Multilevel Approach to the Design of a Freeway Control System

DONALD R. DREW, KENNETH A. BREWER, JOHANN H. BUHR, and ROBERT H. WHITSON, Texas Transportation Institute, Texas A&M University

This report deals with the definition of the functional and operational requirements for freeway control systems and the actual design and installation of a prototype. The specifications proposed for designing the prototype control system are (a) the optimal use of acceptable freeway gaps by merging ramp vehicles and (b) the prevention of congestion. The underlying philosophy is that minimizing inter-vehicular interference at entrance ramps reduces the probability of rear-end collisions in merging areas due to false starts, reduces the tension on a merging driver, and prevents shock waves from developing on the freeway near entrance ramps. The theory behind this is based on utilizing gap availability and gap acceptance models. Another theory suggests that the prevention of congestion ultimately results in moving more traffic faster. Theoretically, congestion is prevented if demand never exceeds some service volume.

The approach taken in the design of the freeway control system prototype is based on the multilevel concept. The freeway is viewed as a single entity with the control law being split into several degrees of sophistication or levels, with the lower levels directed toward recognizing the influence of short-term factors (gap availability for merging ramp vehicles) and the higher levels reserved for factors that influence performance on a long-term basis (freeway capacity reductions due to accidents, incidents, or geometric bottlenecks).

•IF URBAN freeways are to operate at acceptable levels of service during peak traffic periods, the facilities must be controlled. Although projects to control urban freeways have been established in Detroit, Chicago, Houston, Los Angeles, and Seattle, the theory of designing a control system has not reached the stage where a single unified approach has emerged. Accordingly, the Texas Transportation Institute was awarded a research contract by the U. S. Bureau of Public Roads to develop system design specifications for freeway control by computer control functions integrating local merging control parameters with freeway control parameters. This report concerns functional and operational requirements for such a system and the design and installation of a prototype freeway control system.

The control action in early freeway control systems depended on prior calibration, using historical data. This was not a completely satisfactory approach because the system to be controlled, its environment, and the expected inputs could not be completely described beforehand. Therefore, for classification purposes, this single open-ended control system is of "zero" level because it has no feedback and no memory. Examples of this form of control are complete ramp closure and fixed ramp metering.

Paper sponsored by Committee on Freeway Operations and presented at the 48th Annual Meeting.

The use of surveillance—the continual viewing of freeway traffic in time and space—represents one means of replacing crude open-loop controls with a closed-loop control system (one with a higher sensitivity to such parameter changes as surges in traffic demand and such environmental disturbances as a stalled vehicle or incident on the freeway). Certainly, the closed-circuit television systems on the John C. Lodge Freeway in Detroit and the Gulf Freeway in Houston provide feedback—the comparison of actual freeway operation with the desired operation so that appropriate control action can be taken. Thus, this system may be termed first-level because direct feedback is present for control. Traffic-responsive ramp metering, advisory changeable speed message, and lane closure in response to a deterioration in freeway operations are examples of first-level freeway surveillance-oriented control. Although this type of control system can react to some stimulus related to congestion, it cannot make a conditional choice of action and, therefore, may be described as tactical rather than strategic in nature.

During the last decade, various complex military systems incorporating radar, computers, communication networks, and weapons systems have evolved. Rational methods to control system constituents, which may be spread over a whole continent, and to arrive at a smoothly working entity have been developed. The extension of engineering and scientific methods for designing these systems to achieve proper balance, performance, and economy is called systems engineering. Applying systems engineering to the rational improvement of the undesirable characteristics exhibited by given processes or operation through the addition of instrumentation, controllers, or computers gives automatic control the character of a science and appears to hold great promise as an approach to freeway control.

FREEWAY CONTROL MODEL

Control System Theory

A freeway control system is simply the array of surveillance, communication, and control components designed and connected so as to command or regulate traffic operations. Figure 1 shows one general scheme for representing freeway control systems. Blocks C, A, P, and D stand for the "controller" (analog and/or digital device), the "actuator" (entrance ramp traffic signals and other traffic control displays), the "plant" or "process" (freeway traffic operation), and the "detectors" (sensors, surveillance, instrumentation, and measuring subsystems). Two fundamental variables are the system input and system output, denoted by r (for reference inputs such as traffic demand and desired speeds) and c (for controlled variables such as volume and density). Because the freeway operations phenomenon must be characterized by a multiplicity of inputs and outputs, r and c are vectors.

In Figure 1, d is a disturbance vector that represents unintended inputs to the freeway system that cannot be adjusted, such as environmental factors, weather conditions, accidents, and incidents; m represents signals supplied to the controller regarding those disturbance vector components and output vector components that are instrumentable; n stands for signals from the controller to those control devices in the actuator sub-

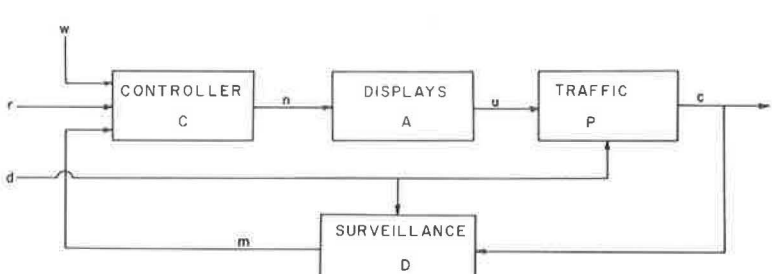


Figure 1. First-level freeway control model.

system, and u is a manipulatable vector that represents those freeway inputs that can be influenced by control. The vector w represents the broad set of operating specifications, restrictions, and hypotheses pertinent to the control problem. Conceivably, r might be considered as a subset of w, except that in some cases it is convenient to distinguish between the two, as shall be seen.

Transfer functions—mathematical descriptions of the ratio of the output of a component to its input—may be written. It is apparent that

$$u = n = f(r, m) = f(r, c, d) \tag{1}$$

In the matter of controlling the system, Eq. 1 may be termed the control law. Ideally a control law is determined so that some performance functional, p, is maximized over a fixed period of time (1). Expressed mathematically, we are trying to accomplish

$$\text{Max. } p = f(u) \tag{2}$$

subject to $c = f(u, d)$ and $f(c, u, r) \geq 0$.

There are several practical problems associated with this development and with system control in general. First, the system vectors and equations developed are rarely known exactly or completely. Second, the mathematical techniques for optimization implied in Eq. 2 are not sufficiently developed to handle the realities and complexities commonly encountered in engineering processes and operations. Third, even where these difficulties can be resolved, the computational problems associated with a single, central controller may make implementation of such a system for on-line control impractical.

