Review of Road Traffic Control Strategies

MARKOS PAPAGEORGIOU, FELLOW, IEEE, CHRISTINA DIAKAKI, VAYA DINOPOULOU,
APOSTOLOS KOTSIALOS, AND YIBING WANG

Contributed Paper

Traffic congestion in urban road and freeway networks leads to a strong degradation of the network infrastructure and accordingly reduced throughput, which can be countered via suitable control measures and strategies. After illustrating the main reasons for infrastructure deterioration due to traffic congestion, a comprehensive overview of proposed and implemented control strategies is provided for three areas: urban road networks, freeway networks, and route guidance. Selected application results, obtained from either simulation studies or field implementations, are briefly outlined to illustrate the impact of various control actions and strategies. The paper concludes with a brief discussion of future needs in this important technical area.

***Keywords**—Driver information, freeway network control, intelligent transportation systems, ramp metering, route guidance, traffic signal control, urban network control.*

I. INTRODUCTION

A. Traffic Congestion and the Need for Traffic Control

Transportation has always been a crucial aspect of human civilization, but it is only in the second half of the last century that the phenomenon of traffic congestion has become predominant due to the rapid increase in the number of vehicles and in the transportation demand in virtually all transportation modes. Traffic congestion appears when too many vehicles attempt to use a common transportation infrastructure with limited capacity. In the best case, traffic congestion leads to queueing phenomena (and corresponding delays) while the infrastructure capacity ("the server") is fully utilized. In the worst (and far more typical) case, traffic congestion leads to a degraded use of the available infrastructure (reduced throughput), thus contributing to an accelerated congestion increase, which leads to further infrastructure degradation, and so forth. Traffic congestion results in excess delays, reduced safety, and increased environmental pollution. The following

impressive statement is included in the European Commission's "White Paper—European Transport Policy for 2010": "Because of congestion, there is a serious risk that Europe will lose economic competitiveness. The most recent study on the subject showed that the external costs of road traffic congestion alone amount to 0.5% of Community GDP. Traffic forecasts for the next 10 years show that if nothing is done, road congestion will increase significantly by 2010. The costs attributable to congestion will also increase by 142% to reach € 80 billion a year, which is approximately 1% of Community GDP."

The emergence of traffic (i.e., many interacting vehicles using a common infrastructure) and subsequently traffic congestion (whereby demand temporarily exceeds the infrastructure capacity) have opened new innovation needs in the transportation area. The energy crisis in the 1970s, the increased importance of environmental concerns, and the limited economic and physical resources are among the most important reasons why a brute force approach (i.e., the continuous expansion of the available transportation infrastructure) cannot continue to be the only answer to the ever increasing transportation and mobility needs of modern societies. The efficient, safe, and less polluting transportation of persons and goods calls for an optimal utilization of the available infrastructure via suitable application of a variety of traffic control measures. This trend is enabled by the rapid developments in the areas of communications and computing (telematics), but it is quite evident that the efficiency of traffic control directly depends on the efficiency and relevance of the employed control methodologies. This paper provides an overview of advanced traffic control strategies for three particular areas: urban road networks, freeway networks, and route guidance and information systems.

B. The Control Loop

Fig. 1 illustrates the basic elements of a control loop. The traffic flow behavior in the (road or freeway or mixed) traffic network depends on some external quantities that are classified into two groups.

Fig. 1. The control loop.

Fig. 2. A traffic network.
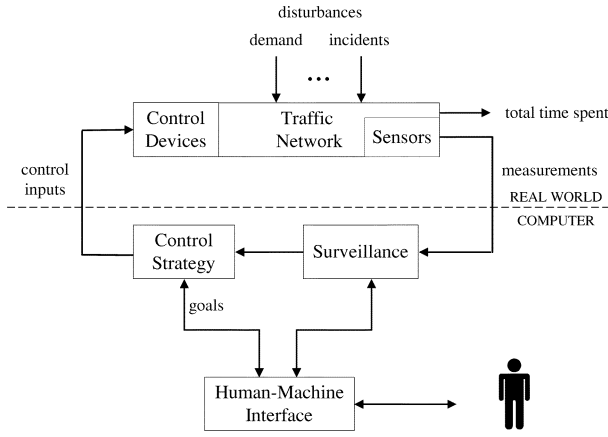
• *Control inputs* that are directly related to corresponding control devices (actuators), such as traffic lights, variable message signs, etc.; the control inputs may be selected from an admissible control region subject to technical, physical, and operational constraints.

