-year period as given in table 25.3, or where local conditions indicate a significantly different value from the prescribed mean speed, then the highway authority may estimate the highest mean speed in a 15-year period.

When the speed is estimated from table 25.3 then the value of speed must be reduced because of the effects of heavy vehicles and gradient. This reduction in speed can be estimated from

$$\Delta V = \left[0.73 + \left(2.3 - \frac{1.15P}{1000}\right)\frac{P}{100}\right] \times G\ \text{km/h}$$

where P and G are the percentage of heavy vehicles and the percentage gradient respectively.

A correction equal to $0.3G$ has also to be made for the additional noise generated by vehicles on a gradient.

TABLE 25.3

Prescribed highest mean speeds in any one year within 15 years

Highway type	*Prescribed speed*	
Special roads (rural) excluding slip roads	(speed limit less than 60 mph)	108 km/h
Special roads (urban) excluding slip roads	(speed limit less than 60 mph)	97 km/h
All-purpose dual carriageways excluding slip roads	(speed limit less than 60 mph)	97 km/h
Single carriageways, more than 9 m wide	(speed limit less than 60 mph)	88 km/h
Single carriageways, 9 m wide or less	(speed limit less than 60 mph)	81 km/h
(Slip roads are to be estimated individually)	(speed limit less than 60 mph)	
Dual carriageways	(speed limit 50 mph)	80 km/h
Single carriageways	(speed limit 50 mph)	70 km/h
Dual carriageways	(speed limit < 50 mph > 30 mph)	60 km/h
Single carriageways	(speed limit < 50 mph > 30 mph)	50 km/h
All carriageways	subject to a speed limit of 30 mph or less	50 km/h

Where carriageways are separated by more than 5 m or where there is one-way traffic then the correction only applies for the upward flow. If there is a single direction down gradient the Memorandum recommends the use of sound-level measurements.

Road surface texture has a number of effects on noise generation depending on whether the texture is randomly distributed as is the case with many bituminous surfaces or transversely aligned as with concrete surfaces. The extent to which water can drain through a bituminous surface also has an effect on noise generation. For roads which are impervious to surface water and where the traffic speed used in figure 25.4 is greater than 75 km/h the following corrections should be made. Correction for concrete surfaces is $10 \log(90TD + 30) - 20$ dB(A), for bituminous surfaces it is $10 \log(20TD + 60) - 20$ dB(A) where TD is the texture depth. Where road surfaces do not meet these requirements because the speed is less than 75 km/h then for impervious surfaces 1 dB(A) should be subtracted from the basic noise level. For roads surfaced with pervious macadams 3.5 dB(A) should be subtracted from the basic noise level.

With the basic noise level determined it is necessary to make corrections for the factors which affect the propagation of sound between the source and the reception position.

First a correction is made for the distance between the source and reception position. The nature of the ground influences distance attenuation and the Memorandum divides the ground over which the sound is propagated into hard ground and grassland. Hard ground is defined as mainly level ground, the surface of which is predominantly (more than 50 per cent) non-absorbent, that is, paved, concrete, bituminous surfaces and water.

The Memorandum gives the distance attenuation in these circumstances as:

$$\text{distance correction} = -10 \log (d'/13.5) \text{ dB(A)}$$

where d' = the minimum slant distance between the effective source position and the reception point.

The effective source line is assumed to be 3.5 m in from the near kerb and at a height of 0.5 m above the road surface. Where the carriageways are considered separately then the source line for the far carriageway is taken as 3.5 m in from the far kerb and the distance from the kerb to be used in figures 25.5 and 25.6 is taken as 7 m in from the far kerb.

The distance correction is obtained graphically from figure 25.5 and in this chart the distance is measured from the edge of the nearside carriageway.

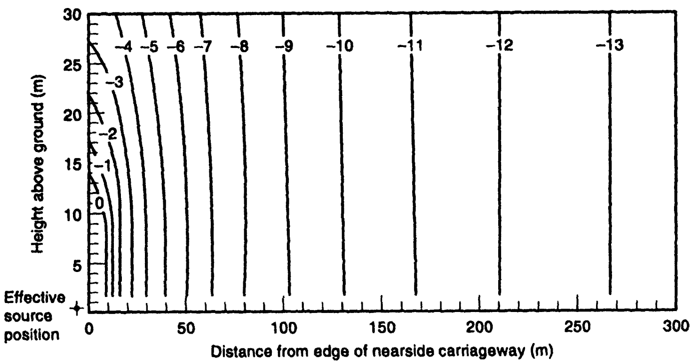

Figure 25.5 *Correction for propagation over hard ground*

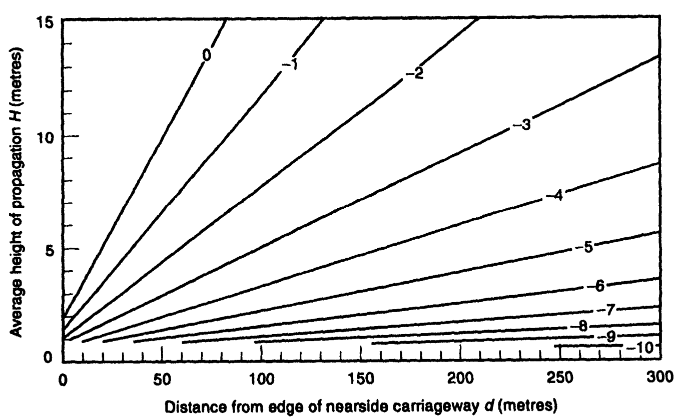

Figure 25.6 *Correction for ground absorption, proportion of absorbent ground 1*

Where the surface between the source line of the noise and the reception point is predominantly of an absorbent nature, such as grass, cultivated or planted land, the effects of ground absorption must be considered in addition to distance attenuation.

Mathematically the correction is given as:

$$\text{Correction} = 5.21 \log\left(\frac{6H - 1.5}{d + 35}\right) \text{ dB(A)}$$

$$(d \geqslant 4 \text{ m}) \qquad 0.75 \leqslant H < \frac{d+5}{6}$$

$$= 5.21 \log \left(\frac{3}{d + 3.5}\right) \text{dB(A)}$$

$$= 0 \text{ when } H \geqslant \frac{d+5}{6}$$

where d is the horizontal distance between the edge of the nearside carriageway and the reception point.

This distance and ground absorption correction is obtained graphically from figure 25.6 and in this chart the distance is measured from the nearside kerb or as previously stated for roads where the two carriageways are being considered separately.

In many practical cases the screening effect of objects between the source line and receiver must be taken into account. Dealing initially with the attenuation due to either a long noise barrier or a continuous obstruction caused by site conditions, the basis of the correction is the path difference between the source line and the reception point as illustrated in figure 25.7.

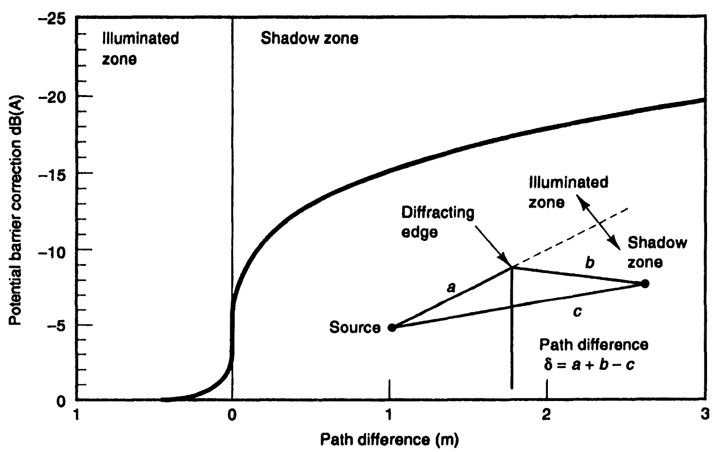

Figure 25.7 *Potential barrier correction*

This correction is applied to the basic noise level which has been corrected for distance using the hard ground correction. Ground absorption is ignored since the near ground rays are obstructed by the barrier.

Where only part of the road is shielded, for example by a short barrier, then a modified correction procedure has to be used which is conveniently illustrated by the following steps:

(a) Let θ_H be the total angle of view of the unscreened section of the source line for which the ground between the road and the reception point is hard. Let θ_S be the similar angle of view for which the ground is soft (grassland). Let θ_B be the total angle of view obtained by barriers. (Normally for straight roads $\theta_B + \theta_H + \theta_S = 180°$.)

(b) Calculate the contribution to the noise at the reception point due to those lengths of line source covered by θ_H using the hard ground propagation correction. Correct this for the restricted angle of view by applying the correction $10 \log (\theta_H/180)$ dB(A).

(c) Repeat using θ_S and the grassland propagation and absorption correction and restricted field of view correction to obtain this contribution at the reception point. Combine with the value calculated in (b) to obtain L_U.
(d) Calculate the contribution from the screened section of road by calculating the unobstructed noise level at the reception point for hard ground and apply the long barrier correction and the restricted angle of view correction to give L_B.
(e) Combine L_U and L_B to give L_{10}.

In many situations reflection of noise from noise barriers or substantial buildings beyond the traffic stream along the opposite side of the road increases the noise level by + 1.5 ($\theta' + \theta$) dB(A) where θ' is the sum of the angles subtended by the reflecting surfaces.

These values calculated by these techniques are 'free field' noise levels. To calculate the noise level 1 m from a facade, as is required by the 1975 Noise Insulation Regulations, a correction of +2.5 dB(A) has to be made to allow for reflection from the facade.

Figure 25.8 shows how exposures from two sources are combined.

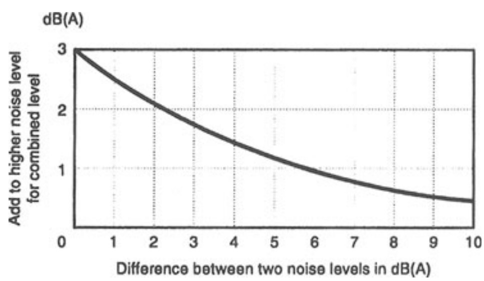

Figure 25.8 *Combining exposures from two sources (adapted from ref. 1)*

Controlling traffic noise by means of screens

When noise exposures from highway traffic are predicted it is frequently found that the predicted exposures are greater than the recommended maximum values. A method of reducing the noise exposure is by means of screens, making possible a reduction in noise exposure averaging 10 dB(A).

A disadvantage of screening is the size of the barriers required because if noise attenuation is to be obtained the facade being protected must be well within the sound shadow formed by the screen. It should not be possible to see over the top or around the ends of the screen if effective insulation is to be achieved. Narrow belts of trees and shrubs are considered relatively unsatisfactory by the Building Research Station for noise attenuation purposes. To be effective the belt should be about 50 m wide, dense and extend to ground level. The foliage should also be evergreen for all-year screening.

To produce an effective sound shadow a noise barrier should either be close to the highway or the building facade being protected. It should be dense enough to create an effective shadow – a recommended value is at least 10 kg/m^2 – and there should not be any sound paths either through or under the barrier. Structural and aesthetic considerations are also very important and these latter factors will obviously influence the design.

Insulation of dwellings does however offer greater possibilities for the reduction of noise nuisance than can usually be obtained in existing situations by distance or noise barriers. It does not of course reduce noise nuisance in gardens and adjacent areas.

The use of thicker glass has only a marginal insulating effect (approximately 1–3 dB(A) improvement) but double windows with a space of at least 150 mm between the panes will give a sound insulation of between 20 and 30 dB(A) when one leaf is sealed and the other is well fitting. Unfortunately the insulation value is reduced when a powered fan is used to provide ventilation.

Controlling vehicle noise

The obvious way in which traffic noise may be reduced is by a reduction in the noise emitted by individual vehicles and, as a consequence, the Quiet Heavy Vehicle Project[8] was initiated under the overall direction of the Transport and Road Research Laboratory. The project was a programme of research and development which aimed to produce a quiet diesel-engined heavy commercial vehicle having external noise levels approximately similar to those of the private car. Initially there was a research phase during which the various noise-producing components of a standard vehicle were quietened experimentally, and later a development phase during which the 'research' vehicle was developed by industry into a commercially viable version for demonstration.

The standard vehicle selected for research was the diesel-engined heavy articulated vehicle as this was considered to be the most difficult to quieten, having the most powerful and noisiest engine; the upper limits of sound levels to be emitted by the vehicle components are given in table 25.4.

TABLE 25.4

Upper limits of sound levels to be emitted by vehicle components (from Transport and Road Research Laboratory Supplementary Report 746)

Source	*Maximum level dB (A)*	
	at 1 m	*at 7.5 m*
Engine and gear box	92	77
Air intake, exhaust system	84	69
Cooling system	84	69
Development of a practical exhaust system	84	69
Cab noise	75	

As a result of this programme, two vehicles were produced which emitted noise levels 10 dB(A) less than the original production vehicles and achieved similar reductions in internal cab noise. One was a standard Leyland Buffalo 4 × 2 tractor rated for operation at 32.5 tonnes and powered by a Leyland 510 158 kW (212 bhp) turbocharged diesel engine. The other was a Foden/Rolls Royce tractor which has been developed by the vehicle and engine manufacturers into a fully engineered, practical and commercially viable vehicle which has met the target projects with a weight penalty of less than 1 per cent of the weight of a fully laden tractor–trailer combination.

The developed Foden quiet heavy vehicle was placed with a haulage contractor for a two year trial period. The object of the trial was to determine the durability of the vehicle's noise reduction features and to assess any effects which the reduction measures might have on the costs and difficulty of operating and servicing the vehicle. Data was collected on journeys made, loads carried and full consumption together with details of maintenance costs and problems of operation.

During the test period the vehicle travelled over 116 000 km and carried over 11 000 tonnes of payload, the maximum payload being carried on approximately 25 per cent of all journeys. The vehicle performed well with fuel consumption within the normally expected range; driver acceptability was good, partly because of the lower cab noise levels. It was reported that the vehicle received consistently good characteristics and the exceptionally quiet idling condition was particularly noted.

References

1. W. E. Scholes and J. W. Sargent, Designing against noise from road traffic, *Bldg Res. Stn Current Paper* CP.20/71 (1971)
2. A. G. McKennel and E. A. Hunt, Noise annoyance in Central London. The Government Social Survey, *C. O. I. Report* SS.332 (1962)
3. Sir Alan Wilson, Chairman, *Noise – Final report of the committee on the problem of noise*, Cmnd 2056, HMSO, London (1963)
4. F. J. Langdon and W. E. Scholes, The traffic noise index: a method of controlling noise nuisance, *Bldg Res. Stn Current Paper* CP.38/68 (1968)
5. Statens Institut for Byggnasforskning, Trafikbuller i Bosladsomraden, *Statens Institut for Byggnasforskning, Rapport* 36/68, Stockholm (1968)
6. D. W. Robinson, An outline guide to criteria for the limitation of urban noise, *Ministry of Technology, NPI, Aero Report Ac. 39* (1969)
7. Department of Transport, Welsh Office, *Calculation of Road Traffic Noise*, HMSO, London (1988)
8. J. W. Tyler and J. F. Collins, TRRL Quiet Heavy Vehicle Project, *Transp. Rd Res. Laboratory Report* LR 1067 (1983)

Problems

(1) The sound pressure in the driving cab of a heavy goods vehicle is 2 Pa. Is the sound-pressure level measured in dB, with reference to 20 μPa: 80, 100, or 150?

(2) A heavy goods vehicle travelling at 60 km/h is found to produce a noise with a very low frequency in the range 100–500 Hz while a high-performance car travelling at 120 km/h is found to produce a noise with a frequency in the range 1000–2000 Hz. If both vehicles produce a noise with the same pressure level, which noise source will register the higher noise pressure level when measured in dB(A)?

(3) Variations of the sound-pressure level with time at a site adjacent to the highway are shown in figure 25.9. Indicate on the diagram which of the broken horizontal lines is likely to represent the L_{10}, L_{50} and L_{90} sound-pressure levels.

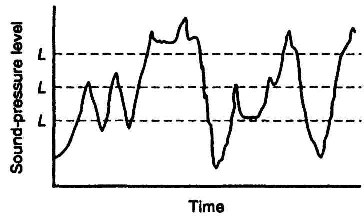

Figure 25.9

(4) Many indices of traffic noise have been developed. From the given list of indices select the appropriate index to meet the requirements listed.

Indices

(a) The traffic noise index
(b) The sound-pressure level in dB(A) exceeded for ten per cent of the time
(c) The noise pollution level
(d) The sound-pressure level in dB(A) exceeded for ninety per cent of the time.

Requirements

(a) An index that has been extensively correlated with the dissatisfaction of human beings with traffic noise
(b) An index that can be used to express the noise nuisance of both aircraft and highway traffic
(c) An index that has limited correlation with dissatisfaction with highway traffic noise but which expresses the variability of sound pressure levels
(d) An index that expresses the background noise level.

(5) It has been suggested that the L_{10} noise index should not exceed a given value at building facades. Is this value in the region of 30 dB(A), 70 dB(A) or 100 dB(A)?

(6) Calculate the noise exposure level at 1 m from the facade of a building 45 m from the nearest edge of a free-flowing traffic lane. The average commercial vehicle content of the traffic stream is 40 per cent; the mean stream speed is 60 km/h; the traffic volume is 15 000 vehicles/day. A second free-flowing traffic stream is 50 m distant from the building facade; there are no commercial vehicles in this stream; the mean stream speed is 80 km/h; the traffic volume is 30 000 vehicles/day. Ignore the effect of ground attenuation, gradient and road surface texture.

(7) Calculate the improvement in noise exposure that will be obtained by the use of a very long barrier adjacent to a motorway where the geometrical layout is as shown in figure 25.10.

(8) Calculate the improvement in noise exposure that will be obtained in the previous example if the barrier has a length of 200 m downstream and 50 m upstream of the reception point at which the noise exposure is to be calculated (figure 25.11).

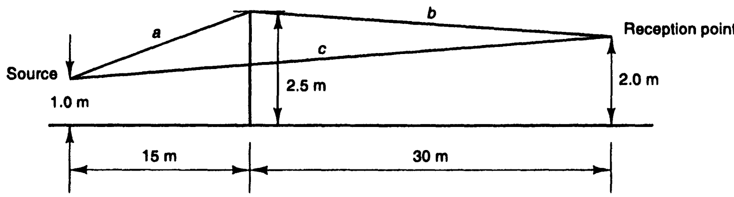

Figure 25.10

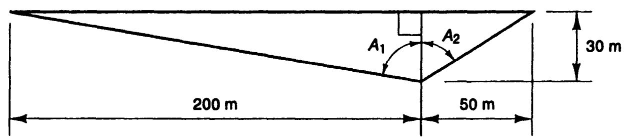

Figure 25.11

Solutions

(1) Sound-pressure level $= 20 \log_{10} \dfrac{\text{pressure measured}}{\text{reference pressure}}$ (dB)

$$= 20 \log_{10} \frac{2}{2 \times 10^{-5}} \text{ dB}$$

$$= 2 \times 5 \text{ dB}$$

$$= 100 \text{ dB}$$

(2) Reference to figure 25.1 shows that sounds with frequencies in the range 100–500 Hz produce less effect on the human ear than sounds in the frequency range 1000–2000 Hz. If both sounds have the same sound-pressure level when measured in dB then the high-performance car with a noise in the frequency range 1000–2000 Hz will register a higher sound-pressure level in dB(A).

(3) The correct placings of the L_{10}, L_{50} and L_{90} sound-pressure levels are marked in figure 25.12.

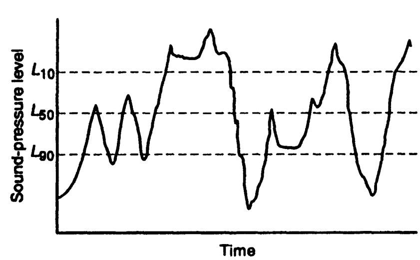

Figure 25.12

(4) The correct combination of indices and requirements is shown in table 25.5.

TABLE 25.5

Index		*Requirement*
(a) traffic noise index	(c)	an index that has limited correlation with dissatisfaction with highway traffic noise but expresses the variability of sound pressure levels
(b) L_{10}	(a)	an index that has been extensively correlated with the dissatisfaction of human beings with traffic noise
(c) noise pollution level	(b)	an index that can be used to express the noise nuisance of both aircraft and highway traffic
(d) L_{90}	(d)	an index that expresses the background noise level

(5) The Noise Advisory Council has recently recommended that the L_{10} index should be used for measuring traffic noise disturbance. It also recommended that as a conscious act of public policy existing residential development should not be subjected to a noise exposure level greater than 70 dB(A) on the L_{10} index unless remedial or compensatory action was taken by the responsible authority (Reported in *Hansard*, 24th June 1971).

(6) Calculation of noise exposure level from stream 1.

From figure 25.3	For a flow rate of 15000 vehicles/18 h day L_{10} (18h) = 70.9 dB(A).
From figure 25.4	Correct for speed and heavy vehicle content. Correction = 4.7 dB(A).
From figure 25.5	Correct for distance over hard ground (assume height of reception point is 2 m). Correction = −5.5 dB(A).

Reflection effect at facade = +2.5 dB(A)
L_{10} (18 h) at facade due to stream 1 = 72.6 dB(A)

Similarly for stream 2.

From figure 25.3	L_{10} (18 h) = 73.9 dB(A).
From figure 25.4	Correction = 0.5 dB(A).
From figure 25.5	Correction = −6.0 dB(A).

Reflection effect at facade = +2.5 dB(A).
L_{10} (18 h) at facade due to stream 2 = 70.9 dB(A).

From figure 25.8 Combined L_{10} (18 h) at facade = 75 dB(A).

(7) The improvement in noise exposure that will be obtained by the use of a very long barrier adjacent to a motorway is calculated from the geometrical layout shown in figure 25.13.

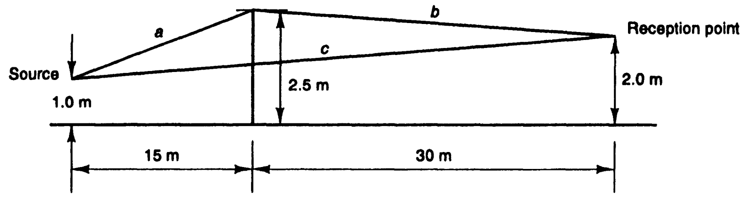

Figure 25.13

The improvement is related to $(a + b) - c$ and

$$
\begin{aligned}
a^2 &= 15^2 + (2.5 - 1.0)^2\\
&= 225 + 2.25\\
&= 227.25\\
\therefore a &= 15.074 \text{ m}\\
b^2 &= 30^2 + (2.5 - 2.0)^2\\
&= 900 + 0.25\\
\therefore b &= 30.004 \text{ m}\\
c^2 &= 45^2 + (2.0 - 1.0)^2\\
&= 2025 + 1.0\\
&= 2026\\
\therefore c &= 45.011 \text{ m}\\
(a + b) - c &= 0.067 \text{ m}
\end{aligned}
$$

From figure 25.7 it can be seen that there is a reduction in the L_{10} noise exposure of approximately 9.5 dB(A).

(8) The noise attenuation due to an unsymmetrical barrier may be calculated from a consideration of the relative geometric layout of the barrier and reception point.

$$\tan A_1 = \frac{200}{30} = 6.6667$$

$$A_1 = 81^\circ\, 28'$$

$$\tan A_2 = \frac{50}{30} = 1.6667$$

$$A_2 = 59^\circ\, 0'$$

$$\theta_B = 140^\circ\, 28'$$

Assume hard ground conditions, then $\theta_S = 0°$, $\theta_H = 39°32'$
Let L_{10} be the noise level at the reception point when the barrier is not present. Correction for angle of view is

$$L_U = L_{10} + 10 \log (39°32'/180°) = L_{10} - 6.6 \text{ dB(A)}$$

Unobstructed noise level at the reception point corrected for barrier and field of view attenuation L_B is: $L_{10} - 9.5 + 10 \log (140°28'/ 180°) = L_{10} - 10.6$ dB(A).
Combining L_U and L_B gives shielded L_{10}. Then attenuation due to barrier is given by: shielded $L_{10} - L_{10} = (L_{10} - 10.6) + (L_{10} - 6.6)$

$$= 10 \log (10^{-1.06} + 10^{-0.66})$$

or, the attenuation of the barrier is 5.1 dB(A).

26

The environmental effects of highway traffic pollution

Air pollution from road traffic

During recent years there has been a widespread attempt to reduce air pollution from all sources. In the United Kingdom the Clean Air Act of 1956 has resulted in a noticeable decrease in coal consumption and a reduction in air pollution from domestic and industrial sources. During this same period there has been a marked increase in the volume of road traffic and consequently an increase in pollution from this source. The National Society for Clean Air[1] estimates that, during 1956 in the United Kingdom, the total coal and oil consumed was equivalent to 276 million tons (coal equivalent) and only 26 million tons of this were used for road or rail transport. It is however an increasing source of pollution, which is emitted in situations close to human activity. Approximately one-third of the carbon monoxide in the atmosphere is produced from vehicle exhausts.

The major sources of atmospheric pollution caused by motor vehicles have been given by Sherwood and Bowers[2] and may be classified as:

(a) exhaust gases;
(b) evaporative losses from the fuel tank and carburettor;
(c) crank case losses;
(d) dust produced by the wearing away of tyres, brake linings and clutch plates.

Considering the exhaust gases, the following compounds are normally present in the discharge from vehicle exhausts:

(a) carbon dioxide;
(b) water vapour;
(c) unburnt petrol;

(d) organic compounds produced from petrol;
(e) carbon monoxide;
(f) oxides of nitrogen;
(g) lead compounds;
(h) carbon particles in the form of smoke.

On occasions these components of the exhaust may react with each other to produce unpleasant secondary products. The most well-known effect of this type is the Los Angeles 'smog', which, because of the bright sunlight and the topography of the region, is formed by the reaction of the oxides of nitrogen and some of the hydrocarbons.

Both petrol and diesel engines give rise to similar products in their exhausts but the relative proportions differ. Diesel engine exhaust gases contain significantly lower proportions of pollutants than do those produced by petrol engines. Unfortunately, an incorrectly operated or maintained diesel engine is liable to emit smoke and produce an offensive smell but even then, apart from carbon particles, the degree of pollution is less than that produced by petrol engines.

The effects of these pollutants have been reviewed at the Transport and Road Research Laboratory[2] and the conclusions will be summarised.

Unburnt fuel and secondary products produced from the fuel

Unburnt fuel is emitted to the atmosphere by evaporation from the fuel tank and carburettor. A high proportion of the hydrocarbons in the crank case blow-by and in the exhaust gases also consists of unburnt fuel. The constituents of petrol are not considered to be toxic, but some of them have slight anaesthetic effects in high concentrations.

Many compounds are found in the gaseous products of the fuel emitted in the exhaust gases. Of these a significant proportion of aldehydes is produced. Aldehydes have an irritant action on the eyes and on the respiratory system and they can be smelt even in very small amountss.

In addition to the gaseous products a number of polynuclear aromatic compounds are also emitted with the exhaust gas in the form of very fine particles, which can persist in the air for lengthy periods. They are important because some of them, such as benzpyrene, are known to be carcinogenic, but the extent of the health hazard for the proportions present is not known. However, half the concentration of polynuclear aromatic hydrocarbons in the urban atmosphere is due to motor vehicles, the exposure level being equal to that produced by smoking one cigarette a day.

Carbon monoxide

The dangers of the absorption of carbon monoxide and its reaction with haemoglobin in the blood are well known. The degree of absorption depends on the carbon monoxide content in the air, the period of exposure and the activity of the individual. A survey by the Transport and Road Research Laboratory of the carbon monoxide content of the air in busy city streets in the United Kingdom has indicated that at the present levels, road users will not be aware of any discomfort from this source, but this may not be true for policemen and others operating in city streets for long periods of time.

While it is believed that carbon monoxide is unlikely to leave any permanent effects or cause acute physical discomfort, its effect cannot be entirely discounted because relatively small concentrations of carboxy-haemoglobin in the blood have been shown to temporarily impair mental ability. Fortunately this is only likely to occur in still weather in traffic jams, and even then only when the subject has been working hard for an hour.

It has been stated that an indication of the scale of the problem of carbon monoxide due to exhaust emissions is that cigarette smoking produces significantly higher exposure to carbon monoxide than that experienced by pedestrians on heavily trafficked roads. It should however be remembered that smoking in public places is likely to face increasing public opposition.

Oxides of nitrogen

Both nitric oxide and nitrogen dioxide are produced by the internal combustion engine, the former in much greater quantities than the latter, but nitric oxide oxidises to nitrogen dioxide. Typically in a city street there is twice as much nitric oxide as there is nitrogen dioxide.

Nitrogen dioxide is considerably more toxic than nitric oxide and a limited amount of data from around the most heavily trafficked roads in the United Kingdom shows that the concentration exceeds the levels recommended in other countries. If an Air Quality Report is prepared then this pollutant should be investigated and likely concentrations exceeding 0.05 parts per million reported.

Lead compounds

In the United Kingdom a great deal of concern has arisen regarding the long-term hazard to health from lead due to vehicle exhaust emission. Airborne lead can be deposited on crops adjacent to highways and then enter the body via the food chain, although in most cases these crops make only a small contribution to the diet. Considering all sources of lead it has been recommended that the mean annual concentration of airborne lead should not exceed 2 micrograms per cubic metre in places where people might be continuously exposed for long periods. To achieve this level the lead content of petrol in the United Kingdom has been progressively reduced from 0.4 grams per litre to 0.14 grams per litre, and it is expected that this will allow the recommended standard of 2 micrograms per cubic metre to be met in virtually all residential areas adjacent to major highways.

Estimation of exhaust pollution levels

It is generally accepted that the assessment of the air pollution due to a highway scheme may be made in terms of estimated levels of carbon monoxide[3]. Where it is desired to consider other pollutants this may be achieved by relationships between the levels of carbon monoxide and the concentrations of hydrocarbons, lead and oxides of nitrogen.

In the initial stages of road design it is usual for several alternative route proposals to be considered but only outline details of road centre lines and estimates of speeds and flows are available. A screening method for estimating air pollution has been developed to predict the annual maximum 8 hour concentration of carbon monoxide arising from traffic at a location near to a road network[4] which consists of straight roads, junctions and roundabouts. This method uses either a series of graphs or mathematical relationships obtained from a computer model which considers vehicle flow, vehicle speed and the distance of the point being considered from the roads. Meteorological and other variables have a large effect on the concentrations of pollutants but they are not considered in this screening process because the method provides an estimate of the maximum concentration likely to occur. The resultant chosen is the highest probable value from a distribution of 8 hour average concentrations which is based on the average hourly value.

The method makes use of four relationships. Firstly the connection between the concentration of carbon monoxide as a function of the distance D (m) between the road and the receptor. For long straight roads this is

$$C = 1.5 \exp(-0.025D)$$

This relationship was derived for A and B roads, a four-lane dual carriageway and a six-lane motorway. The fall of concentration with increasing distance from the road was found to be represented by this relationship when distance was measured from the centre line of each road type. The relationship was derived for a flow of 1000 vehicles per hour and for other flows the concentration should be multiplied by the expected peak hour flow in thousands of vehicles.

The relationship was derived for input weather and traffic conditions and gives the 1-hour average concentration expected for vehicles travelling at 100 km/h. A correction factor for different vehicle speeds can be made from the relationship

$$F = 38.9S^{-0.795}$$

where S is the vehicle speed (km/h),
F is the ratio of the emission rate at speed S to the emission rate at 100 km/h.

A similar relationship was derived for roundabouts with a range of diameters, and a common relationship for all diameters was found when distances D (m) were measured from the centre line of the circulating carriageway. As roundabouts have a wide range of diameters it is suggested that where the central island diameter is 10 m or less then the roundabout should be considered as a straight stretch of road. The relationship is

$$C = 1.55 \exp(-0.033D)$$

There are no United Kingdom recommendations on exposure limits for ambient carbon monoxide and as a consequence United States Federal Air Quality Standards have been suggested as suitable for interpretation of the results of this predictive method. They specify carbon monoxide concentrations of 35 and 9 ppm which should not be exceeded more than once a year for exposure periods of one and eight hours respectively. The Transport and Road Research Laboratory has noted from air pollution survey data that the eight-hour standard is more difficult to meet than the one-hour standard and for this reason the eight-hour standard of 9 ppm was selected to indicate if a more detailed air quality survey was required. The previous relationships give a 1-hour average value and so it was necessary to estimate the 8-hour average value from the previously derived 1-hour average from

$$C_8 = 1.19 + 1.85C_1$$

where C_8 is the 8-hour concentration exceeded once a year,
C_1 is the peak 1-hour concentration.

As this method was derived for very long straight roads it is necessary to divide the road network into as few continuous roads as possible to prevent overestimation of air pollution; only those sections nearest to the reception point should be considered. For example, a four-arm roundabout is divided into three sections, one for each of the two intersecting roads, and the third section is the circulating carriageway of the roundabout. A five-arm roundabout would have one additional section caused by the fifth leg and in all cases the distance would be the shortest distance (not necessarily perpendicular) to the section.

If this preliminary procedure indicates that concentrations of carbon monoxide give rise to concern, then the Department of Transport advises that detailed air quality investigations should be carried out using a suite of computer programs developed by the Transport and Road Research Laboratory[3].

This suite of programs has three major parts. The first and fundamental part predicts the hourly average concentration of carbon monoxide likely at a particular location for given weather and traffic conditions. This result may be used in the remaining two parts for estimating the range of concentrations of carbon monoxide likely for a variety of averaging periods or it may be used to give an indication of likely levels of oxides of nitrogen, hydrocarbons or lead at the site being considered.

References

1. National Society for Clean Air, *Air pollution from road vehicles – a report by the technical committee of the National Society for Clean Air*, London (1967)
2. P. T. Sherwood and P. H. Bowers, Air pollution from road traffic – a review of the present position, *Rd Res. Laboratory Report* 325 (1970)
3. A. J. Hickman and V. H. Waterfield, A user's guide to the computer program for predicting air pollution from road traffic, *Transport Road Research Laboratory Supplementary Report* 806, Crowthorne (1984)
4. V. H. Waterfield and A. J. Hickman, Estimating air pollution from road traffic: a graphical screening method, *Transport Road Research Laboratory Report* 752, Crowthorne (1982)

Problems

Are the following statements correct or incorrect?