There are, moreover, additional problems peculiar to freeway system control. First, there is still a lack of a comprehensive understanding of the traffic stream's behavior. A second area of difficulty regards a suitable analytic description of the system's performance criterion. Ideally, such level of service factors as safety, economy, comfort, and convenience should be included in a system performance criterion.

Failure to understand precisely the complex interactions occurring in freeway system operations is no excuse for proceeding in a haphazard way. The best approach to overcoming these problems lies in separating and resolving the total freeway system control problem into its constituent parts or elements. Two aspects of this approach are (a) the decomposition of the freeway system, and (b) the decomposition of the freeway control function. Both aspects have the objective of reducing the total complex problem into subproblems. The final requirement is a coordination of the subproblem solutions so as to avoid sub-optimization and to achieve the best overall performance of the entire system (2).

Decomposition of Freeway System

After the overall freeway control area has been defined, the system is divided into closed subsystems for analysis. Division points between subsystems should be based on some criteria consistent with the freeway control philosophy. Thus, the division points between subsystems are the freeway "bottlenecks," if the "demand-capacity" philosophy of freeway control is to be employed. If, however, control is to be based on the availability of acceptable gaps in freeway merging areas, each subsystem should be defined so as to contain an entrance ramp, its merging area, and upstream area of influence. For a freeway with substandard entrance ramp geometrics in which the merging areas are in fact the bottlenecks, the subsystem configuration under both control philosophies might very well be the same.

Decomposition of Freeway Control Function

The approach first explained implies that a relatively complex system such as a freeway can be reticulated into a number of independent subsystems, each of which has its own local control law. Another approach, described as the decomposition of the

control function, applies a relatively comprehensive control system to the operation of one of these independent subsystems. The freeway subsystem is viewed as a single entity with the control law being split into several levels or degrees of sophistication.

The general question of control of interacting processes has recently been considered through this new viewpoint as an integral part of a theory of "multilevel systems" (3, 4, 5). The concept is relatively new, with its principal value being the provision of a rational means for developing control configurations for extremely complex systems. The multilevel approach is directed toward establishing a hierarchy of control that results not only in an efficient system, but one that can be implemented in stages. The control hierarchy is established so that lower levels—the zero-level and first-level systems discussed previously—are directed to recognizing the influence of short-term factors, whereas higher levels are reserved for factors that influence performance on a long-term basis. There is also a certain ordering of hierarchical levels based on the degree of complexity of computation, the degree of uncertainty, and the required speed of reaction to a change in operating conditions. The central idea is to share the effort of solution among two or more levels, each of which communicates both with the level directly above and that directly below. Generally, the (n+1)st level influences or even directs the decisions of level n.

Figure 2 shows the conceptual form of a four-level control configuration. These levels are, in ascending order of sophistication, the regulating, the optimizing, the adaptive, and the self-organizing control functions. The basic control activities associated with each layer are identified in the following paragraphs. Most of the variables have been established in previous sections, with superscripts used to denote the respective levels (Fig. 2).

The Regulating Function—This controller accomplished what might be called the basic "subgoal" of the control system. Although the goal of the control system is to provide the best possible level of service on the freeway, its components, and its interfaces, various subgoals have been advanced on which the regulating control subsystem may be based. Implicit is the assumption that optimizing the subgoals will optimize the primary performance criterion. The optimal use of available gaps in the freeway merging process is such a subgoal, and it is accomplished by the regulating controller (Fig. 2). This controller translates the directions of the higher level controllers into direct actions on the operation (the timely release of ramp vehicles by the ramp signal).

The Optimizing Function—The object of this controller (Fig. 2) is the determination of optimum operating conditions based on the appropriate performance criterion and

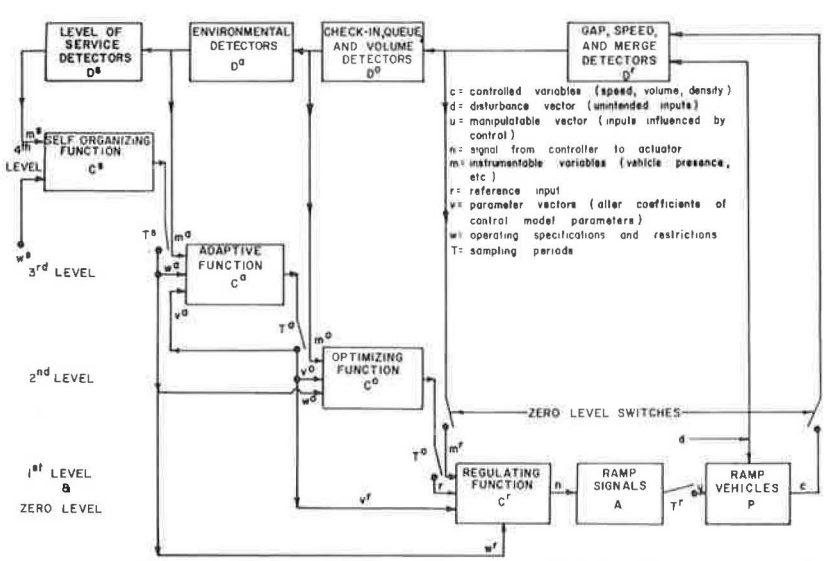


Figure 2. Decomposition of freeway control function.

mathematical model of the process. For example, if the setting on the regulating controller is too high, many marginal gaps are left unfilled; on the other hand, if the setting is too low, many metered vehicles will reject the gaps and be forced to stop in the merging area where their presence, as detected by a loop detector, preempts metering. Obviously the optimum gap setting is somewhere between too high and too low. The form of this optimizing function for a freeway control model will be discussed later.

The Adaptive Function—While the two lowest levels of the control hierarchy are developed on mathematical models approximating the real system, the adaptive controller's function (Fig. 2) is to compensate for the errors introduced by the models by adjusting the parameter values, v^o and v^r. A parameter vector v^a is supplied to the adaptive controller so that, in effect, it can see what it has been doing. The parameter vectors v^r, v^o, and v^a, in effect, alter the coefficients of the control laws that are applicable at the lower control levels, but do not change the laws themselves.

The Self-Organizing Function—This controller (6) determines what the worth or decision vectors w^r, w^o, and w^a should be on the basis of those measurable freeway characteristics m^s and the intervention of humans in the system as represented by w^s. The worth or decision vectors generated by the self-organizing function act to control the overall system to achieve the best total system performance. These decisions are based on the accumulated experience and understanding of the system, and are subject to the specifications, goals, and constraints embodied in the worth vector w^s. These decisions w^r, w^o, and w^a alter the control laws that are applicable at each level in the hierarchy.