• *Disturbances*, whose values cannot be manipulated, but may possibly be measurable (e.g., demand) or detectable (e.g., incident) or predictable over a future time horizon.

The network's output or performance is measured via suitable indices, such as the total time spent by all vehicles in the network over a time horizon. The task of the *surveillance* is to enhance and to extend the information provided by suitable sensors (e.g., inductive loop detectors) as required by the subsequent control strategy and the human operators. The kernel of the control loop is the *control strategy*, whose task is to specify in real time the control inputs, based on available measurements/estimations/predictions, so as to achieve the prespecified goals (e.g., minimization of total time spent) despite the influence of various disturbances. If this task is undertaken by a human operator, we have a manual control system. In an automatic control system, this task is undertaken by an algorithm (the control strategy). The relevance and efficiency of the control strategy largely determines the efficiency of the overall control system. Therefore, whenever possible, control strategies should be designed with care, via application of powerful and systematic methods of optimization and automatic control, rather than via questionable heuristics [1]. Traffic control strategies for urban road and freeway networks is the main focus of this overview paper.

C. A Basic Property

For the needs of this paper we will use a discrete-time representation of traffic variables with discrete time index $k = 0, 1, 2, \ldots$ and time interval (or sampling time) T. A *traffic volume* or *flow* $q(k)$ (in veh/h) is defined as the number of vehicles crossing a corresponding location during the time period $[kT, (k+1)T]$, divided by T. *Traffic density* $\rho(k)$ (in veh/km) is the number of vehicles included in a road segment of length Δ at time kT, divided by Δ. *Mean speed* $v(k)$ (in

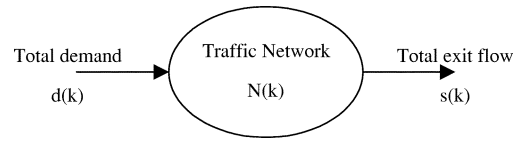km/h) is the average speed at time kT of all vehicles included in a road segment.

We consider a traffic network (Fig. 2) that receives demands $d_i(k)$ (in veh/h) at its origins $i = 1, 2, \ldots$ and we define the total demand $d(k) = d_1(k) + d_2(k) + \ldots$ We assume that $d(k), k = 0, \ldots, K$, is independent of any control measures taken in the network. We define exit flows $s_i(k)$ at the network destinations $i = 1, 2, \ldots$, and the total exit flow $s(k) = s_1(k) + s_2(k) + \ldots$ We wish to apply control measures so as to minimize the total time spent T_s in the network over a time horizon K, i.e.,

$$T_s = T \sum_{k=0}^{K} N(k) \tag{1}$$

where $N(k)$ is the total number of vehicles in the network at time k. Due to conservation of vehicles we have

$$N(k) = N(k-1) + T[d(k) - s(k)] \tag{2}$$

hence

$$N(k) = N(0) + T \sum_{\kappa=0}^{k-1} [d(\kappa) - s(\kappa)]. \tag{3}$$

Substituting (3) in (1) we obtain

$$T_s = T \sum_{k=0}^{K} \left[N(0) + T \sum_{\kappa=0}^{k-1} d(\kappa) - T \sum_{\kappa=0}^{k-1} s(\kappa) \right]. \tag{4}$$

The first two terms in the outer sum of (4) are independent of the control measures taken in the network; hence, minimization of T_s is equivalent to maximization of the following quantity:

$$S = T^2 \sum_{k=0}^{K} \sum_{\kappa=0}^{k-1} s(\kappa) = T^2 \sum_{k=0}^{K-1} (K - k) s(k). \tag{5}$$

Thus, minimization of the total time spent in a traffic network is equivalent to maximization of the time-weighted exit flows. In other words, the earlier the vehicles are able to exit the network (by appropriate use of the available control measures) the less time they will have spent in the network.

*D. Traffic Congestion Revisited*

The above basic property may be used to explain and quantify the degradation of the network infrastructure caused by traffic congestion, as well as to demonstrate via simple mathematics the enormous potential of improvement via suitable traffic control measures.