(1) The major source of pollution adjacent to heavily trafficked motorways is the diesel-engined goods vehicle.
(2) Traffic 'smog' is likely to occur in regions where vehicle mileage is considerable and there is a low incidence of sunlight.
(3) The benzpyrene content of the air in highway tunnels gives cause for alarm because it is greater than is found in industrial areas.
(4) The carbon monoxide produced by vehicle exhausts in busy city streets is not likely to cause ill effects to vehicle drivers under free-flowing traffic conditions.
(5) The concentration of oxides of nitrogen produced by vehicle exhausts as found adjacent to the highway is less than 1 per cent of the maximum allowable concentration for an 8-hour exposure in industrial conditions and has no long-term effects.
(6) The discharge from vehicle exhausts contains a stable aerosol of lead halides, which gives cause for concern because it produces a concentration of lead in the air of city streets which is considerably greater than the average level in rural areas.

Solutions

(1) This statement is incorrect. Diesel-engine exhaust gases contain significantly lower proportions of pollutants than those produced by petrol engines.
(2) This statement is incorrect. Traffic 'smog' is caused by the reaction of oxides of nitrogen and some of the hydrocarbons in the presence of bright sunlight.
(3) This statement is incorrect. The benzpyrene content of the air in highway tunnels is less than that found in industrial areas.
(4) This statement is correct. A survey carried out by the Transport and Road Research Laboratory has indicated that the carbon monoxide content of air in busy city streets in the United Kingdom is not likely to cause discomfort to road users when traffic is flowing freely.
(5) This statement is incorrect. While the concentration of oxides of nitrogen produced by vehicle exhausts is as stated, the long-term effects are unknown.
(6) This statement is correct.

27

Traffic congestion and restraint

The motor car is an invention which, within half a century, has revolutionised our way of life. It has made possible a dispersal of dwellings far exceeding that of the railway age, it has offered a wide choice of employment situations and has increased the scope of recreational activities to a remarkable extent.

At the same time the growth of vehicle ownership and use, together with population increases and the attraction of human activity into urban regions has resulted in considerable problems. The polluting effect of vehicle exhausts, the noise associated with road vehicles and above all the demand for physical space, all make it necessary to control the use of vehicles in urban areas.

The demand for road space, especially in existing central town areas, will always be greater than the supply because even if the necessary financial resources were available there would be conflicting demands for the available land. The fact that demand for road space is greater than the supply results in traffic congestion. While traffic management and urban highway construction have their place in minimising congestion it is now generally accepted that, without the dispersal of town centre activities, the only solution at the present time is a greater emphasis on public transport.

If this transfer from individual to public transport is accepted as one part of a solution to the problems of traffic in towns, then it will be necessary to find some means of traffic restraint. At the present time congestion itself acts as a restraint, causing trips which would take place at congested periods to be made at other times, or by alternative non-congested modes, or the trips may not be made at all. Congestion is however an inefficient form of restraint in that the priority of service is first-come first-served, regardless of the value of the trip to either the tripmaker or the community. It is inefficient in the use of resources and is detrimental to the environment adjacent to the facility.

There are three general ways in which restraint could be applied. Firstly the entry of vehicles to certain areas at certain times could be prohibited by administrative means. On a limited scale this is already frequently employed in the form of pedestrian precincts but its application on a wide scale would involve the entry of specialist service or emergency vehicles and would involve decisions as to who should be allowed entry on the grounds of the value of their trip either to themselves or to the community.

Secondly restraint could be applied by the use of parking regulations. As early as the 1960s the Ministry of Transport report, *Better Use of Town Roads*[1], considered that the most promising method of restraint, at least for the shorter term, would be to intensify control over the location, amount and use of parking space, both on and off the street. The report especially considered it necessary to restrict long-term parking, which is characteristic of car commuting. It was felt that in some places control over the use of publicly available parking space might not be adequate and might have to extend to privately available parking space, even though this would be costly and require new legislation.

In addition to parking charges the Ministry Group examined the possibility of entry charges using a system of supplementary licences, its main advantage over parking charges being that it restrains both non-parking tripmakers and those who are able to park privately. Its disadvantage is that it fails to take into account different amounts of road space within the licensing area, or use at different levels of congestion. The first disadvantage limits the size of the licensing area, and in addition the larger the area, the higher will be the licence fee and the greater the effects at the boundaries. In *Better Use of Town Roads* it is suggested that the maximum practicable area for such a system might be no more than 10 square miles.

Another indirect method of charging for road use which was examined was the employment of differential fuel taxes. A different petrol tax could be levied in different areas according to the congestion in the area. This tax however could be avoided by filling up in low-tax areas or it could be avoided by special fuel-carrying journeys unless the differential was small. This is only a general tax on a large area, including congested and uncongested roads in the same area at congested and uncongested times. The report considered that it would affect individual garages, cause a black market in petrol and encourage people to carry cans of petrol. In addition commuters would find it easy to avoid tax unless the tax area was very large.

The third form of restraint that could be applied is road pricing. This is a form of road user taxation whereby users of congested roads would be charged according to the distance travelled or the time spent on them, at varying rates governed by the degree of congestion. A system of road pricing does however exist; it was introduced by Lloyd George in 1909 and basic features of the system are still unchanged. It is not however an efficient form of road pricing because the annual licence is not related to road use and petrol tax does not discriminate to any significant extent between tripmaking in congested and non-congested conditions. It has been said[2] that, "the issue is not whether we should have a system of road pricing, because we already have one, but whether we could devise a better one".

Road pricing is an attempt to change the principle that highways should be treated as welfare services, that is financed out of taxes, to the principle that they should be treated as public utility services for which charges are made. It has been argued[3] that road pricing is democratic because it is the tripmaker who makes the decision as to whether or not the trip should be made at the given price rather than a government body making a decision that his trip was in the interests of the community.

If the price is placed at a level that reflects the costs a tripmaker imposes on others, it will produce traffic flows that reflect social benefits and costs as well as the tripmaker's private benefits and costs. If alternative means of transport are also priced in this way then the resultant traffic flows would give a better indication of the need for the construction of future transport facilities. This is because the true demand for transport facilities will be known rather than relying on projections of 'free' tripmaking in which the tripmaker does not bear any of the cost imposed on others.

In a real situation however not all tripmakers would attach the same value to the benefit they would obtain from the trip. Because of differences in income and in the nature of the trip as well as differences in the value of getting to the destination the demand for tripmaking may be expected to fall as the cost of the trip increases.

The demand D for tripmaking may be expressed as a function of the private cost or price p of the trip, that is

$$D = f(p) = Kp^{-\gamma} \tag{27.1}$$

where K and γ are parameters. The general form of this relationship is given in figure 27.1.

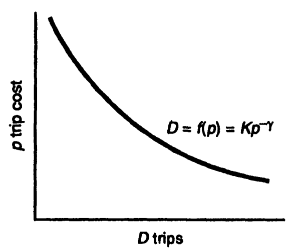

Figure 27.1 *General form of the demand function*

When the effects of varying trip costs are being considered an important factor determining the relationship between a change in trip cost and the resulting change in tripmaking is the price or cost elasticity e_p. It is defined as the percentage change in quantity demanded that results from a 1 per cent change in price.

Then

$$\begin{aligned} e_p &= \frac{\partial D/D}{\partial p/p} \\ &= \frac{p}{D} \times \gamma K p^{-\gamma-1} \\ &= \frac{\gamma K p^{-\gamma}}{D} \\ &= -\gamma \end{aligned}$$

If the demand function is K/p, as is frequently assumed, then the price elasticity is unity.

The cost to the tripmaker has been referred to as the private cost of a journey while the cost that a tripmaker imposes upon other tripmakers because of an increase in congestion is referred to as the congestion cost.

As has been indicated previously the addition to the total costs caused by one extra tripmaker is the marginal cost. It consists of the private costs of the additional trip together with the congestion costs caused by the additional trip.

In addition to these costs there are environmental costs that the trip imposes upon the area adjacent to the highway, and road maintenance costs.

Private costs are divisible into two parts, those proportional to distance, such as fuel, maintenance, depreciation, and those varying with journey speed, which are chiefly associated with the value of time.

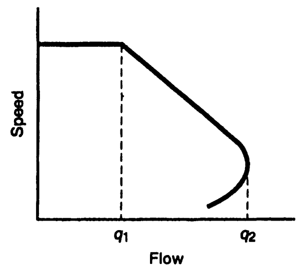

Figure 27.2 *A general form of the speed/flow curve*

Congestion costs arise from the interaction of vehicles and they are illustrated in the typical relationship between speed and flow shown in figure 27.2, where q_1 is the flow at which interaction commences. It may be assumed that there is a linear relationship between q_1 and q_2, the value of maximum flow. This relationship may be assumed to be of the form

$$v = a - bq \tag{27.2}$$

If it is assumed that the total private cost of a trip may be expressed as

$$c + \frac{d}{v} \text{ pence/km} \tag{27.3}$$

where c is the component that is proportional to distance,
d is the component that is proportional to time,

then the private costs of a tripmaker are

$$c + \frac{d}{a - bq} \text{ pence/km}$$

so that the private cost/flow curve has the form shown in figure 27.3. The intersection of the demand curve and the private cost/flow curve at E represents the equilibrium condition at cost c_e and flow q_e. Additional trips above flow q_e will not be made because the private cost of the trip is greater than the benefit of the trip as given by the demand curve.

At this equilibrium condition the net benefit to tripmakers, that is the benefits minus the private costs, is represented by the vertically hatched area.

If however the highway is improved then it will result in a revised private cost/flow curve with a new equilibrium point B and the flow will be q_t at a trip cost c_t. In this case the net benefit to tripmakers will be given by the diagonally hatched area.

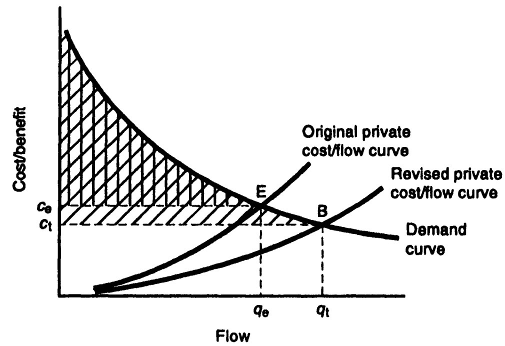

Figure 27.3 *The change in net benefit to tripmakers after highway improvement*

The marginal cost/flow curve may be obtained by differentiating the expression for the total cost of tripmaking:

$$\left(c + \frac{d}{a - bq}\right) q$$

with respect to q, giving the equation of the marginal cost/flow curve

$$c + \frac{d}{v} \times \frac{a}{v} \tag{27.4}$$

It can be seen that this is of the same form as the cost/flow curve with the private cost component multiplied by the factor a/v. From a knowledge of the speed/flow relationship a/v will be greater than unity and the marginal cost/flow curve will have the general relationship to the private cost/flow curve shown in figure 27.4.

In figure 27.4 the equilibrium flow and private cost is given by the intersection of the marginal cost/flow and demand curves at E, a situation which would occur when a decision as to whether a trip should be made is based solely on private cost. The intersection of the marginal cost/flow curve with the demand curve at O however gives the optimum flow. If flow is restricted to below this level then only trips with a marginal cost less than the value of the trip to the tripmaker will be allowed. If the flow is greater than the optimum then a trip is allowed that has marginal costs greater than the value of the trip.

If traffic flow conditions are shifted from the equilibrium position E to the optimum position O then the change in net benefit will be represented by the area A–c_p–P–O minus the area A–c_e–E. This is in fact equal to the area c_e–c_p–P–T minus the area OTE, which represents the gain in net benefit enjoyed by the tripmakers who are not removed due to improved speed, minus the net benefit previously enjoyed by the tripmakers who are now removed.

The movement from E to O can be achieved by the use of a road-pricing charge, represented by PO. Individual tripmakers will however be worse off under the pricing system because the decrease in private costs TP will always be less than the pricing charge OP.

An estimate of the economic benefits to be derived from direct road pricing is contained in *Road Pricing: The Economic and Technical Possibilities*[4] and the

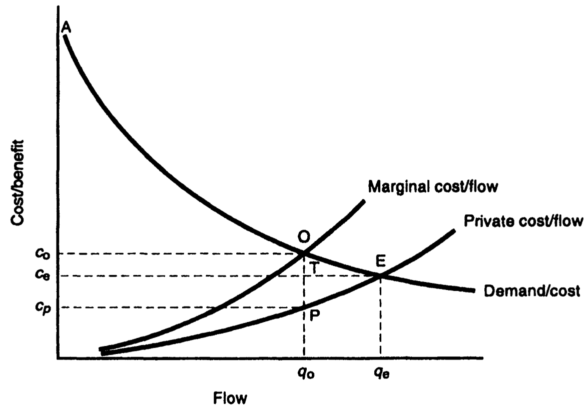

Figure 27.4 *The relationships between marginal and private costs and flow, and between demand and cost*

following illustration of a method which may be used to calculate the net benefits from the imposition of a road-pricing charge is based on that report.

A highway network is considered where the flow is q and the corresponding private costs of a tripmaker are p. The demand curve is of the form $q = f(p)$, which may alternatively be expressed as the inverse relationship $p = f^{-1}(q)$, where $f^{-1} f(p) = p$.

Consider the situation before any price is imposed and let the private cost be α and the flow be Q. From the reasoning given previously, the net benefit to all tripmakers is

$$\int_0^Q f^{-1}(q)\, \mathrm{d}q - \alpha Q \tag{27.5}$$

Consider now the situation in which a road-pricing charge per vehicle/km of β is imposed, causing a decrease in flow to Q^1 and an increase in speed and a reduction in costs from α to α^1 (let $\alpha - \alpha^1 = g$). As before the net benefit to tripmakers is now

$$\int_0^{Q^1} f^{-1}(q)\, \mathrm{d}q - (\alpha^1 + \beta)Q^1$$

The amount βQ^1 is however a transfer payment and represents no real cost to the community, giving a net benefit in the road-pricing situation of

$$\int_0^{Q^1} f^{-1}(q)\, \mathrm{d}q - \alpha^1 Q^1 \tag{27.6}$$

The increase G in net benefits from the imposition of the road charge β is then

$$G = \alpha Q - \alpha^1 Q^1 - \int_{Q^1}^{Q} f^{-1}(q)\, \mathrm{d}q \tag{27.7}$$

If it is assumed that when the flow is of the order of Q or Q^1 the demand function is

$$q = f(p) = \frac{k}{p}$$

then the elasticity of demand is unity and

$$Q = \frac{k}{\alpha}, \quad Q^1 = \frac{k}{\alpha^1 + \beta}, \quad p = f^{-1}(q) = \frac{k}{q}$$

By substituting in equation 27.7 it can be shown that the increase in net benefits is

$$G = \beta Q^1 + \alpha Q \log \frac{Q^1}{Q} \tag{27.8}$$

and therefore

$$\frac{G}{Q} = \beta \frac{Q^1}{Q} + \alpha \log \frac{Q^1}{Q} \tag{27.9}$$

where Q^1/Q is the ratio of the new flow to the old flow.

References

1 Ministry of Transport, *Better Use of Town Roads*, HMSO, London (1967)
2. J.M. Thomson, Case for road pricing, *Traff. Engng Control*, (March 1968), 536–9
3. G.J. Roth, *Paying for Roads: The Economics of Traffic Congestion*, Penguin, Harmondsworth (1967)
4. Ministry of Transport, *Road Pricing: The Economic and Technical Possibilities*, HMSO, London (1967)

Problems

Select the correct completions to the following statements.

(1) In the central areas of existing cities the future use of the private motor vehicle:
  (a) will be made possible by higher parking charges;
  (b) will be prohibited;
  (c) will be assisted by the construction of major highway schemes and parking garages.

(2) The report *Better Use of Town Roads* concluded that in the near future the use of private cars in the central areas of cities would be restrained by:
  (a) an electronic road pricing system;
  (b) the control of all parking facilities and the levying of considerably higher parking charges;
  (c) regulating entry to certain areas by admission charges;
  (d) the use of differential fuel taxes.

(3) The object of road pricing is:
  (a) to ensure that a journey would not be made if the cost the journey imposed on others was less than the benefits obtained by the tripmaker;
  (b) to allow individual tripmakers to make their own decision as to whether to make a trip by private transport after taking into account the costs and benefits of the trip;
  (c) to completely remove traffic congestion from the highway network.

(4) A speed/flow relationship for a highway has the form shown in figure 27.5. When the flow on the highway is:
(i) in the region $O-q_1$
(ii) in the region q_1-q_2
will the additional cost of introducing an extra vehicle into the flow be:
(a) the private costs of the additional vehicle?
(b) the private costs of the additional vehicle together with an addition to the private costs of other vehicles?
(c) negligible?

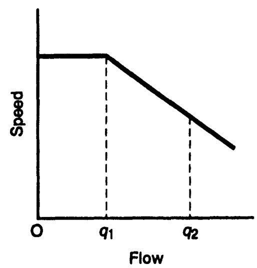

Figure 27.5

(5) Indicate on figure 27.6 which curve is:
(a) a demand curve on which tripmakers all value the benefits of their trip equally;
(b) a demand curve on which some tripmakers value the benefits of their trip at a higher level than others;
(c) a cost/flow curve on which costs are a function of journey distance and time;
(d) the marginal cost/flow curve for the above cost/flow curve.

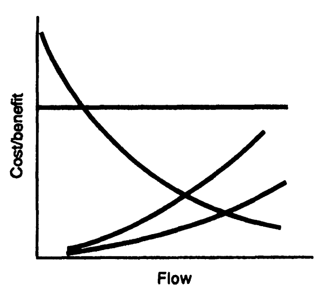

Figure 27.6

(6) A highway link has the following characteristics:
(a) a speed/flow relationship of $v = 50 - q/100$
where v is the stream journey speed (km/h)
q is the stream flow (pcu/h);
(b) a flow/cost relationship of $p - 1.25\ (5 + 300/v)$
where p is the private cost in pcu/km
v is the journey speed in km/h;

(c) a demand relationship of the form

$D = 4 \times 10^4 p^{-1}$

where D is the demand in pcu/h
p is the private cost in km/h.

Determine graphically:

(a) the flow and the speed on the highway when each tripmaker is only aware of his private costs;

(b) the flow and the speed on the highway when a trip is only made if the benefit of a trip to the tripmaker exceeds the additional cost imposed on other tripmakers;

(c) the pricing charge that would result in the flow (b).

Solutions

(1) (a) This statement is correct. The report *Better Use of Town Roads* concluded that for the short term the most promising form of restraint is likely to be the intensification of controls over parking. It is this restraint over the less valued trips that will make it possible to use private motor vehicles for the more valued tips.

(b) This statement is incorrect. It would not be practicable to prohibit the entry of all private motor vehicles into larger town areas and entry by permit would raise administrative difficulties.

(c) This statement is incorrect. Large-scale highway and parking garage construction would be necessary on so large a scale that it is doubtful if the financial resources could be made available. If it were financially possible then it would result in the wholesale reconstruction of central city areas.

(2) (a) This statement is incorrect. The report concluded that this method is potentially the most efficient form of restraint, but that it would take several years to develop.

(b) This statement is correct. The report concluded that the strict control of parking would in the short term be the most promising form of restraint.

(c) This statement is incorrect. The report concluded that the Ministry was far from satisfied that it would be practicable to regulate entry to certain areas by admission charges.

(d) This statement is incorrect. The report did not recommend the use of different fuel taxes.

(3) (a) This statement is incorrect. The correct statement is ... to ensure that a journey would not be made if the cost of the journey imposed on others was greater than the benefits obtained by the tripmaker.

(b) This statement is correct.

(c) This statement is incorrect. Road pricing alone will not produce a cure for congestion because the need for investment in urban highways will still remain.

(4) When the flow on the highway is (i) in the region O – q_1, the additional cost of introducing an extra vehicle will be the private costs of the additional

vehicle, as stated in (a) because increasing flow does not result in a decrease in speed.

When the flow on the highway is (ii) in the region q_1–q_2, the additional cost of introducing an extra vehicle will be the private costs of the additional vehicle together with an addition to the private costs of other vehicles, as stated in (b), because increasing flow results in a decrease in speed.

(5) The appropriate curves are as indicated in figure 27.7.

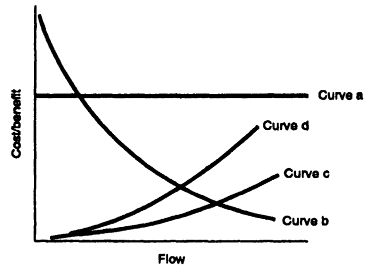

Figure 27.7

(6) From equation (27.2), $v = a - bq$. From the given relationship, $v = 50 - q/100$. Hence

$$a = 50; \qquad b = \frac{1}{100}$$

From equation (27.3), $p = c + d/v$. From the given relationship, $p = 1.25\ (5 + 300/v)$. Hence

$$c = 6.25; \qquad d = 375$$

From equation 27.4

$$\text{marginal cost} = c + \frac{d}{v} \times \frac{a}{v}$$

$$= 6.25 + \frac{18\,750}{v^2}$$

Flows, speeds, private and marginal costs are tabulated in table 27.1.

TABLE 27.1

Flows (pcu/h)	*Speed $50 - q/100$ (km/h)*	*Private cost $1.25(5 + 300/v)$ (p/pcu km)*	*Marginal cost $6.25 + 18750/v^2$ (p/pcu km)*	*Demand $4 \times 10^4 \times p^{-1}$ (pcu/h)*
200	48	14.06	14.39	2844
500	45	14.59	15.51	2742
1000	40	15.63	17.34	2560
1500	35	16.96	21.56	2360
2000	30	18.75	27.08	2134
2500	25	21.25	36.25	1882
3000	20	25.00	53.12	1600
4000	10	43.75	193.75	914

The private cost, marginal cost and demand curves are plotted and the intersection of the private cost and demand curves gives (a) the flow on the highway when each tripmaker is only aware of his private costs (see figure 27.8). The intersection of the marginal cost and demand curves gives (b) the flow on the highway when a trip is only made if the benefit of a trip to the tripmaker exceeds the additional cost imposed on other tripmakers. The ordinate XY gives the pricing charge that will result in flow condition (b). With the flows known the travel speeds can be calculated from equation 27.2, and are given below.

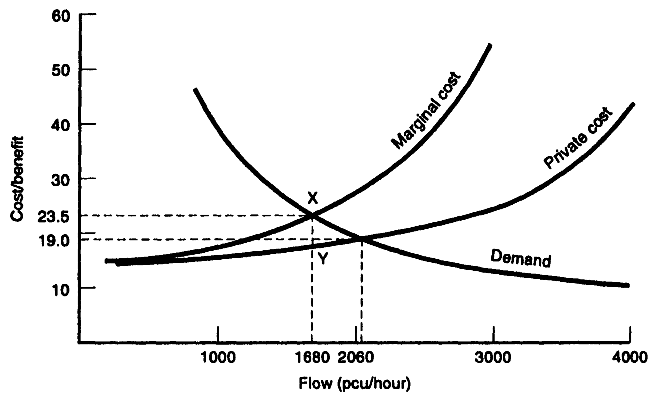

Figure 27.8

From the graphical plot:

Flow condition (a)

$q = 2060$ pcu/h

From equation 27.2

$v = 29.40$ km/h

Flow condition (b)

$q = 1700$ pcu/h

From equation 27.2

$v = 33.00$ km/h

(c) The pricing charge

XY = 5.9 p/km

PART III†

TRAFFIC SIGNAL CONTROL

†See the Appendix, page 388 for a definition of the symbols used in Part III.

28

Introduction to traffic signals

Traffic signals are used at many at-grade junctions, particularly in urban areas, to maximise traffic efficiency and safety by separating conflicting traffic movements in time.

It has been stated that the first traffic signal to be installed in Great Britain was erected in Westminster in 1868. It was illuminated by town gas and unfortunately for the future development of signals of this type was demolished by an explosion. Not until 1918 were signals used again for the control of highway traffic when manually operated three-colour light signals were introduced in New York. Some seven years later manually controlled signals were used in Piccadilly, followed in 1926 by the first automatic traffic signals in Great Britain, which were erected at Wolverhampton.

In the first signals alternate red and green fixed-time periods were automatically timed and while they replaced police manpower they were not as efficient as manual control because they could not respond to changes in the traffic flow. As a refinement controllers were introduced, which were able to vary the relative durations of the fixed-time green and red periods according to a preset timing pattern. It was thus possible to operate the signals with different sequences for the morning, mid-day and evening peak periods.

Where a major highway had several intersections along its length each controlled by traffic signals then means of allowing a nearly continuous progression of traffic along the major route were developed by linking the signals using a master timing device or controller instead of individual timing devices at each intersection.

During the early 1930s there was a further attempt to increase the ability of the signal controllers to deal with varying traffic demand by the incorporation of systems that would allow the signals to respond to individual vehicles. In some situations drivers were requested to sound their horns into microphones at the sides of the highway and later electrical contacts were operated by the passage of vehicles.

The vehicle detection method that became widely adopted in this period was the pneumatic tube detector and this survived in various forms in Great Britain until the 1960s. It had the disadvantages of being easily damaged by vehicles, particularly during snow-ploughing operations, and as these detectors were set in a concrete base obstructions to traffic were caused during installation and maintenance. They have now been replaced by the inductance detector, a cable set into the road surface that detects the passage or presence of a vehicle by a change in the electric field. The standard method of detection in the United Kingdom is known as 'System D' where detection for a green signal indication is carried out by a buried loop positioned 40 metres from the stop line. Other loops to extend the existing green indication are sited between the initial loop and the stop line at distances of 12 metres and 25 metres from the stop line.

Buried loops are particularly vulnerable to roadworks and recent developments have seen the installation of 'above ground' detection using microwave or infra-red detectors, for traffic control purposes. These detectors also offer the opportunity for lower installation costs and less traffic delay during maintenance. The traffic signal equipment at an individual location comprises a traffic signal controller, signal heads (lights) on poles for controlling traffic on each approach, vehicle detectors and associated cabling to the controller and, where provided, pedestrian push buttons and ancillary equipment. The capabilities of traffic signal controllers together with pedestrian facilities are described in Chapter 32.

Signal heads facing drivers have three signal aspects, red, amber and green. The following sequential set of signals occurs in the UK:

(1) red (stop);
(2) red and amber (stop, but prepare to go);
(3) green (go if the way is clear);
(4) amber (stop, unless it would be unsafe to do so);
(5) red (stop).

Green arrows may also be used to assist traffic direction and control as appropriate (such as a left-turn filter arrow). For pedestrians, each signal head has two aspects – a red and a green man. A red man signifies 'do not cross' while a green man signifies 'cross with care'. A flashing green man means 'do not start to cross'.

For drivers, at least two signals are required to be visible from each approach, usually comprising one primary and one secondary signal. Primary signals are located usually one or two metres from the stop line on the junction approach, while the additional signals which are normally sited beyond the junction are known as secondary signals. The layout of a junction and the positioning of signals will vary in each case, dependent on site specific circumstances. Guidance is given by the Department of Transport in a series of Advice Notes 1–5.

References

1. Department of Transport, Requirements for the Installation of Traffic Signals and Associated Control Equipment, *Advice Note* TA13/81, London (1981)
2. Department of Transport, Procedures for the Installation of Traffic Signals and Associated Equipment, *Advice Note* TA14/81, London (1981)
3. Department of Transport, Pedestrian Facilities at Traffic Signal Installations, *Advice Note* TA15/81, London (1981)
4. Department of Transport, General Principles for Control by Traffic Signals, *Advice Note* TA16/81, London (1981)

5. Department of Transport, Junction Layout for Control of Traffic Signals, *Advice Note* TA18/81, London (1981)

Problem

Which of the following signal timing devices or controllers would operate most efficiently when the traffic flows on the intersection–approach highways vary considerably:

(a) an isolated intersection where the signals are operated by a fixed-time controller;
(b) an intersection where the signals are operated by a fixed-time controller, which also controls adjacent intersections so as to assist progression along the major route;
(c) an intersection where the signals operate by vehicle actuation using inductance loops beneath the road surface?

Solution

(a) Incorrect. At isolated intersections where the signals are operated by a fixed-time controller the durations of the red and green periods are fixed with respect to average traffic flows. They are for this reason insensitive to short-term fluctuations in the traffic flow pattern.

(b) Incorrect. At an intersection where the signals are operated by a fixed-time controller which also controls adjacent intersections so as to assist progression along the major route, major road vehicles are not unduly delayed when the traffic flows are similar to those for which the durations of the green and red periods were calculated. When the traffic flows fluctuate on a short-term basis these fixed red and green periods cause excessive delays to both major and minor road vehicles.

(c) Correct. At an intersection where the signals are actuated by vehicles passing over pneumatic or inductance detectors the signals respond to the vehicles on the traffic signal approaches and so are able to control traffic efficiently even when there are short-term variations in the flow.

29

Warrants for the use of traffic signals

A decision to use signal control in urban areas in preference to roundabout control or as a means of increasing traffic capacity at priority intersections may be made from the overall viewpoint of traffic management when urban computer control is being implemented.

A more detailed decision on the installation of traffic signals may also be made on the basis of traffic flow, pedestrian safety, accident experience and the elimination of traffic conflicts. Changes in these traffic characteristics consequent on the installation of signals can be determined and evaluated by cost–benefit methods.

Frequently the decision to install signal control is made by consideration of many issues, not all of which can be evaluated in traffic or monetary terms; some of these will be considered in greater detail.

When signal control is employed in preference to priority or roundabout control, then major road vehicles which previously passed unimpeded through the junction are on occasions required to stop and wait. On the other hand, minor road vehicles suffer reduced delay. Vehicular delay to the minor road stream under priority control can be estimated as described in Chapter 18. Delay at roundabouts can be estimated using the ARCADY program developed by the Transport and Road Research Laboratory (see Chapter 20). The delay to all traffic streams under signal control can be estimated as described in Chapter 43

Accident prediction at priority junctions and roundabouts can be predicted using the methods described in Chapters 16 and 20 respectively. Accidents at signal controlled intersections can be estimated using the method employed in the COBA program.

The accident prediction formulae for traffic signal controlled intersections take the form

$$\text{Annual accidents} = x(f)^y$$

where x and y are constants which depend upon the form of the intersection and f is the product obtained by multiplying the combined inflow of vehicles from the two major opposing links by the sum of the inflows on the other minor links or, alternatively, f is the sum of the total inflow of vehicles into the junction. The first type of formula is referred to as a cross-product (C) model and the second as the inflow (I) model. Vehicles are expressed as thousands of vehicles per annual average day.

The type of formula used depends on the number of arms of the controlled junction and whether the highest standard of link at the intersection has a single or dual carriageway. Details are given in table 29.1.

TABLE 29.1
Formula type for accident prediction at signal-controlled junctions

Junction reference number	*No. of arms*	*Highest link standard*	*Formula type*
1	3	single	I
2	3	dual	C
3	4	single	C
4	4	dual	C
5	5	single	I
6	5	dual	I

The values of the parameters x and y differ with junction form and also according to whether the junction is in an urban or rural situation. Values are given in table 29.2.

TABLE 29.2
Parameters for accident prediction at signal-controlled junctions

Junction reference number	*Rural*		*Urban*	
	x	y	x	y
1	0.223	0.610	0.223	0.610
2	0.494	0.420	0.291	0.510
3	1.378	0.200	1.378	0.200
4	0.494	0.420	0.291	0.510
5	0.254	0.620	0.254	0.620
6	0.238	0.850	0.160	0.970

More detailed accident predictive modelling for four-arm single carriageway urban traffic signals is now possible, following research by the University of Southampton for the Transport Research Laboratory[1]. These new models are incorporated into OSCADY3, a computer program as an aid to the design of signal-controlled junctions. This model is described in Chapter 45.

Accident experience can be worsened if a relatively small number of vehicles on the minor road experience difficulty in entering or crossing a nearly continuous major road traffic stream. In such circumstances the delays caused by the installation of signals are likely to be greater than under the previous priority control. While this type of traffic conflict requires professional judgement, guidance is given for use in the United States. Here for two or more lane approaches on both major and minor roads, with flows of 900 vehicles per hour (total for both major road approaches), and 100 vehicles per hour on the most heavily trafficked minor road, signals are justified when the flows exist for eight hours of an average day. Other values are given for different approach widths.

Signal control, even when pedestrian facilities are not provided, offers considerable assistance to pedestrian movement. In the United Kingdom the Department of Transport[2] advises that a separate pedestrian stage or one combined with a traffic stage may be required when the flow of pedestrians across any one arm of the junction is of the order of 300 per hour or more or when the turning traffic flowing into any arm has an average headway of less than 5 seconds during the time that such traffic can flow and is conflicting with a pedestrian flow of at least 50 pedestrians per hour. These traffic flows to be taken as the average of the four busiest hours over any weekday.

If a pedestrian facility is to be provided it may take the form of a full pedestrian stage where all traffic is stopped when pedestrians are allowed to cross all the arms of the junction. The pedestrian stage is demanded by push button but has the disadvantage of imposing additional delay on vehicular traffic.

A more efficient form of control from the viewpoint of vehicular movement is the use of a parallel pedestrian facility. This is achieved by banning some vehicular turning movements, but to prevent the danger which could be caused by illegal movements particular attention should be given to the kerb layout.

Where road layout permits, a large central island can be provided instead of the usual refuge so that pedestrians negotiate the road in two stages. The Department of Transport[2] recommends a minimum island size of 10 by 2.5 m.

As the provision of any pedestrian facility will normally reduce the proportion of green time available for vehicle movements, it may sometimes be necessary when a junction is close to capacity to install a pedestrian facility away from the junction. It is recommended that the crossing be not more than 50 m from the mouth of the junction; the pedestrian stage is incorporated within the junction signal cycle and the position of the stage chosen to minimise delay to traffic flow.

Overall, it can be seen that traffic signals have a number of advantages over other forms of junction control in urban areas: junction land-take and cost are usually low compared with other forms of control offering similar capacity, and there is inherent flexibility for positive traffic management so that, for example, specific approach arms or categories of road user can be favoured. The linking of signals within an Urban Traffic Control system offers further substantial benefits, as described in Chapter 51. The main disadvantages of signals are the higher delays which can occur in low flow conditions, relative to non-signalised junctions, the increased accident risk for certain types of traffic accident and the operational and maintenance costs of the equipment. Compared with roundabouts, the lack of provision for U-turning manoeuvres is a further disadvantage.

References

1. R.D. Hall, *Accidents at Four Arm Single Carriageway Urban Traffic Signals*, Transport and Road Research Laboratory, Contractors' Report 65 (1986)
2. Department of Transport, Pedestrian facilities at traffic signal installations, *Advice Note* TA/15/81, London (1981)

Problem

Delays at a hypothetical T-junction under priority and signal control are illustrated in figure 29.1. Type A and type B priority junctions have good and poor visibility distances respectively for minor road vehicles entering the major road. Which of the following traffic situations would be likely to warrant a change from priority to signal control?