In conclusion, it is apparent that the higher up the hierarchy one goes, the less rapidly do the environmental and operational conditions pertinent to a given level vary. For this reason, the outputs of the higher levels are considered to be discrete. This is represented in Figure 2 by the samples ("switches" in the logic circuits), which operate with different periods T, in which

$$T^s > T^a > T^o > T^r \tag{3}$$

Control Synthesis

Decomposition is but a means to an end. It remains to put the submodels together. The final requirement is coordination of the subproblem solutions to ensure their compatibility. An example of some of the difficulties involved may be seen in Figure 2 in that the constraint that $c_i = r_{i+1}$ must be recognized. One can envision the problem arising from a controller in subsystem i attempting to minimize c_i at the same time a similar controller in subsystem i + 1 expects an unlimited r_{i+1} to achieve its optimum performance.

To match the conditions between subsystems to guarantee compatibility, some higher level controllers in the hierarchy must be specified to assess information supplied by the lower level controllers and to provide additional worth or decision vectors to the lower level controllers in order to resolve any conflicting requirements and to introduce environmental or policy conditions not directly accessible by the lower level controllers themselves.

In the freeway control model, it is proposed that this task be resolved at the third (adaptive) level, added to its primary control function. In principle, an n level controller could supervise any number of upstream n - 1 level controllers under its influence. This at least suggests the utilization of an n + 1 level control function to help manage the coordination task of the nth, giving rise to a pyramid of control functions.

FREEWAY CONTROL PROTOTYPE

The process of designing a freeway control system may be described by assuming that the design passes through well-defined phases in chronological order, realizing, however, that a phase is often not recognizable until it has passed. The phases are initiation, organization, preliminary design, principal design, prototype installation, testing, training, and evaluation. This section deals principally with the modeling and design of the prototype based on a multilevel approach.

Prototype Specification

Even in the design of a prototype, the designer must have some specification in mind. That is, he must know what the prototype is supposed to do and establish the means for deciding how well the prototype does it. Thus, the specification is composed of four parts: (a) a description of the location (the plant) and the environment, (b) the description of the inputs (traffic demand, speeds, and other variables), (c) specification of outputs, and (d) measures of effectiveness.

The Plant—The Gulf Freeway in Houston was selected as the proving grounds for this research prototype. Operation on this facility is typical of many urban freeways that have been suffering severe congestion and high accident rates. The Gulf Freeway has three 12-ft lanes in each direction, separated by a 4-ft concrete median with a 6-in. barrier-type curb. The section in the study area extends about 6 miles from the Reveille Interchange to the downtown distribution interchange at Dowling Street. Between these interchanges are eight diamond interchanges containing eight entrance ramps. Frontage roads are one-way and continuous except at three railroad crossings. The through lanes of the freeway overpass the intersecting streets at the interchanges, with the effect of this grade line being a tendency to produce bottlenecks at the overpass as well as limiting the sight distance for the entrance ramp merging maneuver.

Description of Inputs—One difficulty in specifying inputs is in separating those factors that are to be treated as inputs and those that are to be treated as actions of the environment on the system to be controlled. This distinction is important for two reasons: (a) an input is included in functional models, whereas an environmental influence is only important because it modifies the properties of elements; and (b) in using the multilevel approach, inputs are acted upon by the regulating and optimizing controllers, whereas the response to environmental factors is usually handled at the adaptive control level.

In the design specifications for a freeway control system, inputs should be limited to such common traffic variables as volume, speed, density, and gaps, leaving such unintended inputs as bad weather, darkness (during the peak hours in winter), accidents, and incidents to be treated as environmental factors. In the past, comprehensive descriptions of the freeway control system inputs have been made. The relevancy of these descriptions will be discussed later.

Specification of Outputs—One does not just install a control system and then "see what can be done with it." There is often a problem, however, in finding a suitable analytic description of what is desired. Although it is generally accepted that the function of the freeway is to present an environment that permits a driver to operate his vehicle economically, safely, and with a minimum amount of anxiety, it is easier to give a qualitative description than to find an analytic one.

The output specification proposed for the design of the Gulf Freeway prototype consists of two objectives: (a) the optimal use of acceptable freeway gaps by merging ramp vehicles, and (b) the prevention of congestion. The underlying philosophy of the first specification is that minimizing intervehicular interference at entrance ramps reduces the probability of rear-end collisions in merging areas due to false starts, reduces the tension on a merging driver, and prevents shock waves from developing on the freeway in the vicinity of entrance ramps. The theory behind this first specification is based on the utilization of gap availability and gap acceptance models (7, 8). Behind the second specification is the idea that the prevention of congestion ultimately results in moving more traffic faster. Theoretically, congestion is prevented if demand never exceeds some service volume.

Measures of Effectiveness—The fourth, and in many ways the most difficult, component of a system specifications is a set of measures by which the success of a system can be evaluated. Various figures of merit have been proposed to evaluate freeway operations. To some extent they can be categorized according to whether they are macroscopic or microscopic in nature, rational or empirical, designer- or user-oriented, and according to their sensitivity and capability of automatic measurement.

Because of the complexity of the freeway phenomenon and the relevancy of most common measures of effectiveness to the two output specifications discussed, not one, but several measures were employed in the design and evaluation of the Gulf Freeway

control system prototype. The actual utilization of the figures of merit in the test and evaluation of the control system prototype are discussed by Whitson et al (9).

System Design Phase

We have identified the output specification of the prototype control system for the Gulf Freeway as consisting of (a) merging control—making optimal use of gaps in merging areas—and (b) freeway control—preventing a breakdown of operation on the freeway between merging areas. A multilevel control model consisting of a regulating, an optimizing, an adaptive, and a self-organizing function has been developed to provide the theoretical basis for implementing the prototype control system and thus fulfilling the stated objectives of the output specification. Roughly, the regulating and optimizing (first and second) levels are used to accomplish the merging control specification, and the adaptive and self-organizing (third and fourth) levels are employed to fulfill the freeway control specifications.