As an example, we consider (Fig. 3) two cases for a freeway on-ramp, (a) without and (b) with metering control.
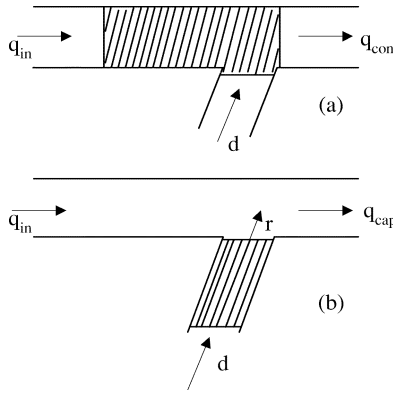
Fig. 3. Two cases, (a) without and (b) with ramp metering; grey areas indicate congestion zones.
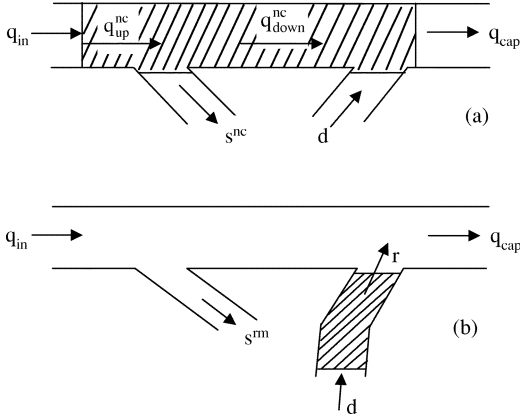

Fig. 4. Two cases, (a) without and (b) with ramp metering.

Let q_{in} be the upstream freeway flow, d be the ramp demand, q_{con} be the mainstream outflow in presence of congestion, and q_{cap} be the freeway capacity. It is well known that the outflow q_{con} at the head of the congestion is lower by some 5%–10% than the freeway capacity q_{cap}. Note that $q_{\text{in}} + d > q_{\text{cap}}$ (else the congestion would not have been created). In Fig. 3(b) we assume that ramp metering may be used to maintain capacity flow on the mainstream, e.g., by using the control strategy ALINEA (see Section III-C). Of course, the application of ramp metering creates a queue at the on-ramp but, because q_{cap} is greater than q_{con} (increased outflow!), ramp metering leads to a reduction of the total time spent (including the ramp waiting time). It is easy to show that the amelioration ΔT_s (in %) of the total time spent is given by

$$\Delta T_s = \frac{q_{\text{cap}} - q_{\text{con}}}{q_{\text{in}} + d - q_{\text{con}}} 100. \tag{6}$$

For example, if $q_{\text{in}} + d = 1.2 q_{\text{cap}}$ (i.e., the total demand exceeds the freeway capacity by 20%) and $q_{\text{con}} = 0.95 q_{\text{cap}}$ (i.e., the capacity drop due to the congestion is 5%) then $\Delta T_s = 20\%$ results from (6), which illustrates the level of achievable improvement.

In addition, we consider (Fig. 4) two cases of a freeway stretch that includes an on-ramp and an off-ramp, namely, (a) without and (b) with metering control. In order to clearly separate the different sources of degradation, we will assume
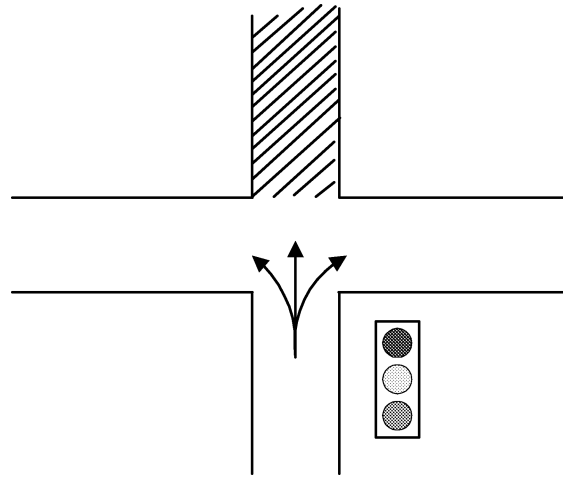
Fig. 5. Detrimental effects of overspilling queues in urban road networks.

