Approximation Methods for Queues with Application to the Fixed-Cycle traffic Light
Author(s): G. F. Newell
Source: *SIAM Review*, Apr., 1965, Vol. 7, No. 2 (Apr., 1965), pp. 223-240
Published by: Society for Industrial and Applied Mathematics

Stable URL: https://www.jstor.org/stable/2027270

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/2027270?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Society for Industrial and Applied Mathematics is collaborating with JSTOR to digitize, preserve and extend access to *SIAM Review*

JSTOR

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

SIAM Review
Vol. 7, No. 2, April, 1965

APPROXIMATION METHODS FOR QUEUES WITH APPLICATION TO THE FIXED-CYCLE TRAFFIC LIGHT*

G. F. NEWELL†

Abstract. Approximations based upon the representation of the queue as a continuous fluid with either deterministic or stochastic properties are applied to the analysis of models of a fixed-cycle traffic light. These approximations are based upon the use of a law of large numbers or a central limit theorem and are not very sensitive to the detailed stochastic structure of the arrival or departure processes. In the applications considered here these approximations give delays correct to within a few percent.

Introduction. The main object of the following paper is to describe and illustrate some approximation methods that can be used to obtain rough estimates of queue lengths, delays, etc., for various queueing problems, particularly highway traffic intersection problems, which may be too difficult to solve exactly, or if solved exactly give formulas that are more detailed than one needs for the purpose of making quick estimates. The types of approximation we have in mind are those that will apply when the average queue lengths are much larger than 1. They will be asymptotic approximations that strictly speaking are valid only in the limit of infinite queues, but ones which still give rough estimates for finite but large queues (perhaps of the order of 10).

There is an extensive literature on queueing theory including at least a half dozen books and several hundred papers, but the vogue in queueing theory has been to obtain exact solutions of highly idealized models of various processes (exact, however, only in the sense that one usually must invert a few generating functions or Laplace transforms to obtain the quantities one really wants). Consequently the practical value of queueing theory has been severely limited by the lack of approximation methods which one can use to analyse more difficult problems, to estimate errors introduced by the model, or even to compute numbers from very cumbersome exact formulas. A few attempts, however, have been made in recent years to break away from the now standard analytic techniques. Particularly noteworthy are the works of Kingman [1], who has shown that properties of nearly saturated queues are rather insensitive to the detailed form of the arrival or service distributions, and of Beneš [2], who has shown that many of the results of queueing theory do not depend upon some of the customary statistical independence assumptions. Even here, however, mathematical elegance takes precedence over intuition or applications.

In the literature on traffic theory there are already at least 20 references relating to the single problem of delays at a fixed-cycle traffic light, three quarters of which are primarily directed toward a fruitless pursuit of simple exact solutions

* Received by the editors October 19, 1964, and in revised form November 19, 1964.

† Division of Applied Mathematics, Brown University, Providence, Rhode Island. Part of this work was done while the author was a Fulbright professor at the University of Adelaide, South Australia. The work was also supported in part by the National Science Foundation at Brown University.

This content downloaded from 130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

of some highly idealized models. There are perhaps a comparable number of papers on closely related problems of bulk service queues. In view of the mathematical difficulties already being encountered with the simplest problems of this type, it seems clear that further progress toward solutions of real intersection problems with elaborate control systems, turning traffic, etc., will not be made until we are willing to abandon the dream of finding useful exact formulas.

The simplest models of traffic flow through intersections were considered by Clayton [3] in 1941 and perhaps by some other people even earlier. Suppose that cars arrive at regularly spaced time intervals with headway $1/q$ (at an average rate of q cars per unit time). They form a queue during some effective red time R of a traffic light, then during a subsequent green time G they leave at regularly spaced intervals with headway $1/s$ until either the end of the green time or until the queue is gone. In the latter case the cars pass the intersection without delay after the queue is discharged. It is a straightforward exercise to evaluate the queue length at any time for such a model, but even here one encounters a variety of trivial algebraic complications (which we will not describe here) resulting from the discrete nature of the cars and depending upon whether or not $1/q$, $1/s$, R, and G are rational multiples of each other, plus some other things equally irrelevant in the real physical process.

The algebra is greatly simplified, however, if one disregards the discrete nature of the cars and thinks of the traffic as a continuous fluid which arrives at a uniform rate q, is dammed for a time R, and then released at a rate s until the dam is empty. Thereafter it leaves at the arrival rate q provided the green time is long enough.

If $A(\tau)$ is the cumulative number of arrivals in time τ and $D(\tau)$ the cumulative number of departures during a time τ of the green phase in which the queue is nonempty, then

$$A(\tau) = q\tau, \qquad D(\tau) = s\tau. \tag{1}$$

Let $Q(t)$ be the queue length at time t and let $t = 0$ be the start of the red time. The curve of $Q(t)$ in this continuum model (the stored fluid at time t) is the piecewise linear curve shown in Fig. 1, i.e.,

$$Q(t) = \begin{cases} Q(0) + A(t), & 0 < t < R, \\ Q(0) + A(t) - D(t - R), & R < t < t_0, \\ 0, & t_0 < t < R + G, \end{cases} \tag{2}$$

where t_0 is the first time the queue vanishes; that is,

$$Q(0) + A(t_0) - D(t_0 - R) = 0, \tag{3}$$

$$t_0 = \frac{Q(0) + qR}{s - q} + R \tag{4}$$

This content downloaded from 130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

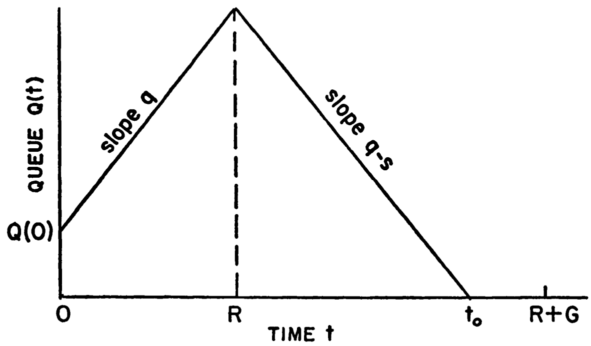

FIG. 1. *The evolution of queue length for continuum approximation*

provided this is less than $R + G$. If $t_0 > R + G$, (2) still applies until the time $t = R + G$.