First-Level Control

To meet the merging control specifications one must be able (a) to detect acceptable gaps on the outside freeway lane, (b) to predict when these acceptable gaps will reach the merging point, and (c) to arrange for the speed adjustment of the merging ramp vehicle so it hits the gap at the merge point. A sensor is required to measure the time interval between successive freeway vehicles (gap detection) and vehicular speed (gap projection). A standard traffic signal installation on the ramp offers a conventional means of communicating with the ramp driver to standardize the required speed adjustment in the merging maneuver. The effect is to stop all ramp vehicles at some point on the ramp far enough from the merge point to allow these vehicles to accelerate to the speed of the freeway traffic system. A call for the green signal is made when the projected gap reaches the position in time (designated the decision point) at which the travel time of the gap to the merge area is the same as the travel time of the ramp vehicle from the signal to the merge area. If the gap is equal to or greater than the designated acceptable gap size for more than one vehicle, the controller holds the green signal until the gap passes the decision point. A loop detector is placed in the pavement of the ramp just upstream from the merge area. All vehicles entering the freeway from

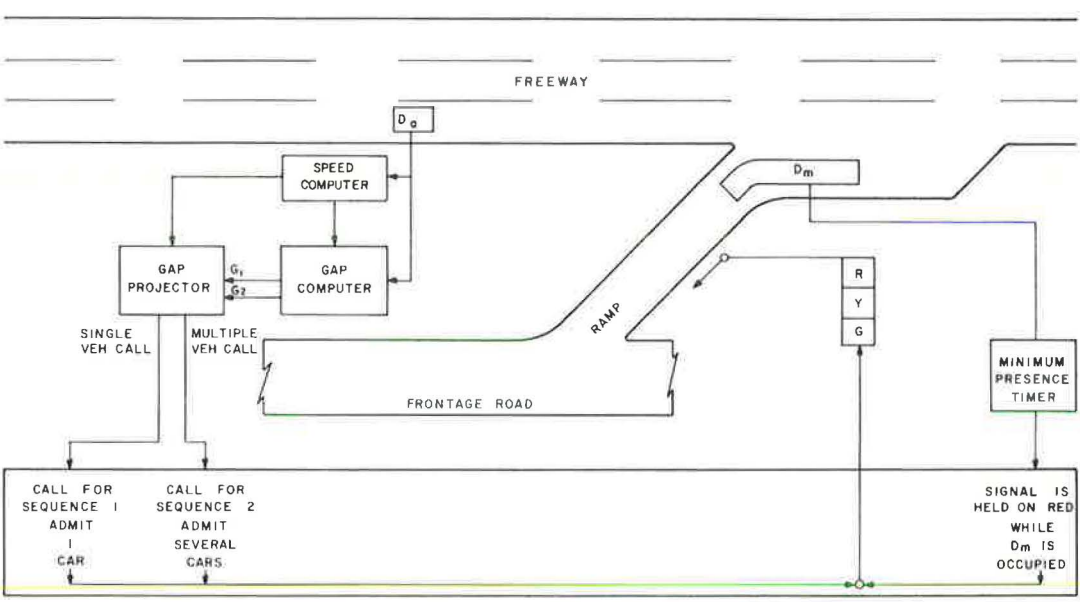


Figure 3. First-level freeway control showing regulating function components.

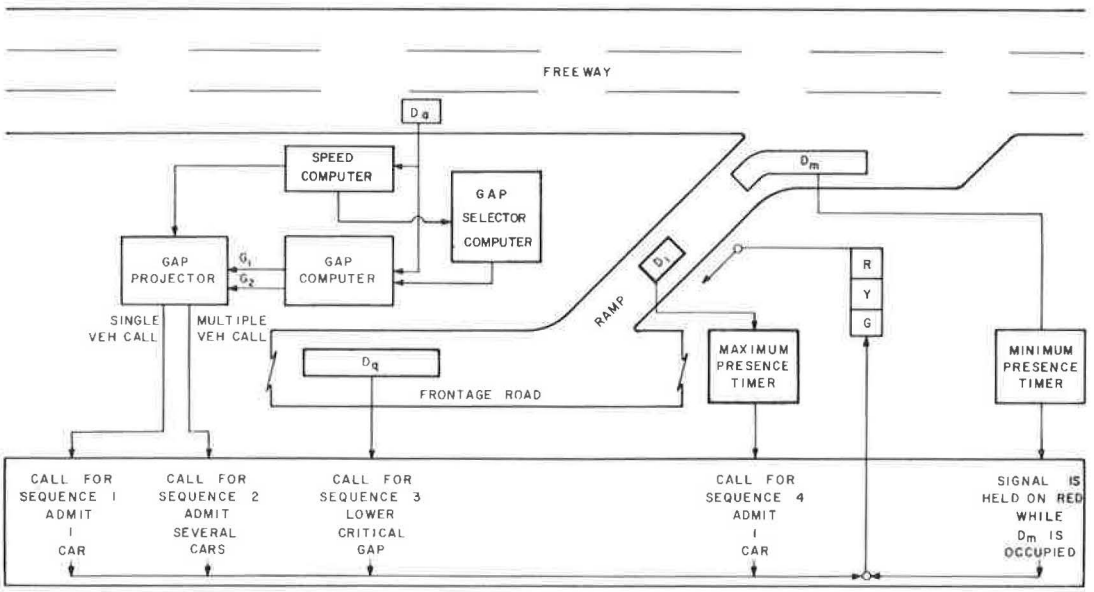


Figure 4. Second-level freeway control with optimizing function components identified.

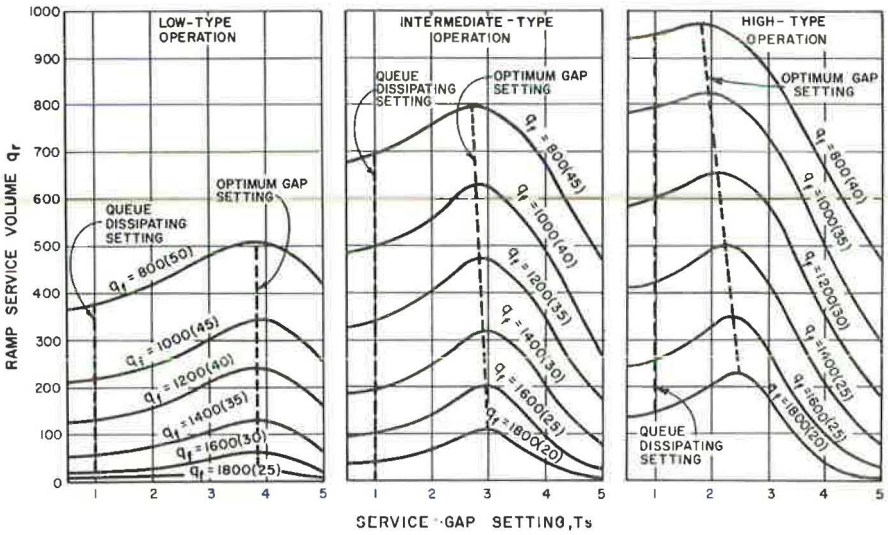
the ramp will actuate the detector. If a vehicle stops on the ramp in this area, blocking the entrance to the freeway, the detector will time out and the signal controller will hold on red until the detector is cleared.