here that $q_{\text{con}} = q_{\text{cap}}$, i.e., no capacity drop due to congestion. Defining the exit rate γ ($0 \le \gamma \le 1$) as the portion of the upstream flow that exits at the off-ramp, it is easy to show that the exit flow without control is given by

$$s^{\text{nc}} = \frac{\gamma}{1 - \gamma}(q_{\text{cap}} - d) \tag{7}$$

while with metering control we have

$$s^{\text{rm}} = \gamma \cdot q_{\text{in}}. \tag{8}$$

Because $(1 - \gamma)q_{\text{in}} + d > q_{\text{cap}}$ holds (else the congestion would not have been created), it follows that s^{nc} is less than s^{rm}; hence, ramp metering increases the outflow, thus decreasing the total time spent in the system. It is easy to show that the amelioration of the total time spent in this case amounts to

$$\Delta T_s = \gamma \cdot 100. \tag{9}$$

For example, if the exit rate is $\gamma = 0.05$, then the amelioration is $\Delta T_s = 5\%$. If several upstream off-ramps are covered by the congestion in absence of ramp metering (which is typically the case in many freeways during rush hours) then the amelioration achievable via introduction of ramp metering is accordingly higher.

Summing up these effects in a freeway network, an overall amelioration of total time spent by as much as 50% (i.e., halving of the average journey time) may readily result (see Section III-D) due to the increased throughput enabled by ramp metering application.

Similar effects may be observed in saturated signal-controlled urban traffic networks (Fig. 5). A saturated link prevents the traffic movements at the upstream intersection to cross, even though they have the right of way (green signal). This is a waste of resources (waste of green time) that contributes to an accelerated increase of congestion due to vehicles trapped in the upstream links, which leads to blocking of further upstream intersections, increased waste of green time, and so forth [2]. This vicious circle frequently leads to gridlocks in network cycles with devastating effects for the traffic flow in extended urban areas.
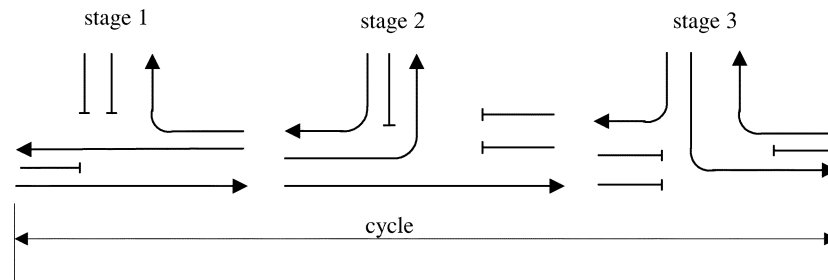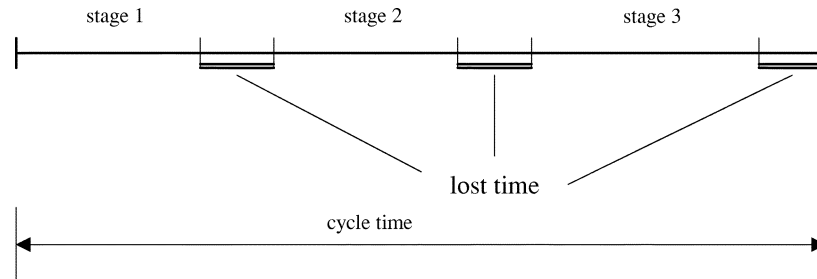
Fig. 6. Example of signal cycle.

Fig. 7. Cycle time and lost times.

The outlined phenomena make clear that the extended congestion encountered daily in modern freeway and urban road networks are not only due to excessive demand exceeding the network capacity. As a matter of fact, demand may temporarily exceed the capacity of specific links leading to limited congestion. The infrastructure degradation, however, caused by the initially limited congestion, leads to an unstable escalation when no suitable control systems are employed to counter this devastating evolution. In conclusion, the observed extended (in both space and time) congestion in modern metropolitan areas is indeed triggered by a temporarily and locally excessive demand, but it is expanded and maintained due to the lack of suitable control actions that would prevent the corresponding infrastructure degradation.