The total delay for all cars in the queue during a time interval t to $t + dt$ is $Q(t)\,dt$. The total delay for all cars during the cycle $R + G$ is therefore the area under the curve of Fig. 1,

$$W = \int_0^{R+G} Q(t)\,dt; \tag{5}$$

$$W = \begin{cases} \dfrac{s[qR + Q(0)]^2}{2q(s-q)} - \dfrac{Q^2(0)}{2q} & \text{if } t_0 \leqq R + G, \\ GQ(R+G) + RQ(0) + \tfrac{1}{2}qR^2 + \tfrac{1}{2}(s-q)G^2 & \text{if } t_0 \geqq R + G, \end{cases} \tag{6}$$

where for $t_0 > R + G$ the queue $Q(R + G)$ at time $R + G$ is

$$Q(R + G) = Q(0) + qR - (s - q)G. \tag{7}$$

These formulas represent only the simple origin from which a variety of much more difficult mathematical problems arise. First of all, these formulas give only the crudest description of the growth and decay of a queue and most of the more realistic descriptions of this process lead to rather complicated formulas. Secondly, these formulas are not really an end in themselves but only represent one of the elementary processes that enter into the description of the long-time evaluation of one or more traffic streams. Even if one uses these crude formulas to describe a single queue, many of the practical problems—for example, the determination of optimal signal settings for intersecting streams—lead to non-trivial algebra.

A simple application of (1)–(7) is to the fixed-cycle traffic light in which the above pattern of red and green times is repeated periodically. If

$$q(R + G) < sG, \tag{8}$$

then for any finite $Q(0)$ the queue will after finitely many cycles reach a state in which $Q(t)$ vanishes at the start of some red periods and thereafter $Q(t)$ will

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

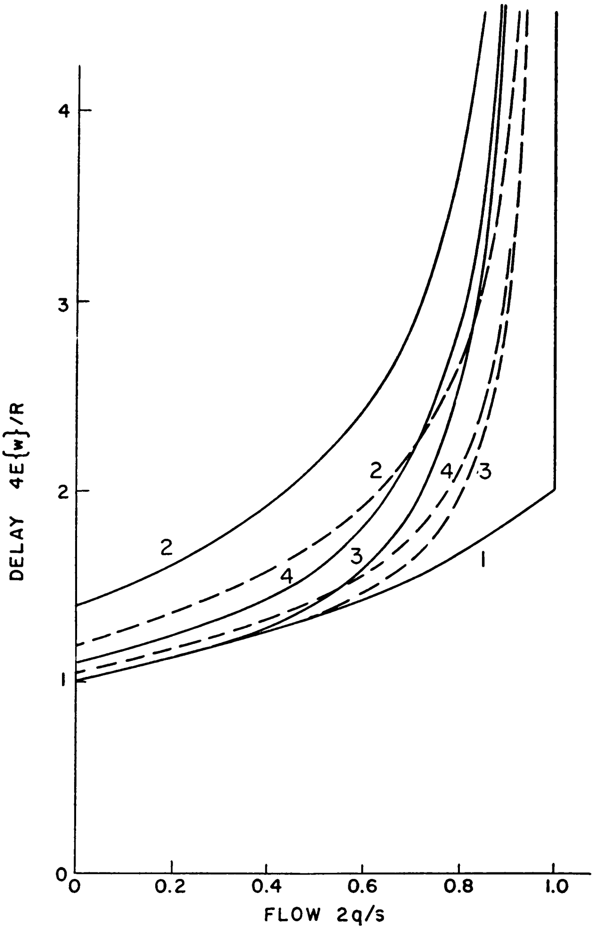

FIG. 2. *The average delay per car vs. the degree of saturation for $R = G$ and $I = 1$. Curve* 1 *is the continuum approximation* (10). *Curves* 2, 3, *and* 4 *represent* (26), (33), *and* (35), *respectively, with the solid lines for $sR = 10$, the broken lines for $sR = 20$.*

also be periodic in time. The wait per cycle in this equilibrium state is, according to (6), given by

$$W = \frac{sqR^2}{2(s - q)}. \tag{9}$$

The average wait per car is this divided by the number of arrivals per cycle, $q(R + G)$, or

$$w = \frac{R^2}{2(R + G)(1 - q/s)}. \tag{10}$$

A graph of delay per car w vs. the arrival rate q is shown in Fig. 2.

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

For $q > sG/(R + G)$, the queues grow and the delay per cycle becomes infinite. Equation (10) is essentially the formula derived by Clayton [3].

For two intersecting streams of traffic, the red time for one stream can be represented as the sum of the green and amber times for the other stream. The total delay for both streams per cycle is then the sum of two expressions obtained from (9) by appropriate identification of the R and G for the two streams. Wardrop [4] uses such a formula to determine the cycle times and red-green splits which minimize the average delay per car.

The continuum approximation has also been used by Gazis and Potts to study oversaturated intersections [5], by Dunne and Potts [6] to study various vehicle-actuated controls in which the red and green times are determined by the lengths of the queues in two or more streams, and by Grace, Pak-Poy, and Morris [7] to study optimal extensions of green times for vehicle-actuated lights.

Appropriate generalization of these approximations for time-dependent arrivals could also be used to study optimal synchronization of a series of traffic lights but this problem has not been studied in any detail yet, at least not from the point of view of delays.

Stochastic approximations for fixed-cycle traffic lights. The desirability of treating traffic as a stochastic process was recognized long ago (at least as early as the 1930's) but Wardrop [4] seems to have been the first to report any calculations of random delays at signal controlled intersections. By computer simulation he obtained delays for a fixed-cycle traffic light with Poisson arrivals and regularly spaced departures during the green. His results agreed with Clayton's formula for low flows but gave much larger average delays as q approached the limit of saturation. The most extensive work on this problem, however, was done by Webster [8] who obtained formulas for the average delay by fitting curves to data obtained by simulation. Webster's work is used by many traffic engineers for the design of interesection controls and his formulas are still among the best available for quick numerical evaluation despite many more recent attempts to improve upon or derive them.