Second-Level Control

Figures 3 and 4 show the functions and components of the first-level and second-level control systems. Figure 3 shows that the control of the ramp signal is accomplished at the first level basically through the detection and projection of gaps. As a provision for keeping the ramp area from the ramp signal to the freeway clear, the presence of a vehicle in the merge area precludes a green signal indication, thus preventing a queue from forming at the merge point and reducing driver anxiety and the potentiality of accidents.

Whereas the first-level regulating controller compares measured gaps in the outside lane of the freeway with an arbitrary gap setting and then meters ramp vehicles into acceptable gaps, there is no assurance that the resulting operation has been optimized with respect to any criteria. The objective of the second-level optimizing controller is to adjust the gap settings on the first-level regulating controller in response to the outside lane freeway operation (volume and speed) so as to maximize the ramp service volume. This is accomplished by the gap selector computer component shown in the second-level freeway control block diagram in Figure 4 according to a family of curves plotted in Figure 5 developed from a mathematical model described by Brewer et al (10). Reference to high-, intermediate-, and low-type operation is based on the criteria of relative speed of merging vehicles with respect to the freeway traffic stream under stable flow conditions.

In the control of an entrance ramp with an hourly demand of q_r, if the freeway outside lane volume q_f can be sensed, the service gap setting T_s can be readily determined. However, this assumes that the freeway is operating under stable and free-flow conditions, and although the control system is designed to maintain such flow conditions, there will be occasions in practice when the freeway reverts to unstable and forced-flow conditions. Any control system should, of course, be versatile enough to accommodate such occurrences. Based on the volume sensed, the control system is unable to distinguish between free flow and forced flow unless speed is also sensed. Therefore, it has been found most practical to sense speed only because this is also required for project-



NOTE: q_f = 1200(40) IS THE OUTSIDE LANE UPSTREAM FREEWAY DEMAND WITH THE SPEED OF THE TRAFFIC STREAM IN PARENTHESES

Figure 5. Service volume control curves.

ing gaps. To convert the volumes in Figure 5 to speeds, use was made of the speed-flow relationship measured on the Gulf Freeway. These appear in the parentheses on the curves of Figure 5. For unstable and forced-flow conditions, gap selection may be described by

$$T_s = T_{min} + (k/v); \; v \leq 25 \text{ mph} \tag{4}$$

where T_s is the service gap for freeway speed v, T_{min} is the minimum service gap at high freeway speed, and k is a constant assumed to equal 15.

Two additional functions of the second-level controller are depicted by sequences 2 and 3 in Figure 4. The ramp may be so located that an excessively long queue at the ramp signal will back into an intersection of the frontage road with a cross-street, thus adversely influencing an adjacent traffic system. To minimize this interference with off-freeway traffic systems, it is necessary to detect such an occurrence with a suitably placed presence detector D_q, which, if continuously occupied for longer than a certain period, will cause vehicles to be metered at a faster rate by reducing the service gap to a minimum gap setting (Fig. 5).

Another loop detector D_i is placed in the vicinity of the ramp signal. If a vehicle is delayed at the signal for longer than a certain period, chances are that the driver will assume the signal to be out of order and proceed past the signal anyway, thus violating the control. This period varies among drivers, of course, but is considerably shorter than at a traffic signal on a regular surface street intersection probably because of the somewhat unconventional location of the signal and the absence of any immediate danger in violating the signal. The violating driver is then more often than not forced to stop in the merge area. It is therefore necessary to have a maximum red phase, insuring that the signal will turn green at least once every so often. In practice, on the Gulf Freeway a 20-second maximum waiting time is used.

Third-Level Control

In designing a freeway control system, the automatic detection and location of a reduction in capacity must be given high priority. This reduction may result from either a bottleneck caused by a deficiency in design or from an incident blocking one or more lanes.

In the vicinity of each entrance ramp, six detectors were installed to monitor the accumulation of traffic in an entrance ramp subsystem. Three detectors, located from 1000 to 1500 ft upstream from the entrance ramp nose, are used to determine freeway demand. The speeds at the location of these detectors are also monitored. Three more detectors are located from 500 to 1000 ft downstream from the ramp nose (past the end of the acceleration lane) and are used primarily to detect reduced capacity operation. Low speeds in this area indicate congestion from a downstream bottleneck, with volume counts at this location used to estimate the capacity of the critical bottleneck.

A downstream capacity reduction, other than a geometric bottleneck, may be caused by either the prevailing ambient conditions or by an incident on the freeway. When this effect is detected and the degree of capacity reduction is measured, using the third-level adaptive function components (Fig. 6), these new parameters must be fed to the optimizing controller, which in turn will modify the critical gap setting on the regulating controller. It should be pointed out that a capacity reduction due to the ambient conditions is predictable, whereas an incident can only be ascertained after the fact. The ambiance components are envisioned ultimately as containing a light-meter, thermometer, and rain gage so as to describe driving conditions as evidenced by visibility and the condition of the pavement.

The adaptive controller's function is to adjust the parameters of the lower controllers to compensate for deviations from the assumptions inherent in the mathematical models governing the lower control functions. Another way of looking at this is that the third-level, or adaptive, controller handles the unexpected inputs, often referred to as environmental factors. Ambient conditions and operational incidents represent two such environmental factors; trucks on the entrance ramp may be interpreted as a third.

The two detectors D_i and D_t will be used to classify vehicles. Classification, although normally thought of as distinguishing between trucks and normal passenger vehicles, is not that simple. The significant difference between the two classes of vehicles as inputs to a control system is not their size, shape, weight, etc., but rather their accelerating characteristics. A fast-accelerating, empty truck may well be placed in the same category as an ordinary passenger vehicle. On the other hand, a slow-accelerating passenger vehicle may have the same effect as a truck. Consequently, the purpose of the classification really is to distinguish between fast- and slow-accelerating vehicles and, once slow-accelerating vehicles have been detected, revise the operation of the metering equipment accordingly. Use of the two detectors should also provide an estimate of the vehicle's length.

Each of the controllers at each entrance ramp should be operated in a manner that results in optimal operation of that particular system. Until now, no stipulation has been made regarding controller configuration (central or local) or type of hardware (all three levels of computer controllers described could be either analog or digital or a combination of both). In the discussions that follow, a central digital computer is envisioned in the role of either a third- and fourth-level controller or both, monitoring the first- and second-level control functions for each entrance ramp subsystem as well as guaranteeing the compatibility between subsystems.

The controlled portion of the Gulf Freeway is divided into 6 closed subsystems that are being monitored by an IBM 1800 digital computer. Since the shock wave resulting from a capacity reduction travels back through the system at a speed of from 15 to 25 mph, it is important to devise some means of early detection so that the entrance ramp metering rates may be adjusted. Using the system of upstream and downstream detectors (D_{1V} - D_{6V}) shown in Figure 6, a point of reduced capacity can be pinpointed to within a few hundred feet.