(a) An intersection where the priority control flow/delay curve is represented by curve B of figure 29.1, where the traffic signal flow/delay curve is as shown, and where the minor road flow is 200 veh/h.
(b) As above, but where the delay to minor road vehicles is 10 s and traffic volumes are not expected to increase rapidly in the future.
(c) As above, but where the priority control flow/delay relationship is illustrated by curve A, the major road flow is 800 veh/h, and the intersection is the scene of frequent traffic accidents.

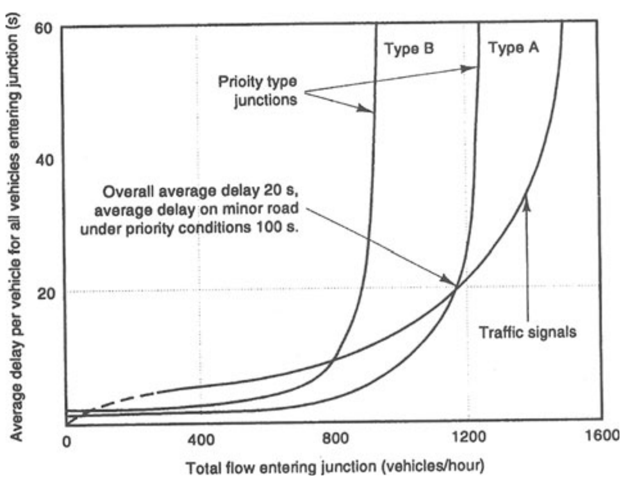

Figure 29.1

Solution

(a) Very likely. The minor road flow is 200 veh/h, and since the ratio of major/minor road flow is 4:1 the total flow entering the intersection is 1000 veh/h. From figure 29.1 the delay with traffic signal control is very considerably less than with priority control.
(b) Not likely. The delay to minor road vehicles is 10 s and since the delay to major road vehicles under priority control is zero the average delay to all vehicles entering the intersection is 2 s, which occurs when the total flow

entering the intersection is approximately 400 veh/h. At this volume the delay at traffic signals is approximately 5 s.

(c) Likely. The major road flow is 800 veh/h and since this represents 0.8 of the total flow entering the intersection the delay with priority control for conditions of good visibility is less than with traffic signal control. The intersection is, however, the scene of frequent accidents and the increase in average delay due to the change to traffic signal control would probably be compensated by the decrease in accidents. As the traffic volumes are increasing rapidly, the change to signal control would be justified on delay considerations after a short period of time.

30

Staging and phasing

In the control of traffic at signal-controlled intersections, conflicts between streams of vehicles are prevented by a separation in time. Green signals are therefore shown to different sets of movements (or streams), so that conflicting ones do not receive a green signal simultaneously unless permitted in some circumstances (such as opposed right-turning traffic). Traffic streams therefore have to be divided into separate sets so that all streams in each set always receive identical signal indications. This procedure by which the streams are separated is known as 'phasing'.

A phase describes a set of movements which can take place simultaneously or the sequence of signal indications received by such a set of movements. The term 'stage' describes that part of the cycle during which a particular set of phases receive green. The term cycle here is used to refer to one repetition of the sequence in which the various phases receive green.

These terms are illustrated with an example of a conventional crossroads where the major conflicts are between the north–south traffic stream and the east–west traffic stream, but where there is a heavy right-turn movement from the east. The junction is illustrated in figure 30.1, together with its corresponding phasing and staging diagrams. The figure illustrates how a phase can run in more than one stage (phase b in this case). Note that it is convention to denote phases by letters and stages by numbers.

For the purposes of specification writing for a signal controller, the individual streams of traffic are considered separately and each stream is allocated a phase letter. Thus, the streams shown in the staging diagram figure 30.1 might be allocated phases a to e as shown. This allows for a full specification of the junction control to include:

(1) the allocation of phases to stages;
(2) the minimum phase length (often 7 seconds);
(3) the maximum stage length (often varied by time of day);
(4) the 'intergreen' time between stages;
(5) the normal order of stages and 'permitted moves' – this specifies alternative stages which could occur if a stage is completed and there is no demand (traffic or pedestrian) for the stage which would normally follow.

Traffic signal control can be either stage based on phase based, according to the method of control. For example, isolated junctions operating under D-system vehicle actuation now operate with phase-based microprocessor control, whereas Urban Traffic Control systems incorporating TRANSYT or SCOOT (see Chapter 51) are stage based.

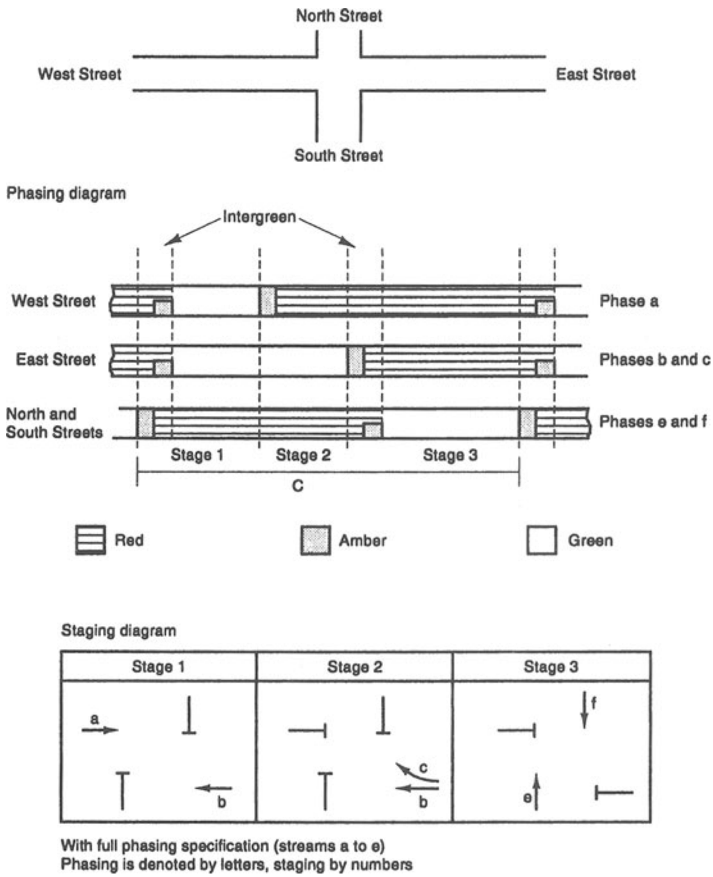

Figure 30.1 *A typical phasing and staging diagram*

31

Signal aspects and the intergreen period

The indication given by a signal is known as the signal aspect. The usual sequence of signal aspects or indications in Great Britain is red, red/amber, green and amber. The amber period is standardised at 3 s and in all new signal installations the red/amber at 2 s.

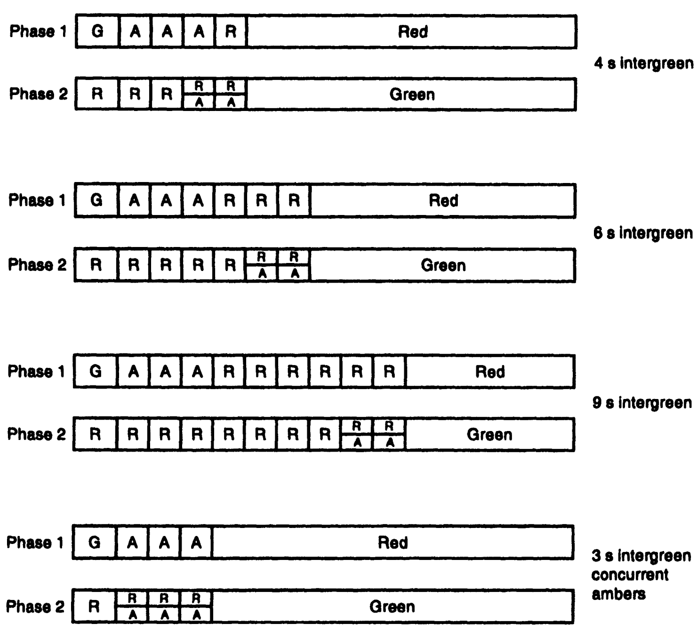

Figure 31.1 *Examples of intergreen periods at a two-phase traffic signal*

In some older installations the amber indication on the first phase is shown concurrently with the red/amber indication on the second phase: in this case the red/amber indication has a duration of 3 s.

The period between one phase losing right of way and the next phase gaining right of way, that is the period between the termination of green on one phase and the commencement of green on the next phase, is known as the intergreen period.

With modern controllers the minimum intergreen period is 4 s but this may be exceeded in particular circumstances. When the distance across the intersection is excessive the length of the intergreen period must be based on the time required for a vehicle which passes over the stop line at the start of the amber period to clear a potential collision point with a vehicle starting at the onset of green of the following stage and travelling at the normal speed for the intersection. The Department of Transport[1] recommends intergreen periods based on this distance ranging from 5 to 12 seconds for distances of 9 to 74 metres for straight ahead movements. For turning movements the corresponding distances are 9 to 50 metres. The recommended values may have to be modified when a pedestrian stage is losing right of way, when the following stage may have to be delayed until the pedestrian area is clear.

When signals are located on higher speed roads a longer intergreen period provides a margin of safety for vehicles which are unable to stop on the termination of green. Controllers on high speed roads have the ability to extend the intergreen period on a maximum green termination.

Intergreen periods provide a convenient time during which right-turning vehicles can complete their turning movement after waiting in the centre of the intersection. Whilst the number of vehicles turning in this way is limited, it is useful when a separate right turning phase is not justified.

Referring to figure 31.1 it can be seen that within an intergreen period there is a period of time when all vehicle movement is prohibited, because the signal indication is either red or red/amber, equal to the intergreen period minus the 3 second amber period. For this reason an increase in the length of the intergreen period will result in a loss of traffic capacity and intergreen periods should have a minimum value consistent with safety.

The period of time lost to traffic flow is referred to as 'lost time' in the intergreen period and should not be confused with start and end lost times associated with green periods and dealt with subsequently in Chapter 35.

Reference

1. Department of Transport, General principles of control by traffic signals, *Advice Note* TA/16/81, London (1981)

Problems

(a) State the factors which are considered in the selection of an appropriate intergreen period.

(b) When the intergreen period is 7 s what is the time between the last possible vehicle crossing the stop line on the approach losing right of way and the first possible vehicle crossing the stop line on the approach gaining right of way?

Solutions

(a) Factors considered in determining intergreen periods are:
   (i) the lengths of the travel paths within the intersection between the stop lines and potential collision points;
   (ii) the speeds of vehicles on the approaches and dangers which may exist on a maximum green termination;
   (iii) the time required for pedestrians to cross the carriageway when a pedestrian stage terminates;
   (iv) the clearance of right-turning vehicles waiting within the intersection.

(b) A vehicle may cross the stop line at amber if it cannot stop with safety and hence the period between a vehicle crossing the stop line on the approach losing right of way and a vehicle crossing the stop line on an approach gaining right of way is $7 - 3 = 4$ s.

32

Signal control strategies

Signal control strategies are implemented at the junction by the traffic signal controller which may be isolated (stand alone) or linked either locally to one or more adjacent junctions or, in the case of Urban Traffic Control, to a central computer which controls all the traffic signals. Recently manufactured signal controllers use electronic timing to determine the length of the signal stages and electromechanical interlocking relays to switch the signals on and off in the specified sequence. Signal timings can be adjusted according to vehicle detection, remote computers (UTC) or nearby linked controllers, or varied according to time of day.

Microprocessor control

The most modern controllers incorporate microprocessors which use digital timing and solid-state switching to perform similar tasks with greater reliability and flexibility. Increased flexibility is achieved because microprocessor controllers are software based rather than hardwired as were their predecessors. For most controllers the 'firmware' (the program) and data are held on EPROM (expandable programmable read only memories). Users can change timing data within defined limits but configuration data (such as phasing structures) can only be altered by changing EPROM. Controllers can also be accessed via a portable computer or handset, to check or alter timings on the clock, read the fault log, monitor detection operation and so on. Remote monitoring is also possible (for example via telephone lines) and is often used for fault monitoring.

Microprocessor controllers can function in a variety of modes:

(1) manual mode (usually for emergency use only);
(2) vehicle actuated, responding to direct demands;
(3) fixed-time mode, but with certain phases being retained as demand dependent, if necessary;
(4) under co-ordination with a neighbouring junction or pelican crossing using a cableless linking facility (CLF) or hardwired linking;
(5) in a priority mode this includes a 'hurry' call, where the controller is forced to run a particular stage (such as for bus priority);

(6) in remote input mode, which is used with UTC systems or with other new control strategies such as MOVA, described below.

When a green signal is displayed it is desirable for the green indication to be shown for an initial fixed period which cannot be overridden by other demands. Such a period is built into traffic signal controllers, its value normally being 7 s.

Exceptions are when pedestrians walk with green on a stage that does not allow vehicles to cross their path, when the minimum green must be related to the pedestrian crossing time. When the proportion of heavy vehicles on an approach is high and they accelerate slowly owing to the gradient, then the minimum green time may be increased. On late start and early cut-off stages, shorter minimum green times are frequently used.

Isolated, vehicle-actuated control

The most common form of vehicle-actuated strategy still in use in the United Kingdom is known as D-system VA (vehicle actuation). With this system, a series of buried loops is placed on the approaches with the initial detector some 40 metres distant from the stop line. This method of control may be summarised as follows, for stage-based control.

A vehicle detected on an approach during the display of the green indication will normally extend the period of green so that a vehicle can cross the stop line before the expiry of green. Usually there are three loops on an approach, each one of which extends the green time by 1.5 s. If the approach has a steep gradient, a longer extension period will be required.

When a vehicle is detected approaching a red signal indication, the demand for the green signal is stored in the controller which serves stages in cyclic order and omits any stages for which a demand has not been received.

The demand for the green stage is satisfied when the previous stage that showed a green indication has exceeded its minimum green period and there has not been a demand for a green extension on the running stage, or the last vehicle extension on the running stage has elapsed and there has not been a further demand. Alternatively, the demand for the green stage is satisfied if, after the demand is entered in the controller, the running stage runs to a further period of time known as the maximum green time. This would occur if there were continuous demands for green on the running stage. The first type of change is known as a gap change and is safer than the second which is known as a maximum green change.

MOVA

A new form of control for isolated intersections has been developed recently by the TRL, called MOVA[1] (Microprocessor Optimised Vehicle Actuation). It takes advantage of the flexibility of modern microprocessor controllers and evaluation has shown that it controls junctions more efficiently than D-system VA. A recent pilot study of 20 sites carried out by TRL[1] showed an average reduction in delay of 13 per cent, while large heavily-used sites performed even better with up to 25 per cent improvement.

The MOVA system uses a new detector arrangement with two detectors per approach lane, typically at 40 metres and 100 metres before the stop line.

Vehicles queued between the 40 metre detector and the stop line are initially discharged, after which the time intervals between vehicle detections are monitored to determine when queues have cleared. Finally, an on-line optimisation process is undertaken which balances the benefits of extending a green against the disbenefit to other vehicles stopped by red signals. This is an advantage over D-system VA, in which the 'gap-seeking' operation tends to extend green times even when the flow rate has dropped considerably and does not take account of queueing on other approaches. It also means that the setting of maximum green times is no longer critical.

Optimisation is based on delay and stops minimisation although when an approach becomes oversaturated (that is retaining a significant queue at the end of green), MOVA switches to a capacity-maximising routine to clear the congested approaches as quickly as possible.

Other improvements of MOVA over conventional control include new facilities for optimising right-turn movements and signal timings on 'flared' approaches, where local widening gives extra lanes near to the stop line, with a resulting higher saturation flow early in the green.

Bus priority at traffic signals

Signal timings are usually optimised to minimise vehicle delays. However, when high bus flows are present, the optimum movement of people, rather than vehicles, requires the implementation of some form of bus priority. Some priority can be achieved by providing a higher green time proportion to those traffic streams containing a high proportion of buses.

Alternatively it is possible to use selective detection techniques. With this system, individual buses are detected on the approach to a junction, usually using special loops which either 'recognise' a bus because of its size or 'signature' or using buses equipped with a special transponder which interacts with a bus detector. On detection of a bus, the signal aspect is either held on green (extended) until the bus has cleared the stop line, or a priority recall is registered if the signals are on red, to recall the green aspect as soon as possible.

In such a system in SELKENT[2], London, where some 900 buses and 56 isolated junctions were equipped, a single bus detector was used, typically some 70 metres upstream of the stop line. Buses benefited from reduced delay, averaging 9 s per approach, and reduced delay variability. Disbenefits to non-priority traffic were minimised by retaining existing stage ordering and by giving compensation (additional green times) to non-priority stages following a priority call. At congested sites with high bus flow, a facility was also used to 'inhibit' priority calls in alternate cycles.

Systems for selective detection within UTC systems such as TRANSYT and SCOOT (Chapter 51) are also being developed. A key issue here is the potential adverse effects of a loss of signal co-ordination, and some system constraints are therefore required.

Pedestrian facilities

When a traffic signal installation is being designed or modified, it is necessary to consider the nature and extent of pedestrian flows as well as vehicular flows. Pedestrian facilities are provided to assist pedestrians to cross the road in safety, with minimum delay to traffic.

The justification for pedestrian facilities is given in Department of Transport *Advice Note* TA 15/81[3]. A separate pedestrian stage, or one combined with a traffic stage may be required if (i) the flow of pedestrians across any one arm of the junction is around 300 per hour or more, or (ii) the turning traffic flow into any arm has an average headway of less than 5 s during the time that such traffic can flow and is conflicting with a pedestrian flow of at least 50 pedestrians per hour. Other special circumstances can also warrant pedestrian facilities, such as an above average number of infirm or handicapped pedestrians.

Even with no red and green man displays, the presence of traffic signals offers some assistance to pedestrians, owing to the existence of the red signal. Pedestrian refuges can be particularly helpful.

Pedestrian facilities include:

(1) 'Full pedestrian stages', where all traffic is stopped for the stage. This method causes delay to traffic and should only be called by demand from push buttons which should be provided at all points where pedestrians may cross. The duration of the green man indication is typically 6–10 s, dependent on carriageway width, followed by a clearance period and an 'all-red' period. The green man may also be accompanied by an audible signal.

(2) 'Parallel pedestrians', where, by prohibiting some vehicular turning movements, it is safe to signal pedestrians to cross an arm simultaneously with a specified traffic stage. No conflicts exist between pedestrians and turning traffic. (Note that such conflicts can occur at some non-United Kingdom installations, with pedestrians having priority over turning traffic.)

(3) 'Staggered pedestrian facility', where a large island is placed in the centre of the carriageway allowing separate pedestrian signalling of the two sides of the carriageway. Pedestrians negotiate one-half of the carriageway at the normal stop line when traffic on that approach is held on red. Normal pedestrian signals are shown during this period. The other half of the carriageway is controlled by separate signals which are located at the opposite end of the island, and pedestrian right of way is demanded by push button. The pedestrian stage starts concurrently with the commencement of the green on the other traffic stage, so that sufficient reservoir space is necessary only to accommodate turning vehicles.

(4) 'Displaced pedestrian facility', where the pedestrian facility is placed away from the junction, because the junction itself is operating close to capacity. Pedestrians are diverted towards the crossing by means of guard railing.

Pelican crossings

Signal-controlled pedestrian crossings sited away from traffic junctions are called pelican crossings. These provide safe crossing points on roads carrying considerable volumes of traffic. These pedestrian facilities have become increasingly common in urban areas, either at new crossing sites or as an enhancement to existing zebra crossings.

Justification for a pelican crossing is usually based on the degree of conflict between pedestrians and vehicles. Counts are taken of the two-way total hourly flow of pedestrians (P) crossing the road within 50 m on either side of the site at busy times and the two-way total hourly flow of vehicles (V). A pelican crossing is then numerically justified if $PV^2 > 10^8$ for an undivided road or 2×10^8 for a divided road. Reference 4 lists a range of other circumstances when a pelican crossing may be desirable, if the PV^2 criterion is not met (such as crossings near a hospital, school or busy shopping area, where the pedestrian flow is intermittently heavy or where the traffic stream contains in excess of 300 heavy goods vehicles per hour). Design aspects are described in Department of Transport Standard TD 28/87[5] and also in *Advice Note* TA 52/87[4], including proximity of junctions, visibility, crossing width, guard-rails, road surfacing requirements (such as skid resistance), lighting, signing and facilities for disabled persons.

Pelican crossings are normally installed only at sites where neither the 85th percentile speed, nor the speed limit on any approach exceed 50 mph. Where the speed limit is 30 mph, it is acceptable to use fixed-time operation or microwave vehicle detection. Alternatively inductive loop detection (for example, system 'D') can be used. This form of vehicle-actuated detection is mandatory at higher speed sites, with the additional requirement for speed discrimination loops, so that high-speed vehicles are given a sufficient green extension to ensure clearance of the crossing.

Pelican crossings have a fixed operational cycle within which timings are selected according to site conditions and to constraints specified in TD 28/87[5]. Where a pelican crossing is sited within 100 m of a signal-controlled junction, it has to be linked to the junction controller to ensure efficient operation. Where it is sited within an Urban Traffic Control (UTC) system, the pelican is often controlled directly by the UTC system. When a pedestrian stage is demanded, the UTC system co-ordinates the timing of this stage with that for adjacent junctions, thus retaining efficient traffic co-ordination while still providing the required pedestrian facility. Where possible a complete pelican sequence can be operated twice with the area cycle time (that is, it 'double cycles') to minimise pedestrian delays.

A new form of signalised pedestrian facility has been introduced recently, known as the PUFFIN[6] (Pedestrian User-Friendly Intelligent Crossing). With the PUFFIN[6], the presence of pedestrians is automatically detected on the kerb side, using pressure-sensitive mats or other detection technique, and while crossing using infra-red sensors. A pedestrian phase is only called if intending crossers are still waiting on the kerb and the crossing time is extended where slow walkers on the crossing are detected. The red/green man signal is moved to the near-side to enable late-comers to be discouraged from crossing, by showing a red man signal without alarming those already on the crossing. The signals for traffic are then held on red until pedestrians have crossed, so there is no need for a flashing signal with its potential for misunderstanding or misuse.

Signal co-ordination

Where traffic signals are closely spaced, it has been found beneficial to co-ordinate their timings, to help the efficient progression of vehicle platoons through the junctions. Co-ordination between signals at adjacent intersections can be achieved by the use of controllers which are synchronised by the mains supply frequency and which also incorporate a solid-state memory store containing stage timings, cycle times and offset periods. These plans are selected by time of day, or day of the week, according to changes in the traffic flow. It is also possible to incorporate one demand-dependent stage either by vehicle detection or by a pedestrian demand.

An older system, which is being superseded by cableless linking, is cable linking. Here information is passed between controllers so that the commencement of a selected stage at a key intersection controls the beginning or end of other selected stages at other intersections. It is usually necessary in linked systems such as these for a common cycle time to be imposed so that co-ordination with the key intersection may be maintained.

Traffic signal control over a wider area can also be controlled by instructions from a central computer. Cycle time, the start and the length of the green stage together with the offsets between green stages at adjacent junctions are controlled by a central computer using plans determined either by historic records of traffic flow or by reacting to the changes in detected traffic flows in a dynamic manner. The subject of computer control of signals on an area wide basis is considered in detail in Chapter 51.

References

1. J.R. Peirce and P.J. Webb, MOVA: The 20 Site Trial, *Transport and Road Research Laboratory Research* Report RR 279 (1990)
2. N.B. Hounsell, Bus Priority: Techniques and Evaluation, *PTRC 18th Summer Annual Meeting*, University of Sussex (September 1990)
3. Department of Transport, Pedestrian Facilities at Traffic Signal Installations, *Advice Note* TA 15/81, London (1981)
4. Department of Transport, Design Considerations for Pelican and Zebra Crossings, *Advice Note* TA 52/87, London (1987)
5. Department of Transport, Pedestrian Crossings: Pelican and Zebra Crossings, *Standard* TD 28/87, London (1987)
6. H.E.H. Davies, The 'Puffin' Pedestrian Crossing: Experience with the First Experimental Sites, *Transport Research Laboratory Research Report* RR 364 (1992)

33

Geometric factors affecting the capacity of a traffic signal approach

The capacity of a signal-controlled intersection is limited by the capacities of individual approaches to the intersection. There are two types of factors which affect the capacity of an approach: geometric factors which are considered in this chapter, and traffic and control factors which are discussed subsequently.

The capacity of a traffic signal approach is the sum of the capacities or saturation flows of the individual lanes comprising the approach. The saturation flow is independent of traffic and control factors and is the maximum flow, expressed in equivalent passenger car units, that can be discharged from a traffic lane when there is a continuous green indication and a continuous queue on the approach.

Geometric factors affecting lane saturation flow are: the position of the lane (nearside or non-nearside), the width of the lane and its gradient, and the radius of any turning movements. Research[1] has indicated the following values as being appropriate for the United Kingdom.

For unopposed streams in individual traffic lanes, the saturation flow S_1, is given by

$$S_1 = (S_0 - 140d_n)/(1 + 1.5 f/r) \text{ pcu/h}$$

where $S_0 = 2080 - 42\, d_g \times G + 100\,(w - 3.25)$

and d_n is 1 for nearside lanes and 0 for non-nearside lanes,
d_g is 1 for uphill entries and 0 for downhill entries,
G is the percentage gradient for the entry,
w is the lane width at entry (m),
f is the proportion of turning vehicles in the lane,
r is the radius of curvature of vehicle paths (m).

For opposed streams containing opposed right-turning traffic in individual lanes the saturation flow S_2 is given by

$$S_2 = S_g + S_c$$

where S_g is the saturation flow in lanes of opposed mixed turning traffic during the effective green period (pcu/h)

S_c is the saturation flow in lanes of opposed mixed turning traffic after the effective green period (veh/h or pcu/h)

$$S_g = (S_0 - 230)/(1 + (T - 1)f)$$

and
$$T = 1 + 1.5/r + t_1/t_2$$
$$t_1 = 12X_0^2/(1 + 0.6(1 - f)N_s)$$
$$t_2 = 1 - (fX_0)^2$$
$$S_c = p(1 + N_s)(fX_0)^{0.2}\, 3600/\lambda c$$
$$X_0 = q_0/\lambda n_t S_0$$

where X_0 is the degree of saturation on the opposing arm, that is, the ratio of the flow on the opposing arm to the saturation flow on that arm,

N_s is the number of storage spaces available inside the intersection which right turners can use without blocking following straight ahead vehicles,

λ is the proportion of the cycle time effectively green for the phase being considered, that is, the effective green time divided by the cycle time,

c is the cycle time (s),

q_0 is the flow on the opposite arm expressed as vehicles per hour of green time and excluding non-hooking right turners,

p is a conversion factor from vehicles to pcu and is expressed as

$$1 + \sum_i (\alpha_i - 1)p_i$$

where α_i is the pcu value of vehicle type i,

p_i is the proportion of vehicles of type i in the stream,

n_t is the number of lanes on the opposing entry,

S_0 is the saturation flow per lane for the opposite entry (pcu/h),

T is the through car unit of a turning vehicle in a lane of mixed turning traffic, each turning vehicle being equivalent to T straight ahead vehicles.

Most traffic signal approaches are marked out in several lanes and the total saturation flow for the approach is then the sum of the saturation flows of the individual lanes.

Reference

1. R. M. Kimber, M. McDonald and N. B. Hounsell, The prediction of saturation flows for road junctions controlled by traffic signals, *Transport and Road Research Laboratory Research Report* 67, Crowthorne (1986)

Problems

(1) A nearside unopposed lane of a traffic signal approach has a width at entry of 2.4 m and an uphill gradient of 5 per cent; 25 per cent of vehicles turn left with a turning radius of 20 m. Calculate the saturation flow for this lane.

(2) A non-nearside lane of a traffic signal approach has a width at entry of 3.0 m and a downhill gradient of 3 per cent; 40 per cent of vehicles turn right with a turning radius of 25 m. The cycle time is 60 s and the effective green time is 40 s. The right-turning vehicles are opposed by a straight ahead lane with a degree of saturation of 0.85. Two right-turning vehicles may wait within the intersection without obstruction to following straight ahead vehicles and the ratio of pcu/vehicles is 1.5. Calculate the saturation flow for this lane.

Solutions

(1) For a non-opposed lane the saturation flow is calculated from

$$S_l = (S_0 - 140\ d_n)/(1 + 1.5\,f/r) \text{ pcu/h}$$

where $S_0 = 2080 - 42\ d_g \times G + 100(w - 3.25)$

$$= 2080 - 42 \times 1 \times 5 + 100(2.4 - 3.25)$$

$$= 1770.85$$

$$S_l = (1770.85 - 140 \times 1)/(1 + 1.5 \times 0 \times 25/20)$$

$$= 1601 \text{ pcu/h}$$

(2) For an opposed lane the saturation flow is calculated from

$$S_2 = S_g + S_c$$

$$S_g = (S_0 - 230)/(1 + (T - 1)f)$$

and $T = 1 + 1.5/r + t_1/t_2$

$$t_1 = 12X_0^2/(1 + 0.6(1 - f)N_s)$$

$$t_2 = 1 - (fX_0)^2$$

$$t_1 = 12 \times 0.85^2/(1 + 0.6(1 - 0.4)2)$$

$$= 5.04$$

$$t_2 = 1 - (0.4 \times 0.85)^2$$

$$= 0.88$$

$$T = 1 + 1.5/25 + 5.04/0.88$$

$$= 6.79$$

$$S_0 = 2080 - 42\ d_g \times G + 100(w - 3.25)$$

$$= 2080 - 42 \times 0 \times 3 + 100(3.0 - 3.25)$$

$$= 2055$$

$$S_g = (2055 - 230)/(1 + (6.79 - 1)0.40)$$
$$= 550 \text{ pcu/h}$$
$$S_c = p(1 + N_s)\,(fX_0)^{0.2}\,3600/\lambda c$$
$$= 1.5\,(1 + 2)\,(0.4 \times 0.85)^{0.2}\,3600/40$$
$$= 326 \text{ pcu/h}$$
$$S_2 = S_g + S_c$$
$$= 550 + 326 \text{ pcu/h}$$
$$= 876 \text{ pcu/h}$$

34

The effect of traffic factors on the capacity of a traffic signal approach

The effect of traffic factors on the capacity of a traffic signal approach is usually allowed for by the use of weighting factors, referred to as 'passenger car units', assigned to differing vehicle categories. As a consequence the saturation flow of a signal approach or of a single approach lane is expressed in passenger car units per hour (pcu/h).

Whilst the effect of vehicle type on capacity might depend upon the geometric design of the junction, that is, large vehicles might be affected by turning radius or heavy vehicles by gradient, it has been usual in the United Kingdom to use constant factors independent of junction geometry.

The passenger car unit values recommended for use in United Kingdom junction design, with the exception of those for two-wheel vehicles, have been determined using observations of headway ratios[1].

In this method the time headways between individual vehicles crossing the signal approach stop line are measured and time headways between individual pairs of vehicles of interest are grouped. For instance, if it was required to determine the passenger car equivalent of a medium commercial vehicle (defined as a vehicle with 2 axles but more than 4 wheels) then pairs of headways between these vehicles and light vehicles (defined as 3 or 4 wheel vehicles) would be grouped as follows:

(1) a light vehicle following a light vehicle;
(2) a light vehicle following a medium commercial vehicle;
(3) a medium commercial vehicle following a light vehicle;
(4) a medium commercial vehicle following a medium commercial vehicle.

Vehicles crossing the stop line within 3 seconds of the commencement and termination of the green period should not be included in the observations because of the effects of acceleration and deceleration.

The passenger car unit for a medium commercial vehicle can be found by dividing the mean headway for a medium goods vehicle following a medium goods vehicle by the mean headway for a light vehicle following a light vehicle. This is true if the effect of a medium commercial vehicle is independent of whether the vehicles preceding it and following it are light or medium commercial vehicles. The necessary and sufficient condition for this is that the sum of the average headways for a light vehicle following a light vehicle and a medium commercial vehicle following a medium commercial vehicle should be equal to the sum of the average headways for a light vehicle following a medium commercial vehicle and a medium commercial vehicle following a light vehicle.

If this condition is not met then corrected values of the mean headways can be calculated as follows.

Corrected value of mean headway for a light vehicle following a light vehicle equals the uncorrected value (w) minus the correction factor divided by the number of headways in this group (a).

Corrected value of mean headway for a light vehicle following a medium commercial vehicle equals the uncorrected value (x) plus the correction factor divided by the number of headways in this group (b).

Corrected value of mean headway for a medium commercial vehicle following a light vehicle equals the uncorrected value (y) plus the correction factor divided by the number of headways in this group (c).

Corrected value of mean headway for a medium commercial vehicle following a medium commercial vehicle equals the uncorrected value (z) minus the correction factor divided by the number of headways in this group (d).

The correction factor is given by

$$\frac{abcd\,(w - x - y + z)}{bcd + acd + abd + abc}$$

As a result of investigations carried out by Martin and Voorhees Associates, Southampton University and the Transport and Road Research Laboratory, values of passenger car equivalents have been proposed for use in United Kingdom signal design[2] and are given below.

Light vehicles (3 or 4 wheeled vehicles)	1.0
Medium commercial vehicles (2 axles but more than 4 wheels)	1.5
Heavy commercial vehicles (vehicles with more than 2 axles)	2.3
Buses and coaches	2.0
Motorcycles	0.4
Pedal cycles	0.2

The number of vehicles crossing the stop line in a given period of time depends not only on the compaction of the traffic and the saturation flow, but also on the proportion of time during which the signal is effectively green (λ).

A cycle is a complete sequence of signal indications, that is, a green period and a red period for a two-phase system, and the time during which the signal is effectively green during a cycle, known as the 'effective green time'. The maximum number of vehicles crossing the stop line per hour is then

$$\frac{\text{saturation flow} \times \text{effective green time}}{\text{cycle time}}$$

$$= \text{saturation flow} \times \lambda$$

where the saturation flow is in vehicles per hour. It is however usual to express saturation flow in terms of passenger car units per hour, and the flow across the stop line during one hour is then expressed in passenger car units.

Alternative methods for calculating pcu values have been suggested by some authors, based on 'synchronous' and 'asynchronous' regression techniques. With synchronous regression, the number of vehicle departures n_i of each class i are recorded over time periods t, beginning and ending with the departure of a vehicle (the first vehicle departing is excluded from n_i, so that t contains n_i headways). t is regressed on the n_i to obtain estimates of a_i of the coefficients in a linear model. For example, the model of a traffic stream containing cars, buses and heavy commercial vehicles would be:

$$t = an_c + bn_b + cn_h$$

where t = counting period,
n_c = number of cars recorded in a counting period,
n_b = number of buses recorded in a counting period,
n_h = number of heavy commercial vehicles recorded in a counting period.