II. ROAD TRAFFIC CONTROL

A. Basic Notions

Traffic lights at intersections is the major control measure in urban road networks. Traffic lights were originally installed in order to guarantee the safe crossing of antagonistic streams of vehicles and pedestrians; with steadily increasing traffic demands, it was soon realized that, once traffic lights exist, they may lead (under equally safe traffic conditions) to more or less efficient network operations, hence there must exist an optimal control strategy leading to minimization of the total time spent by all vehicles in the network.

Although the corresponding optimal control problem may be readily formulated for any road network, its real-time solution and realization in a control loop like the one of Fig. 1 faces a number of apparently insurmountable difficulties.

- The red–green switchings of traffic lights call for the introduction of discrete variables, which renders the optimization problem combinatorial.
- The size of the problem for a whole network is very large.

- Many unpredictable and hardly measurable disturbances (incidents, illegal parking, pedestrian crossings, intersection blocking, etc.) may perturb the traffic flow.
- Measurements of traffic conditions are mostly local (via inductive loop detectors) and highly noisy due to various effects.
- There are tight real-time constraints, e.g., decision making within 2 s for advanced control systems.

The combination of these difficulties renders the solution of a detailed optimal control problem infeasible for more than one intersection. Therefore, proposed control strategies for road traffic control introduce a number of simplifications of different kinds or address only a part of the related traffic control problems. Unfortunately, most proposed simplifications render the corresponding control strategies less suitable to address traffic saturation phenomena.

An *intersection* consists of a number of approaches and the crossing area. An *approach* may have one or more lanes but has a unique, independent queue. Approaches are used by corresponding *traffic streams* (veh/h). A *saturation flow s* is the average flow crossing the stop line of an approach when the corresponding stream has right of way (r.o.w.), the upstream demand (or the waiting queue) is sufficiently large, and the downstream links are not blocked by queues. Two *compatible* streams can safely cross the intersection simultaneously, else they are called *antagonistic*. A *signal cycle* is one repetition of the basic series of signal combinations at an intersection; its duration is called *cycle time c*. A *stage* (or *phase*) is a part of the signal cycle, during which one set of streams has r.o.w. (Fig. 6). Constant *lost* (or *intergreen*) *times* of a few seconds are necessary between stages to avoid interference between antagonistic streams of consecutive stages (Fig. 7).

There are four possibilities for influencing traffic conditions via traffic lights operation.

- *Stage specification*: For complex intersections involving a large number of streams, the specification of the optimal number and constitution of stages is a nontrivial task that can have a major impact on intersection capacity and efficiency.
- *Split*: This is the relative green duration of each stage (as a portion of the cycle time) that should be optimized according to the demand of the involved streams.
- *Cycle time*: Longer cycle times typically increase the intersection capacity because the proportion of the constant lost times becomes accordingly smaller; on the other hand, longer cycle times may increase vehicle delays in undersaturated intersections due to longer waiting times during the red phase.
- *Offset*: This is the phase difference between cycles for successive intersections that may give rise to a "green wave" along an arterial; clearly, the specification of offset should ideally take into account the possible existence of vehicle queues.

Control strategies employed for road traffic control may be classified according to the following characteristics.

- *Fixed-time strategies* for a given time of day (e.g., morning peak hour) are derived off-line by use of appropriate optimization codes based on historical constant demands and turning rates for each stream; *traffic-responsive strategies* make use of real-time measurements (typically one or two inductive loops per link) to calculate in real time the suitable signal settings.
- *Isolated strategies* are applicable to single intersections while *coordinated strategies* consider an urban zone or even a whole network comprising many intersections.
- Most available strategies are only applicable to *undersaturated* traffic conditions, whereby vehicle queues are only created during the red phases and are dissolved during the green phases; very few strategies (see Section II-D) are suitable also for *oversaturated* conditions with partially increasing queues that in many cases reach the upstream intersections.

B. Isolated Intersection Control

1) Fixed-Time Strategies: Isolated fixed-time strategies are only applicable to undersaturated traffic conditions. *Stage-based strategies* under this class determine the optimal splits and cycle time so as to minimize the total delay or maximize the intersection capacity. *Phase-based* strategies determine not only optimal splits and cycle time but also the optimal staging, which may be an important feature for complex intersections.