A capacity reduction can be caused by three different situations. The most common and frequent stoppage is an incident such as a stalled vehicle or accident, either of which could affect the freeway up to several minutes. Adverse ambient conditions and geometric bottleneck represent two other capacity mitigating factors that can produce shock waves. The effects on freeway operation can be just as dramatic. Whereas capacity reductions produced by any of the three causes can be detected, it is apparent that the effects of environmental factors and geometric bottlenecks can be anticipated.

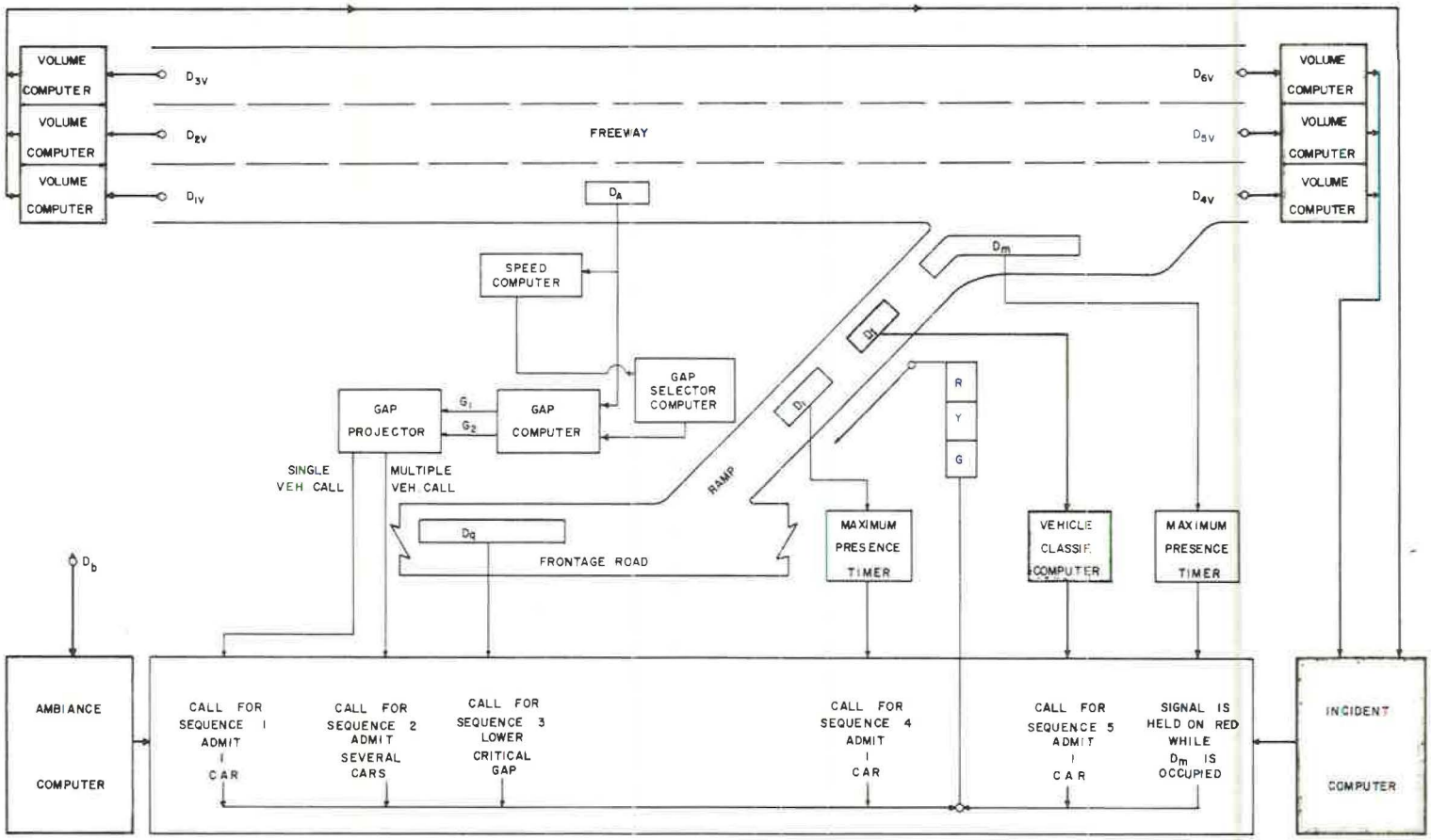


Figure 6. Third-level freeway control with adaptive function components identified.

Fourth-Level Control

In automatic control technology, a number of expressions have been coined to designate the various control systems, such as regulating, optimizing, adaptive, and self-organizing. The fundamental property of a self-organizing or learning control system, as it is often called, is its ability to perform better as time progresses. Using the notation of Figure 2, learning might be implemented as follows: Suppose the optimum performance with respect to a given output specification is accomplished for the parameter settings v_i^j (j = r, o, a) when the input is r_i. Corresponding to a given input r_i, for example, the optimizing system and adaptive systems previously discussed would ultimately settle on the vectors v_i^r, v_i^o, and v_i^a, with the search procedure carried out by C^a always being the same. However, in the fourth-level self-organizing control system, the results of previous computations are stored in memory, which makes it unnecessary that the same lengthy process of attaining the optimum settings v_i^r, v_i^o, and v_i^a be repeated each time the command input r_i is observed by D^a (Fig. 2). The memory of this simple self-organizing level would ultimately consist of a table such that for each r_i there would be a corresponding value of v_i^r, v_i^o, and v_i^a. Let us see how this concept may be used in the optimization of a freeway control system.

The fourth-level computer can be programmed to automatically update the parameters used in the three lower control levels. Capacity reductions offer an example of its application to third-level control. The capacities of geometric bottlenecks, an icy pavement, a section of freeway being paint-striped by a maintenance crew, etc., can be "learned". The curves in Figure 5 afford an example of the utility of the self-organizing controller to second-level control. Since the classification of the merging operation at a given ramp as to high, intermediate, and low is based on relative speed, one function of the fourth-level computer is to measure these relative speeds to be sure that the assumption of a given ramp's classification is, in fact, correct.

Complex control algorithms also may be devised. One such approach utilizes a linear programming model (11). Briefly, this model maximizes the output of the freeway system subject to constraints that keep the demand less than the capacity or some specified service volume through each subsystem and that maintain the feasibility of the solution. Additional constraints applicable to the freeway control system proposed in this paper are the control of ramp queues and maintaining the sum of the merging volume in the outside freeway lane and the entrance ramp less than or equal to a specified merging service volume. The translation of these constraints into local controller gap-settings for the Gulf Freeway control system is treated in another project report (10). One special subroutine in this algorithm particularly dependent on this fourth-level concept is the procedure for learning freeway trip origin and destinations as inputs into the linear programming model.