In this model, which is constrained through the origin (there is no constant term), the coefficients a, b and c represent the mean headway of vehicles of the class. Estimates of pcu values can be obtained from the ratios b/a and c/a. The saturation flow is estimated from $1/a$. Research[3] has shown that pcu factors found from synchronous regression give results which are very close to those calculated from the headway ratio method described earlier in the chapter.

In 'asynchronous' regression, vehicle departures are recorded over time periods T, which begin and end at arbitrary instants. The number of passenger cars is then regressed on T and on the number of vehicles of other classes. However, research[3] has also shown the use of multiple linear regression in this context gives biased (underestimated) pcu values and the method is therefore not recommended.

References

1. D. A. Scraggs, The passenger car equivalent of a heavy vehicle in single lane flow at traffic signals, *Road Research Laboratory Report* LN/573/DAS, Transport and Road Research Laboratory, Crowthorne (1964)
2. R. M. Kimber, M. McDonald and N. B. Hounsell, The prediction of saturation flows for road junctions controlled by traffic signals, *Research Report* 67, Transport and Road Research Laboratory, Crowthorne (1986)
3. R.M. Kimber, M. McDonald and N.B. Hounsell, Passenger Car Units in Saturation Flows: Concept, Derivation, Deviation, *Transportation Research B*, 19B (1985), 1, pp. 39–61

Problems

(1) The hourly traffic flow on an unopposed non-nearside 3.8 m level lane of an approach where turning vehicles follow a path with a radius of curvature of 20 m is 400 passenger cars, 100 medium goods vehicles and 40 motor cycles; one-fifth of all vehicles turn. If the approach is just able to pass all the traffic, should the ratio of effective green time to cycle time be 0.27, 0.33 or 0.55?

(2) The following headways were obtained on a traffic signal approach. Calculate the passenger car equivalent of a medium commercial vehicle.

	Light vehicle following light vehicle	*Light vehicle following medium commercial*	*Medium commercial following light vehicle*	*Medium commercial following medium commercial*
Number of headways	66 (*a*)	33 (*b*)	32 (*c*)	9 (*d*)
Mean headway (seconds)	1.9 (*w*)	2.8 (*x*)	2.3 (*y*)	3.0 (*z*)

Solutions

(1) Converting the flow to passenger car units

$$\begin{aligned}
\text{400 passenger cars} &= 400 \times 1 = 400 \text{ pcu}\\
\text{100 medium goods vehicles} &= 100 \times 1.5 = 150 \text{ pcu}\\
\text{40 motor cycles} &= 40 \times 0.4 = 16 \text{ pcu}\\
& 566 \text{ pcu}
\end{aligned}$$

$$\begin{aligned}
\text{Saturation flow} &= (S_0 - 140\, d_n)/(1 + 1.5\, f/r)\\
S_0 &= 2080 - 42\, d_g\, G + 100(w - 3.25)\\
&= 2080 + 100(3.8 - 3.25)\\
&= 2135\\
\text{Saturation flow} &= 2135/(1 + 1.5 \times 0.20/20)\\
&= 2103 \text{ pcu/h}
\end{aligned}$$

$$\text{Maximum flow over the stop line} = \frac{\text{saturation flow} \times \text{effective green}}{\text{cycle time}}$$

$$566 = \frac{2103 \times \text{effective green}}{\text{cycle time}}$$

$$\frac{\text{effective green}}{\text{cycle time}} = \frac{566}{2103} = 0.27$$

(2) An initial check must be made to determine if a correction is necessary. This is not required if 1.9 + 3.0 equals 2.8 + 2.3; this is not so and hence correction is therefore necessary.

$$\text{Correction factor} = \frac{abcd\,(w - x - y + z)}{bcd + acd + abd + abc}$$

$$= \frac{66 \times 33 \times 32 \times 9(1.9 - 2.8 - 2.3 + 3.0)}{33 \times 32 \times 9 + 66 \times 32 \times 9 + 66 \times 33 \times 9 + 66 \times 33 \times 32}$$

$$= \frac{-125\,452.8}{117\,810}$$

$$= -1.0649$$

Corrected value of mean headway for a light vehicle following a light vehicle is given by

$$1.9 - (-1.0649/66)$$
$$= 1.9 \text{ s}$$

Corrected value of mean headway for a medium commercial vehicle following a medium commercial vehicle is given by

$$3.0 - (-1.0649/9)$$
$$= 3.1 \text{ s}$$

The passenger car equivalent of a medium commercial vehicle is given by 3.1/1.9 or 1.6.

35

Determination of the effective green time

In Chapter 34 the concept of effective green time was introduced as a means of determining the number of vehicles that could cross a stop line over the whole of the cycle comprising both red and green periods.

It is obvious that in practice the flow across the stop line cannot commence or terminate instantly because at the end of the green indication this flow is slowly reduced to zero.

A study of the discharge of vehicles across the stop line allows the effective green time to be determined. Figure 35.1 shows the variation of discharge with time, the area beneath the curve representing the number of vehicles that cross the stop line during the green period. The area beneath the curve is not easily determined and for convenience a rectangle of equal area to that under the curve is

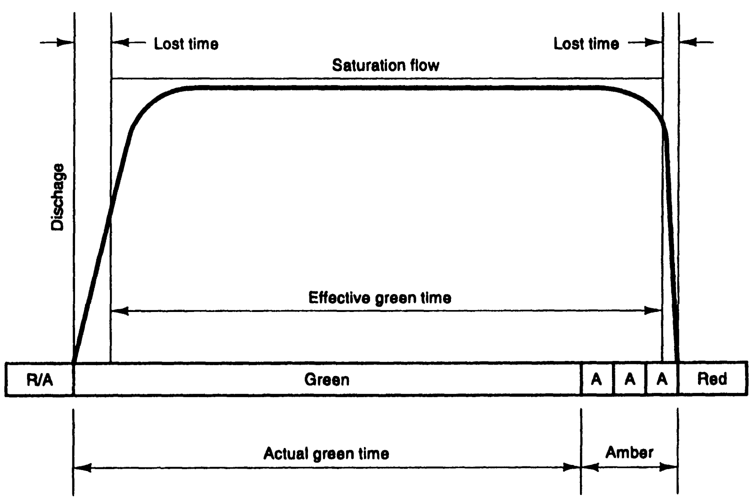

Figure 35.1 *Variations in the discharge across the stop line*

superimposed upon the curve. The height of the rectangle is equal to the saturation flow and the base of the rectangle is the effective green time.

Note that figure 35.1 applies to a single lane or to an entry where the width is constant on the approach. Where an entry is flared (for example the number of lanes is greater at the stop line than upstream on the approach), the saturation flow is usually higher during the first part of the green, until the vehicles in the flare have discharged.

As the amber indication is a period during which, under certain circumstances, vehicles may cross the stop line the discharge across the stop line commences at the beginning of the green period and terminates at the end of the amber period. The time intervals between the commencement of green and the commencement of effective green and also between the termination of effective green and the termination of the amber period are referred to as starting (or initial) and end lost times respectively.

From figure 35.1 it can be seen that the actual green time plus the amber period is equal to the effective green time plus the lost time due to starting and end delays.

In actual practice the lost time due to starting and end lost times is often taken as 2 s and so the effective green time is equal to the actual green time plus the 3 s amber period minus the 2 s lost time.

Determination of the lost time on a traffic-signal approach

Data from a traffic signal-controlled intersection in the City of Bradford will be used to illustrate the determination of lost time due to starting delays. The observations are given in table 35.1.

TABLE 35.1

Time (minutes)	0		0.1		0.2		0.3		0.4		0.5
No. of vehicles crossing stop line		60		76		71		78		79	
No. of saturated intervals observed		32		32		32		32		32	
Discharge per 0.1 minute		1.88		2.48		2.22		2.44		2.47	

No. of last saturated intervals 24, average duration 5.9 s.

Total duration of the last saturated intervals = 142 seconds

Total number of vehicles crossing stop line = 41

Discharge per 0.1 minute during last saturated interval = $41 \times 6/142 = 1.74$ vehicles

The saturation flow for this data is 2.40 vehicles/0.1 minute or 1670 pcu/h.

The lost time at the beginning and end of the green period may be calculated by reference to figure 35.2. By definition the number of vehicles represented by the rectangle efij is equal to the number of vehicles represented by the original histogram (because total flow during effective green time is equal to total flow during green plus amber period). The number of vehicles represented by the area dghk is also equal to the number of vehicles represented by the four 0.1 minute periods of saturated flow between d and k.

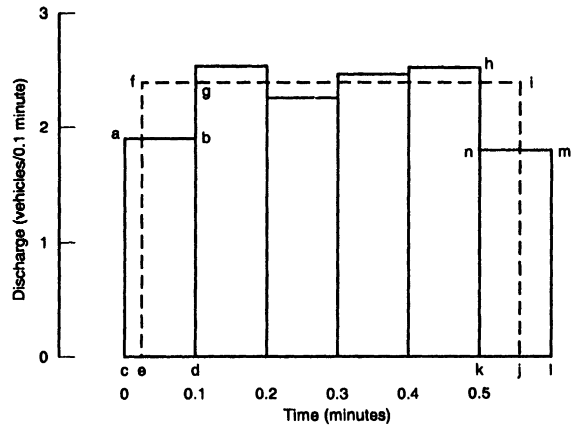

Figure 35.2 *Observed discharge across the stop line*

This means that:

(1) The number of vehicles represented by abdc is equal to the number of vehicles represented by efgd.

(2) The number of vehicles represented by hijk is equal to the number of vehicles represented by nmlk.

That is

$$ed \times 2.40 = 1.88 \times 0.1$$

giving

$$ed = 0.08 \text{ minute}$$

or

$$cd = 0.02 \text{ minute}$$

similarly

$$jl = 0.03 \text{ minute}$$

and

$$jl + ce = \text{lost time during green phase} = 2.9 \text{ s}$$

This value is higher than the accepted value of 2 s and the saturation flow was also noted to be lower than the usual design value. These departures from the accepted values were considered to be caused by a high proportion of elderly drivers in the traffic flow.

Problem

The lost time due to starting and ending delays on a traffic signal approach is noted to be 3 s; the actual green time is 25 s. Is the effective green time 24 s, 25 s or 26 s?

Solution

$$\text{Effective green time} = 25 \text{ s} + 3 \text{ s (amber)} - 3 \text{ s (lost time)}$$
$$= 25 \text{ s}$$

36

Optimum cycle times for an intersection

The length of the cycle time under fixed-time operation is dependent on traffic conditions. Where the intersection is heavily trafficked cycle times must be longer than when the intersection is lightly trafficked.

One definition of the degree of trafficking of an approach is given by the y value, which is the flow on the approach divided by the saturation flow.

For any given traffic-flow conditions with the signals operating under fixed-time control, the duration of the cycle must affect the average delay to vehicles passing through the intersection. Where the cycle time is very short, the proportion of the cycle time occupied by the lost time in the intergreen period and by starting delays is high, making the signal control inefficient and causing lengthy delays.

When on the other hand the cycle time is considerably longer, waiting vehicles will clear the stop line during the early part of the green period and the only vehicles crossing the stop line during the latter part of the green period will be those that subsequently arrive, often at extended headways. As the discharge rate or saturation flow across the stop line is greatest when there is a queue on the approach this also results in inefficient operation.

As a result of the computer simulation of flow at traffic signals carried out by the Road Research Laboratory (*Road Research Technical Paper* 56), it was possible to show these variations of average delay with cycle time that occur at any given intersection when the flows on the approaches remain constant. They are illustrated in figure 36.1 and can be explained for most practical purposes by the total lost time and the y value found at the intersection.

It is shown in *Road Research Technical Paper* 39 that a sufficiently close approximation to the optimum cycle time C_0 could be obtained by the use of the following equation

$$C_0 = \frac{1.5L + 5}{1 - Y}$$

where L is the total lost time per cycle,

Y is the sum of the maximum y values for all the phases comprising the cycle.

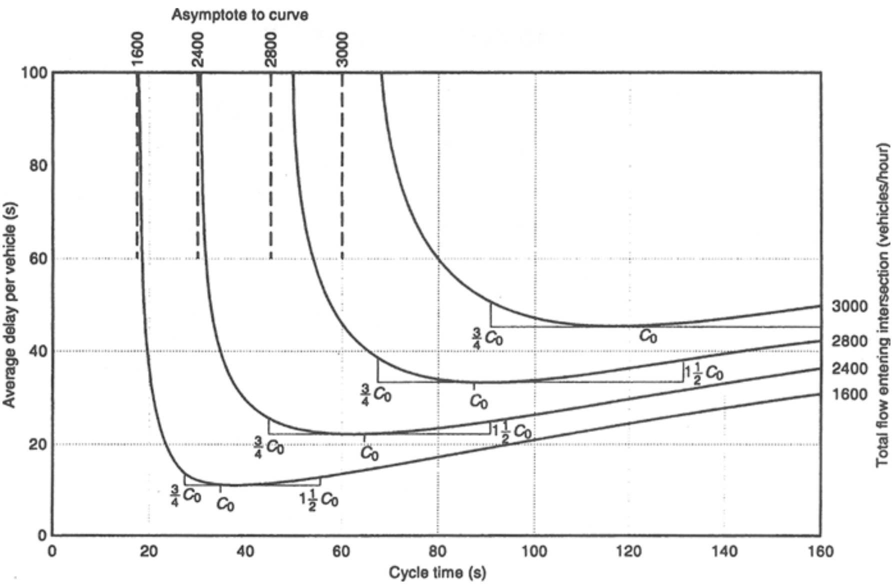

Figure 36.1 *Effect on delay of variation of the cycle length (based on Road Research Technical Paper 56); 2-phase, 4-arm intersection, equal flows on all arms, equal saturation flows of 1800 vehicles/hour, equal green times, total lost time/cycle 10 s*

The application of this formula may be illustrated from figure 36.1 and the calculation of the optimum cycle time C_0 is tabulated in table 36.1 for each of the four flow rates illustrated. An inspection of figure 36.1 shows that the values by the approximate formula are similar to those obtained by simulation.

TABLE 36.1

(1) *Total flow, (veh/h)*	(2) *Flow/approach (veh/h)*	(3) *y value* (2)/1800	(4) *L*	(5) *Y* ($n \times$ (3)*)	(6) C_0 (s)
3000	750	0.42	10	0.84	125
2800	700	0.39	10	0.78	91
2400	600	0.33	10	0.66	59
1600	400	0.22	10	0.44	36

*n = number of phases.

There is a minimum cycle time of 25 s fixed from safety considerations while a maximum cycle time of 120 s is generally considered desirable. Calculated optimum cycle times should normally be limited to this range of values.

The calculation of the optimum cycle time consists of several steps, best illustrated by the flowchart given in figure 36.2.

Some alternative methods have also been derived for determining optimum cycle times at signal-controlled junctions. Notable among these are those developed at University College, London which use techniques such as linear programming to calculate signal settings which minimise total junction delays or, for overloaded junctions, maximise junction reserve capacity (or minimise the overload). These methods are used within signal calculation programs such as OSCADY, which are described in Chapter 45.

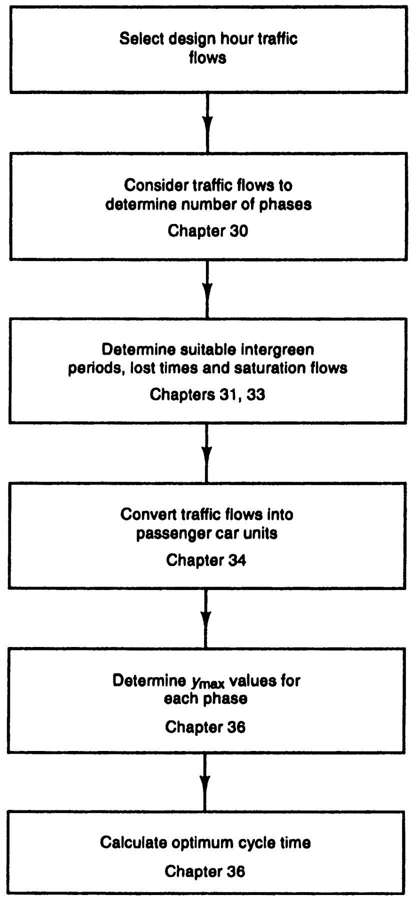

Figure 36.2

The procedures for optimum cycle time calculation described here assume that saturation flow remains constant with increasing green time. While there is evidence to support this assumption at 'good' sites, a reduction in saturation flow with increasing green has been observed in some situations where there is interference to traffic (such as parking/loading activities, high pedestrian flows, etc.). In these circumstances, a lower cycle time may be preferable to the theoretical optimum cycle time calculated above. Lower cycle times also offer lower pedestrian delays and a higher capacity for right-turning traffic owing to the greater number of cycles per hour giving increased opportunity for discharge during the intergreen period. On the negative side, there is evidence that frequencies for some accident categories increase with reduced cycle time, owing to the increased number of stage changes and hence potential conflicts.

Chapter 43 describes a procedure for determining optimum signal settings when saturation flow falls during the green period. Alternatively, this feature can be accounted for in most of the signal calculation programs described in Chapter 45.

Problem: Optimum cycle times for an intersection

Design-hour traffic flows at a 4-arm 3-phase intersection are given in table 36.2; left-turn vehicles comprise 20 per cent of the total lane flow. Two right-turn vehicles can wait within the intersection without delay to following vehicles. The intergreen period is 5 s, start and end lost times are 2 s per green period. All approaches are level and the radius of curvature of all turning paths is 20 m. Is the optimum cycle time 60 s, 90 s or 120 s?

TABLE 36.2
Design-hour traffic flows

	Light vehicles	*Medium goods*	*Buses*	*Motor cycles*	*Approach width (m)*
North approach, straight ahead and left turning	500	100	10	20	3.0
North approach, right turning	50	10	0	10	3.0
South approach, straight ahead and left turning	400	150	0	30	3.0
South approach, right turning	40	30	0	5	3.0
West approach, straight ahead and left turning	400	50	5	10	3.65
West approach, right turning	300	60	0	20	3.65
East approach, straight ahead and left turning	200	180	4	15	3.65
East approach, right turning	260	20	0	10	3.65

Solution

For the traffic flows given in table 36.2 the three major traffic movements are:

north/south, all directions of movement;
east/west, straight ahead and left-turning movements;
east/west, right-turning movements.

These three major traffic movements will each be given a separate phase.

TABLE 36.3
Design-hour traffic flow in pcus

	Light vehicles	*Medium goods*	*Buses*	*Motor cycles*	*Total*
North approach, ahead and left	500	150	20	8	678
North approach, right	50	15	0	4	69
South approach, ahead and left	400	225	0	12	637
South approach, right	40	45	0	2	87
West approach, ahead and left	400	75	10	4	489
West approach, right	300	90	0	8	398
East approach, ahead and left	200	270	8	6	484
East approach, right	260	30	0	4	294

The traffic flow for each of these phases will now be converted to passenger car equivalents (table 36.3).

The saturation flows of the approaches will now be calculated (see Chapter 33).
For non-opposed flows.
North approach, ahead and left (nearside lane).

$$S_0 = 2080 + 100\,(3.0 - 3.25)$$
$$= 2055$$
$$S_1 = (2055 - 140)/(1 + 1.5 \times 0.2/20)$$
$$= 1887 \text{ pcu/h}$$

South approach, ahead and left (nearside lane).

$$S_1 = 1887 \text{ pcu/h}$$

West approach, ahead and left (nearside lane).

$$S_0 = 2080 + 100\,(3.65 - 3.25)$$
$$= 2120$$
$$S_1 = (2120 - 140)/(1 + 1.5 \times 0.2 / 20)$$
$$= 1951 \text{ pcu/h}$$

East approach, ahead and left (nearside lane).

$$S_1 = 1951 \text{ pcu/h}$$

West approach, right.

$$S_0 = 2080 + 100\,(3.65 - 3.25)$$
$$= 2120$$
$$S_1 = 2120/(1 + 1.5 / 20)$$
$$= 1972 \text{ pcu/h}$$

East approach, right.

$$S_1 = 1972 \text{ pcu/h}$$

For opposed flows.

North approach, right and South approach, right.

An inspection of the right-turning flows and the opposing flows indicates that the south approach right-turn flow is the critical one to be used in design.

$$S_2 = S_g + S_c$$
$$S_g = (S_0 - 230)/(1 + (T - 1)f)$$
$$S_0 = 2055$$
$$t_1 = 12X_0^2/(1 + 0.6(1 - f)N_s)$$

Assume $X_0 = 0.75$

$$t_1 = 12 \times 0.75^2/(1 + 0.6(1 - 1)2)$$
$$= 6.75$$
$$t_2 = 1 - (fX_0)^2$$
$$= 1 - (1 \times 0.75)^2$$
$$= 0.44$$
$$T = 1 + 1.5/r + t_1/t_2$$
$$= 1 + 1.5/20 + 6.75/0.44$$
$$= 16.42$$

$S_g = (2055 - 230)/(1 + (16.42 - 1)1)$
$= 111$ pcu/h
$S_c = p(1 + N_s)(fX_0)^{0.2}\ 3600/\lambda c$
$p = 87/75$ pcu/vehicle
$= 1.16$ pcu/vehicle
Assume $\lambda = 0.3$, $c = 100S$ so $\lambda c = 30$ (initial assumption)
$S_c = 1.16(1 + 2)(1 \times 0.75)^{0.2}\ 3600/30$
$= 1.16 \times 3 \times 0.94 \times 120$
$= 393$ pcu/h
$S_2 = 111 + 393$ pcu/h
$= 504$ pcu/h

The flows, saturation flows and y values, the ratio of flow to saturation flow, for each approach are now tabulated in table 36.4.

TABLE 36.4
Flows, saturation flows and y values

	Flow	*Saturation flow*	*y*
North approach, ahead and left	678	1887	0.36
North approach, right	69	–	–
South approach. ahead and left	637	1887	0.34
South approach, right	87	504	0.17
West approach, ahead and left	489	1951	0.2 5
West approach, right	398	1972	0.20
East approach, ahead and left	484	1951	0.25
East approach, right	294	1972	0.15

Considering the three major traffic movements, the critical approaches to be used in design (the y_{max} values) are:

for the north/south movements the maximum value of y occurs on the north approach, ahead and left, and is equal to 0.36

for the east/west, straight ahead and left, the maximum value of y occurs on the west approach and is equal to 0.25

for the east/west right-turning movements the maximum value of y occurs on the west approach and is equal to 0.20

The total lost time per cycle (L) is composed of the lost time during the green period and the lost time during the intergreen period.

The lost time during the 5 s intergreen period is 2 s and the intergreen period occurs three times with a three-phase system (see Chapter 31). Start and end lost times total 2 s for each of the three green periods of the three-phase cycle (see Chapter 33). The total lost time (L) is $3 \times 2 + 3 \times 2 = 12$ s. The optimum cycle time C_0 is given by

$$C_0 = \frac{1.5L + 5}{1 - Y}$$

$$Y = \Sigma y_{max} = 0.36 + 0.25 + 0.20 = 0.81$$

$$C_0 = \frac{1.5 \times 12 + 5}{1 - 0.81} = 121 \text{ s}$$

It is usual to limit cycle times to a maximum value of 120 s and in this example this maximum value will be adopted.

37

The timing diagram

Previously that optimum cycle time has been calculated which would result in minimum overall delay when employed with fixed-time signals. The intergreen periods have also been selected previously and the remaining calculation, before the whole sequence of signal aspects can be described, is to calculate the duration of the green signal aspects.

Once the duration of these green aspects has been calculated, they can be employed with vehicle-actuated signals, and are then the maximum green times at the end of which a phase change will occur regardless of any demands for vehicle extensions.

The first step is to calculate the amount of effective green time available during each cycle from

available effective green time/cycle = cycle time − lost time/cycle (L)

This available effective green time is then divided between the phases in proportion to the y_{max} value for each phase.

For the particular value of lost time due to starting delays associated with each green period, the actual green time can be calculated. This is the indication shown on the signal and depicted on the timing diagram.

These steps will now be illustrated by the following example.

C_0 82 s
L 12 s
y_{max}, north/south phase (all movements) 0.21
y_{max} east/west phase (straight ahead and left turn) 0.26
y_{max} east/west phase (right turning) 0.25

available effective green time/cycle = 82 − 12 = 70 s

This is divided between the phases in the ratio

0.21: 0.26: 0.25:

or

20.0 s, 25.0 s and 25.0 s

When the lost time due to starting delays is 2 s then the actual green time is equal to the effective green time minus 1 s (see Chapter 35).

The actual green times used are in practice

north/south traffic flow	19 s
east/west straight ahead and left turning	24 s
east/west right-turning flows	24 s

The actual green times have been rounded to the nearest second.

Figure 37.1 is a timing diagram showing these signal indications.

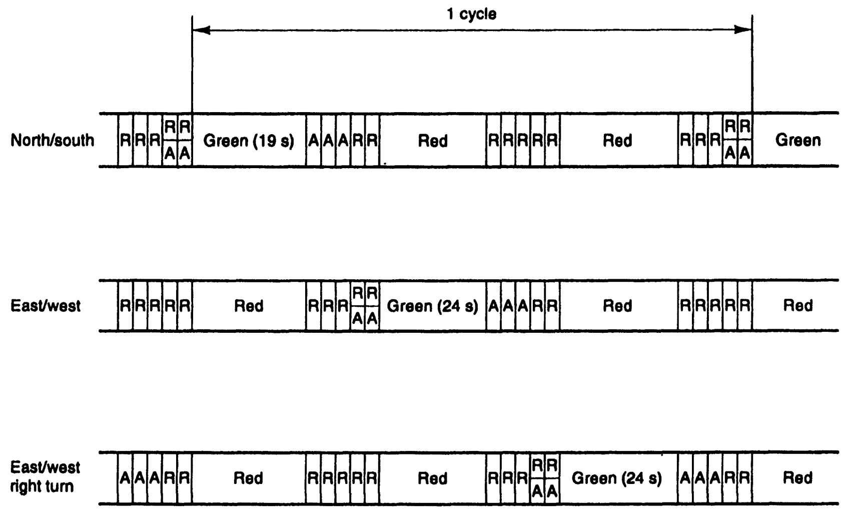

Figure 37.1 *A three-phase timing diagram*

Although frequently used in practice, the allocation of green times by equalising the degree of saturation between competing phases may not produce delay-minimising settings. This is particularly true where flows and capacities are markedly different between phases. Delay-minimising settings are more complex to derive and are therefore mostly used within traffic-signal calculation computer programs such as OSCADY (see Chapter 45).

Problems

An intersection is controlled by four-phase traffic signals, with a cycle time of 100 s. The minimum intergreen period is employed and a value of lost time per green period of 3 s is assumed. Saturation flows on all approaches are identical, but the maximum traffic flows on two of the phases are twice the maximum traffic flows on the remaining two phases. Which of the following series of actual green times would be appropriate?

(a) 28 s, 28 s, 14 s and 14 s;
(b) 30 s, 15 s, 30 s and 15 s;
(c) 30 s, 15 s, I5 s and 15 s;
(d) 14 s, 14 s, 30 s and 30 s.

Solutions

The minimum intergreen period that can be employed is 4 s (see Chapter 31), and when it has this value the lost time in the intergreen period is 1 s.

The total lost time L per cycle is equal to the number of phases multiplied by the lost time in the intergreen period plus the lost time caused by starting delays, that is $4(1 + 3) = 16$ s (see Chapter 35).

The y value for an approach is the flow divided by the saturation flow, and the maximum flows for each phase are in the ratio: 2:2:1:1. As the saturation flows on each approach are equal, this is also the ratio of the y values.

$$\begin{aligned}\text{available effective green time/cycle} &= \text{cycle time} - \text{lost time/cycle}\\ &= 100 - 16\text{ s}\\ &= 84\text{ s}\end{aligned}$$

This available effective green time is divided in the ratio of the y values, that is

28 s, 28 s, 14 s and 14 s

$$\begin{aligned}\text{actual green time} &= \text{effective green time} + \text{starting delays} - \text{amber period}\\ &= \text{effective green time} + 3\text{ s} - 3\text{ s}\end{aligned}$$

The actual green times are 28 s, 28 s, 14 s and 14 s.

38

Early cut-off and late-start facilities

Where the number of right-turning vehicles is not sufficient to justify the provision of a right-turning phase but where right-turning vehicles have difficulty in completing the traffic movement, then an early cut-off or a late-start of the opposing phase is employed.

An early cut-off of the opposing flow allows right-turning vehicles to complete their traffic movement at the end of the green period when the opposing flow is halted. To allow straight ahead vehicles on the same approach to flow without interruption during the early part of the green period, it is necessary for there to be sufficient room in the intersection for right-turning vehicles to wait.

In contrast a late-start facility discharges the right-turning vehicles at the commencement of the green period and for this reason storage space is not as important as with the early cut-off facility.

The calculation of optimum cycle times for fixed-time operation and hence the determination of maximum green times for vehicle-actuated operation is as follows.

The traffic flows are as shown in figure 38.1 and it is necessary to introduce a late-start or early cut-off facility to permit vehicles to turn right from the west approach.

These traffic flows are to be controlled by a two-phase system and it is first necessary to determine the maximum value of the ratio of flow to saturation flow for each phase.

For the north/south phase the maximum y value will be denoted by $y_{\max N/S}$. For the west/east phase, however, it is necessary to determine the greater of

$$y_w \quad \text{or} \quad y_{w_r} + y_e$$

Whichever is the greater will be the $y_{\max E/W}$.

The reason for this combination of y values is that during the west/east phase the flow w continues for the whole of the green time, while flows w_r and e share this green time.

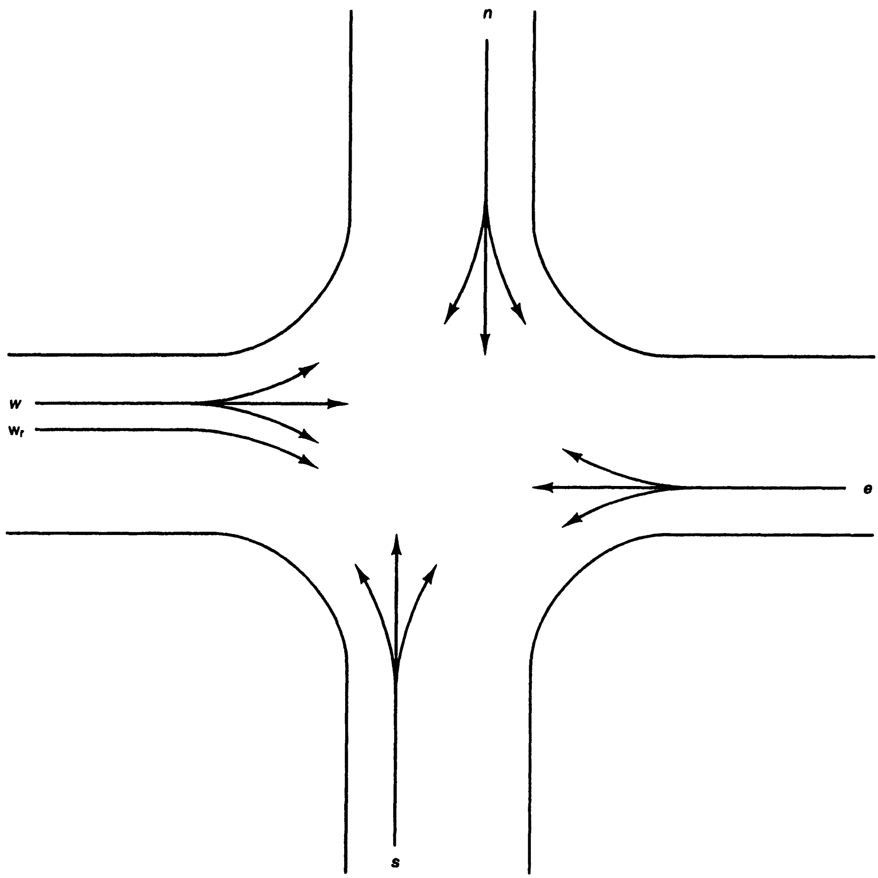

Figure 38.1 *Traffic flows at an intersection with a heavy right-turning movement*

If $y_{w_r} + y_e$ is greater than y_w, then the y value for the first stage will be y_e and that for the second stage y_{w_r}. If however y_w is the greater, then the green time that flow w requires should be divided in proportion to the y values of streams w_r and e. Then the y value for the first stage is

$$\frac{y_e y_w}{y_e + y_{w_r}}$$

and for the second stage

$$\frac{y_{w_r} y_w}{y_e + y_{w_r}}$$

The same procedure may be used with a late start for the opposing flow.

Problem

The following hourly flows (table 38.1) and saturation flows relate to an intersection to be controlled by two-phase signals incorporating a late-start feature. Minimum intergreen periods are employed and lost times are 2 s, for each green plus amber period.

Is the period that right-turning vehicles from the west approach require to complete their turning movement without obstruction from the straight ahead flow on east approach 5 s, 11 s or 16 s?

TABLE 38.1

Approach	*Flow (pcu/h)*	*Saturation flow (pcu/h)*
west, straight ahead and left-turning	400	1900
west, right-turning	200	1600
east, all movements	700	1900
north, all movements	500	1900
south, all movements	600	1900

Solution

To calculate the optimum cycle time for the intersection it is first necessary to calculate the y values for the approaches. They are tabulated in table 38.2.

TABLE 38.2

	Approach	*Flow (pcu/h)*	*Saturation flow (pcu/h)*	*y value*
1	west, straight ahead, left-turning	400	1900	0.21
2	west, right-turning	200	1600	0.13
3	east, all movements	700	1900	0.37
4	north, all movements	500	1900	0.26
5	south, all movements	600	1900	0.32

$y_{\text{max west/east}}$ is 0.21 or 0.13 + 0.37, whichever is the greater (see page 314).

$y_{\text{max north/south}}$ is 0.26 or 0.32, whichever is the greater.

The total lost time is equal to the sum of the lost time in the intergreen period which is 1 s (see pages 283 and 284) plus starting and end lost times of 2 s (see Chapter 35) multiplied by the number of phases, that is 2(1 + 2) = 6 s.

The optimum cycle time C_0 is given by (see Chapter 36)

$$C_0 = \frac{1.5L + 5}{1 - Y}$$

$$= \frac{1.5 \times 6 + 5}{1 - (0.50 + 0.32)}$$

$$= 78 \text{ s}$$

The effective green time available for distribution between the phases is the cycle time minus the total lost time per cycle (see Chapter 35).