The first attempts at analytic solutions for models of a fixed-cycle traffic light were by Beckmann, McGuire, and Winsten [9] and Newell [10]. The former considered a simple discrete time model with a binomial distribution of arrivals and obtained a formula for the average delay per cycle in terms of an unknown average queue at the start of the red time. The latter proposed a much more elaborate model in which arrival headways were independent identically distributed random variables with a more or less arbitrary distribution and departures were again regularly spaced during the green. Delays could be evaluated approximately for such a model provided that the average queue length at the start of the red period was small compared with the average queue length at the start of the green. Clayton's formula was shown to be a good approximation for any reasonable arrival pattern provided the flow was not too close to the critical value given by (8). The method of approximation failed, however, for nearly critical flows.

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

Except for some papers on special aspects of the traffic light problem by Uematu [11], Haight [12], and Little [13], most of the theoretical work since these early papers resulted from the recognition that values of $Q(t)$ at the start of successive red times represent, in essence, a bulk service queue which had been studied originally by Bailey [14]. Bailey's methods have been exploited by Bisi [15], Newell [16], Meissl [17], [18], [19], and Darroch [20] and found to be particularly well suited for the evaluation of the average delay when q is very close to the critical flow. From this it was shown that the curve of w vs. q in Fig. 2 should have a simple pole at the critical flow rather than an infinite discontinuity.

All papers of this last group deal with very special types of arrival patterns (usually either Poisson or binomial) but it is now quite clear that the average delay is not very sensitive to the detailed stochastic properties of the model. That this is true for low flows was demonstrated in the early papers [4], [8], [9], [10]. That it is true for nearly critical flows was made clear by Kingman's work [1]. (We will show below that it is also true for the intermediate flows between these two limits.) Miller [21] has recently tried to exploit this fact and has produced some fairly simple formulas of accuracy comparable to Webster's for the special case of Poisson arrivals and some generalizations to other types of arrival patterns. Here we shall try to pursue this further by defining more precisely in what sense the delays are independent of the stochastic properties of the system and what sort of limit behavior one should obtain for the delay curve when the cycle times are very long.

In practical applications, values of R and G are typically chosen so that R and G are of comparable size and the saturation flow per cycle sG is of the order of 10. We will approach this problem, however, by imagining a hypothetical situation in which sG is arbitrarily large. We will obtain asymptotic approximations for delays when sG becomes infinite and hope that the formulas are sufficiently accurate as to apply for values of sG as small as about 10.

Normally one would think of the time headways as being more or less fixed and the cycle time as being variable. The above limit would then correspond to $R \to \infty$, $G \to \infty$ with R/G fixed. If, however, we imagine times as being measured in units of the cycle time $R + G$, then R and G will be of order 1 by definition, but the flows q and s, which are now interpreted as number of arrivals or departures per cycle time, become infinite. In any nonzero time interval (in the new units) we would have infinitely many arrivals or departures.

If the arrivals (or departures during the green time) form a stationary stochastic process of any reasonable stochastic structure, the number of arrivals (departures) during any interval of time should satisfy a law of large numbers, i.e., for any limiting process in which the average number of arrivals becomes infinite, then with probability 1 the random number of arrivals N, say, divided by $E\{N\}$ should have the limit 1. In essence, the traffic behaves like a continuous fluid in this limit and the formulas of the last section are the correct limit formulas.

Specifically, we would now interpret $A(\tau)$ and $D(\tau)$ as random variables whose expectations satisfy (1), i.e.,

$$E\{A(\tau)\} = q\tau, \qquad E\{D(\tau)\} = s\tau, \tag{11}$$

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

and which in the limit $sG \to \infty$ satisfy (1) in the sense that

$$\frac{A(\tau)}{sG} \to \frac{q\tau}{sG}, \quad \frac{D(\tau)}{sG} \to \frac{\tau}{G} \qquad \text{with probability 1.} \tag{12}$$

For this to be true it is sufficient that the arrivals (departures) define renewal processes, i.e., the time intervals between arrivals (departures) are independent identically distributed random variables, but this is by no means a necessary condition. Similarly we can interpret $Q(t)$, t_0, W, and w in (2)–(10) as random variables which still satisfy these equations with probability 1 when measured in suitable units; for example, $Q(t)$ in units of sG, t_0 in units of G, W in units of sG^2 and w in units of G. In particular the Clayton formula (10) is an exact limit formula.

The above argument is somewhat academic because in practical situations the cycle time is nowhere near large enough so that one can apply a law of large numbers to the number of arrivals or departures in any small fraction of a cycle time as is implied here, but this is only the crudest of several approximations or interpretations. Furthermore, if we are primarily interested in the total delays and $Q(t)$ is some random function of time, it is not important that (12) be accurate for every value of τ as long as the random area under the curve of $Q(t)$ as in Fig. 1 has small fluctuations. As a practical matter this is still not a very good approximation but at least it begins to make some sense because the maximum queue length is likely to be much larger than 1.

If the fluctuations in queue length cannot be neglected our first goal would probably be to evaluate the expectations of queue lengths, delays, etc., and there is still the possibility that the formulas of the last section yield good approximations for these expectations.

For a fixed-cycle traffic light (in contrast to a vehicle-actuated light) R and G are not random and so we have from (5) that

$$E\{W\} = \int_0^{R+G} E\{Q(t)\}\, dt. \tag{13}$$

Over a long period of time, n cycles say, the number of arrivals should be approximately $nq(R + G)$ (provided the arrivals satisfy a law of large numbers). The total wait during this time is this number of arrivals times the average wait per car, $E\{w\}$. It is also equal to n times the average wait per cycle $E\{W\}$. Thus for almost any reasonable stationary arrival pattern,

$$E\{w\} = \frac{E\{W\}}{q(R + G)}. \tag{14}$$

If we divide W into the sum of two parts W_R and W_G, the waits during the red and green times respectively, the former can be evaluated using (11).

$$E\{W_R\} = \int_0^R E\{Q(0) + A(t)\}\, dt = RE\{Q(0)\} + \int_0^R qt\, dt.$$

Thus

$$E\{W\} = R\, E\{Q(0)\} + \tfrac{1}{2}qR^2 + E\{W_G\}. \tag{15}$$

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

If q/s is small compared with 1 and R/G is comparable with 1, the average time to discharge the queue will be of order qR/s which is small compared with the total green time G. The probability that the queue in its equilibrium cycle will fail to be discharged during the green should be extremely small in this case and furthermore the average wait during the green $E\{W_G\}$ should be small compared with $E\{W_R\}$. Except for highly unusual arrival patterns (for example, arrivals in batches containing of order sG cars even for arbitrarily low flows q) we conclude that for the fixed-cycle traffic light,

$$E\{W\} \sim \frac{1}{2} qR^2 \qquad \text{for } \frac{q}{s} \ll 1. \tag{16}$$

This result agrees with (9) but is now valid for almost any reasonable pattern of arrivals. It does not require the existence of long queues or long cycle times. In fact it is exact in the limit $q \to 0$ for which $E\{Q(t)\} \to 0$.