Well-known examples of stage-based strategies are SIGSET and SIGCAP proposed in [3] and [4]. Assuming m prespecified stages, SIGSET and SIGCAP specify the splits $\lambda_1, \ldots, \lambda_m$ and the cycle time c. Note that

$$\lambda_0 + \lambda_1 + \cdots + \lambda_m = 1 \qquad (10)$$

holds by definition, where $\lambda_0 = L/c$, and L is the total lost time in a cycle. In order to avoid queue building, the following capacity constraint must hold for each stream j:

$$s_j \sum_{i=1}^{m} \alpha_{ij} \lambda_i \geq d_j \quad \forall j \qquad (11)$$

where s_j and d_j are the saturation flow and the demand, respectively, of stream j; α_{ij} is one if stream j has r.o.w. at stage i, and zero else. Inequality (11) requires that the demand d_j of stream j should not be higher than the maximum possible flow assigned to this stream. Finally, a maximum-cycle and m minimum-green constraints are also taken into account.

A nonlinear total delay function derived by Webster [5] for undersaturated conditions is used in SIGSET as an optimization objective. Thus, SIGSET solves a linearly constrained nonlinear programming problem to minimize the total intersection delay for given stream demands d_j. On the other hand, SIGCAP may be used to maximize the intersection's capacity as follows. Assume that the real demand is not d_j as in (11) but $\mu \cdot d_j$ with $\mu \geq 1$. SIGCAP replaces d_j in (11) by $\mu \cdot d_j$ and maximizes μ under the same constraints as SIGSET, which leads to a linear programming problem.

Note that, for reasons mentioned earlier, capacity maximization always leads to the maximum allowable cycle time. Clearly, SIGCAP should be used for intersections with high demand variability in order to prevent oversaturation, while SIGSET may be used under sufficient capacity margins by replacing d_j in (11) by $p_j d_j$, where $p_j \leq 1$ are prespecified margin parameters.

Phase-based approaches [6] solve a similar problem, suitably extended to consider different staging combinations. Phase-based approaches consider the compatibility relations of involved streams as prespecified and deliver the optimal staging, splits, and cycle time, so as to minimize total delay or maximize the intersection capacity. The resulting optimization problem is of the binary-mixed-integer-linear-programming type, which calls for branch-and-bound methods for an exact solution. The related computation time is naturally much higher than for stage-based approaches, but this is of minor importance, as calculations are performed offline.

2) Traffic-Responsive Strategies: Isolated, traffic-responsive strategies make use of real-time measurements provided by inductive loop detectors that are usually located some 40 m upstream of the stop line, to execute some more or less sophisticated vehicle-actuation logic. One of the simplest strategies under this class is the *vehicle-interval method* that is applicable to two-stage intersections. Minimum-green durations are assigned to both stages. If no vehicle passes the related detectors during the minimum green of a stage, the strategy proceeds to the next stage. If a vehicle is detected, a critical interval (CI) is created, during which any detected vehicle leads to a green prolongation that allows the vehicle to cross the intersection. If no vehicle is detected during CI, the strategy proceeds to the next stage, else a new CI is created, and so forth, until a prespecified maximum-green value is reached. An extension of the method also considers the traffic demand on the antagonistic approaches to decide whether to proceed to the next stage or not.

A more sophisticated version of this kind of strategies was proposed by Miller [7] and is included in the control tool MOVA [8]. Miller's strategy answers every T seconds (e.g., $T = 2$) the question: *Should the switching to the next stage take place now, or should this decision be postponed by T?* To answer this question, the strategy calculates (under certain simplifying assumptions) the time gains and losses caused in all approaches if the decision is postponed by $\kappa \cdot T$ seconds. The corresponding net time gains $J_\kappa, \kappa = 1, 2, \ldots$, are combined in a single criterion $J = \max\{J_\kappa; \kappa = 1, 2, \ldots\}$, and if $J < 0$, the switching takes place immediately, else the decision is postponed until the next time step.

A comparative field evaluation of these simple algorithms is presented in [9].

C. Fixed-Time Coordinated Control

The most popular representatives of this class of strategies for urban networks are outlined below. By their nature, fixed-time strategies are only applicable to undersaturated traffic conditions.