$= 78 - 6$

$= 72$ s

This must be divided between the phases in the ratio of the y_{max} for each phase, that is 0.50 and 0.32 (see Chapter 37), giving effective green times of 44 s and 28 s respectively.

The actual green time may be obtained from the effective green times by the relationship, actual green time = effective green time + lost time − amber period (see Chapter 35).

The actual green times are thus:

north/south phase	27 s
east/west phase	43 s

The east/west phase is divided into two stages in the proportion of 0.13 to 0.37 giving a late start to green on the east approach of 11 s. The timing diagram is shown in figure 38.2.

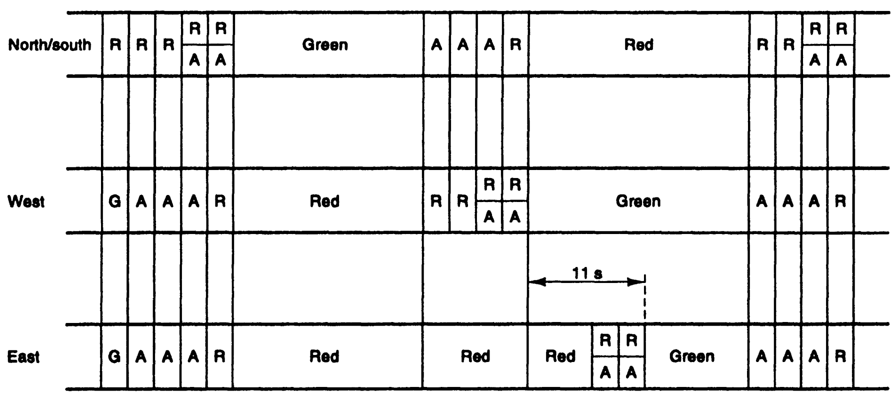

Figure 38.2

39

Opposed right-turning vehicles and gap acceptance

Right-turning vehicles present particular problems at signal-controlled intersections. On a signal approach, right-turn vehicles may be given an exclusive right-turn lane or they may be mixed with straight ahead vehicles. If vehicles do not have to give way, that is, the flow is unopposed, then the maximum discharge or saturation flow can be calculated using the relationships given in Chapter 33. When the flow is opposed and the right-turn vehicles are mixed with straight ahead vehicles then once again the relationships given in Chapter 33 can be used to calculate the saturation flow. If however the right-turn flow is opposed and is not mixed with straight ahead vehicles then the following simplified approach to the calculation of the right-turn discharge can be used. In this case some right-turn vehicles are able to turn through gaps in the opposing straight ahead flow and the remainder must turn in an early cut-off intergreen period at the end of the green period.

At the beginning of the green period the opposing flow is discharged at the saturation rate but later in the green period when the queue has been discharged, opposing vehicles are discharged across the stop line in the random or unsaturated manner in which they arrive on the approach.

If the length of the initial saturated green time can be calculated then the length of the unsaturated green time is also known. The discharge of right-turning vehicles through gaps in the opposing flow can then be determined from the analytical work carried out by Tanner (see Chapter 18) on the discharge of vehicles at priority intersections.

This unsaturated green period during which right-turning vehicles can be discharged through the opposing flow can be determined by assuming that no vehicles remain in the opposing queue at the end of the green period. Such an assumption is correct in the flow conditions being considered.

When the saturated green time for the opposing flow is denoted by g_s then the number of vehicles discharged during this time is

$$(r_e + g_s)q$$

where r_e is the effective red period, that is the cycle time minus the effective green period,
q is the opposing flow.

These vehicles discharge during the saturated green time g_s at the saturation flow s. Then

$$g_s = \frac{(r_e + g_s)q}{s}$$

giving

$$g_s = \frac{r_e \times q}{s - q}$$

$$= \frac{(c - g)q}{s - q} \tag{39.1}$$

where c is the cycle time,
g is the effective green time.

Let g_u be the unsaturated green time. Then

$$g_u = g - g_s$$

Substituting from equation 39.1

$$g_u = \frac{gs - qc}{s - q}$$

The number of vehicles turning right during this period n_r is then

$$n_r = s_r\left(\frac{gs - qc}{s - q}\right) \tag{39.2}$$

where s_r is the maximum theoretical right-turning flow passing through gaps in the opposing flow.

Tanner has given the maximum intersection flow s_r as

$$\frac{q_0(1 - B_0q_0)}{\exp\left[q_0\left(\alpha - B_0\right)\right]\left[1 - \exp\left(- B_rq_0\right)\right]}$$

where q_0 is the rate of arrival of the opposing flow in vehicles per second;
B_0 is the mean minimum time interval between opposing vehicles passing through the intersection;
B_r is the mean minimum time interval between right-turning vehicles passing through the intersection;
α is the average gap accepted by right-turning vehicles in the opposing flow.

Plotted values of the right-turning flow for values of α, B_r and B_0 typical of those noted when the opposing flow is in one or two lanes are shown in figure 39.1, which is reproduced from *Road Research Technical Paper* 56.

The difference between the average number of right-turning vehicles arriving per cycle and n_r calculated from equation 39.2 gives the number of vehicles that must turn right during an early cut-off or intergreen period.

Right-turning vehicles are assumed to discharge at headways of 2.5 s past a point in the centre of the intersection, the first right-turning vehicle passing this point at the end of the amber period and subsequent vehicles passing this point at intervals of 2.5 s.

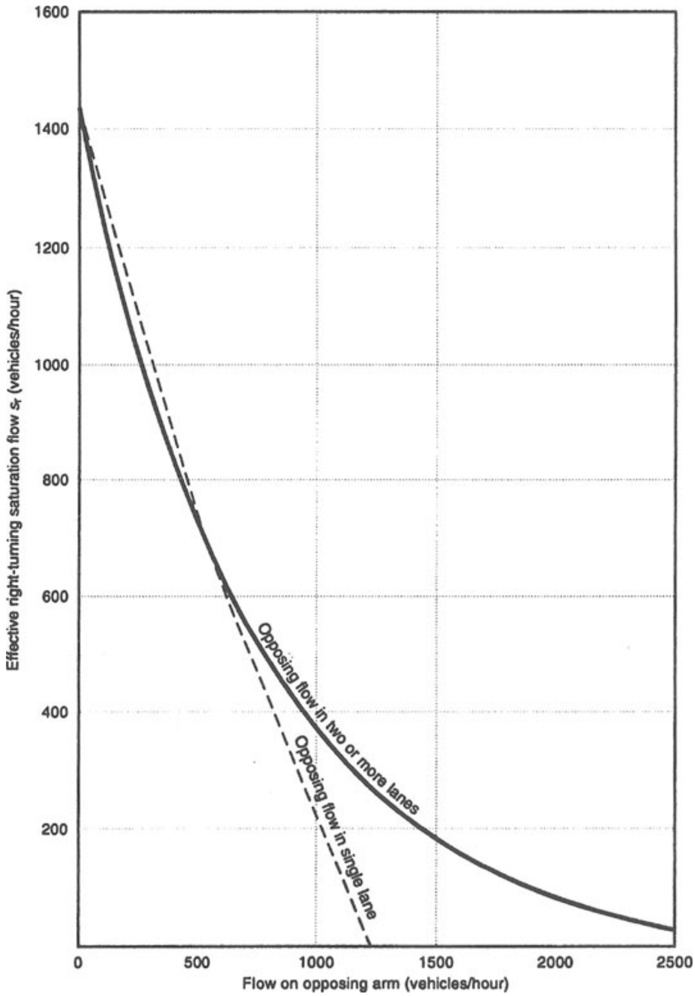

Figure 39.1 *Right-turning saturation flows (based on Road Research Technical Paper 56)*

	α	B_r	B_0
		(seconds)	
Single-lane opposing flow	6	2.5	3
Two (or more)-lane opposing flow	6	2.5	1

α is the minimum gap required in the opposing flow for 1 right-turner
B_r is the minimum headway between successive right-turners ($1/\beta_r$ is the saturation flow)
B_0 is the minimum headway between successive vehicles in the opposing flow

If it is further assumed that the first vehicle on the cross phase needs 3 s from the start of the green indication on the cross phase to reach the mid-point of the intersection then the length of the intergreen period will be 2.5 × the number of vehicles waiting to turn right.

The same procedure may be used to estimate the duration of any early cut-off period.

Problem

At an intersection with traffic flows as given in Chapter 38 the effective green time on the west/east approach is 35 s, and the cycle time is 78 s. The right-turning saturation flow may be estimated from figure 39.1 for single-lane flow. The pcu/h vehicle ratio is 1.2. Is the average number of vehicles waiting to turn right at the end of the green period approximately 2, 5 or 8?

Solution

Using equation 39.2

$$n_r = s_r\left(\frac{gs - qc}{s - q}\right)$$

Details of the traffic flows as given in Chapter 38 and in this chapter are:

q (flow on the opposing east approach) = 700 pcu/h
or 700/1.2 = 583 vehicles/h

g (effective green time) = 35 s
c (cycle time) = 78 s
s (saturation flow on cast approach) = 1900 pcu/h
or 1900/1.2 = 1583 vehicles/h

From figure 39.1, when q = 583 vehicles/h
s_r = 650 vehicles/h

Then

$$n_r = \frac{650}{3600}\left(\frac{35 \times 1583 - 583 \times 78}{1583 - 583}\right)$$
$$= 1.8$$

40

The ultimate capacity of the whole intersection

It was explained in Chapter 35 that the capacity of an approach is dependent on the lost time during the cycle. When the whole intersection is considered the capacity is also dependent on the total lost time on all the phases because the remainder of the time is shared equally between the phases and used as running time.

When the cycle time is calculated as described in Chapter 36 and the green times apportioned as described in Chapter 37, the approaches on each phase that have the highest ratio of flow to saturation flow (y value) will all reach capacity simultaneously as traffic growth takes place.

The ultimate capacity of the intersection is then the maximum flow that can pass through the intersection with the same relative flows and turning movements on the approaches.

As traffic growth takes place the capacity of the intersection can be increased by increasing the cycle time, because the ratio of lost time to cycle time decreases. There is however a practical limit beyond which there is little gain in efficiency as the cycle time is increased and at which drivers may become impatient at not receiving a green indication for the approach at which they are queueing. This practical limit is set at 120 s and the ultimate capacity of the intersection could be calculated as the flow that could just pass through the intersection when the signals were set at this cycle time.

This capacity is however the maximum capacity and is associated with long delays. It is more usual in traffic engineering work to use a practical capacity of 90 per cent of the maximum and this will result in shorter delays.

When traffic flows are uniform, the cycle time C_m, which is just long enough to pass all the traffic that arrives in one cycle, is easily calculated. Traffic flow is however semi-random or random and time is wasted because of variability of arrival times so that the minimum cycle time C_m is associated with extremely long delays.

The minimum cycle time is given by

$$C_m = L + \frac{q_1}{s_1} C_m + \frac{q_2}{s_2} C_m + \frac{q_3}{s_3} C_m + \ldots + \frac{q_n}{s_n} C_m$$

where q_n/s_n is the highest ratio of flow to saturation flow for phase n. Then

$$\begin{aligned} C_m &= L + y_1 C_m + y_2 C_m + \ldots + y_n C_m \\ &= L + C_m \times Y \end{aligned}$$

or

$$Y = 1 - \frac{L}{C_m}$$

The maximum possible ratio of flow to saturation flow Y that can be accommodated over all the phases when C_m is at the practical limit and Y practical is employed is

$$Y_{\text{pract.}} = 0.9\left(1 - \frac{L}{120}\right)$$

where $Y_{\text{pract.}}$ is 90 per cent of Y, or

$$Y_{\text{pract.}} = 0.9 - 0.0075L \tag{40.1}$$

As the Y value is the sum of the maximum ratios of flows divided by saturation flows for each phase and as the saturation flow is fixed, then as the flow increases Y increases from any given present-day value of Y existing, to a future maximum value of Y practical. The reserve capacity at the existing flow is thus

$$\frac{100\,(Y_{\text{pract.}} - Y_{\text{exist.}})}{Y_{\text{exist.}}} \text{ per cent} \tag{40.2}$$

Problem

For the intersection with the flows given in Chapter 38 and controlled by a two-phase system incorporating a late-start feature on the east approach, is the reserve capacity of the intersection 4 per cent, 14 per cent or 24 per cent?

Solution

The approaches and their y values are reproduced from Chapter 38 in table 40.1.

TABLE 40.1

	Approach	*Y value*
1	west, straight ahead and left-turning	0.21
2	west, right-turning	0.13
3	east, all movements	0.37
4	north, all movements	0.26
5	south, all movements	0.32

The y values selected as y_{max} values are

east/west phase $0.13 + 0.37 = 0.50$

north/south phase $= 0.32$

From equation 40.1

$$Y_{pract.} = 0.9 - 0.0075L$$

and $L = 6$ s (see page 310)

$$Y_{pract.} = 0.9 - 0.0075 \times 6$$
$$= 0.85$$

From equation 40.2

$$\text{reserve capacity} = \frac{100(0.85 - 0.82)}{0.82}$$
$$= 4 \text{ per cent}$$

41

The optimisation of signal-approach dimensions

The procedures previously outlined have optimised cycle time and green times so as to produce minimum overall delay for an intersection where the physical dimensions of the approach highways were fixed.

When an intersection is being redesigned or a new intersection being constructed, then it is often desirable to modify the widths of the signal approaches. The necessity for this is not difficult to see, for if the intersection is to have the capacity of the approach highways, then it must have approximately twice the width of the highway if it has a green time of approximately half the real time.

Once the approach widths have been determined, then the optimum cycle time and the green times can be calculated for minimum delay conditions. Variations in approach width will result in differing optimum cycle times and thus there is theoretically an infinite number of variations possible. In practice the choice is less because the designer must normally produce a design with a standard lane width or multiples of this width.

Recently developed signal calculation programs, described in Chapter 45, are now widely used as an aid to geometric design. Where they are not available, another approach to this problem is that made by Webster and Newby[1]. They assumed that the maximum possible rate of flow across the stop line was proportional to the widths of the approach, w_1 and w_2, as shown in figure 41.1, and also that the widened sections of the approaches had lengths d_1 and d_2 which were just long enough to accommodate the queues that could pass through the intersection during fully saturated green periods.

For the intersection shown in figure 41.1 operating under two-phase control, q_1 and q_2 are the design flows of the first and second phase, the approach widths of which are w_1 and w_2 and the effective green times of which are g_1 and g_2.

Then

$$cq_1 = sw_1g_1$$

and

$$cq_2 = sw_2g_2$$

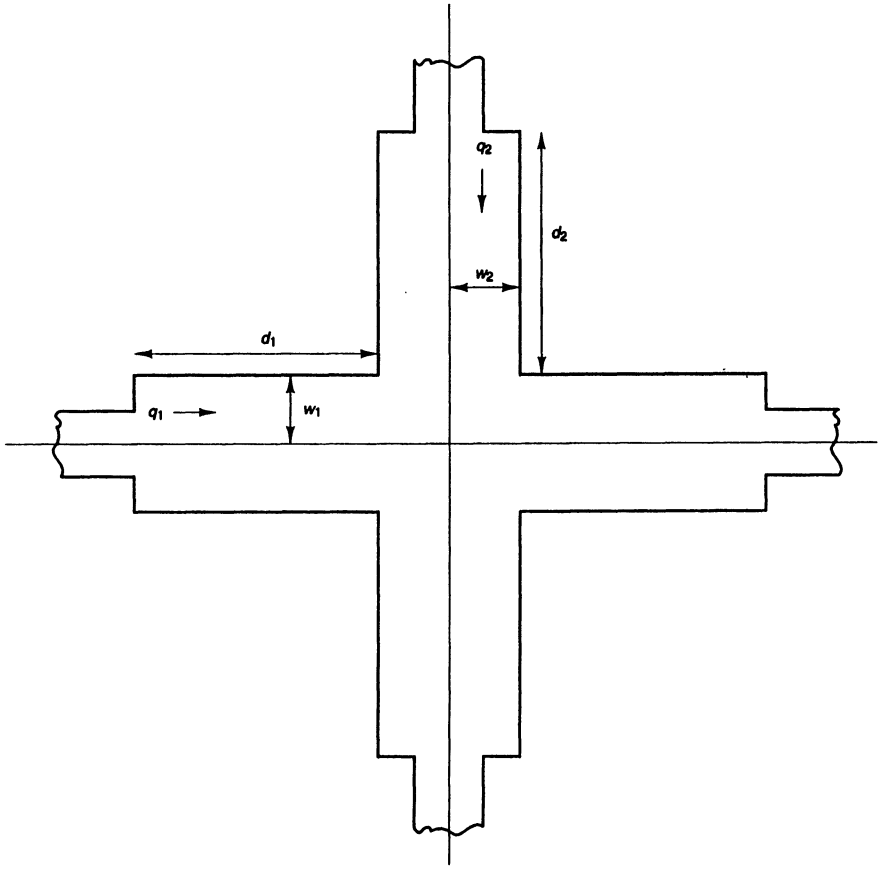

Figure 41.1 *Idealised geometric layout for a signal-controlled junction with widened approaches*

where s is the saturation flow per unit width of road and c is the cycle time.

Since the sum of the effective green periods is constant

$$g_1 + g_2 = cq_1/sw_1 + cq_2/sw_2$$

Differentiating with respect to w gives

$$\frac{dw_2}{dw_1} = -\frac{q_1}{q_2(w_1/w_2)^2}$$

It is required to minimise the total width W ($W = w_1 + w_2$) at the intersection. For minimum total width

$$\frac{dW}{dw_1} = 0$$

or

$$1 + \frac{dw_2}{dw_1} = 0$$

Substituting for dw_2/dw_1 gives

$$1 - \frac{q_1}{q_2(w_2/w_1)^2} = 0$$

or

$$\frac{w_1}{w_2} = \sqrt{\frac{q_1}{q_2}}$$

and

$$\frac{q_1}{q_2} = \frac{w_1 g_1}{w_2 g_2} = \frac{g_1}{g_2}\sqrt{\frac{q_1}{q_2}}$$

or

$$\frac{g_1}{g_2} = \sqrt{\frac{q_1}{q_2}}$$

Also

$$\frac{d_1}{d_2} = \frac{g_1}{g_2} = \sqrt{\frac{q_1}{q_2}}$$

These rules can be summarised as

$$\frac{d_1}{d_2} = \frac{g_1}{g_2} = \sqrt{\frac{q_1}{q_2}} \qquad (41.1)$$

This means that a major road carrying four times as much traffic as its minor cross-road should have approaches which are twice as wide as the minor road approaches and have green times which are twice as long.

If a width obtained from this expression results in a minor road theoretically having a width less than the practical minimum carriageway width then the practical minimum should be employed and a green time less than that theoretically necessary used. The green time saved can then be allocated to the other phases so allowing a reduction in their width.

It is often found that the flows on the differing arms of the same phase are approximately equal even though they may occur at differing times of the day. If this is not so then the highest flow should be used in the formula to obtain the width and the green time. Next the width of the approach on the same phase with the lower flow can be obtained.

The same rule can be applied to multi-phase intersections, when

$$w_1 : w_2 \ldots w_n$$

$$g_1 : g_2 \cdots g_n$$

$$d_1 : d_2 \ldots d_n$$

are proportional to

$$\sqrt{q_1} : \sqrt{q_2} \ldots \sqrt{q_n}$$

For T-junctions with 2-phase control the ratios of widths, green times and widened lengths should be

$$\frac{w_1}{w_2} = \sqrt{\frac{q_1}{2q_2}}$$

and

$$\frac{g_1}{g_2} = \frac{d_1}{d_2} = \sqrt{\frac{2q_1}{q_2}} \qquad (41.2)$$

where q_1 etc., refer to the major road,
q_2 etc., refer to the minor road.

Reference

1. F. V. Webster and R. F. Newby, Research into the relative merits of roundabouts and traffic-signal intersections, *J Instn Civ. Engrs*, 27 (Jan. 1964), 47–76

Problem

The traffic flows at an intersection are given in table 41.1.

TABLE 41.1

Approach	*Flow (pcu/h)*
north	3600
south	3300
east	800
west	900

Show that the relative proportions of the approach widths and effective green times on the widest approaches of each phase are 2:1 when two-phase control is used and the design minimises the area of land required.

Show also that when

(a) the cycle time is 120 s;
(b) the total lost time per cycle is 6 s;
(c) the intergreen period is 4 s;
(d) the minimum approach width on the east/west approach gives a saturation flow of 3300 pcu/h.

The actual green settings are 32 s and 80 s.

Solution

The two phases used for control purposes will be the north/south flow and the east/west flow (see Chapter 37).

Select the maximum flow on each phase

north	3600	pcu/h
west	900	pcu/h

From equation 41.1

$$\frac{g_1}{g_2} = \frac{w_1}{w_2} = \sqrt{\frac{3600}{900}} = 2$$

The relative proportion of widths and effective green times for the widest approaches on each phase is thus 2:1.

When the cycle time is 120 s and the total lost time per cycle is 6 s, then the available effective green time is (see Chapter 36)

$$120 - 6 = 114 \text{ s}$$

When this is divided between the phases in the ratio of 2:1 then

effective green time, north/south phase = 76 s

effective green time, east/west phase = 38 s

For the west approach the minimum width is stated to give a saturation flow of 3300 pcu/h.

The maximum flow that can be passed with the signal settings as calculated is

$$\frac{\text{effective green time}}{\text{cycle time}} \times 3300 \text{ pcu/h} = \frac{3300}{120}$$
$$= 1045 \text{ pcu/h}$$

This maximum flow is greater than the design flow on the west approach, which is given as 900 pcu/h. It is therefore possible to reduce the green time from the value calculated using equation 41.1. The effective green time required can be calculated from

$$\text{actual flow} = \frac{\text{saturation} \times \text{effective green time}}{\text{cycle time}}$$

or

$$\text{effective green time} = \frac{900 \times 120}{3300} = 33 \text{ s (approx.)}$$

The effective green time as calculated from equation 41.1 is 38 s and this leaves $38 - 33 = 5$ s available for use on the north/south phase. The total effective green time on the north/south phase is then $76 + 5 = 81$ s.

As the intergreen period is 4 s and the total lost time per cycle is 6 s then starting delays are 2 s of each green period.

$$\begin{aligned}\text{actual green time} &= \text{effective green time} + \text{lost time due to starting delays} \\ &\quad - \text{amber period} \\ &= \text{effective green time} - 1 \\ &= 80 \text{ s and } 32 \text{ s.}\end{aligned}$$

42

Optimum signal settings when saturation flow falls during the green period

On some traffic-signal approaches it cannot be assumed that saturation flow will remain constant throughout the greater part of the green period. The effect of blocked right-turning movements and of approaches that are wider at the stop line

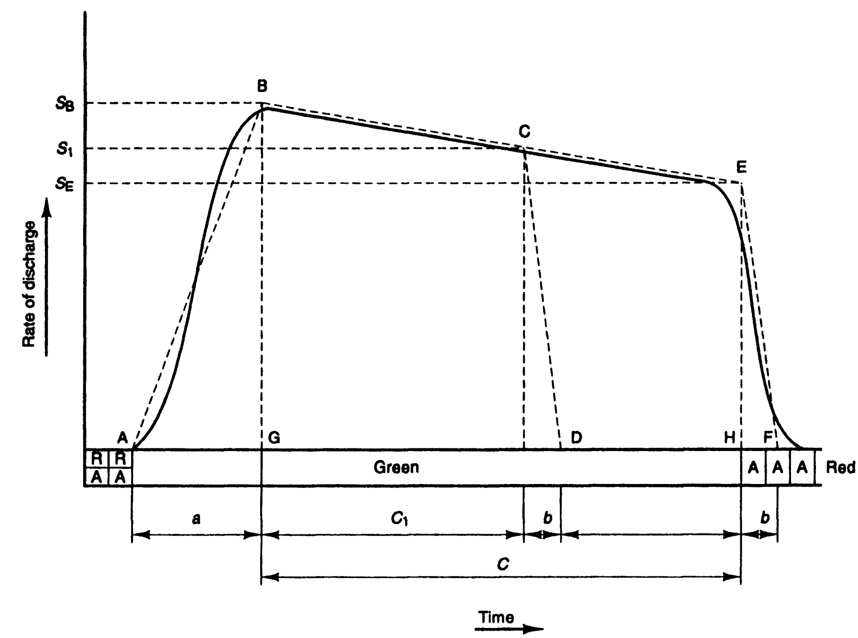

Figure 42.1 *Variation of discharge across a traffic signal stop line where saturation flow decreases during the green period (based on Road Research Technical Paper 56)*

than on the remainder of the approach is to reduce the flow rate over the stop line as the green period proceeds.

In this problem it is the saturation flow just as the amber period commences that is important because it is this value that determines the variation in the number of vehicles crossing the stop line under saturated conditions when there is a small change in green time. It is thus necessary to find the length of green time that corresponds to the saturation flow just as amber begins and that, when substituted into the formula for optimum setting, produces the original green time.

The change in flow rate with time is illustrated for a generalised case in figure 42.1.

The rate of discharge against time curve may be replaced by a quadrilateral with the same area; this is shown by dashed lines. Then the areas ABG and EHF are equal to the areas under the corresponding portions of the curves. The saturation flow at B is S_B and at E is S_E. If the green time ends earlier at C then the saturation flow at the commencement of amber is S_1, and the reduction in flow during the amber period can be denoted by the effective line CD. From figure 42.1

$$S_1 = S_B - \frac{C_1}{C}(S_B - S_E) \tag{42.1}$$

The effective green time g_1 when the green time terminates at C can be calculated from the area of the quadrilateral ABCD. which is

$$\frac{1}{2} S_B\, a + \frac{1}{2}(S_B + S_1)C_1 + \frac{1}{2} S_1\, b$$

From the original definition of effective green time this area is equal to $g_1 S_1$. Hence

$$g_1 = \frac{S_B(a + C_1) + S_1(b + C_1)}{2S_1} \tag{42.2}$$

If G is the combined green plus amber period, the lost time is

$$l_1 = G - g_1 \tag{42.3}$$

The value of g_1 is not necessarily the optimum green time. To obtain a first approximation to the optimum green line which can then be compared with the originally assumed value of g_1, it is assumed that S_1 and l_1, calculated from equations 42.1 and 42.3, are the appropriate values of saturation flow and lost time. Then

$$C_0 = \frac{1.5L + 5}{1 - Y} \tag{42.4}$$

Also when the effective green is divided between the phases in proportion to their respective y values

$$g_1 = \frac{y_1}{Y}(C_0 - L) \tag{42.5}$$

Substituting equation 42.4 in equation 42.5

$$g_1 = \frac{y_1}{Y}\left(\frac{1.5L + 5}{1 - Y} - L\right)$$

$$= \frac{y_1}{Y(1 - Y)}(5 + L(Y + 0.5)) \qquad (42.6)$$

The value of g_1 obtained from equation 42.6 is compared with the value originally obtained from equation 42.2. A new estimate of g_1 may then be made by taking a second approximation midway between the previous two values. The iteration may then be repeated until there is no significant difference between successive values of g_1.

Problem

The discharge/time curve at a traffic-signal stop line for the critical approach of phase (a) may be approximated by the curve shown in figure 42.2.

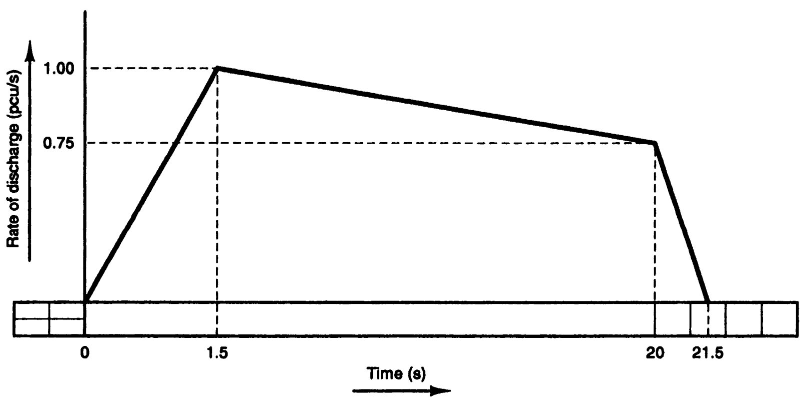

Figure 42.2

On the other approaches the saturation flow is uniform with time. The design-hour flows and saturation flows for the approaches on each phase with the maximum y values are given in table 42.1.

TABLE 42.1

Approach	*Design-hour flow (pcu/h)*	*Saturation flow (pcu/h)*
phase 1	600	shown graphically
phase 2	500	2000

Intergreen period 4 s, start and end lost time on phase 2 is 2 s.

Calculate the optimum cycle time and the corresponding actual green times.

Solution

The iterative procedure to arrive at the optimum cycle time will be illustrated by carrying out the calculation in steps.

(1) Assume the actual green plus amber period on phase 1 is 18 s (point C, figure 42.1)

(2) $S_1 = 1.0 - \dfrac{13.5}{18.5}(1.00 - 0.75)$

$= 0.82$ pcu/s (see equation 42.1)

(3) $g_1 = \dfrac{1.0(1.5 + 13.5) + 0.82(1.3 + 13.5)}{2 \times 0.82}$

$= 16.55$ s (see equation 42.2)

(4) $l_1 = 18 - 16.55$

$= 1.45$ s (see equation 42.3)

(5) $L = l_1 + l_2 +$ lost time in the intergreen periods

$= 1.45 + 2.00 + 1.00 + 1.00$

$= 5.45$ s

(6) $y_1 = \dfrac{q_1}{S_1} = \dfrac{600}{0.82 \times 3600} = 0.30$

$y_2 = \dfrac{500}{2000} = 0.25$

(7) $g_1 = \dfrac{0.30}{0.55(1 - 0.55)} 5 + 5.45(0.55 + 0.5)$

$= 12.86$ s (see equation 42.6)

(8) The first calculation of actual green plus amber period on phase 1 is

12.86 s + 1.45 s = 14.31 s (equation 42.3)

The initial assumption of actual green plus amber period produced a value of $g_1 = 16.55$ s in Step 3 and finally in Step 7 it was calculated as 12.86 s.

A second iteration may now be performed using as the initial assumption that the actual green plus amber period is

$$18 - \frac{16.55 - 12.86}{2} = 16.06 \text{ s.}$$

Successive iterations are tabulated in table 42.2.

$$\text{effective green time phase 1} = 13.52 \text{ s}$$

$$\text{effective green time phase 2} = \frac{13.52 \times 0.25}{0.30}$$

$$= 11.27 \text{ s (see Chapter 37)}$$

$$\text{actual green time phase 1} = 13.52 + 1.86 - 3.00 = 12.38 \text{ s}$$

$$\text{actual green time phase 2} = 11.27 - 1.00 = 10.27 \text{ s (see Chapter 35)}$$

$$\text{cycle time} = 12.38 + 10.27 + 8.00 = 30.65 \text{ s}$$

(see Chapter 37)

TABLE 42.2

Step	*Description*	*Second iteration*	*Third iteration*
1	initial value of green plus amber period	16.06 s	15.72
2	$S_1 = 1.0 - \frac{11.56}{18.5}(1.00 - 0.75)$	0.84 pcu/s	—
3	$g_1 = \frac{1.0(1.5 + 11.56) + 0.84(1.3 + 11.56)}{1.68}$	14.20 s	the values of the green plus amber period in the second and third iterations are sufficiently close to make further iterations unnecessary.
4	$l_1 = 16.06 - 14.20$	1.86 s	
5	$L = l_1 + l_2 +$ intergreen lost time	5.86 s	
6	$y_1 = \frac{600}{0.84 \times 3600}$	0.30	
	$y_2 = \frac{500}{2000}$	0.25	
7	$g_1 = \frac{0.30}{0.55(1 - 0.55)}(5 + 5.86(0.55 + 0.5))$	13.2 s	

43

Delay at signal-controlled intersections

Two approaches are in common use for calculating traffic delays at signal-controlled junctions. The first method, based on steady-state queueing theory, is relevant to situations of relatively constant traffic demand where demand does not exceed about 90 per cent of capacity. More recently, time-dependent queueing theory has been developed, applicable to all traffic states including time-varying demand and 'oversaturation'.

Steady-state queueing and delays

Research by F.V. Webster, reported in *Road Research Technical Paper* 39, used a combination of queueing theory and digital computer simulation. It was shown that the average delay per vehicle on a particular intersection is given by

$$d = \frac{c(1 - \lambda)^2}{2(1 - \lambda x)} + \frac{x^2}{2q(1 - x)} - 0.65\left(\frac{c}{q^2}\right)^{1/3} x^{(2 + 5\lambda)}$$

where d = average delay per vehicle,
c = cycle time,
λ = proportion of the cycle that is effectively green for the phase under consideration (that is, effective green time/cycle time),
q = flow,
s = saturation flow,
x = degree of saturation, which is the ratio of actual flow to the maximum flow that can be passed through the approach (that is $q/\lambda s$).

The relative values of the three terms are illustrated in figure 43.1, reproduced from *Road Research Technical Paper* 39. The first term in this expression is the delay due to a uniform rate of vehicle arrival, the second term is the delay due to the random nature of the vehicle arrivals. The third term was derived from the simulation of traffic flow.

To assist in the calculation of delays on traffic-signal approaches, values of

$$\frac{(1-\lambda)^2}{2(1-\lambda x)} = A$$

$$\frac{x^2}{2(1-x)} = B$$

$$\text{correction term} = C$$

are given in *Road Research Technical Paper* 39 and are tabulated in tables 43.1, 43.2 and 43.3.