If $sG \gg 1$ there will be (for most reasonable arrival patterns) a range of q values in which the wait during the green time, $E\{W_G\}$, is not negligible (q is too large for (16) to be valid) but still q is sufficiently small so that there is a negligible probability that in equilibrium the queue fails to clear during the green time, i.e., the term $RE\{Q(0)\}$ can be neglected in (15).

An accurate evaluation of $E\{W_G\}$ is complicated by the fact that the middle expression in (2) is valid only so long as $t < t_0$, but t_0 is itself a random variable determined by (3); the time t_0 is a "first passage time", the first time that $Q(t)$ vanishes. There is an extensive literature on first passage time problems in probability theory [22] and for various types of processes one can evaluate the relevant quantities exactly, but it suffices for the present to make only some very crude estimates. The difficulty stems from the fact that W is not a linear functional of $A(t)$ and $D(t)$. Consequently $E\{W\}$ cannot be exactly evaluated in terms of the expectations of $A(t)$ and $D(t)$ alone.

For $sG \gg 1$, the delays during the green time will be significant only if it requires on the average a significant fraction of the green time to clear the queue. This implies that the average number of arrivals per cycle must be large compared with 1 and under such conditions we would postulate that the relative fluctuations in the number of arrivals or departures should be small.

If we were to replace $Q(t)$ as given by (2) by a random function $Q^*(t)$ which is equal to the first two lines of (2) for all t including $t > t_0$, i.e., $Q^*(t)$ can be negative, then $E\{Q^*(t)\}$ would be given exactly by the curve shown in Fig. 1 for t less than the value t_0 given by (4), i.e., the time at which $E\{Q^*(t)\}$ vanishes. Equation (6) thus represents the average area under the curve of $Q^*(t)$ during this time.

In Fig. 3 are drawn two possible realizations of the random curve $Q^*(t)$, one above the average and one below. For the upper curve, the area under the curve for $Q(t)$ until $Q(t) = 0$ exceeds that under the curve for $Q^*(t)$ until $E\{Q^*(t)\} = 0$ by the shaded area to the right. For the lower curve, $Q^*(t)$ and $Q(t)$ differ in that $Q(t) = 0$ when $Q^*(t) < 0$, and thus the area under $Q(t)$ exceeds that under $Q^*(t)$ by the shaded area to the left.

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

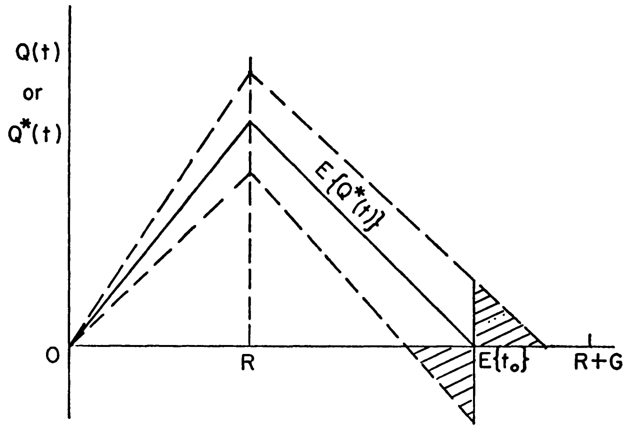

FIG. 3. *Realizations of random queue*

If the fluctuations are small so that these shaded regions have an area small compared with the total area, we can approximate these regions by triangles in which $Q^*(t)$ or $Q(t)$ have slope approximately $(q - s)$ near t_0. To obtain $E\{W\}$ we must therefore add to (6) a term $E\{W_1\}$ equal to the expected area of these triangles, i.e.,

$$E\{W_1\} = \frac{E\{Q^{*2}(t_0)\}}{2(s-q)} = \frac{\text{Var}\,\{A(t_0) - D(t_0 - R)\}}{2(s-q)}$$

if $Q(0) = 0$.

The formulas for $E\{W\}$ or $E\{w\}$ should therefore be larger than the expressions given in (9) and (10) by a fraction

$$\frac{E\{W_1\}}{E\{W\}} = \frac{\text{Var}\,\{A(t_0) - D(t_0 - R)\}}{sqR^2}. \tag{17}$$

For most stationary arrival or departure patterns the coefficient of variance (variance/mean) for the cumulative arrivals and/or departures is bounded for all time, and for large t_0 is nearly constant. (For certain processes this is essentially a consequence of the central limit theorem.) If I_A and I_D are the coefficients of variance for the arrivals and departures and the processes $A(t_0)$ and $D(t_0 - R)$ are uncorrelated, then the numerator of (17) is $(I_A + I_D)qt_0$ for t_0 as given in (4) with $Q(0) = 0$, and

$$\frac{E\{W_1\}}{E\{W\}} = \frac{(I_A + I_D)}{sR(1 - q/s)} = O\left(\frac{1}{sG}\right). \tag{18}$$

For this model of intersection flow we have not given a precise operational definition of red and green times. In the discussion of the continuum model, for example, we did not worry about whether the green time should start at the instant of the first departure, a time $1/s$ before the first departure or perhaps $1/(2s)$ before the first departure. It is actually the last of these which is probably

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

the most reasonable since this guarantees that in the fluid approximation the mean time of the departure for the first unit of fluid occurs at the departure time of the first car, but if we worry about these details we must also consider a number of other effects associated with the discrete nature of traffic, for example, how close sG is to an integer. The relative errors introduced by neglect of such things are all of order $1/(sG)$.