System Hardware

The hardware required to implement a multilevel freeway control system can be categorized into six basic subsystems: sensors, controllers, traffic signals, transmission subsystem, digital computer, and displays. Sensors (devices embedded in or placed above the roadway to detect vehicles) may be of one of the following types: pressure sensitive, inductive loop, ultrasonic, radar, or magnetic. Controllers transform the computer commands into controls for the signals on the entrance ramps. The traffic signals simply present the traffic control indications to the ramp drivers. The transmission, or communication, subsystem provides a means of transferring information from the sensors and controllers to the computer, and transferring command information from the computer to the ramp controllers (if local controllers are used). The digital computer accumulates the incoming data, performs analyses, makes decisions, and sends commands up and down the four-level hierarchy. Displays are incorporated so that the operator and other observers can monitor the status of the computer, the individual ramps, and the overall freeway traffic operation.

Figure 7 shows the Gulf Freeway Surveillance and Control Center, located on the frontage road south of the Wayside Interchange. Analog controllers, built to the first-

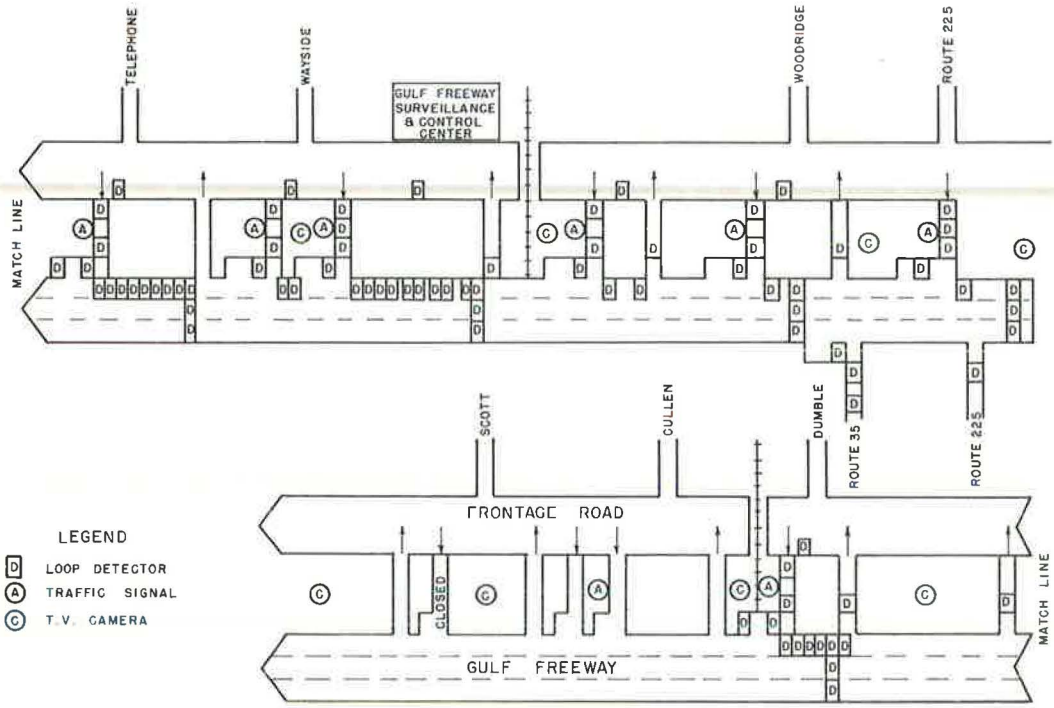


Figure 7. Gulf Freeway prototype surveillance and control system (inbound).

and second-level functional requirements described, have been installed for the control of eight inbound ramps. The new analog computer equipment is shown in Figure 8. In addition, an IBM 1800 digital computer has been installed in the control center (Fig. 9). This computer can be used to either perform the third- and fourth-level functions in conjunction with the local analog controllers, or to perform all four levels as a central computer controller. With the analog and digital computers now installed and operating in the Surevillance and Control Center, a wide range of control measures can be effected—from the simplest to the most sophisticated. This redundancy is not recommended for operational projects, but in a research project this flexibility is needed to compare various control system configurations, to establish control warrants, and to perform cost-effectiveness analyses.

As the name implies, the Center contains surveillance equipment as well as control components. Figure 7 shows the location of a portion of the camera stations for the closed circuit television surveillance system. Closed circuit television is not a part of a freeway control system in the context of the multilevel approach advanced in this paper, it is merely a useful display device to complement the control system. Strategically placed cameras give the observer a view of critical merging areas and bottlenecks for observing the results of decisions changing system operational characteristics. There are other display devices in the control center, such as the digital computer peripheral equipment (plotter, keyboard, etc.), meters in conjunction with the analog controllers, and various event recorders.

PERSPECTIVES

The Enigma of Freeway Control

Freeway traffic control projects are being conducted on sections of the John C. Lodge Freeway in Detroit, the Eisenhower and Dan Ryan Expressways in Chicago, and the Gulf

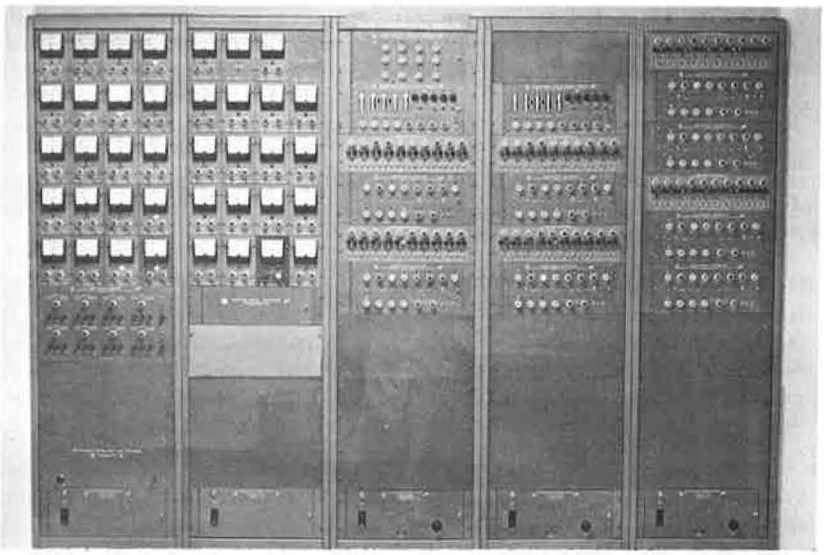


Figure 8. Functional analog computer-controller equipment in Gulf Freeway surveillance and control office.