1) MAXBAND: The first version of MAXBAND was developed by Little [10]; see also [11]. MAXBAND considers a two-way arterial with n signals $S_1, \ldots, S_n$ (intersections) and specifies the corresponding offsets so as to maximize the number of vehicles that can travel within a given speed range without stopping at any signal (green wave); see Fig. 8. Splits are considered in MAXBAND as given (in accordance with the lateral street demands); hence, the problem consists in placing the known red durations (see the horizontal lines of each signal S_i in Fig. 8) of the arterial's signals so as to maximize the inbound and outbound bandwidths $\bar{b}$ and b, respectively. For an appropriate problem formulation, it is necessary to introduce some binary decision variables, which leads to a binary-mixed-integer-linear-programming problem. The employed branch-and-bound solution method benefits from a number of nice properties of this particular problem to reduce the required computational effort. Attempts to farther reduce the computational effort required by the method are reported in [12]. Little [10] extended the basic MAXBAND method via incorporation of some cycle constraints to render it applicable also to networks of arterials; see also [15].

MAXBAND has been applied to several road networks in North America and beyond. A number of significant extensions have been introduced in the original method in order to consider a variety of new aspects (see [13]) such as: time of clearance of existing queue, left-turn movements, and different bandwidths for each link of the arterial (MULTIBAND) [14], [15].

2) TRANSYT: TRANSYT was first developed by Robertson [16] but was substantially extended and enhanced later. It is the most known and most frequently applied signal control strategy, and it is often used as a reference method to test improvements enabled by real-time strategies. First field implementations of TRANSYT-produced signal plans indicated savings of some 16% of the average travel time through the network.
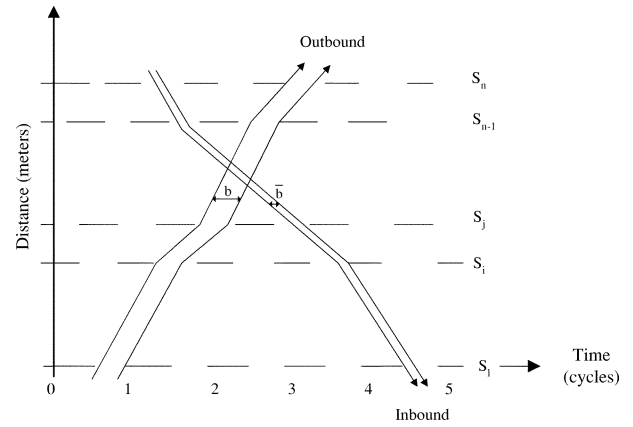

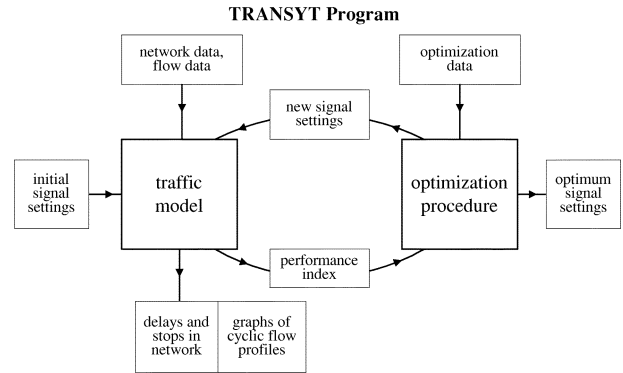**Fig. 8.** A maximum band along an arterial (after [10]).

Fig. 9. Structure of TRANSYT (after [16]).

Fig. 9 depicts the method's basic structure: TRANSYT is fed with the initial signal settings including the prespecified staging, the minimum green durations for each stage of each intersection, and the initial choice of splits, offsets, and cycle time. A unique cycle time c or $c/2$ is considered for all network intersections in order to enable offset coordination. The network and traffic flow data comprise the network's geometry, the saturation flows, the link travel times, the constant and known turning rates for each intersection, and the constant and known demands. The traffic model consists of nodes (intersections) and links (connecting streets). The concept of "platoon dispersion" (dynamic first-order time-delay system) is used to model flow progression along a link. Oversaturated conditions cannot be described, although some improvement has been achieved in this respect in a recent enhanced release of the program [17]. The method proceeds in an iterative way: For given values of the decision variables (control inputs), i.e., of splits, offsets, and cycle time, the dynamic network model calculates the corresponding performance index, e.g., the total number of vehicle stops. A heuristic "hill-climb" optimization algorithm introduces small changes to the decision variables and orders a new model run, and so forth, until a (local) minimum is found.