It is then possible to calculate the delay from

$$d = cA + \frac{B}{q} - C \tag{43.1}$$

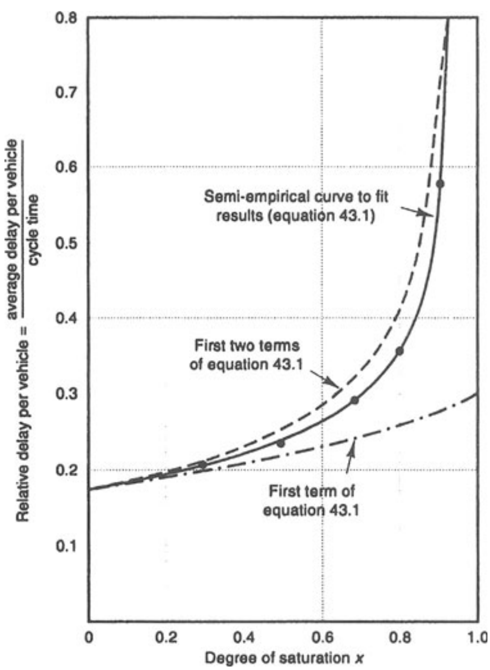

Figure 43.1 *Illustration of delay on traffic-signal approaches (based on Road Research Technical Paper 39)*

TABLE 43.1

Tabulation of $A = \dfrac{(1-\lambda)^2}{2(1-\lambda x)}$

x/λ	0.1	0.2	0.3	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.80	0.90
0.1	0.409	0.327	0.253	0.219	0.188	0.158	0.132	0.107	0.085	0.066	0.048	0.022	0.005
0.2	0.413	0.333	0.261	0.227	0.196	0.166	0.139	0.114	0.091	0.070	0.052	0.024	0.006
0.3	0.418	0.340	0.269	0.236	0.205	0.175	0.147	0.121	0.098	0.076	0.057	0.026	0.007
0.4	0.422	0.348	0.278	0.246	0.214	0.184	0.156	0.130	0.105	0.083	0.063	0.029	0.008
0.5	0.426	0.356	0.288	0.256	0.225	0.195	0.167	0.140	0.114	0.091	0.069	0.033	0.009
0.55	0.429	0.360	0.293	0.262	0.231	0.201	0.172	0.145	0.119	0.095	0.073	0.036	0.010
0.60	0.431	0.364	0.299	0.267	0.237	0.207	0.179	0.151	0.125	0.100	0.078	0.038	0.011
0.65	0.433	0.368	0.304	0.273	0.243	0.214	0.185	0.158	0.131	0.106	0.083	0.042	0.012
0.70	0.435	0.372	0.310	0.280	0.250	0.221	0.192	0.165	0.138	0.112	0.088	0.045	0.014
0.75	0.438	0.376	0.316	0.286	0.257	0.228	0.200	0.172	0.145	0.120	0.095	0.050	0.015
0.80	0.440	0.381	0.322	0.293	0.265	0.236	0.208	0.181	0.154	0.128	0.102	0.056	0.018
0.85	0.443	0.386	0.329	0.301	0.273	0.245	0.217	0.190	0.163	0.137	0.111	0.063	0.021
0.90	0.445	0.390	0.336	0.308	0.281	0.254	0.227	0.200	0.174	0.148	0.122	0.071	0.026
0.92	0.446	0.392	0.338	0.312	0.285	0.258	0.231	0.205	0.179	0.152	0.126	0.076	0.029
0.94	0.447	0.394	0.341	0.315	0.288	0.262	0.236	0.210	0.183	0.157	0.132	0.081	0.032
0.96	0.448	0.396	0.344	0.318	0.292	0.266	0.240	0.215	0.189	0.163	0.137	0.086	0.037
0.98	0.449	0.398	0.347	0.322	0.296	0.271	0.245	0.220	0.194	0.169	0.143	0.093	0.042

TABLE 43.2

Tabulation of $B = \dfrac{x^2}{2(1-x)}$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.1	0.006	0.007	0.008	0.010	0.011	0.013	0.015	0.017	0.020	0.022
0.2	0.025	0.028	0.031	0.034	0.038	0.042	0.046	0.050	0.054	0.059
0.3	0.064	0.070	0.075	0.081	0.088	0.094	0.101	0.109	0.116	0.125
0.4	0.133	0.142	0.152	0.162	0.173	0.184	0.196	0.208	0.222	0.235
0.5	0.250	0.265	0.282	0.299	0.317	0.336	0.356	0.378	0.400	0.425
0.6	0.450	0.477	0.506	0.536	0.569	0.604	0.641	0.680	0.723	0.768
0.7	0.817	0.869	0.926	0.987	1.05	1.13	1.20	1.29	1.38	1.49
0.8	1.60	1.73	1.87	2.03	2.21	2.41	2.64	2.91	3.23	3.60
0.9	4.05	4.60	5.28	6.18	7.36	9.03	11.5	15.7	24.0	49.0

TABLE 43.3
Correction term of equation 43.1 as a percentage of the first two terms

x	λ/M*	2.5	5	10	20	40
0.3	0.2	2	2	1	1	0
	0.4	2	1	1	0	0
	0.6	0	0	0	0	0
	0.8	0	0	0	0	0
0.4	0.2	6	4	3	2	1
	0.4	3	2	2	1	1
	0.6	2	2	1	1	0
	0.8	2	1	1	1	1
0.5	0.2	10	7	5	3	2
	0.4	6	5	4	2	1
	0.6	6	4	3	2	2
	0.8	3	4	3	3	2
0.6	0.2	14	11	8	5	3
	0.4	11	9	7	4	3
	0.6	9	8	6	5	3
	0.8	7	8	8	7	5
0.7	0.2	18	14	11	7	5
	0.4	15	13	10	7	5
	0.6	13	12	10	8	6
	0.8	11	12	13	12	10
0.8	0.2	18	17	13	10	7
	0.4	16	15	13	10	8
	0.6	15	15	14	12	9
	0.8	14	15	17	17	15
0.9	0.2	13	14	13	11	8
	0.4	12	13	13	11	9
	0.6	12	13	14	14	12
	0.8	13	13	16	17	17
0.95	0.2	8	9	9	9	8
	0.4	7	9	9	10	9
	0.6	7	9	10	11	10
	0.8	7	9	10	12	13
0.975	0.2	8	9	10	9	8
	0.4	8	9	10	10	9
	0.6	8	9	11	12	11
	0.8	8	10	12	13	14

*M is the average flow per cycle = q_c.

Time-dependent queueing theory

When demand reaches capacity, the formulae described above cannot be used, because an infinite delay is predicted, as illustrated in figure 43.2. This figure shows an alternative deterministic queueing model for situations where demand exceeds capacity, in which queue length is directly proportional to the excess of demand over capacity. However, deterministic queueing models have been shown to underestimate queues where demand is at or around capacity, as no account is taken of the random elements in the arrival and discharge processes. The transformed curve, produced from time-dependent queueing theory shown in figure 43.2, has been shown to predict queues most accurately over a range of conditions, and is incorporated into models such as COBA and OSCADY. This generalised queueing theory is applicable to all junction types, with some 'calibration' parameters to reflect some operational differences between the junction types.

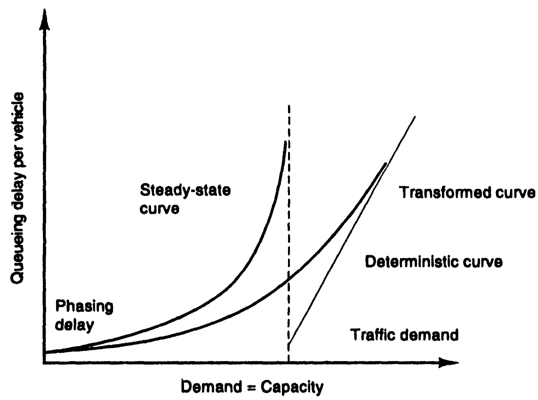

Figure 43.2 *Queueing delay curves*

COBA uses 'low definition' queueing theory with flow/delay relationships based on a co-ordinate transformation technique which smoothes the steady-state relationship for vehicle delays into the over-capacity deterministic results (see figure 43.2). The queueing delay formulae in COBA are divided into two components:

(1) For steady-state queuing,

$$\text{Delay} = \frac{Rq}{\mu(1-\rho)} + L \tag{43.2}$$

where ρ = traffic intensity,
q = demand flow,
μ = capacity,
R = randomness factor (sometimes termed C),
L = low flow queueing delays = $C/2\ (1-\lambda)^2$,
where C = cycle time,
λ = green time/cycle time.

(2) For time-dependent queueing

$$\text{Delay} = 1/2\ [(F^2 + G)^{1/2} - F] + e + L$$

where

$$F = \frac{1}{\mu_0 - q_0}\left\{1/2T\,(\mu - q)\left(1 - \frac{h}{q}\right) + 2R\left[1 - h\left(\frac{1}{q} + \frac{1}{\mu}\right)\right]\right\} + e$$

$$G = \frac{2T}{\mu_0 - q_0}\left[2R\left(\frac{q}{\mu}\right) - (\mu - q)\,e\right]\left(1 - \frac{h}{q}\right)$$

$$h = \mu - \mu_0 + q_0$$

$$e = \frac{2Rq_0}{\mu_0(\mu_0 - q_0)}$$

where q_0 = demand for the preceding time interval,
μ_0 = capacity for the preceding time interval,
T = length of the modelled period.
Other parameters are as above.

At give-way junctions it is usually assumed that the vehicle arrival and departure processes are random, and R is taken as 1.0. At signalised junctions, however, the cyclic operation causes traffic to be released in platoons with substantial gaps between. Within each platoon, vehicles are spaced almost regularly. For this case, R is not equal to unity and a value of 0.5 has been found to be most appropriate.

A more detailed ('high definition') queueing theory is applied in OSCADY, to cater for demand variability within a peak period. This is described further in Chapter 45.

Problems

(1) An intersection controlled by two-phase traffic signals has the design-hour flows and saturation flows tabulated in table 43.4.

TABLE 43.4

Approach	*Flow (vehicles/h)*	*Actual green (s)*	*Saturation flow (vehicles/h)*
north	2000	80	3350
south	1750	80	3350
east	660	32	2750
west	750	32	2750

Lost time per green period 2 s; intergreen period 4 s; cycle time 120 s.

For traffic flow on the north approach, is the delay element due to the random arrival of vehicles greater than the element due to the regular arrival of vehicles?

For the traffic flow on the east approach, is the average delay per vehicle approximately 3.0 s, 60 s or 9.0 s?

(2) An approach to a signalised junction has the following time-varying traffic characteristics. Using the queueing delay formulae in COBA, calculate the average traffic delay in the two time periods.

	Time period 1 (0700–0800 hrs)	*Time period 2* (0800–0900 hrs)
Traffic classification	off-peak	peak
Demand flow, q (veh/s)	0.4	0.75
Capacity, μ (veh/s)	0.8	0.8
Traffic intensity, p	0.5	0.94
Green time (s)	40	65
Cycle time (s)	60	90

Solutions

(1) First calculate the parameters of equation 43.1. For the north approach

$c = 120$ s

λ = effective green/cycle time

$= (80 + 1)/120$

$= 0.675$

$q = 2000/3600$

$= 0.556$ vehicles/s

$s = 3350/3600$

$= 0.931$ vehicles/s

$x = q/\lambda s$

$= 0.556/(0.675 \times 0.931)$

$= 0.885$

The delay element due to the uniform rate of arrival of vehicles is given by the first term of equation 43.1, that is cA.

Tabulated values of A are given in table 43.1 and referring to this table and using the approximate values $\lambda = 0.70$, $x = 0.90$

$A = 0.122$

and

$d = 120 \times 0.122 = 14.64$ s

The delay element due to the random rate of arrival of vehicles is given by the second term of equation 43.1, that is B/q. Tabulated values of B are given in table 43.2. Referring to this table and using the approximate value for x of 0.89

$B = 3.60$

Then

$d = 3.60/0.556$ s

$= 6.47$ s

Note. The delay element due to the uniform arrival of vehicles (14.64 s) is greater than the delay element due to the random arrival of vehicles (6.47 s).

First calculate the parameters of equation 43.1 for the east approach

$$
\begin{aligned}
c &= 120 \text{ s}\\
\lambda &= (32 + 1)/120\\
&= 0.275\\
q &= 660/3600\\
&= 0.183 \text{ vehicle/s}\\
s &= 2750/3600\\
&= 0.764 \text{ vehicle/s}\\
x &= 0.183/(0.275 \times 0.764)\\
&\doteq 0.871\\
M &= qc\\
&= 0.183 \times 120\\
&= 21.96 \text{ vehicles / cycle}
\end{aligned}
$$

From equation 43.1

$$d = cA + \frac{B}{q} - C$$

From table 43.1, $A = 0.329$.
From table 43.2, $B = 2.91$.
From table 43.3, $C = 11$ per cent of first two terms
and

$$d = 49.3 \text{ s/vehicle}$$

(2) (i) For off-peak period, use steady-state queueing formula (equation 43.2)

$$\text{Delay} = \frac{0.5 \times 0.4}{0.8\,(1 - 0.5)} + \frac{60}{2}\left(1 - \frac{40}{60}\right)^2$$

$$= 0.5 + 3.3 = 3.8 \text{ s/vehicle}$$

(ii) For peak period, use time-dependent queueing equations

$$h = 0.8 - 0.8 + 0.4 = 0.4$$

$$e = \frac{2 \times 0.5 \times 0.4}{0.8\,(0.8 - 0.4)} = 1.25$$

$$F = \frac{1}{0.8-0.4}\left\{1/2 \times 3600 \times 0.05 \times 0.47 + 1\left[1-0.4\left(\frac{1}{0.75}+\frac{1}{0.8}\right)\right]\right\} + 1.25$$

Therefore

$$F = \frac{1}{0.4}\,[42.3 + (1 - 1.03)] + 1.25$$

$$= 107$$

$$G = \frac{7200}{0.8 - 0.4}\left[1 \times \frac{0.75}{0.8} - (0.8 - 0.75) \times 1.25\right]\left[1 - \frac{0.4}{0.75}\right]$$

Hence

$$G = 18\,000\,(0.938 - 0.063) \times 0.467$$

$$= 7355$$

$$L = \frac{90}{2}\left(1 - \frac{65}{90}\right)^2 = 3.5$$

Therefore

$$\text{Delay, } D = 1/2\,[(107^2 + 7355)^{1/2} - 107] + 1.25 + 3.5$$

$$= 15.06 + 1.25 + 3.5 = 19.81 \text{ s/vehicle}$$

44

Average queue lengths at the commencement of the green period

In the design of traffic signals it is often desirable to be able to estimate the queue length at the beginning of the green period. The queue length at this period of the cycle will normally be the greatest experienced because during the green period the queue is being discharged at a rate which for practical purposes must be greater than the flow on the approach.

Of the several cases that may be considered, the simplest one is the unsaturated approach where the queue at the beginning of the green period disappears before the signal aspect changes to amber. In this case the maximum queue is simply the number of vehicles that have arrived during the preceding effective red period, that is

$$N_u = qr \tag{44.1}$$

where N_u is the initial queue at the beginning of an unsaturated green period,
q is the flow,
r is the length of the effective red period; the effective red period is equal to the cycle time minus the effective green period.

The next simplest case is when the approach is fully saturated. The queue length will have gradual variations over the interval considered, on which are superimposed sudden increases and decreases caused by the red and green periods. The range of the short variations will be qr and the average queue over the whole of the interval considered (assuming no great variation between beginning and end) is equal to the product of the flow and the average delay per vehicle.

The average queue at the beginning of the green period is thus equal to the average queue throughout the interval plus half the average range of the cyclic fluctuations; that is

$$N_s = qd + \frac{1}{2} qr$$

where N_s is the initial queue at the beginning of a saturated green period,
d is the average delay per vehicle on a single approach.

Note that when the approach is only just saturated then d is $qr/2$ and $N_s = qr$, the same as for an unsaturated approach. The value of N_s cannot be less than qr so that

$$N_s = qd + \frac{1}{2}qr \quad \text{or} \quad qr \tag{44.2}$$

whichever is the greater.

In most practical cases intersections will operate under a mixture of saturated and unsaturated conditions. The initial queue at the commencement of the green period for these conditions will now be discussed.

Assume there are n fully saturated cycles and m unsaturated cycles. The flow is q and the average delay per vehicle d over $(m + n)$ cycles. These are composed of flow q_s and delay d_s for n cycles and flow q_u with delay d_u for m cycles.

From the previous discussion the average queue at the beginning of the green period for the n saturated cycles will be

$$\frac{q_s r}{2} + q_s d_s$$

Similarly for the m unsaturated cycles. The average queue at the beginning of the green period for the m unsaturated cycles will be

$$q_u r$$

Therefore the average queue at the beginning of the green period over $(m + n)$ cycles will be

$$N = \frac{1}{(m + n)}(nq_s r/2 + nq_s d_s + mq_u r)$$

The average flow is given by

$$q = \frac{nq_s + mq_u}{(n + m)}$$

so that

$$N = qr/2 + \frac{nq_s d_s + mq_u r/2}{n + m}$$

The second expression on the right-hand side presents difficulties in evaluating a solution so it must be converted into a more easily manipulated form.

Considering the unsaturated cycles, the net rate of discharge of the queue is $s - q_u$, the average queue at the beginning of the green period is N and the time taken for this initial average queue to disperse is $N/(s - q_u)$. The average queue throughout the cycle is

$$\frac{N}{2c}\left(\frac{N}{s - q_u} + r\right)$$

as illustrated in figure 44.1.

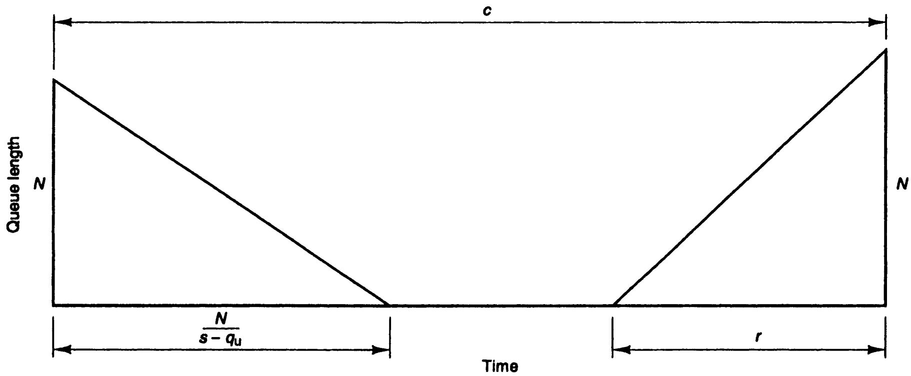

Figure 44.1 *Variation of queue with time when the approach is operating under a mixture of saturated and unsaturated conditions*

$$\text{Average delay} = \text{average queue} \times \text{cycle time/number of vehicles per cycle}$$

$$d_u = \text{average queue}/q_u$$

$$d_u = \frac{r}{2c}\left(\frac{rs}{c(s - q_u)}\right)$$

This is approximately $r/2$, an approximation that underestimates delay, But d_u is very much smaller than d_s and m is also much smaller than n, so that

$$N = \frac{qr}{2} + \frac{nq_s d_s + mq_u d_u}{n + m}$$

$$N = \frac{qr}{2} + qd$$

We know also that the average queue at the beginning of the green period cannot be less than qr so that the value of N is usually taken as the greater of these two values.

This theoretical approach neglects the finite extent of the queue and in practice vehicles join the queue earlier and so have an increased delay. This increase can be calculated from

length of queue when the green period begins $= Nj/a$

where a = number of lanes in queue,

j = average spacing of vehicles in the queue.

Therefore

$t = Nj/av$

where v = free running speed of the traffic,

t = time for a vehicle to travel the length of the queue at the running speed.

The number of vehicles that arrive in this time is

$$\frac{qNj}{av}$$

and this correction should be added to the previously derived values of N giving

$$N = q\left(\frac{r}{2} + d\right)\left(1 + \frac{qj}{av}\right) \quad \text{or} \quad qr\left(1 + \frac{qj}{av}\right) \tag{44.3}$$

whishever is the greater.

Problems

The flows and saturation flows, given in Part A below, are observed on traffic-signal approaches. Select from Part B the appropriate value of the queue at the beginning of the green period for each of the approaches described in Part A. In all cases the intergreen period is 4 s, starting delays may be taken as 2 s of each green period and the spacing of queued vehicles may be taken as 6 m. Average speed on the approach is 6 m/s.

Part A

(a) A two-lane approach with a width of 7.65 m carries a flow of 1000 pcu/h; the cycle time is 100 s and the actual green time 40 s. On this approach it may be assumed that vehicles arrive with a uniform headway distribution and that the saturation flow is 4015 pcu/h. An average vehicle is equal to 1.2 pcu.

(b A single right-turning lane with a saturation flow of 1600 pcu/h carries a flow of 800 pcu/h; the actual green time is 60 s and the cycle time is 100 s. An average vehicle is equal to 1.3 pcu.

Part B

(a) Approximately 8 vehicles.
(b) Approximately 14 vehicles.
(c) Approximately 20 vehicles.

Solutions

(a) The saturation flow on this approach is 4015 pcu/h.

$525 \times w$ pcu/h $= 525 \times 7.65$ (see Chapter 33)

$= 4015$ pcu/h

The capacity of the approach is given by

$$\frac{\text{effective green time} \times \text{saturation flow}}{\text{cycle time}} \text{ pcu/h}$$

and

effective green time = actual green time + amber period − lost time (see Chapter 37)

The capacity of the approach

$$= \frac{41 \times 4015}{100}$$
$$= 1646 \text{ pcu/h}$$

The flow on the approach is 1000 pcu/h.

The flow on the approach is less than the capacity and because the vehicles arrive at regular intervals the maximum queue at the commencement of the green period is equal to the number of vehicles arriving during the preceding red and red-plus-amber periods.

The average queue at the commencement of the green period is then

$$N_u = qr \qquad \text{(see equation 44.1)}$$
$$= \frac{1000 \times (100 - 41)}{3600} \text{ pcu}$$
$$= 16.4 \text{ pcu}$$
$$= 16.4/1.2 \text{ vehicles}$$
$$= 13.6 \text{ vehicles}$$

(b) The capacity of the approach $= \dfrac{61 \times 1600}{100}$ pcu/h

$= 976$ pcu/h

The flow on the approach $= 800$ pcu/h

The approach is nearing saturation so that the average number of vehicles waiting on the approach at the commencement of the green period is given by equation 44.3

$$N = q\left(\frac{r}{2} + d\right)\left(1 + \frac{qj}{av}\right) \text{or} \quad qr\left(1 + \frac{qj}{av}\right)$$

whichever is the greater.

The average delay on the approach d can be calculated using the method described in Chapter 43

$$d = cA + \frac{B}{q} - C \quad \text{(equation 43.1)}$$

and

$\lambda = 0.61$

$x = 800/976 = 0.82$

$$M = \frac{800 \times 100}{3600 \times 1.3} \text{ vehicles/cycle}$$
$$= 17.1 \text{ vehicles/cycle}$$

By interpolation from table 43.1

$$A = 0.153$$

By interpolation from table 43.2

$$B = 1.87$$

By interpolation from table 43.3

$$C = 13 \text{ per cent}$$

$$d = 100 \times 0.153 + \frac{1.87}{0.17} - 13 \text{ per cent of previous two terms}$$

$$= 22.9 \text{ s}$$

Then

$$N = 0.17\left(\frac{39}{2} + 22.9\right)\left(1 + \frac{0.17 \times 6}{1 \times 6}\right)$$

$$\text{or } 0.17 \times 39\left(1 + \frac{0.17 \times 6}{1 \times 6}\right) \qquad \text{(equation 44.3)}$$

$$= 6.2 \quad \text{or} \quad 7.8$$

that is

$$N = 7.8 \text{ vehicles}$$

45

Programs for traffic signal design

A number of computer programs have become available in recent years as an aid to the design of signal-controlled junctions. These can be used to help determine the optimum geometric design of the junction, calculate appropriate signal settings for control purposes and predict queues and delays at the junction. Such programs are widely used and are a valuable aid to the traffic engineer, but do not remove the need for experienced design engineers. For example, the choice of stage sequencing, phase combinations, geometric details and so on still require the skill of the designer; the signal programs then provide a powerful tool for evaluating different alternatives and for optimisation. Other issues outside the scope of signal programs, but vital in the design process, include:

(1) 'designing for safety' (for example, correct intergreen/minimum green settings and minimisation of conflicts);
(2) adequate provision for all road users, including heavy vehicles (for example minimum turning circles), buses (for example, bus stops, priority facilities) and cyclists/pedestrians (for example, special protection/signalling facilities);
(3) other design issues, such as junction location, signing, lighting and environmental impact.

For isolated signal-controlled junctions, the main design programs in use in the United Kingdom are OSCADY[1], LINSIG[2] and SIGSIGN[3]. A program known as SIDRA, produced by the Australian Road Research Board, is also becoming widely used in some countries. The main program used for analysing linked signals is TRANSYT, described in Chapter 51. A number of other network models also contain signal modelling capabilities, notably SATURN, CONTRAM and TRAFFICQ. SATURN and CONTRAM can be particularly useful in that, being assignment models, they can be used to predict the effects of signal improvements on drivers' route choice and hence link flows, for feedback into the junction design process.

OSCADY

OSCADY[1] (Optimised Signal Capacity and Delay) is a computer program developed by the Transport and Road Research Laboratory as an aid to the analysis and design of isolated signal-controlled junctions. It is a companion program to ARCADY2 and PICADY2, which model roundabouts and priority intersections respectively. OSCADY is a stage-based model in which the stage sequence is input by the user. It uses high-definition time-dependent queueing theory for the prediction of queues and delays in periods of varying demand and includes signal optimisation routines developed at University College, London and incorporated in the SIGSET model developed there. These routines produce a theoretical minimum overall delay solution when the junction has spare capacity and overload minimising settings when the junction is overloaded. The signal optimisation process is illustrated in figure 45.1. OSCADY makes use of flow/delay relationships rather than microscopically modelling individual vehicle movements. It does not therefore explicitly model the operation of vehicle-actuated signals although, in peak periods, which are usually of most interest, OSCADY can reasonably mimic vehicle actuation by allowing signal optimisation in small time-intervals.

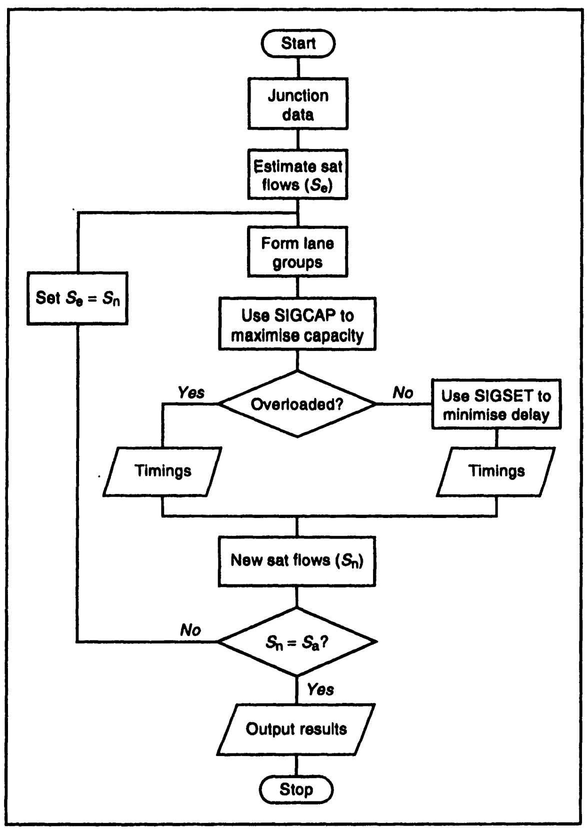

Figure 45.1 *The OSCADY optimisation process (ref. 1)*

Input requirements for OSCADY start with general information such as junction layout (number of arms, etc.), modelling time period and time segment length. Geometric data is then required for each approach, covering those items which affect capacity (see Chapter 33). Alternatively measured, site-specific saturation flows can be used if available. Where appropriate OSCADY combines lanes with similar discharge characteristics and staging into lane groups, with traffic allocated to lanes so that the degree of saturation is equalised between such lanes. Lanes with dedicated or mixed turning movement traffic subject to opposing flows are also modelled using the formulae given in Chapter 33.

Signal information input to OSCADY includes the specification of stage sequence, minimum or fixed green and amber times, maximum or fixed cycle times, lost times and intergreen times. OSCADY can optimise cycle times and green times or green times only. Fixed signal timings can also be supplied if queueing characteristics are of most interest.

Traffic demand data, including turning movements and vehicle type proportions, can be specified in a variety of ways, ranging from detailed data for each time segment, to average values for the modelled period, from which OSCADY estimates a suitable peak profile.

OSCADY outputs include, for each time segment, the demand and capacity information, the queue and delay predictions and, if required, the geometric delay. This latter calculation, based on recent research carried out by the Transport and Road Research Laboratory, is similar in concept to that described for other junctions in Chapter 24. A specification of the optimised signal timings is also given. Summary tables give overall delay for each traffic movement, and for the whole junction, for the completed modelled period.

The latest version of OSCADY, version 3, includes improved data specification and modelling for flared approaches and new facilities for predicting accident frequencies. These predictions are either junction based, using a 'simple' flow-based model combined with other junction and control details, or disaggregated for the different streams, for which a much more detailed model is used. These facilities should significantly aid designers when evaluating junction alternatives in terms of capacity, delay and safety.

LINSIG

LINSIG[2] is a phase-based computer program developed to help with the design and specification of traffic signal installations and enable an assessment to be made of traffic performance. A key feature of LINSIG is its ability to model junction controllers as realistically as possible, providing a computerised method of checking many items when specifying microprocessor controllers. For this reason, LINSIG has become increasingly popular with engineers for designing and writing signal specifications.

LINSIG uses traffic links which are defined as movements of traffic either from a single lane or a combination of lanes similar to the grouping of traffic lanes when calculating y values. All traffic links are related to traffic phases, with the phases then grouped into stages, this process having the same definition as in the signal controller specification (for example, MCE 0141). Hence the model exactly reproduces stage sequencing, minimum constraints and intergreens.

The input requirements for LINSIG start with a junction layout and geometric data, sufficient for establishing intergreen times and saturation flows, using the formulae given in Chapter 33. Alternatively, observed saturation flows may be used. Detailed signal data specific to the controller is required, including phase minimums, intergreens, phase delays, allocation of phases to stages and permitted moves. Three sets of traffic flows are available for each traffic link, for which the time period for each set is variable. Outputs are via tables and figures showing all the results for links, stages and phases.

LINSIG incorporates two optimisation procedures: (i) optimisation of reserve capacity for a given cycle time and (ii) optimisation of cycle time for a given reserve capacity. In both cases, stage sequence is set by the user and it is possible to run a stage more than once to enable split cycling or alternate pedestrian demands to be investigated. In optimisation option (i), the maximum degree of saturation on each stage is equalised, giving timings which may differ from delay minimising timings. In optimisation option (ii), the lowest cycle time for a specified reserve capacity is calculated.

Other facilities in LINSIG include the ability to cater for flared approaches and for right-turning traffic stored in the middle of the junction, which discharge through gap acceptance or clearance in the intergreen period. LINSIG is gradually being enhanced with new facilities and companion programs. For example, one program is for interstage design, allowing maximum use of phase delays in calculating the most efficient way of closing and starting phases to maximise traffic green time.

SIGSIGN

SIGSIGN[3] is a phase-based optimisation program for isolated signal-controlled junctions, developed at University College, London and distributed by the Consultants Steer Davies Gleave. SIGSIGN provides signal settings which are optimal for a given pattern of traffic according to the criteria of (i) critical cycle time, (ii) maximum practical capacity and (iii) minimum total junction delay. This latter facility is unique for a phase-based program. Calculations are based on fixed-time analysis, although the results of optimisations (i) and (ii) are stated to apply to both fixed-time and vehicle-actuated operation, whether the junction is isolated or in a linked system. Results of optimisation (iii) apply only to isolated junctions. The calculated green times provide either settings for fixed-time operations or maxima for vehicle-actuated operation. Signal optimisation uses mathematical programming techniques developed over several years at University College, London. As with LINSIG, SIGSIGN allows full use to be made of existing phase-based microprocessor controllers.

The phase-based modelling in SIGSIGN is particularly useful for analysing complex junctions where the optimal stage sequence and interstage structure are not apparent, and can provide benefits in increased capacity and reduced delay. Flexibility is particularly apparent in phase-to-phase intergreens, which can be set directly in the signal controller without having to be grouped into stages. At each stage change, the start and end of each phase can be staggered, having regard to the intergreen time required between conflicting phases. This can lead to more efficient junction control compared with constraining phases to start and end simultaneously according to the stage change.

SIGSIGN incorporates a sophisticated user interface for the input of data, sensible default values and data checking facilities. The input is all phase-based (phase intergreen matrix, minimum phase durations, phase constraints, etc.). Additional facilities include the provision for streams to receive two periods of green in a cycle, allowing, for example, double green periods at staggered junctions. Opposed turning streams can also be included, providing optimal timings for both 'early cut offs' and 'late starts', depending on the input stage sequence and taking account of gap-seeking.

References

1. I. J. Burrow, OSCADY: A Computer Program to Model Capacities, Queues and Delays at Isolated Traffic Signal Junctions, *Transport and Road Research Laboratory Research Report* 105, Crowthorne (1987)
2. B. Summonite, Traffic Signal Design Using LINSIG, *PTRC Course on Programs for Traffic Signal Design*, London (1993)
3. J. P. Silcock and A. Sang, SIGSIGN: A Phase Based Optimisation Program for Individual Signal Controlled Junctions, *Traffic Engineering and Control*, May 1990

46

The co-ordination of traffic signals

When several traffic signal-controlled intersections occur along a major traffic route, some form of co-ordination is necessary to prevent, so far as is possible, major road vehicles stopping at every intersection. Alternatively, or in addition, the signals may be co-ordinated to minimise delays to vehicles. Sometimes linking between signals is carried out to prevent queues stretching back from one intersection to the preceding signals.

Where only two sets of signals are linked, co-ordination is often achieved by calculating offsets according to the cruise time of vehicles from one intersection to the next. The offset specifies the difference in the start of the green time at the junctions for the phases to be co-ordinated. However, offsets are most commonly calculated using the TRANSYT program described in Chapter 51, particularly where a network of signals makes manual calculations impractical.

Three examples of signal linking which are possible are the simultaneous system, the alternate system and the flexible progressive system. In these systems co-ordination between intersections is usually achieved by means of a master controller.

In the simultaneous system all signals along the co-ordinated length of highway display the same aspect to the same traffic stream at the same time. Some local control is possible using vehicle actuation but a master controller keeps all the local controllers in step and imposes a common cycle time. An obvious disadvantage of this control system is that drivers are presented with several signals each with a green aspect, and there is a tendency to travel at excessive speed so as to pass as many signals as possible before they all change to red. Where turning traffic is light and intersections are closely spaced, then this system may have advantages for pedestrian movement.

The alternate system allows signal installations along a given length of road to show contrary indications. This means that if a vehicle travels the distance between intersections in half the cycle time, then a driver need not stop. The cycle time must be common to all the signals and must be related to the speed of progression. Major roads with unequal distances between intersections present difficulties for this reason.

With the flexible progressive system the green periods at adjacent intersections are offset relative to each other according to the desired speed on the highway. Progression along the highway in both directions must be considered and this usually results in a compromise between the flow in both directions and also between major and minor road flows.

A master controller keeps the local controllers, which may be either fixed-time or vehicle-actuated, in step. Vehicle-actuated control is possible when there is no longer a continuous demand from all detectors, the signals changing in accordance with traffic arriving at the isolated intersection.