The correction term (18) is of the same order of magnitude as the things which we have, by implication at least, previously discarded. Even if the approximations used to derive (18) may not be very accurate, the order of magnitude should be correct for a wide class of models. The correction (18) is small essentially because to a first approximation the fluctuations in $A(t)$ or $D(t)$ contribute to W a term that may be either positive or negative and that has zero expectation. The first correction to $E\{W\}$ is, therefore, relatively of the order of the square of the relative fluctuations of $A(t)$ and $D(t)$.

We conclude that for typical stochastic flows, Clayton's formula (10) is correct to within a relative error of order $1/(sG)$ except possibly for effects due to the overflow of traffic from one cycle to the next which can arise only for nearly saturated flow.

Delay due to overflow. We consider now the modification in $E\{W\}$ that results when the flow is so close to saturation that statistical fluctuations in the flow cause some cars to be delayed into the next light cycle.

The average total delay $E\{W\}$ during a cycle (13) depends only upon the queue lengths $Q(t)$ during the cycle but not on the order in which the cars are served. To compute $E\{W\}$ it is more convenient to imagine a hypothetical queue discipline of "last come first served" in which the original $Q(0)$ cars are kept in the queue until any new arrivals have been served. For this type of service, however, the delay suffered by the new arrivals is essentially the same as that which would exist for $Q(0) = 0$ as evaluated in the last section. If the flow is nearly saturated, any of the original $Q(0)$ cars that might be served during this cycle will wait nearly the whole cycle, since it will take most, if not all, of the green time to serve the new arrivals. Any of the $Q(0)$ cars which cannot be served during this cycle will, of course, wait exactly the whole cycle.

The average total wait per cycle for all cars is therefore given approximately by

$$E\{W\} \sim \frac{qR^2}{2(1 - q/s)} + (R + G)\, E\{Q(0)\}, \tag{19}$$

the wait for the new arrivals as given by (9) plus that for the queued cars. A formula essentially of this form was given by Beckmann, McGuire and Winsten [9] for a rather special model.

The errors in (19) are again of the same order of magnitude as terms which we have previously discarded. The first term of (19) representing the delay to the new arrivals is in error because the cycle time may end before these cars have been served. This means that in Fig. 3 the point $(R + G)$ may lie in the shaded area on the right (if it lies to the left of this region the flow is oversaturated).

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

We have already shown, however, that this area is negligible, and so it will not make much difference if we cut some of it off by evaluating areas only until time $R + G$. The second term of (19) is in error because some of the $Q(0)$ cars might be served before the end of the cycle (but after the new arrivals are served). In statistical equilibrium, however, there would not be any $Q(0)$ cars unless some new arrivals had failed to clear in the green time of some previous cycle. The time $(R + G)$ must, with a nontrivial probability, overlap the right shaded triangle of Fig. 3. The time which the $Q(0)$ cars can save by being served before the cycle ends is, therefore, of the same order of magnitude as the base of the triangles in Fig. 3. The most cars that could be served in this time is of the same order of magnitude as the height of the triangles. The total time saved by all $Q(0)$ cars is, therefore, at most of the same order of magnitude as the area of the triangles.

Since in the asymptotic approximations for $sG \to \infty$ with R/G fixed we have been discarding contributions to $E\{W\}$ of order R, we should also discard the second term of (19) if $E\{Q(0)\} = O(1)$. The two terms of (19) will actually be of the same order of magnitude only if $E\{Q(0)\} = O(sG)$, i.e., the average queue length at the start of the red is of the same order as the number served in an entire cycle. In any case, we are only interested in queues with $E\{Q(0)\} \gg 1$ and it will still suffice to think of the queue as a random amount of continuous fluid disregarding errors of one unit (one car) or less.

To estimate the distribution function for $Q(0)$, we assume that the arrivals minus departures in one cycle are statistically independent of those in previous cycles (at least to within an error of about one car). Let

$$F_Q(z) = P\{Q(0) \leqq z\} \tag{20}$$

be the distribution function (d.f.) for $Q(0)$ and

$$F_{A-D}(x) = P\{A(R + G) - D(G) < x\} \tag{21}$$

be the d.f. for the arrivals less departures during a complete cycle. If the queue vanishes during the cycle we interpret the latter random variable as the hypothetical one that would have existed had the queue not vanished.

At time $R + G$, the queue length is

$$Q(R + G) = \max\{0, Q(0) + A(R + G) - D(G)\}.$$

In statistical equilibrium the d.f. for $Q(R + G)$ must be the same as for $Q(0)$ and satisfy the equation

$$F_Q(z) = P\{Q(R + G) < z\} = \int_0^\infty F_Q(x)\, dF_{A-D}(z - x). \tag{22}$$

This equation expresses the fact that the event $Q(R + G) < z$ results if $Q(0) < x$ and $A(R + G) - D(G)$ lies in a differential interval at $z - x$ for some $x \geqq 0$. This type of integral equation (Wiener-Hopf integral equation) occurs frequently in a wide variety of physical processes [23] including many aspects of queueing theory where its application was recognized in some of the earliest papers ([24],

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

[25]). There is an extensive literature on the solution of such equations but in keeping with the style which we wish to continue here we shall study this equation in a heuristic way as long as possible.

To obtain a first estimate of $F_Q(z)$ we consider a flow so close to saturation that we can anticipate $E\{Q(0)\}$ will be large compared with $(sG)^{1/2}$ and also that the random variable $Q(0)$ has only a small probability of having a value of order $(sG)^{1/2}$ so that $F_Q((sG)^{1/2}) \ll 1$.

If as in (18) we assume that

$$\operatorname{Var}\{A(R+G) - D(G)\} = Iq(R+G)$$

for some appropriate coefficient of variance I (which for uncorrelated arrivals and departures is the $I_A + I_D$ of (18)), then we expect $Q(R+G) = Q(0) + O(sG)^{1/2}$ with probability close to 1. Thus for sufficiently large sG, $Q(R+G)/Q(0)$ is only slightly different from 1 and the random process $Q(t)/E\{Q(0)\}$ is in essence a Brownian motion on a scale of time in which the cycle time itself is small.