Freeway in Houston. The results are not only smooth-running freeways and substantial reduction in peak-period accidents, but the aggregate delay to all traffic in the respective corridors has been reduced. Freeway control furnishes the highway administrators with the answer to the creeping innuendo that "freeways won't work."

Figure 9. IBM 1800 digital computer in Gulf Freeway surveillance and control office.

One wonders why, then, if freeway control is such a good thing, there are less than a half-dozen freeways in the country being controlled. Certainly, there are scores of urban freeways throughout the United States experiencing extremely poor operation during the peak hours.

The reason there are not more controlled freeways is the misconception that the design, implementation, and operation of a freeway control system is so complex as to border on the impossible. For example, in a report called "Analysis and Projection of Research on Traffic Surveillance, Communication, and Control" prepared for the Highway Research Board (12), Moskowitz writes:

> Highway administrators will never be able to buy a [freeway control] system, install it, and let it run itself. They will have to face up to the fact that there is such a thing as Operation, and create an organization to operate completed highways. It appears that the most practical way of propagating freeway operational techniques would be to create teams of experts who could act as resident engineers during the formative stages of operating departments.

One can only speculate as to whether or not the same fatalistic prediction was made regarding process control, i. e., that industrialists "will never be able to buy a control system, install it, and let it run itself." Thousands of steel mills, hydroelectric plants, refineries, and chemical plants would never have been automatically controlled if they had required "teams of experts who could act as resident engineers during the formative stages of the operating departments."

The Reality of Freeway Control

Certainly the design and operation of a freeway control system is a challenge; it is not, however, a mystery. Well-meaning researchers have contributed to the confusion surrounding freeway control by not distinguishing between the requirements for an operational control system and a research facility. Almost by tradition it has been assumed that a complete description of a freeway's operating characteristics was needed before the control system could be designed and installed. The implication is that all bottlenecks and thus capacities must be determined, that trip origins and destinations for the freeway and its ramps must be known, and that gap availability and gap acceptance characteristics must be established—all before the control system can be designed. The procedure has been to (a) perform manual system input-output studies requiring as many as 30 people, (b) use time-lapse aerial photography, (c) conduct questionnaire studies to obtain trip information, and (d) employ moving-vehicle study techniques. As it turns out, this part of the design specification for the actual design of a freeway control system is unnecessary.

Basically, the components in the freeway control system consist of one traffic signal per entrance ramp, one merge detector, one gap-speed detector on the outside lane of the freeway, and a regulating controller (Fig. 3). These are the minimum requirements for controlling a single entrance ramp. For controlling the entire freeway, one check-in detector, one vehicle-classification detector, and one queue detector can be added per entrance ramp; one detector per exit ramp; one detector per freeway lane between entrance ramps; and preferably a real-time central computer controller (Fig. 6). This is what is needed and all the studies and measurements explained in the previous paragraph will not change these needs once the output specification proposed in this report, the optimal use of gaps and the prevention of congestion, is chosen. If, however, characteristics are desired, they may be obtained by using the same detection system rather than the manual or aerial techniques.

The purpose of this report is to provide some rationale for the design of a freeway control system based on the multilevel systems approach used in process control. A freeway is just another street and its control is not too different from that of any other street. On a major arterial, intersections become signalized one by one as the control is warranted. In the beginning the traffic engineer responsible for the operation does not have to know anything about network theory or control; he merely installs a detector on each intersection approach, the traffic signals, and the controller, and worries about

calibration of the intersection's operation after the system is working. Eventually, as more intersections are signalized, it may evolve into a complex network problem necessitating a central computer controller to coordinate the local controllers on the arterial and neighboring streets; but even then the detection system, the signal system, and the local controller requirements do not change. Only the problem of calibration and coordination changes, and if this has been anticipated at the time of the signalization of the first intersection (first-level control), the network problem becomes much more tractable.

Many freeway control systems will evolve in the same way, entrance ramp by entrance ramp. Eventually, a collection of controlled entrance ramps must be regulated by some higher order of control just as in the case of a collection of intersections. Yet, the components employed in the control of the individual entrance ramps are the same ones needed for the control of the entire freeway. The control functions will be built up level by level as more and more entrance ramps and freeways are brought under control. While this is being accomplished, those responsible for the operation of the freeway facilities will be gaining expertise. Although the multilevel systems approach has its basis in the theory of controlled processes, it is compatible with the practicalities of the stage implementation of freeway traffic control.

Ideally, new freeways should be designed for control with the detection and transmission systems built in. In this way, surveillance using these sensors would begin the day the freeway is opened to traffic, and freeway control would be implemented as needed—certainly before the demand had built up to a point where it exceeded the capacity. After more than a decade of experimentation with surveillance and control, progress in the implementation of freeway control systems on existing facilities is still lagging because of the lack of a rational, unified approach to the design of such systems. It is believed that one answer lies in the application of the multilevel control theory advanced in this report.

REFERENCES

1. Brosilow, C. B., Lasdon, L. S., and Pearson, J. D. Feasible Optimization Methods for Interconnected Systems. Jour. Automatic Control Preprints, p. 79-84, 1965.
2. Lasdon, L. S., and Schoeffler, J. D. Decentralized Plant Control. ISA Trans., Vol. 2, p. 175-183, April 1966.
3. Sanders, J. L. Multilevel Control. Trans. IEEE, Applications and Industry, p. 473-479, Nov. 1964.
4. Lefkowitz, I. Multilevel Approach Applied to Control System Design. Jour. Automatic Control Prymois, p. 100-109, 1965.
5. Rekoff, M. G. Multilevel Systems. 1966, unpublished.
6. Mesarovic, M. D. Self Organizing Control Systems, Descrete Adaptive Processes. Trans. IEEE, p. 265-269, Sept. 1964.
7. Drew, D. R., McCasland, W. R., Pinnell, C., and Wattleworth, J. A. The Development of an Automatic Freeway Merging Control System. Texas Transportation Institute Research Rept. 24-19, 1966.
8. Drew, D. R. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
9. Whitson, R. H., Buhr, J. H., Drew, D. R., and McCasland, W. R. Real Time Evaluation of Freeway Quality of Traffic Service. Texas Transportation Institute Research Rept. RF 504-4, 1968.
10. Brewer, K. A., Buhr, J. H., Drew, D. R., and Messer, C. J. Ramp Capacity and Service Volume as Related to Freeway Control. Paper presented at 48th Annual Meeting and included in this Record.
11. Wattleworth, J. A. Peak Period Analysis and Control of a Freeway System. Texas Transportation Institute Research Rept. 24-15, 1965.
12. Moskowitz, Karl. Analysis and Projection of Research on Traffic Surveillance, Communication, and Control. NCHRP Rept. 87, unpublished.