Any change of right of way must however not interfere with the progressive plan so that at certain periods a change of right of way cannot take place because there would be insufficient time in which to regain right of way as required by the progressive plan.

Under very light traffic conditions, the flexible progressive system is likely to produce greater delays than an unlinked system because of the overriding priority of a small number of major road vehicles. In some circumstances traffic density measuring devices bring the master controller into action to impose an overall flexible progressive system as traffic volume increases, while at lower volumes the signals at each intersection act in an independent manner.

At key intersections vehicle actuation is often allowed and the control of the preceding or following signals linked to the operation of the key intersection.

Where traffic leaving the key intersection is assisted to pass through the next intersection then this is referred to as forward linking. If however traffic arriving at the key intersection on a particular approach is given preference this is referred to as backward linking. Forward linking prevents a key intersection from becoming blocked while backward linking will prevent a queue stretching back from the key intersection causing it to be blocked.

Problem

A major traffic route through a central city area has frequent regularly spaced traffic signal intersections. It is proposed to co-ordinate these signals so as to minimise delay to all vehicles passing through the intersections. Arrange the following control systems in order of their suitability for the above situation.

(a) A flexible progressive control system with fixed-time operation at each intersection.
(b) A simultaneous-control system with a major road green time of 90 s and a cycle time of 120 s.
(c) An alternate-control system with vehicle actuation at each intersection.
(d) A control system that imposes flexible progression during periods of heavier traffic flow and allows isolated-vehicle actuation at other times.

Solution

The order of merit of these control systems is:

(d) This system would allow progression during periods of heavier flow but at other times would allow the intersections to function under isolated vehicle-actuated control to deal with fluctuations in the traffic flow.

(c) This system would allow progression provided the cycle time was correctly chosen and at periods of low demand vehicle actuation would allow flexibility.

(a) This system would offer progression at one level of flow but would be inflexible under changing traffic conditions.

(b) This system would not form a satisfactory method of control being unable to meet changing traffic demands and being likely to produce long delays at low traffic flows.

47

Time and distance diagrams for linked traffic signals

When the flexible progressive system of co-ordinating traffic signals is employed, then it is frequently desirable to construct a time-and-distance diagram to estimate the best offset or difference in the start of the green time, between adjacent signals.

The time-and-distance diagram is simply a graph on which time and hence the signal settings are plotted horizontally.

Distance along the major route between intersections is plotted vertically. The slope of any line plotted on this diagram represents the speed of progression along the major route.

Before the diagram can be prepared it is necessary to examine each of the intersections which it is desired to co-ordinate and calculate the optimum cycle times for the expected traffic flows using the relationship

$$C_0 = \frac{1.5L + 5}{1 - Y} \qquad \text{(see page 304)}$$

From these calculations it is possible to determine the intersection that is most heavily loaded and that therefore requires the longest cycle time. This intersection is then referred to as the key intersection and because in any linked system it is necessary for each cycle time to be the same, or a multiple of the same value, the cycle time of the key intersection is adopted as the cycle time for the whole system.

As traffic flows vary throughout the day and also throughout the week, it may be necessary to employ differing key intersection and common cycle times and hence differing progression plans for differing days and times.

With the cycle time c_1 for the system determined, it is possible to calculate the effective green times for the differing phases at the key intersection from the relationship

$$\text{effective } g_1 = \frac{y_1}{Y}(c_1 - L) \qquad \text{(see Chapter 37)}$$

The actual green time may then be determined if the lost time within the green period is known (see Chapter 35).

The actual green time at the key intersection gives the minimum actual green time at the other intersections for the main route with progression.

To obtain the maximum green times for the intersections other than the key intersection it is necessary to determine the shortest acceptable green times for side road phases. The shortest acceptable effective green time for a side road phase is obtained from

$$\frac{y_{\text{side}} \times c_1}{0.9}$$

and from a knowledge of the lost time in the green period the shortest acceptable actual green time can once again be calculated.

The longest actual green for the route with progression is then the linked cycle time c_1 minus the shortest acceptable actual green minus the intergreen periods.

With the calculated value of green times at the key intersection and the maximum and minimum green times at the other intersections known, it is possible to arrange the relative timing or offsets of the signals. In the preparation of this dia-

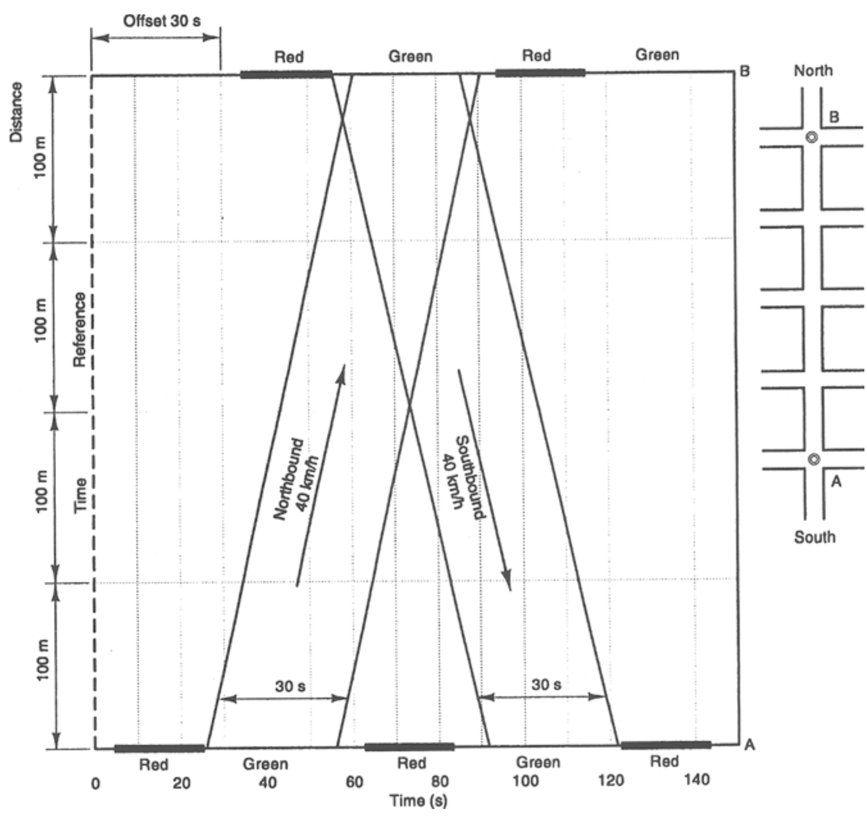

Figure 47.1 *A time/distance diagram showing how a time offset is arranged between intersections A and B to give progression through two signals*

gram the speed of progression between intersections should be chosen taking into account the known or likely speed/volume relationship for the highway and such physical features as horizontal and vertical curvature and pedestrian activity. Turning movements and critical queue lengths also should be considered in the arrangement of the progression through the intersection. A typical time/distance diagram is shown in figure 47.1.

Problem

Four two-phase traffic signal-controlled intersections along a major north/south traffic route are spaced at distances of 0.5 km apart. Details of the evening peak hour traffic flows at these intersections are given in table 47.1. Lost times in all cases may be taken as 2 s of each green period.

TABLE 47.1

Intersection	*Approach*	*Flow*	*Saturation flow*	*Lost time per cycle (s)*
A	north	1250	4015	6
	south	1450	4015	
	east	1000	2250	
	west	800	1950	
B	north	1350	4015	8
	south	1550	4015	
	cast	1200	2700	
	west	650	2250	
C	north	1100	4015	8
	south	1500	4015	
	east	900	2250	
	west	550	2250	
D	north	1300	4015	8
	south	1400	4015	
	east	1000	2700	
	west	600	1950	

Prepare a time and distance diagram showing how the offsets for these signals may be arranged to produce progression for major road vehicles.

Solution

It is first necessary to calculate the optimum cycle time for each intersection, so that the maximum value can be found and adopted for the linked system (table 47.2). The longest optimum cycle time is required at the intersection B and this is the key intersection. The cycle time for the linked system is therefore 100.0 s.

The effective green time for the north/south traffic stream at intersection B is given by

$$\frac{yN/S}{Y}(c_1 - L) = \frac{0.39}{0.83}(100.0 - 8.00)$$
$$= 43.2 \text{ s}$$

and actual green time = effective green time + lost time per green period − amber period

= effective green time + 2 − 3

= 42.2 s

This is also the minimum actual green time for the north/south traffic stream at the remaining intersections.

TABLE 47.2

Intersection	*Approach*	*y value*	C_0
A	north	0.31	
	south	0.36	$\frac{1.5 \times 6 + 5}{1 - (0.36 + 0.44)} = 70.0$ s
	east	0.44	
	west	0.41	
B	north	0.34	
	south	0.39	$\frac{1.5 \times 8 + 5}{1 - (0.39 + 0.44)} = 100.0$ s
	east	0.44	
	west	0.29	
C	north	0.27	
	south	0.37	$\frac{1.5 \times 8 + 5}{1 - (0.37 + 0.40)} = 73.9$ s
	east	0.40	
	west	0.25	
D	north	0.32	
	south	0.35	$\frac{1.5 \times 8 + 5}{1 - (0.35 + 0.37)} = 60.7$ s
	east	0.37	
	west	0.31	

The maximum actual green times at the remaining intersections are then calculated from a consideration of the minimum effective green time required for side road traffic using the relationship

$$\frac{y_{\text{side}} \times c_1}{0.9}$$

With minimum and maximum actual green times determined it is possible to plot a time-and-distance diagram by a trial-and-error process. As the intersections are regularly spaced, it is not too difficult to arrange for progression of the major road traffic streams in both directions. The diagram allows for progression, so that vehicles travel between intersections in half a cycle giving a speed of 36 km/h. The timing diagram for this very simple situation is shown in figure 47.2.

(1) Intersection	(2) $y_{side\ max}$	(3) Minimum effective green on minor route $\frac{y_{side} \times c_1}{0.9}$ (s)	(4) Minimum actual green on minor route (3) − 1 (s)	(5) Maximum actual green on major route c_1 − (40) − intergreen time* (s)
A	0.44	48.9	47.9	44.1
C	0.40	44.4	43.4	46.6
D	0.37	41.1	40.1	49.9

$$\text{*Intergreen time} = \frac{(\text{lost time} - \text{starting delays} \times 2)}{2} + 3\ \text{s per phase (see Chapter 31).}$$

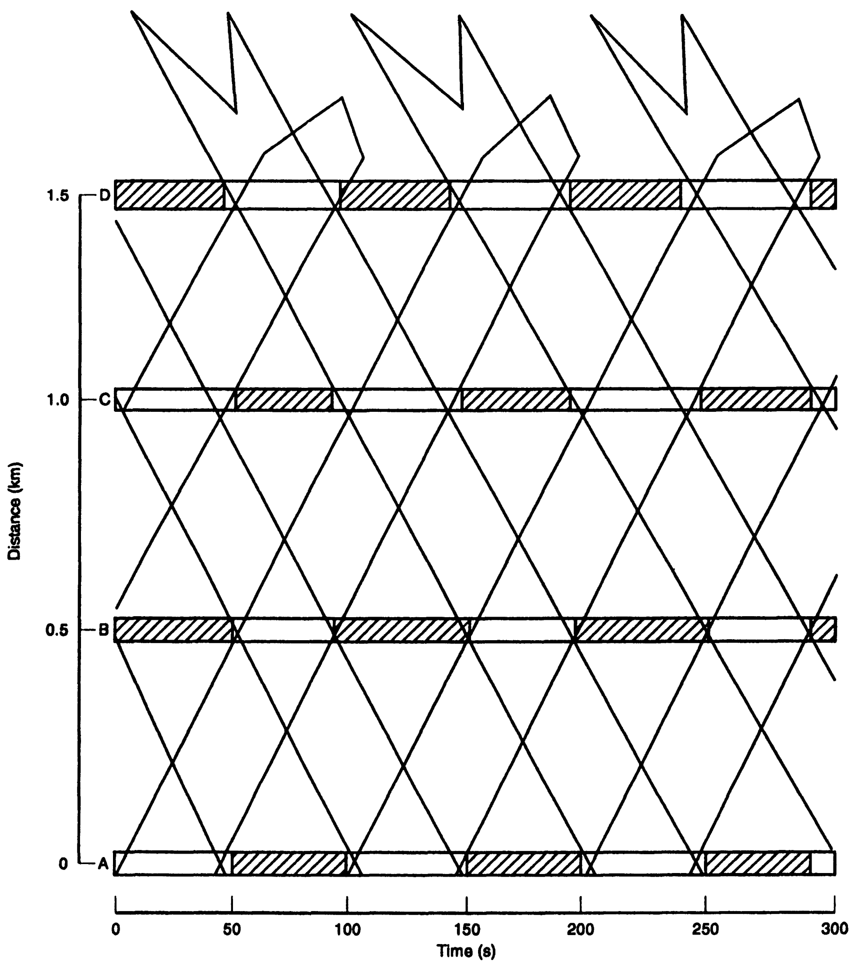

Figure 47.2

48

Platoon dispersion and the linking of traffic signals

A trial-and-error approach such as is involved in the preparation of a time-and-distance diagram can produce reasonable progression along a major traffic route. Where however it is desired to minimise delay at signal-controlled intersections in a network a more rigorous approach is desirable.

An alternative method of linking traffic signals so as to minimise delay is to consider the one-way stream of vehicles which after release from one traffic sig-

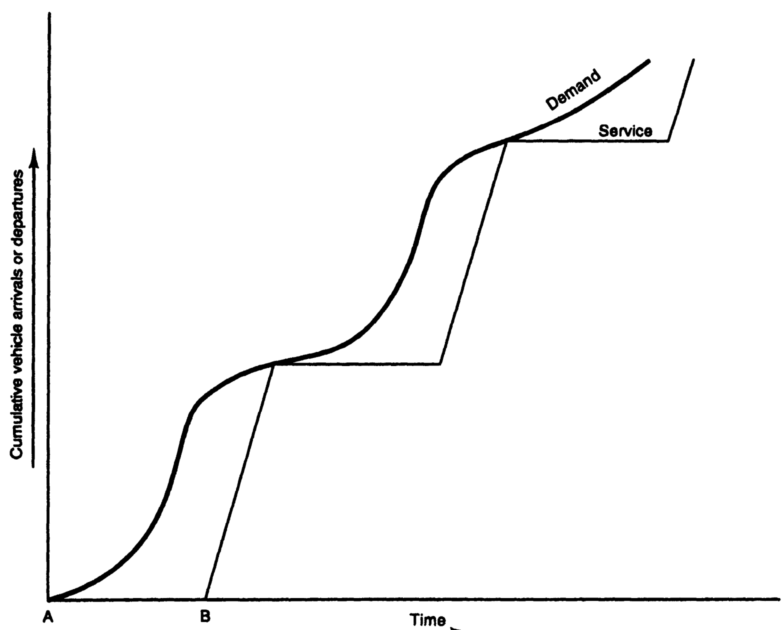

Figure 48.1 *Cumulative demand and service volumes on a traffic-signal approach*

nal travel to the next controlled intersection and are then released when the signals change to green.

If the cumulative demand function, or the number of vehicles arriving at a traffic signal approach, and the cumulative service function, or the number of vehicles discharging from a traffic-signal approach, are known then the area between the curves gives the delay on the approach. These functions are illustrated in figure 48.1.

The vertical distance between the curves represents the number of vehicles delayed at the stop line at any instant while the horizontal distance between the curves represents the duration of delay.

If the time interval between the arrival of the first vehicle at the stop line (A) and the departure of the first vehicle from the stop line (B) is varied then the area between the curves and hence the total delay is varied. By adjustment of the time interval A–B the delay may be minimised.

Such an approach can be used to minimise delay at intersections provided the forms of the demand and service functions are known. There is little difficulty in defining the service function for it is simply illustrated by variations in the flow across the stop line (see Chapter 35).

It is the form of the demand function that must be defined and in a linked-signal system the problem is to predict how platoons of vehicles that are released from a traffic-signal approach disperse as the vehicles travel down the highway towards the next stop line.

Problem

The cumulative demand function and the cumulative service functions for a traffic-signal approach are given in figure 48.2. Is the optimum time *t* for the commencement of green so as to minimise delay 0, 1.5, 2.5 or 4.5 s?

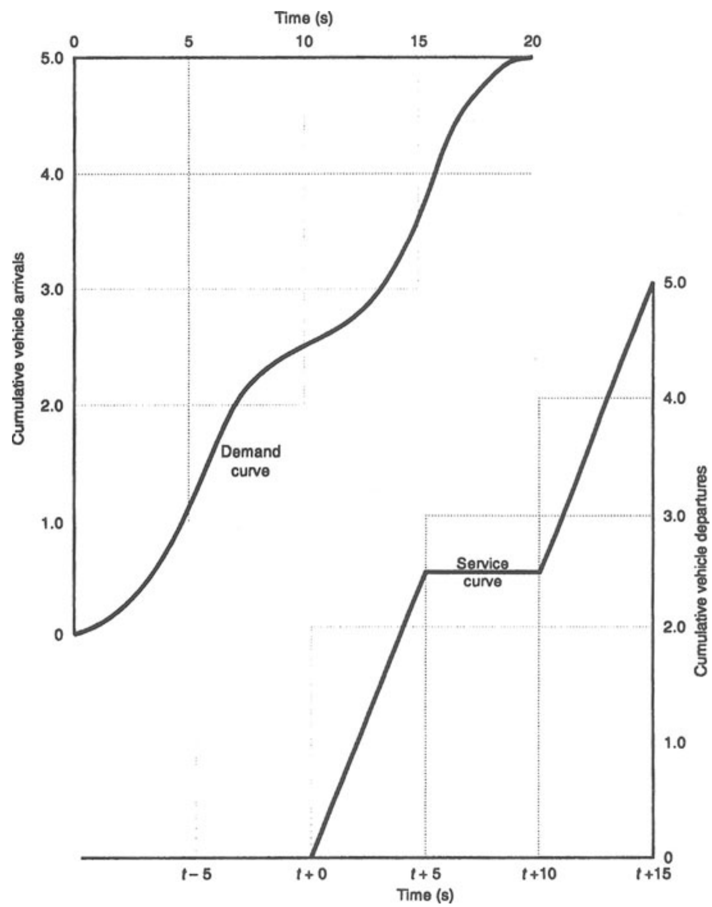

Figure 48.2

Solution

The offset of the green signal in relation to the arrival of the first vehicle at the stop line so as to minimise delay can be obtained by the graphical superposition of the service curve on the demand curve. In doing this it is necessary to minimise the area between the curves.

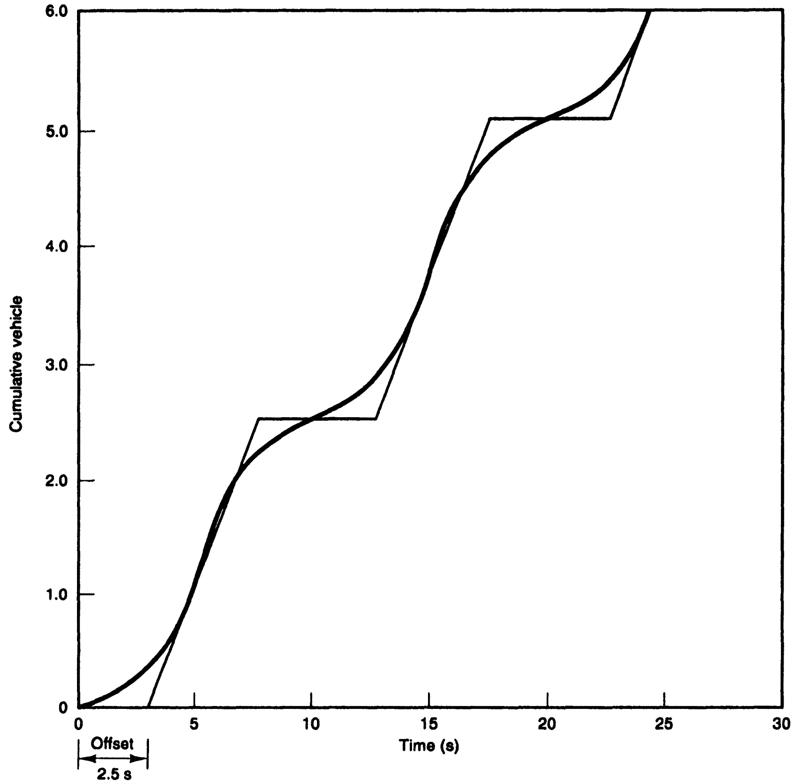

Figure 48.3

This is shown in figure 48.3 where it can be seen that for minimisation of the area between the curves the time offset should be 2.5 s.

49

The prediction of the dispersion of traffic platoons downstream of signals

The minimisation of delay by the adjustment of the offset of the green signal depends on the ability to predict the arrival rate of vehicles at the stop line. In a network controlled by traffic signals the platoons of vehicles travelling towards the stop line will have been discharged initially from a traffic signal upstream. For this reason the prediction of the dispersion of vehicles downstream from traffic-signal approaches is of considerable importance.

A considerable amount of research has been carried out on the prediction of platoon dispersion downstream of traffic signals but probably the technique that has received the greatest application is that given by D. I. Robertson[1].

Platoon dispersion is easy to predict using the following recurrence relationship

$$q_2(i + t) = Fq_1(i) + (1 - F)q_2(i + t - 1) \qquad (49.1)$$

where $q_2(i)$ is the derived flow in the ith time interval of the predicted platoon at a point along the road;
$q_1(i)$ is the flow in the ith time interval of the initial platoon at the stop line;
t is 0.8 times the average journey time over the distance for which the platoon dispersion is being calculated (measured in the same time intervals used for $q_1(i)$ and $q_2(i)$, which are usually 1/50th of the cycle time);
F is a smoothing factor.

Using data from four highway sites in London an expression for F has been given as

$$F = 1/(1 + 0.5t) \qquad (49.2)$$

Reference

1. D. I. Robertson, TRANSYT: A Traffic Network Study Tool, *Road Research Laboratory Report* 253 (1969)

Problem

A highway link commences and terminates with traffic signal-controlled intersections. The discharge from the first intersection may be assumed to have a uniform rate of 1 vehicle/unit time and commences at zero time. The green time at the first intersection is 5 time units and at the second intersection 8 time units. The average travel time of the platoon between intersections may be taken as 10 time units and the travel time of the queue leader as 8 time units.

Using the recurrence relationship derived by Robertson, is the percentage of vehicles delayed at the end of the green period, when the offset of the second signal from the first signal is 8 time units, approximately 25 per cent, 27.5 per cent or 35 per cent?

Solution

The average journey time is given as 10 time units. Then

$$t = 0.8 \times \text{average journey time}$$
$$= 8 \text{ time units}$$

and

$$F = 1/(1 + 0.5t) \qquad \text{(see equation 49.2)}$$
$$= 0.2$$

Using the recurrence relationship given in equation 49.1

$$q_2(i + t) = Fq_1(i) + (1 - F)\, q_2(i + t - 1)$$

See table 49.1

TABLE 49.1

i	$q_1(i)$	$i+t$	$q_2(i+t)$	Cumulative Demand	Cumulative Service
1	1.0	9	$0.2 \times 1.0 + 0.8 \times 0 = 0.20$	0.20	1.0
2	1.0	10	$0.2 \times 1.0 + 0.8 \times 0.2 = 0.36$	0.56	2.0
3	1.0	11	$0.2 \times 1.0 + 0.8 \times 0.36 = 0.49$	1.05	3.0
4	1.0	12	$0.2 \times 1.0 + 0.8 \times 0.49 = 0.59$	1.64	4.0
5	1.0	13	$0.2 \times 1.0 + 0.8 \times 0.59 = 0.67$	2.31	5.0
6		14	$0.2 \times 0 + 0.8 \times 0.67 = 0.54$	2.85	6.0
7		15	$0.2 \times 0 + 0.8 \times 0.54 = 0.43$	3.28	7.0
8		16	$0.2 \times 0 + 0.8 \times 0.43 = 0.34$	3.62	8.0

Total platoon content = initial discharge $\times$ time

= 1.0 vehicles/time unit $\times$ 5 time units

= 5.0 vehicles

Cumulative demand during green period at the second intersection is 3.62 vehicles.

Number of vehicles delayed = 5.0 − 3.62 = 1.38 vehicles

= 27.5 per cent (approximately)

50

The delay/offset relationship and the linking of signals

By the use of the technique of calculating the delay to vehicles on a traffic-signal approach from the difference between the demand and service distributions, as discussed in Chapter 48, and the prediction of the demand distribution from a knowledge of platoon diffusion relationships as described in Chapter 49, it is possible to obtain a relationship between the offset of one signal relative to another and the delay to vehicles passing through both signals.

Using this approach it is possible to calculate the offset or difference in time between the commencement of green for two linked signals likely to result in minimum delay. The process can be extended using a method proposed by the Transport and Road Research Laboratory to a highway network. In this way a series of signal offsets can be obtained for different sets of traffic flow conditions. If signal indications are changed by a central controller or computer then when changes in traffic flow take place it will be possible to select the appropriate offset sequence to minimise delays to vehicles travelling through the network.

These calculations are relatively time-consuming and are usually undertaken within the computer program TRANSYT, described in Chapter 51. However, it is useful also to follow a method of manual calculation for the linking of signals, to demonstrate the key issues involved. Such a method, called the 'combination method' was developed at the Transport and Road Research laboratory and is described in this chapter.

The basic assumptions made in the combination method are:

(i) The settings of the signals do not affect the amount of traffic or the route used.
(ii) All signals have a common cycle (or have a cycle that is a submultiple of some master cycle).
(iii) At each signal the distribution of the available green time among the phases is known.

(iv) The delay to traffic in one direction along any link of the network depends solely on the difference between the settings of the signals at each end of the link; it is not affected by any other signals in the network.

(The term 'link' is used to refer to the length of road between signalised intersections, because this is the more familiar usage when discussing networks. Thus 'link' and 'section' are broadly synonymous.) It is assumed unless otherwise stated that the following basic data relating to the network is available:

(i) the cycle time to be used for the network;
(ii) the delay/difference-of-offset relation in each direction for each link. The delay/difference-of-offset relation comprises N delays, one for each of the N possible differences of offset between the signals at each end of the link.

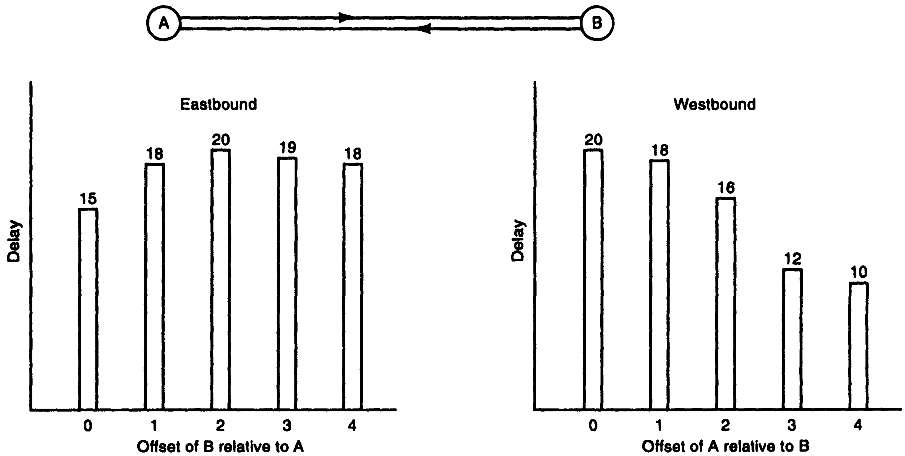

Figure 50.1 *Delay/offset relationships for link AB*

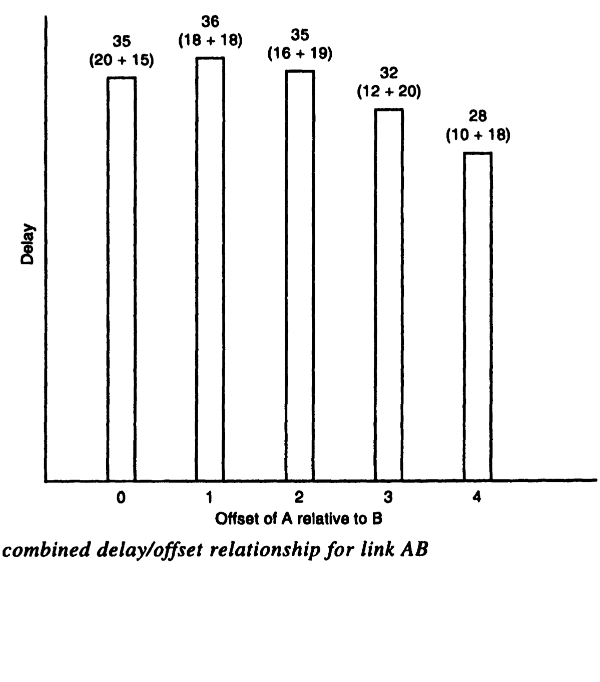

Figure 50.2 *The combined delay/offset relationship for link AB*

Delays in links can be combined in two ways:

(1) in parallel,
(2) in series.

Consider two one-way links in parallel between signalised intersections as illustrated in figure 50.1 for which the delay/offset relationships are also given.

The combined delay/difference of offset relationship for the link can then be obtained by addition, taking care to add delays so that they are relative to the same signal. The total relationship is shown in figure 50.2.

It can be seen from this combined relationship that minimum delay occurs when the offset of A relative to B is 4.

When links are arranged in series, the delay difference in offset relationship for each difference of offset between the extremities has to be calculated and the minimum value selected. Consider the two links in series AB and BC shown in figure 50.3 for which the delay/offset relationships are also given.

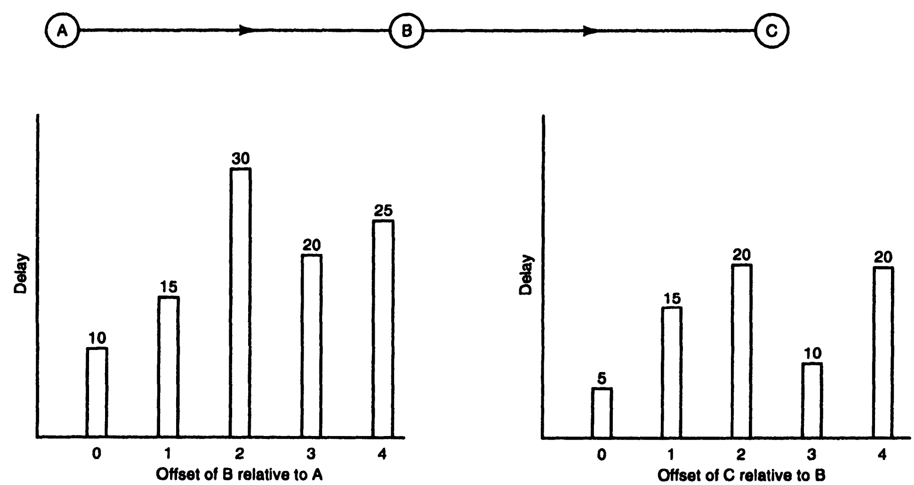

Figure 50.3 *The combination of links in series*

Consider initially the offset of C relative to A to be 0. Then the combined delay/offset relationship is shown in table 50.1.

TABLE 50.1

Offset of B relative to A	*Offset of C relative to B*	*Total delay*
0	0	10 + 5 = 15*
1	4	15 + 20 = 35
2	3	30 + 10 = 40
3	2	20 + 20 = 40
4	1	25 + 15 = 40

When the offset of C relative to A is 0, minimum delay (marked *) occurs when the offset of B relative to A is 0.

Consider next the offset of C relative to A to be 1. Then the combined delay/offset relationship is as shown in table 50.2.

TABLE 50.2

Offset of B relative to A	*Offset of C relative to B*	*Total delay*
0	1	10 + 15 = 25
1	0	15 + 5 = 20*
2	4	30 + 20 = 50
3	3	20 + 10 = 30
4	2	25 + 20 = 45

When the offset of C relative to A is 1, minimum delay (marked *) occurs when the offset of B relative to A is 1.

Consider next the offset of C relative to A to be 2. Then the combined delay/offset relationship is as shown in table 50.3.

TABLE 50.3

Offset of B relative to A	*Offset of C relative to B*	*Total delay*
0	2	10 + 20 = 30*
1	1	15 + 15 = 30*
2	0	30 + 5 = 35
3	4	20 + 20 = 40
4	3	25 + 10 = 35

When the offset of C relative to A is 2, minimum delay (marked *) occurs when the offset of B relative to A is 0 or 1.

Consider next the offset of C relative to A to be 3. Then the combined delay/offset relationship is as shown in table 50.4.

TABLE 50.4

Offset of B relative to A	*Offset of C relative to B*	*Total delay*
0	3	10 + 10 = 20*
1	2	15 + 20 = 35
2	1	30 + 15 = 45
3	0	20 + 5 = 25
4	4	25 + 20 = 45

When the offset of C relative to A is 3, minimum delay (marked *) occurs when the offset of B relative to A is 0.

Finally when tile offset of C relative to A is 4, the combined delay/offset relationship is as shown in table 50.5.

TABLE 50.5

Offset of B relative to A	*Offset of C relative to B*	*Total delay*
0	4	10 + 20 = 30
1	3	15 + 10 = 25*
2	2	30 + 20 = 50
3	1	20 + 15 = 35
4	0	25 + 5 = 30

When the offset of C relative to A is 4, minimum delay (marked *) occurs when the offset of B relative to A is 1.

It is now possible to produce a combined delay/difference-of-offset relationship, as shown in figure 50.4.

It can be seen that minimum delay for the combined link occurs when the offset of C relative to A is 0 and when the offset of B relative to A is 0 respectively.

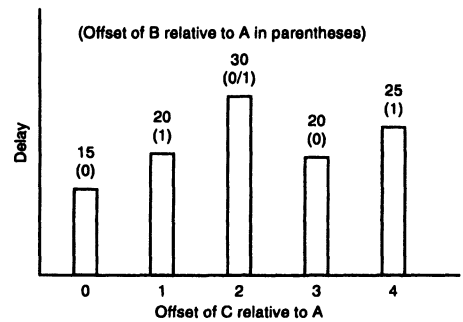

Figure 50.4 *The combined delay/offset relationships for links AB and BC*

Problem

For the network and delay/offset relationships shown in figure 50.5 determine the offsets for minimum delay in the network.