The d.f. $F_Q(z)$ can be approximated by the solution of a diffusion-type equation (Fokker-Planck equation) as occurs in Brownian motion. Since the queue length changes only slightly (in the appropriate relative sense) in one cycle, we expand $F_Q(x)$ in (22) to get

$$F_Q(x) = F_Q(z + (x - z)) = F_Q(z) + (x - z)\frac{dF_Q(z)}{dz} + \frac{(x-z)^2}{2}\frac{d^2F_Q(z)}{dz^2} + \cdots .$$

Thus (22) becomes

$$F_Q(z) \sim F_Q(z)\int_{-\infty}^{z} dF_{A-D}(u) - \frac{dF_Q(z)}{dz}\int_{-\infty}^{z} u\, dF_{A-D}(u) + \frac{1}{2}\frac{d^2F_Q(z)}{dz^2}\int_{-\infty}^{z} u^2\, dF_{A-D}(u).$$

The distribution of $A - D$ is expected to be narrow compared with that of Q so for any z larger than the effective range of $A - D$ we can replace z by $+\infty$ as the upper limit of integration and obtain the equation

$$-\frac{dF_Q(z)}{dz}E\{A - D\} + \frac{1}{2}\frac{d^2F_Q(z)}{dz^2}E\{(A-D)^2\} = 0.$$

The solution of this equation which satisfies the boundary conditions

$$F_Q(\infty) = 1 \quad \text{and} \quad F_Q(0) = 0$$

is

$$F_Q(z) = 1 - e^{-Bz}, \tag{23}$$

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

where

$$B = -2\frac{E\{A - D\}}{E\{(A - D)^2\}}.$$

These approximations are valid only if B is small compared with $(sG)^{-1/2}$ in which case $E\{(A - D)^2\}$ can be replaced by Var $(A - D)$ and

$$B = -\frac{2E\{A - D\}}{\text{Var}\{A - D\}} = \frac{2s}{Iq}\cdot\left[\frac{G}{R+G} - \frac{q}{s}\right]. \tag{24}$$

We thus conclude that for flows sufficiently close to saturation, specifically for $B(sG)^{1/2} \ll 1$, the queue length is approximately exponentially distributed with

$$E\{Q(0)\} \sim B^{-1} = \frac{Iq}{2s}\left[\frac{G}{R+G} - \frac{q}{s}\right]^{-1}. \tag{25}$$

$E\{Q(0)\}$, and consequently also $E\{W\}$ (see (19)), has a simple pole as a function of q as q approaches the saturation flow. Substitution of (25) into (19) and (14) gives

$$E\{w\} = \frac{R^2}{2(R+G)(1-q/s)} + \frac{I}{2s}\left[\frac{G}{R+G} - \frac{q}{s}\right]^{-1}. \tag{26}$$

This is plotted in Fig. 2 for $I = 1$, $R = G$, and $sR = 10$ or 20.

The second term of (26) is valid only near saturation, but for other values of q it is of order $(sG)^{-1}$ compared with the first term and contributes nothing in the limit as $sG \to \infty$. However, even for $q = 0$, the two terms differ by a factor $I(R + G)^2/(sGR^2)$ which for $I = 1$ and $R = G$ is $4/(sG)$. We have the unfortunate situation that "order $1/(sG)$" is not really small for values of sG of interest, namely, $sG \sim 10$. We could have avoided this difficulty by forcing a more rapid decay of the second term of (26) as the flow departs from saturation (we could have made it worse also by doing the opposite) but we did otherwise in order to illustrate the potential dangers of asymptotic approximations.

The methods used to obtain (25) mimic those used in the theory of Brownian motion (see, for example, the review papers by Chandrasekhar, Ornstein and Uhlenbeck [26]) and the theory of Markov processes. Asymptotic results of this type were obtained for queues by Smith [25], whereas Kingman [1] has shown that they are not very sensitive to the stochastic assumptions. Miller [21] has also applied similar approximations to the problem of the fixed-cycle traffic light.

From a practical point of view as applied to the traffic light problem, (25) is still not as accurate as one would like for typical flows with sG of order 10. Prior to the present evaluation of $E\{Q(0)\}$ we had been discarding terms that were of relative order $(sG)^{-1}$ in the formula for $E\{W\}$, equivalent to a unit error in the value of $E\{Q(0)\}$, but (25) is valid only for queues large compared with $(sG)^{1/2}$ or for flows within a fraction, of order less than $(sG)^{-1/2}$, of the limit of saturation. In comparison with errors of traffic measurements, relative errors of order $(sG)^{-1}$ (hopefully about 10% or less) are tolerable but errors of order $(sG)^{-1/2}$ are apt to be a bit too large.

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

To obtain a second approximation for $E\{Q(0)\}$ we return to (22). For nearly saturated flow we have about sG cars arriving per cycle, which should be enough to justify the use of a central limit approximation so that $A(R+G) - D(G)$ should be almost normally distributed.

Suppose we measure arrivals and queue lengths in units of the standard deviation of $A - D$ at saturation, i.e., in units of $(IsG)^{1/2}$. We let

$$\mu = -\frac{E\{A(R+G) - D(G)\}}{(IsG)^{1/2}} = \frac{sG - q(R+G)}{(IsG)^{1/2}}, \tag{27}$$

$$Z = \frac{A(R+G) - D(G)}{(IsG)^{1/2}}, \qquad Q^* = \frac{Q(0)}{(IsG)^{1/2}}. \tag{28}$$

Z is now assumed to be approximately normal with $E\{Z\} = -\mu$ and Var $Z = 1$. The distribution of the rescaled queue length Q^* satisfies the equation

$$F_{Q*}(z) = \int_0^\infty F_{Q*}(x) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z - x + \mu)^2\right] dx \tag{29}$$

analogous to (22), and

$$E\{Q(0)\} = (IsG)^{1/2} E\{Q^*\}. \tag{30}$$

The exact solution of (29) is unfortunately not elementary but this rescaling has at least reduced the problem to the evaluation of a function $E\{Q^*\}$ that depends only upon one parameter μ. There are a variety of methods one can use to solve Wiener-Hopf integral equations but (29) is not one of the simple examples and the best one can do is to obtain various integral representations of the solution. Essentially the same limit behavior arises also for more specialized models and various properties of $E\{Q^*\}$ can be inferred by comparison with results in [16]. Suffice it to say here that the following properties can be derived by appropriate manipulation of (29) and (30).

If we let

$$E\{Q^*\} \equiv \frac{H(\mu)}{2\mu}, \tag{31}$$

then

$$\begin{aligned} H(\mu) &= \frac{2\mu^2}{\pi} \int_0^{\pi/2} \tan^2\theta \left\{-1 + \exp\left[\frac{\mu^2}{2\cos^2\theta}\right]\right\}^{-1} d\theta \\ &= \begin{cases} 1 - (1.164\cdots)\mu + \frac{1}{2}\mu^2 + \cdots & \text{for } \mu \ll 1, \\ \dfrac{\exp\left(-\frac{1}{2}\mu^2\right)}{\sqrt{2\pi}\,\mu^2}\left\{1 - \dfrac{3}{\mu^2} + \dfrac{3\cdot 5}{\mu^4} - \cdots + O\left(\exp\left(-\frac{1}{2}\mu^2\right)\right)\right\} & \text{for } \mu \gg 1. \end{cases} \end{aligned} \tag{32}$$

The function $H(\mu)$ has been defined so as to be a correction factor for (25). The final form of (26) is therefore

This content downloaded from 130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

$$(33)\qquad E\{w\} = \frac{R^2}{2(R+G)(1-q/s)} + \frac{IH(\mu)}{2s}\left[\frac{G}{R+G} - \frac{q}{s}\right]^{-1}.$$

The function $H(\mu)$ (obtained by numerical integration) is shown in Fig. 4. It is everywhere less than 1 and drops off very rapidly for μ larger than about 1. This makes the second term of (33) very small for low flows. The graph of $E\{w\}$ with $sG = 10$ or 20 is also shown in Fig. 2 for comparison with (10) and (26).

Equation (33) is now correct to within a relative error of order at most $(sG)^{-1}$ throughout the entire range of q below the limit of saturation. Furthermore one can estimate that the coefficient of the order $(sG)^{-1}$ error should not be large (probably less than 2). The derivation did not require detailed assumptions about the stochastic properties of the arrivals or departures. The strongest assumption was that they were approximately normally distributed for a complete cycle. Any further refinement of (33) would require a more detailed description of the model and would necessarily involve the introduction of parameters other than the first and second moments of the arrivals and departures.

Comparison with Webster's formula. In the notation of this paper, Webster's formula is

$$(34)\qquad E\{w\} = \frac{R^2}{2(R+G)(1-q/s)} + \frac{1}{2s}\left[\frac{G}{R+G} - \frac{q}{s}\right]^{-1} x - 0.65\,\frac{(R+G)}{(sG)^{2/3}}\,x^{\alpha},$$

where x is the degree of saturation

$$x = \frac{q}{s}\,\frac{(R+G)}{G}$$

and

$$\alpha = \frac{4}{3} + 5\,\frac{G}{(R+G)}.$$

The first term of (34) is the same as the first term of (33). The second terms differ in that (33) has an $IH(\mu)$ where (34) has an x but for $I = 1$ they are at least equal when $x = 1$ which corresponds to $\mu = 0$, $H(\mu) = 1$. The third term of (34) Webster obtained by fitting curves to results of simulation.

Equations (33) and (34) agree numerically to within the expected 10% or 15% accuracy of (33) for the examples shown in Fig. 2 with $sR = 10$ or 20. One can also account for most of the difference between these two formulas.

In the computer simulation Webster used a Poisson distribution of arrivals for which $I_A = 1$ and defined the effective red time R so that the first departure was randomly distributed over the time interval R to $R + 1/s$. The latter has the effect of smoothing out variations in delay due to the discrete nature of the cars so that Webster's red time is essentially the same as that used here. It also causes a nonzero I_D but this is not large enough to be of much significance.

The main difference between Webster's evaluations and those described above must arise from the neglect of the delay represented by (18) and the additional errors introduced in (19). The latter, however, occur only for flows very close

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

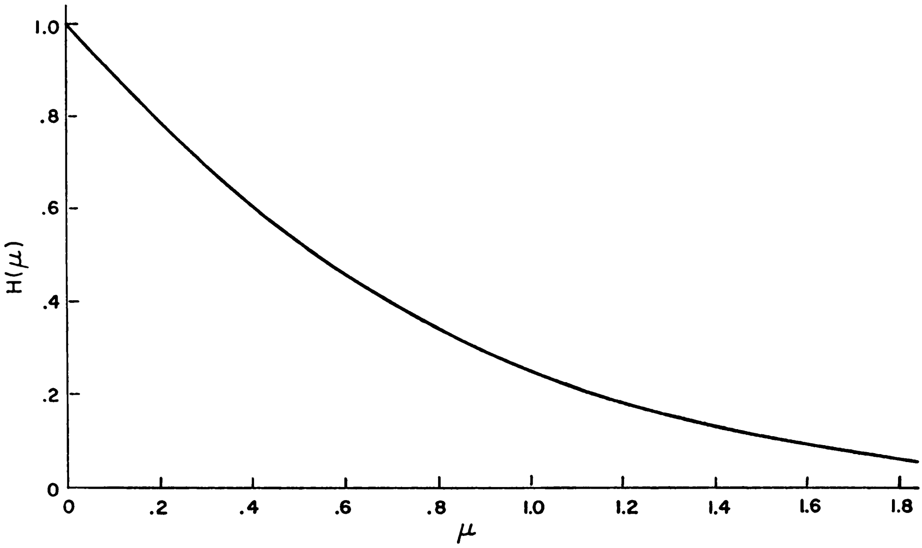

FIG. 4

to saturation and here the delay varies so rapidly with q that small errors in the delay would not be of much practical importance. (There is a limit to the accuracy with which one can measure q). The most important difference between (33) and (34) must therefore come from (18).