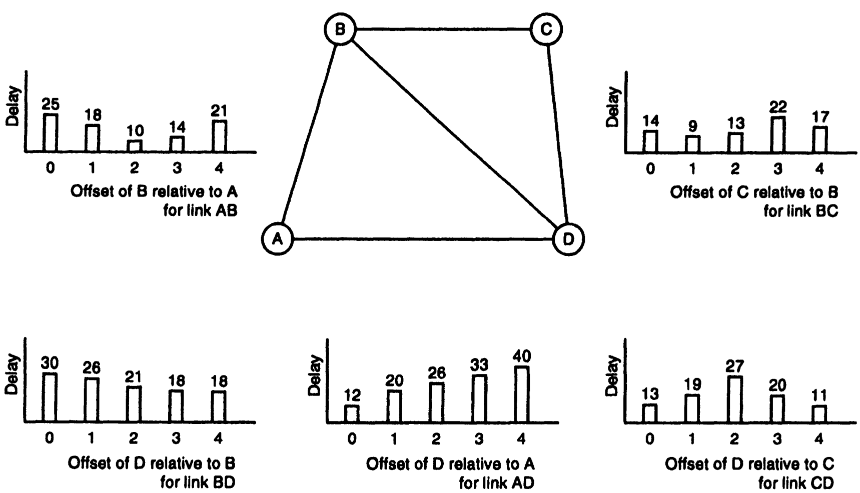

Figure 50.5

Solution

Combination of links BC and CD (a series combination)

1. *Difference of offset between D and B of 0*

Offset of C relative to B	*Offset of D relative to C*	*Total delay*
0	0	14 + 13 = 27
1	4	9 + 11 = 20*
2	3	13 + 20 = 33
3	2	22 + 27 = 49
4	1	17 + 19 = 36

2. *Difference of offset between D and B of 1*

Offset of C relative to B	*Offset of D relative to C*	*Total delay*
0	1	14 + 19 = 33
1	0	9 + 13 = 22*
2	4	13 + 11 = 24
3	3	22 + 20 = 42
4	2	17 + 27 = 44

3. *Difference of offset between D and B of 2*

Offset of C relative to B	*Offset of D relative to C*	*Total delay*
0	2	14 + 27 = 41
1	1	9 + 19 = 28
2	0	13 + 13 = 26*
3	4	22 + 11 = 33
4	3	17 + 20 = 37

4. *Difference of offset between D and B of 3*

Offset of C relative to B	*Offset of D relative to C*	*Total delay*
0	3	14 + 20 = 34
1	2	9 + 27 = 36
2	1	13 + 19 = 32
3	0	22 + 13 = 35
4	4	17 + 11 = 28*

5. *Difference of offset between D and B of 4*

Offset of C relative to B	*Offset of D relative to C*	*Total delay*
0	4	14 + 11 = 25*
1	3	9 + 20 = 29
2	2	13 + 27 = 40
3	1	22 + 19 = 41
4	0	17 + 13 = 30

The minimum delay values are marked * and for the varying differences of offset between D and B the delay histogram can be plotted when BC and CD are combined (figure 50.6).

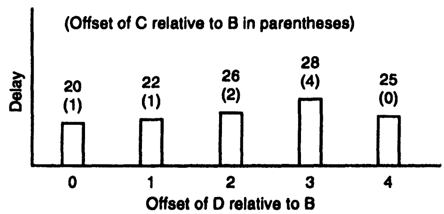

Figure 50.6 *The series combination of links BC and CD*

Combination of above and link BD (a parallel combination). The delay for link BD is known for the offset of D relative to B and so the combination of the two links is a straightforward addition, giving figure 50.7.

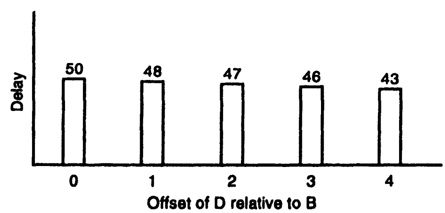

Figure 50.7 *The parallel combination of link BD and the previous combination*

Combination of above and AB (a series combination)

6. *Difference of offset between D and A of 0*

Offset of D relative to B	*Offset of B relative to A*	*Total delay*
0	0	50 + 25 = 75
1	4	48 + 21 = 69
2	3	47 + 14 = 61
3	2	46 + 10 = 56*
4	1	43 + 18 = 61

7. *Difference of offset between D and A of 1*

Offset of D relative to B	*Offset of B relative to A*	*Total delay*
0	1	50 + 18 = 68
1	0	48 + 25 = 73
2	4	47 + 21 = 68
3	3	46 + 14 = 60
4	2	41 + 10 = 51*

8. *Difference of offset between D and A of 2*

Offset of D relative to B	*Offset of B relative to A*	*Total delay*
0	2	50 + 10 = 60
1	1	48 + 18 = 66
2	0	47 + 25 = 72
3	4	46 + 21 = 67
4	3	43 + 14 = 57*

9. *Difference of offset between D and A of 3*

Offset of D relative to B	*Offset of B relative to A*	*Total delay*
0	3	50 + 14 = 64
1	2	48 + 10 = 58*
2	1	47 + 18 = 65
3	0	46 + 25 = 71
4	4	43 + 21 = 64

10. *Difference of offset between D and A of 4*

Offset of D relative to B	*Offset of B relative to A*	*Total delay*
0	4	50 + 21 = 71
1	3	48 + 14 = 62
2	2	47 + 10 = 57*
3	1	46 + 18 = 64
4	0	43 + 25 = 68

The minimum values are marked *. For varying offsets between D and A, the delay histogram can be plotted (figure 50.8) when AB is added to the previous delay combinations.

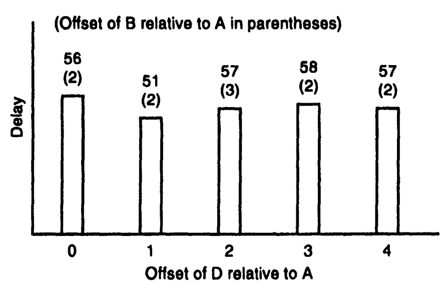

Figure 50.8 *The series combination of link AB and the previous combination*

Finally the previous combinations can be combined with the link AD (a parallel combination).

11. The combination is a straightforward addition giving the histogram of figure 50.9.

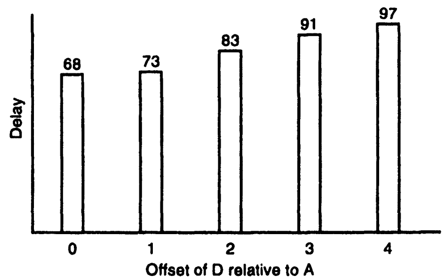

Figure 50.9 *The parallel combination of link AD and the previous combination*

Minimum delay for the total combination thus occurs when the offset of D relative to A is 0 (para. 11).

When the offset of D relative to A is 0 minimum delay occurs when the offset of D relative to B is 3 and the offset of B relative to A is 2 (para. 6).

When the difference of offset between D and B is 3 minimum delay occurs when the offset of C relative to B is 4 and the offset of D relative to C is 4 (para. 4).

51

Urban traffic control systems

During the late 1950s, proposals were made that the linking and co-ordination of traffic signals, which had already been applied to produce green waves of traffic along major routes, could be extended to a network of highways over an area. Originally referred to as the area 'control of traffic', co-ordination over a network is now more usually termed 'urban traffic control'.

Urban traffic control strategies often seek to minimise delay over a network, in contrast to the linking of signals along specified traffic routes which minimises delay to major route traffic at the expense of vehicles on the remainder of the network. The actual co-ordination of signals will depend upon the traffic and road pattern and the objective which is to be attained.

One of the first reported applications of signal control by computer was in Toronto where a pilot project was commenced in 1959; this was judged to be effective and was later extended. Subsequently, the use of computers in signal control was considered for use in the United Kingdom during the 1960s, where the extensive use of vehicle actuation and differences in the road pattern initially raised doubts as to the most effective application of computer control.

Two experimental systems were set up to assess the value of computer control in the United Kingdom; one was in West London, covering an area of 6 square miles, while the other consisted of an extensive series of trials carried out in Glasgow. Within the London area all the signals were already vehicle actuated, and some were already linked. In particular, a linked system already operated along Cromwell Road, which at that time carried 38 000 vehicles per day towards central London. In a description of this scheme, Mitchell[1] stated that the main intention was to establish progression over the greatest route mileage practicable whenever traffic conditions were favourable.

During the Glasgow experiment a number of systems were implemented to test their effectiveness in minimising delay. These systems could be divided into three main categories: fixed-time plans based on historical data and calculated off-line by a computerised optimising technique, co-ordinated systems with local traffic response at each signal, and fully responsive systems. In the first group were the Combination method, the TRANSYT and SIGOP systems. In the second group were the FLEXIPROG and EQUISAT systems; and in the third group the systems Dynamic plan generation and PLIDENT were tested.

Results of these experiments have been summarised by Holroyd and Robertson[2]. It was found that the Combination method, when compared with the vehicle-actuated system previously operating in Glasgow, gave consistently shorter journey times, averaging 12 per cent. The TRANSYT system gave a 4 per cent saving in journey time compared with the Combination method; the TRANSYT optimising program is discussed later in greater detail. The SIGOP program was compared with the TRANSYT system during the morning and evening peaks and the off-peak period, but did not indicate any improvement over the TRANSYT program.

In the second group of programs, the FLEXIPROG program, which used the Combination method and the TRANSYT program for basic linking, together with local response to change in traffic conditions, was not found to offer any improvement in journey times. Isolated vehicle actuation was also compared with the TRANSYT program under very light flow conditions and no measurable difference in journey times was observed. A further program in this group was the EQUISAT program, which used the Combination method for basic linking and, once again, no measurable difference in journey times was found.

The third group of control programs, which were tested in the Transport and Road Research Laboratory, comprised dynamic plan generation and the PLIDENT program. The former program was tested in Madrid and compared with an optimised fixed-time system but produced longer journey times. The PLIDENT program was compared with the Combination method but, once again, gave longer journey times.

As a result of these experiments, the TRANSYT program was adopted during the early 1970s as a means of optimising traffic flow through urban networks.

The TRANSYT program

The TRANSYT[3] program, which stands for Traffic Network Study Tool, is an off-line program for calculating optimum co-ordinated signal timings in a network of traffic signals. It has become the most widely used program of its type in the world. After the first program was developed in 1967 a number of versions have been produced, all of which have two main elements: a traffic model and a signal optimiser, as illustrated in figure 51.1.

The traffic model represents traffic behaviour in a highway network in which most junctions are controlled by traffic signals. The model predicts the value of a 'Performance Index' for the network for a given fixed-time plan and an average set of flows on each link. The Performance Index measures the overall cost of traffic congestion and is usually a combination of the total delay and the number of stops made by vehicles.

The optimisation process adjusts the signal timings and, using the traffic model, calculates if the adjusted timings reduce the Performance Index. By successive adoption of beneficial timings, an optimum is reached.

TRANSYT assumes that all major junctions in the network are signal or priority controlled, that all the signals have a common cycle time, and that all signal stages and their minimum periods are known. For each distinct traffic stream it is assumed that, for traffic flowing between junctions, or turning at junctions, the flow rate, averaged over a specified period, is known and assumed to be constant.

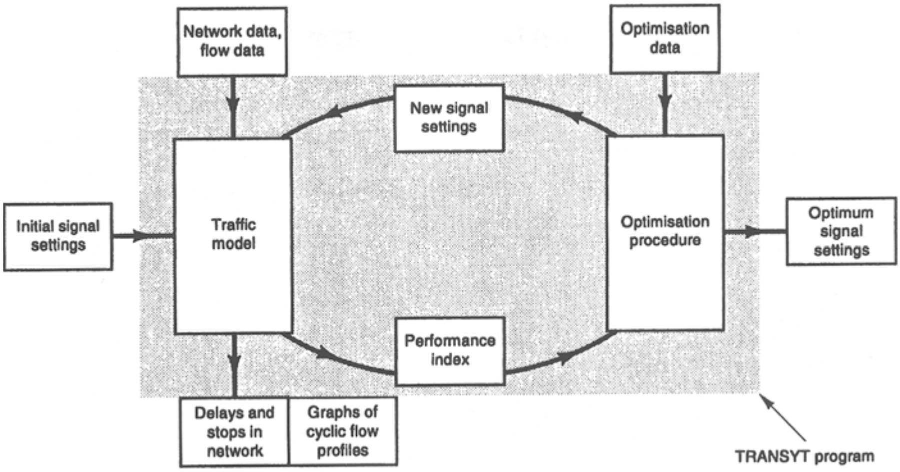

Figure 51.1 *Structure of the TRANSYT program (ref. 3)*

In TRANSYT, the network is represented by nodes for the signal-controlled intersections and each distinct one-way traffic stream by links. A link may represent one or more traffic streams and one signal approach may be represented by several links. It is a matter of judgement which traffic streams warrant separate representation; two approach lanes where vehicles are equally likely to join either lane can be represented by one link whereas, if one of the lanes contains separately signalled right-turn vehicles, for example, then a separate link is necessary.

The common cycle time necessary for optimisation is divided into a number of equal intervals or steps which are usually of duration from 1 to 3 seconds. All calculations are made on the basis of the average flow rates during each step of the cycle, using a histogram representing the traffic arrivals and departures at and from nodes; the dispersion of a departure stream before it arrives at the next downstream node is estimated by means of the smoothing function described in Chapter 49.

Delay in the TRANSYT program is obtained from a consideration of these histograms, called cyclic flow profiles. Two further elements of delay have to be added to this value of uniform average delay. If the arrival traffic flow exceeds capacity, then the approach is saturated and an element of over-saturation delay must be added. In addition, there is an element of delay associated with the random nature of traffic flow. TRANSYT uses the following formula for the random and over-saturated delay rate, d, in pcu hours per hour:

$$d = \frac{T}{4}\left\{\left[(f-F)^2 + \frac{4f}{T}\right]^{0.5} + (f-F)\right\}$$

where f is the average arrival flow on the link (pcu per hour), F is the maximum flow able to discharge from the link (pcu per hour) and T is the duration of the flow condition being modelled (hours).

Vehicle stops are included in the Performance Index and, as with the calculation of delay, TRANSYT calculates the total rate at which vehicles are forced to stop on a link as the sum of uniform, random and over-saturation flow rates. The uniform component is calculated from the cyclic flow profiles and the random and over-saturation elements from simple equations.

In most road networks there are a number of give-way junctions in addition to the signal-controlled junctions. The TRANSYT program allows priority junctions to be represented by referring to the major road flow as the controlling link; the flow from the non-priority link depends on the controlling link flow, so producing a flow profile from the give-way link.

TRANSYT allows five separate classes of vehicles to be represented in any one queueing situation where the vehicles may in reality be mixed together. These classes may not be different types of vehicles but vehicles which have entered the road system from differing origins. It is thus possible to represent buses on different routes with different bus stops. This facility, referred to as a 'shared stopline', also allows the development of special 'green-wave' fixed-time plans for emergency vehicles.

Fuel consumption estimates can also be output by the TRANSYT program for a particular set of signal timings. This estimate of fuel consumption is composed of three elements: fuel consumed while travelling at constant cruise speed between stoplines, extra fuel used during delay, and the extra fuel used in making each fuel stop and having to resume cruise speed.

The optimisation model in TRANSYT seeks the best offsets which minimise the Performance Index, a measure of vehicle stops and delays in the network. The index is evaluated in monetary terms by assuming appropriate values for delay and stops. Optimisation of green times is also possible using TRANSYT but optimisation of cycle times is a more complex problem. However, TRANSYT has the option of automatically selecting a range of cycle times, and the cycle time with the best Performance Index output with a warning that the choice is made on the basis of unco-ordinated signal calculations.

Before TRANSYT can carry out the off-line optimisation process, it is necessary to select the network which is to be optimised, with the links and nodes being numbered in the order in which traffic enters the road system. A decision will have to be made of the number of optimised fixed-time plans which are necessary. As plan changes impose additional delay on network traffic, the number of plan changes should be kept to a minimum consistent with changes in the level of flow; a recommended number is in the region of ten. Traffic data which must be collected is journey time and speed, saturation flows, demand flows and turning movements.

This data is then input to the TRANSYT program which is run off-line to produce an optimised plan. Because these optimised plans are based on past or historical records of traffic plan, the best form of control may not be achieved if there are large random variations in flow or if an accident changes the traffic flow pattern. Because of the cost and the time required to collect traffic data, it is often out of date. Nevertheless, the use of fixed-time plans produced by such optimisation programs as TRANSYT is effective provided the required traffic engineering resources are available.

A recent application of TRANSYT is for the optimisation of signal timings on signalised roundabouts. Such signalisation is becoming increasingly common in the United Kingdom, part-time (for example, peak hour) signalling on grade-separated roundabouts, to increase roundabout capacity and reduce the incidence of traffic queues extending back onto a major highway. Techniques for applying TRANSYT in this context have been produced by Lines and Crabtree[4].

Improvements to TRANSYT currently being addressed by the Transport Research Laboratory are the modelling of flared approaches and the modelling of congested short links; TRANSYT's vertical queueing model does not allow queueing interactions between junctions to be represented, the usual technique being to apply a 'stop penalty' to a congested link to give it preferential signalling and hence minimise the incidence of 'blocking back'.

To overcome the main problem with fixed-time optimised plans – that they do not respond to changing traffic conditions – a fully responsive Urban Traffic Control system which does not require the off-line pre-calculation was developed in the late 1970s by the Transport and Road Research Laboratory in collaboration with Ferranti, GEC and Plessey traffic signal system companies. This is known as SCOOT.

The SCOOT traffic model

As a result of this co-operative effort, the SCOOT[5] (Split Cycle and Offset Optimising Technique) has been developed and put into service. It has been stated[5] that the structure of SCOOT is similar to that of the TRANSYT method of calculating fixed-time plans. Both use a traffic model which predicts delays and stops with given signal settings, but while TRANSYT does this off-line, predicting delays and stops from average flows, SCOOT makes re-calculations every few seconds from the latest measurements of traffic behaviour. An outline of the SCOOT system is illustrated in figure 51.2.

Both SCOOT and TRANSYT signal optimisers automatically make systematic trial alterations to current signal timings and implement only those alterations which improve the Performance Index. In the case of TRANSYT, any new fixed-time plan must be stored in the library of the Urban Traffic Control computer.

The SCOOT traffic model uses vehicle presence measurements obtained from road detectors which are located as far as possible from the signal stopline, ideally just downstream of the adjacent signal-controlled junction provided a detector in this position can detect all vehicles approaching the stopline. Being sited upstream, these detectors rarely become covered with stationary traffic, and therefore provide 'continuous' demand data for the SCOOT traffic model to estimate flows, queues and delays accurately, and hence optimum signal timings. This is not possible when detectors are sited close to a stopline. The main requirement is that the traffic crossing a detector should accurately reflect that which will subsequently pass through the junction under control. Appropriate detector siting is therefore important with mid-link 'sinks' and/or 'sources' (for example car parks) requiring particular consideration.

Information from the detectors is stored in the SCOOT computer in the form of cyclic flow profiles for each signal approach. The cyclic flow profile represents the rate of traffic flow against time and the data is repeated with each cycle of traffic flow, hence the term 'cyclic flow profile'. Unlike the flow profiles which have to be calculated in TRANSYT, SCOOT uses a cyclic flow profile in which the more recent data on traffic flow is combined with existing values so that large ran-

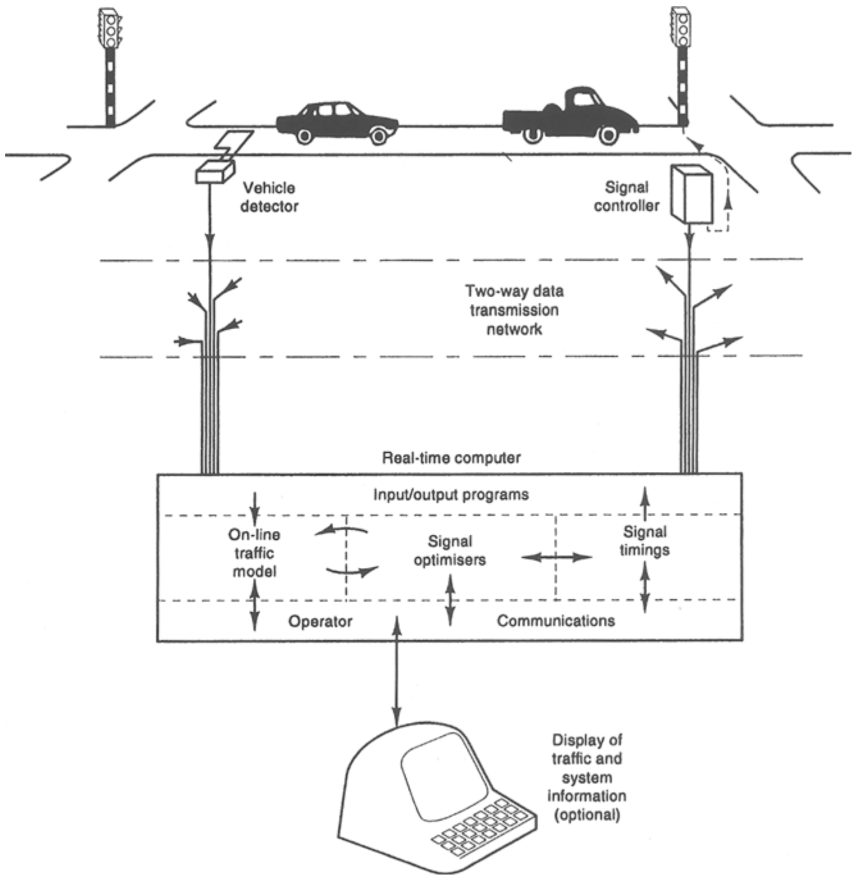

Figure 51.2 *Information flow in a SCOOT UTC system (ref. 5)*

dom fluctuations in the profiles are averaged out so as to be typical of the existing traffic situation.

These cyclic flow profiles are used by the signal optimiser to find the signal timings which give the best Performance Index for the network. As traffic flows change in the network, the cyclic flow profiles will also change and cause the signal optimiser to search for new timings.

Using information from the cyclic flow profiles and the cruise times between the detectors and the downstream stoplines, the SCOOT traffic model predicts queue lengths using the green and red signal timings. This is illustrated in figure 51.3. As in the TRANSYT system, the SCOOT optimiser calculates a Performance Index which includes a weighted aggregation of delays and stops. In a system where every vehicle received a green indication, then the sum of the queues would be zero; this is not normally possible but the optimiser seeks to adjust the signal timings so that the value is as small as possible. In addition, SCOOT predicts the number of vehicle stops, the proportion of the cycle time during which vehicles are stationary over the SCOOT detector, and the degree of saturation on the signal approaches. These values are employed by the optimiser to control the cycle time and green durations.

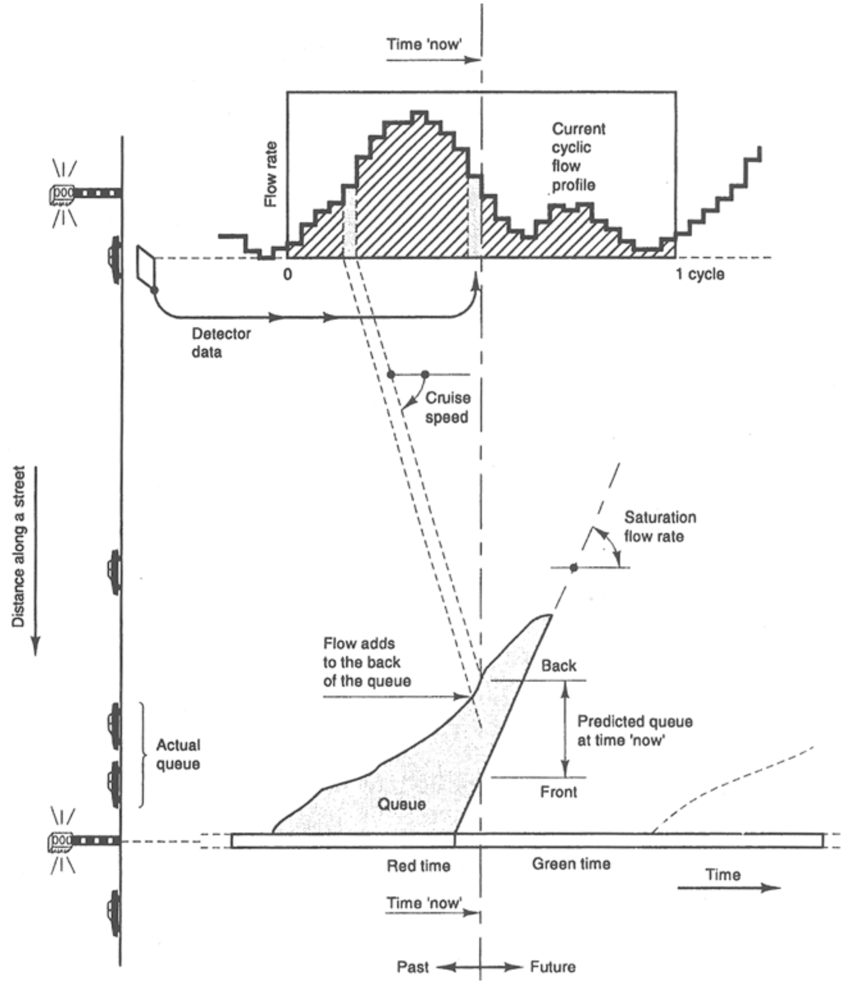

Figure 51.3 *Principles of the SCOOT traffic model (ref. 5)*

The SCOOT optimiser makes frequent small changes to the signal timings as the traffic demand changes, rather than large changes which are made infrequently for systems with several fixed-time plans. It has been stated[5] that the advantages of the SCOOT method are considered to be as follows.

Firstly, that there are no large, sudden changes in signal timings. In a comparison of alternative methods of changing from one timing plan to another, even the best methods caused significant increases in vehicle delay during the transition period, which may last for two or more signal cycles. This means that a new plan must operate for ten to fifteen minutes to ensure that total traffic benefits exceed the initial disbenefits during the change.

Secondly, there is no need to predict average traffic behaviour for several minutes into the future. If, as discussed in the previous paragraph, a new plan must remain in operation for at least ten minutes, then the average flows for the subsequent ten minutes must be predicted in the previous period so that the new plan can be calculated on-line. There are difficulties in making accurate predictions because of large random variations in traffic flow which disguise longer-term trends. The use of small frequent timing alterations allows SCOOT to follow changes in traffic behaviour without requiring longer-term predictions of average flows.

Thirdly, the sensitivity of SCOOT to faulty information from vehicle detectors is reduced because new signal timings result from the accumulation of a large number of small changes, hence a few poor decisions by the optimiser are of limited importance.

When SCOOT adjusts green durations the 'split' optimiser estimates, a few seconds before each stage change, whether it is better to change the stage earlier, as scheduled or later. Each decision may alter the stage change by only a few seconds based on the desire to minimise the maximum degree of saturation on the junction approaches, taking into account current estimates of queue length, congestion on approaches and input minimum green times. What are referred to as 'temporary changes' are made because of cycle-by-cycle random variations in traffic flow, and for each temporary variation a smaller permanent variation is made to stored values of green times. The result is that green times at a junction can be completely revised to meet new traffic problems.

Offset optimisation is achieved in the SCOOT program because, once every cycle time, the optimiser makes a decision on whether or not to alter all the schedule stage change times at a junction. Because all stages are changed by the same amount, it is the offset which is altered relative to other junctions. Any decisions alter the offset by a few seconds so that the offset between adjacent junctions may be altered twice every cycle. At a pre-determined stage in each signal cycle, the offset optimiser considers the stored cyclic profiles and estimates whether a change in the offset will improve traffic progression on the streets immediately upstream or downstream of the junction being considered. Performance Indictors on adjacent streets for offsets a few seconds earlier or later are compared and the most beneficial implemented.

The SCOOT optimiser also optimises the cycle time, which is common within sub-areas of the signal-controlled network, at intervals of 2.5 or 5 minutes. Each sub-area can be varied independently of other sub-areas with limits set by normal signal design for minimum and maximum cycle times. A decision to change the cycle time is made on the basis of maintaining, if possible, at the most heavily loaded junction in the sub-area a degree of saturation of approximately 90 per cent. This means that, in periods of low flow, the cycle times will shorten and, in periods of heavy flow, cycle times will lengthen up to a maximum. Less heavily trafficked junctions in a sub-area may have a low degree of saturation and, in this case, a decision to operate these signals by double cycling may take place. Pelican signals for pedestrians are often included in SCOOT networks and these frequently double cycle so as to minimise delay to pedestrians.

When installed, a SCOOT database has to be established which describes the road network under control (links, nodes, etc.) and the key geometric, traffic and control characteristics. For the signals, this includes minimum and maximum cycle times and stage lengths, stage order, intergreen times and other special data (such as for filters). Following installation, it is necessary to validate certain parameters in the model to reflect on-street traffic performance for each link, before the system becomes operational. These parameters include: (i) the cruise speed between the detector and the stopline for a free-flowing platoon, (ii) saturation occupancy (similar to saturation flow) and (iii) the maximum queue of vehicles which can normally be accommodated between the detector and the stopline. This latter information allows SCOOT to model the effects of full links on the capacity of upstream links.

The SCOOT model produces a wealth of traffic information on flows, delays, congestion and associated parameters, which is useful in its own right as well as for signal control. An automatic traffic information database, ASTRID[6], has been developed to process and analyse this data according to user requirements, which can include real-time traffic information systems.

SCOOT developments in recent years have focused on providing enhanced facilities, particularly for traffic and congestion management. For example, SCOOT version 2.3 contains facilities for weighting or biasing signal timings in favour of specific traffic streams. Thus, where a junction has spare green time, split weighting can be used to allocate most of this spare green to a favoured link, such as one with a high flow of buses. Similarly, offset weighting can be used to benefit particular movements/routes (such as outbound routes in the evening peak). SCOOT (2.4) contains improved modelling facilities as well as new facilities for congestion management. These include 'congestion offsets' where offsets take account of congestion on a specified link, and 'gating' where flow entering the network can be restricted, using appropriate signal timings, to maintain smooth flow further downstream. SCOOT (3.0) includes bus priority.

Traffic surveys in Glasgow and Coventry have shown that the use of SCOOT is likely to reduce average delay at traffic signals compared with good fixed-time plans by about 12 per cent. Surveys in Southampton indicated delay savings of up to 40 per cent compared with the existing isolated vehicle-actuated control. It is considered that reductions in delay are likely to vary considerably from junction to junction, depending on the magnitude and variability in traffic flow, the street layout and the previous standard of co-ordination between signals. The greatest benefits are likely as flows approach capacity, where flows are variable and difficult to predict and where signal-controlled junctions are relatively close.

References

1. G. Mitchell, Control Concepts of the West London Experience, *Symposium on Area Control Road Traffic*, Institution of Civil Engineers, London (1967)
2. J. Holroyd and D.I. Robertson, Strategies for Area Traffic Control Systems: Present and Future, *Transport and Road Research Laboratory Report* LR 569, Crowthorne (1973)
3. R.A. Vincent, A.I. Mitchell and D.I. Robertson, User Guide to TRANSYT Version 8, *Department of Transport TRRL Report* LR666, Transport and Road Research Laboratory, Crowthorne (1980)
4. C.J. Lines and M.R. Crabtree, The Use of TRANSYT at Signalised Roundabouts, *Transport and Road Research Laboratory Research Report* 274, Crowthorne (1990)
5. P. B. Hunt, D. I. Robertson, R. D. Bretherton and R. I. Winton, SCOOT: A Traffic Responsive Method of Co-ordinating Signals, *Transport and Road Research Laboratory Report* 1014, Crowthorne (1981)
6. N.B. Hounsell and F. N. McLeod, ASTRID: Automatic SCOOT Traffic Information Database, *Department of Transport TRRL Report CR235*, Transport and Road Research Laboratory, Crowthorne (1990)

APPENDIX

Definition of symbols used in Part III

a = number of traffic lanes on a signal approach

$A = \dfrac{(1-\lambda)^2}{2(1-\lambda x)}$ = first term in the equation for average delay on a traffic-signal approach; due to the uniform rate of vehicle arrivals

$B = \dfrac{x^2}{2(1-x)}$ = second term in the equation for average delay on a traffic-signal approach; due to the uniform rate of vehicle arrivals

B_r = mean minimum time headway between right-turning vehicles as they pass through a traffic signal-controlled intersection

B_0 = mean minimum time headway between opposing vehicles as they pass through a traffic signal-controlled intersection

B_1 = mean minimum time headway between bunched vehicles on the major road

B_2 = mean minimum time headway between vehicles as they discharge without conflict from the minor to the major road

c_1 = common cycle time of linked signals

C = correction term in the equation for average delay on a traffic-signal approach

C_0 = optimum cycle time

C_m = minimum cycle time

d = average delay per vehicle on a traffic-signal approach

d_g = dummy variable for gradient

d_n = dummy variable for nearside lane

f = proportion of turning vehicles

F = smoothing factor used in the prediction of platoon dispersion

g = effective green time

g_s = saturated green time

G = percentage gradient on a signal approach

j = average spacing of vehicles in a queue on a traffic-signal approach

L = total lost time per cycle
n_t = number of lanes in an opposing entry
n_r = number of right-turning vehicles
n_s = number of right-turn storage spaces within a junction without blocking following straight ahead vehicles
N = average queue at the beginning of the green period
N_s = initial queue at the beginning of a saturated green period
N_u = initial queue at the beginning of an unsaturated green period
p = conversion factor from vehicle to pcu
q = flow on a traffic-signal approach
q_0 = flow on an opposing traffic-signal approach
q_1 = flow on the major road
q_2 = flow on the minor road
$q_1(i)$ = flow in the ith time interval of a predicted platoon at the stop line
$q_2(i)$ = derived flow in the ith time interval of a predicted platoon at a point along the road
r = turning radius
r_0 or r = effective red period (cycle time – effective green time)
S_0 = saturation flow of an opposing entry
S_1 = saturation flow of an individual lane
S_2 = saturation flow of an opposed lane
S_c = saturation of an opposed lane after the effective green period
S_g = saturation of an opposed lane during the effective green period
T = the through car unit of a turning vehicle
v = free running speed of vehicles on a traffic-signal approach
w = width of approach
x = degree of saturation; that is, the ratio of actual flow to the maximum flow
x_0 = degree of saturation on an opposing arm
y = ratio of flow to saturation flow $= \dfrac{q}{\lambda s}$
y_{max} = maximum value of y
Y = sum of the maximum y values for all phases comprising the cycle
$Y_{pract.}$ = 90 per cent of Y
α = average gap accepted in the opposing flow by right-turning vehicles, or the mean lag or gap in the major-road stream that is accepted by minor-road drivers
λ = ratio of effective green time to cycle time

Index