Equation (18) was a very crude approximation that certainly is not accurate for flows of about one car per cycle or less and it is not even accurate for higher flows except for those with rather special stochastic structure, but it should be fairly good if the arrivals are Poisson distributed. We would expect therefore that the agreement between (33) and (34) would be improved by adding to (33) the additional term from (18) giving

$$E\{w\} = \frac{R^2}{2(R + G)(1 - q/s)} + \frac{IH(\mu)}{2s}\left[\frac{G}{R + G} - \frac{q}{s}\right]^{-1} + \frac{RI}{2s(R + G)(1 - q/s)^2}. \tag{35}$$

For the examples of Fig. 2 this now agrees with Webster's formula to within 3% or 4% except possibly for low flows, $qR < 1$. This is well within the errors expected in (35). It is probably also within the error between (34) and the simulation calculations. It is certainly within the accuracy necessary for any experimental verification.

Webster's formula, Miller's and those derived here are probably equally useful for estimation in typical applications. The main virtue of the present formula is that it is based upon systematic estimations and the formula is certain to be more accurate for sufficiently large values of sG.

Conclusion. We have presented here a review of the literature on the fixed-

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

cycle traffic light, a new formula for the average delay and a discussion or illustration of some asymptotic approximation methods which can also be applied to a variety of other traffic problems. It is the last of these which should be emphasized most.

Exact solutions of various models of traffic flow are mathematically interesting and are often a useful first step in the study of some practical problem but such solutions are of value only insofar as the results are either insensitive to the details of the model or the model is in fact an accurate representation of the physical process. The ultimate goal is to obtain formulas that apply to as broad a range of processes as possible and contain as few variables as possible, preferably only parameters representing physical quantities that can be easily measured experimentally.

If it is necessary to treat traffic as a stochastic process, it is desirable to do so in such a way that one need only consider the crudest description of its stochastic properties (the details can not be experimentally verified anyway without considerable effort). In many cases one need only consider the mean values of flows in the first approximation and perhaps variances in a second approximation.

The methods used here for the fixed-cycle light can with some modifications be applied also to vehicle-actuated lights, synchronization, etc. It is certainly possible to obtain estimates for a wide variety of traffic problems which one could not readily solve exactly. The present example of the fixed-cycle light is not one of the simplest applications, but in order to obtain accurate estimates of the errors we have made it more complicated than necessary for practical applications.

REFERENCES

[1] J. F. C. Kingman, *On queues in heavy traffic*, J. Roy. Statist. Soc. Ser. B, 24 (1962), pp. 383–392.

[2] Václav E. Beneš, *General Stochastic Processes in the Theory of Queues*, Addison-Wesley, Reading, Massachusetts, 1963.

[3] A. J. H. Clayton, *Road traffic calculations*, J. Inst. Civil Engrs., 16 (1941), pp. 247–284, 558–594.

[4] J. G. Wardrop, *Some theoretical aspects of road traffic research*, Proc. Inst. Civil Engrs. II, 1 (2) (1952), pp. 325–362.

[5] Denos C. Gazis and Renfrey B. Potts, *The oversaturated intersection*, Second International Symposium on Traffic Flow Theory, London, 1963, (to appear).

[6] Michael Dunne and Renfrey B. Potts, *Algorithm for traffic control*, Operations Res., 12 (1964), pp. 870–881.

[7] M. G. Grace, R. W. J. Morris, and P. G. Pak-Poy, *Some aspects of intersection capacity and traffic signal control by computer simulation*, Second Conference Australian Road Research Board, 1964.

[8] F. V. Webster, *Traffic signal settings*, Road Research Laboratory Technical Paper, Her Majesty's Stationery Office, London 1958.

[9] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*, Yale University Press, New Haven, 1956.

[10] G. F. Newell, *Statistical analysis of the flow of highway traffic through a signalized intersection*, Quart. Appl. Math., 13 (1956), pp. 353–369.

[11] Tosio Uematu, *On the traffic control at an intersection controlled by a repeated fixed-cycle traffic light*, Ann. Inst. Statist. Math., 9 (1958), pp. 87–107.

[12] F. A. Haight, *Overflow at a traffic light*, Biometrika, 46 (1959), pp. 420–424.

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms

[13] JOHN D. C. LITTLE, *Approximate expected delays for several maneuvers by a driver in Poisson traffic*, Operations Res., 9 (1961), pp. 39–52.

[14] NORMAN T. J. BAILEY, *On queueing processes with bulk service*, J. Roy. Statist. Soc. Ser. B, 16 (1954), pp. 80–87.

[15] W. BISI, *Controllo del traffico agli incroci stradali*, Statistica, 19 (1959), pp. 422–433.

[16] G. F. NEWELL, *Queues for a fixed-cycle traffic light*, Ann. Math. Statist., 31 (1960), pp. 589–597.

[17] P. MEISSL, *Stochastisches Modell einer festzeitgesteuerten Rot-grün-signalanlage*, Math. Tech.-Wirtschaft, 9 (1962), pp. 11–18, 65–69.

[18] ———, *Zufallsmodell einer Signalanlage mit mehrspurigem Stauraum*, Ibid., 10 (1963), pp. 1–4, 63–68.

[19] ———, *Uber ein Wartezeitenproblem an einem Schalter mit regelmassigen Offnungs und Sperrzeiten*, Österreich. Ing.-Arch., 18 (1964), pp. 79–93.

[20] J. N. DARROCH, *On the traffic-light queue*, Ann. Math. Statist., 35 (1964), pp. 380–388.

[21] A. J. MILLER, *Settings for fixed-cycle traffic signals*, Operational Res. Q., 14 (1963), pp. 373–386.

[22] J. H. B. KEMPERMAN, *The Passage Problem for a Stationary Markov Chain*, The University of Chicago Press, Chicago, 1961.

[23] B. NOBLE, *Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations*, Pergamon Press, New York, 1958.

[24] D. G. KENDALL, *Some problems in the theory of queues*, J. Roy. Statist. Soc. Ser. B., 13 (1951), pp. 151–173.

[25] W. L. SMITH, *On the distribution of queueing times*, Proc. Cambridge Philos. Soc., 49 (1953), pp. 449–461.

[26] N. WAX, *Selected Papers on Noise and Stochastic Processes*, Dover Publications, New York, 1954.

This content downloaded from
130.195.253.8 on Fri, 09 Jun 2023 06:15:23 +00:00
All use subject to https://about.jstor.org/terms