3) Drawbacks of Fixed-Time Strategies: The main drawback of fixed-time strategies is that their settings are based on historical rather than real-time data. This may be a crude simplification for the following reasons.

• Demands are not constant, even within a time-of-day.

• Demands may vary at different days, e.g., due to special events.

• Demands change in the long term leading to "aging" of the optimized settings.

• Turning movements are also changing in the same ways as demands; in addition, turning movements may change due to the drivers' response to the new optimized signal settings, whereby they try to minimize their individual travel times [18].

• Incidents and farther disturbances may perturb traffic conditions in a nonpredictable way.

For all these reasons, traffic-responsive coordinated strategies, if suitably designed, are potentially more efficient, but also more costly, as they require the installation, operation, and maintenance of a real-time control system (sensors, communications, central control room, local controllers).

D. Coordinated Traffic-Responsive Strategies

1) SCOOT: SCOOT was first developed by Robertson's team [19] and has been extended later in several respects. It is considered to be the traffic-responsive version of TRANSYT and has been applied to over 150 cities in the United Kingdom and elsewhere. SCOOT utilizes traffic volume and occupancy (similar to traffic density) measurements from the upstream end of the network links. It runs in a central control computer and employs a philosophy similar to TRANSYT. More precisely, SCOOT includes a network model that is fed with real measurements (instead of historical values) and is run repeatedly in real time to investigate the effect of incremental changes of splits, offsets, and cycle time at individual intersections (functionally decentralized operation). If the changes turn out to be beneficial (in terms of a performance index), they are submitted to the local signal controllers; see [20] for a comparative field evaluation. SCOOT's performance deteriorates in case of saturated traffic conditions.

2) Model-Based Optimization Methods: More recently, a number of more rigorous model-based traffic-responsive strategies have been developed: *OPAC* [21], *PRODYN* [22], *CRONOS* [23], *RHODES* [24]. These strategies do not consider explicitly splits, offsets, or cycles. Based on prespecified staging, they calculate in real time the optimal values of the next few switching times $\tau_i, i = 1, 2, \ldots$, over a future time horizon H, starting from the current time t and the currently applied stage. To obtain the optimal switching times, these methods solve in real time a dynamic optimization problem employing realistic dynamic traffic models with a sampling time of 2–5 s, fed with traffic measurements. The models include discrete variables to reflect the impact of red/green phases on traffic flow. Several constraints, e.g., for maximum and minimum splits, are included. The typical performance index to be minimized is the total time spent by all vehicles.

The *rolling horizon* procedure (similar to model predictive control approaches [25]) is employed for real-time application of the results. Hereby, the optimization problem is solved in real time over a time horizon H (e.g., 60 s) using measurement-based initial traffic conditions and demand predictions over H, but results are applied only for a much shorter roll period h (e.g., 4 s), after which new measurements are collected and a new optimization problem is solved over an equally long time horizon H, and so forth. The rolling horizon procedure avoids myopic control actions while embedding a dynamic optimization problem in a traffic-responsive (real-time) environment.

The basic problem faced by these strategies is due to the presence of discrete variables that require exponential-complexity algorithms for a global minimization. In fact, OPAC employs complete enumeration (assuming integer switching times) while PRODYN and RHODES employ dynamic programming. Due to the exponential complexity of these solution algorithms, the control strategies (though conceptually applicable to a whole network) are not real-time feasible for more than one intersection. Hence, we end up with a number of decentralized (by intersection) optimal strategies, whose actions may be coordinated heuristically by a superior control layer (see, e.g., [26], [27]). On the other hand, CRONOS employs a heuristic global optimization method with polynomial complexity which allows for simultaneous consideration of several intersections, albeit for the price of specifying a local (rather than the global) minimum.

3) Store-and-Forward Based Approaches: Store-and-forward modeling of traffic networks was first suggested by Gazis and Potts [28], [29] and has since been used in various works notably for road traffic control [30]–[39]. The main idea when using store-and-forward models for road traffic control is to introduce a model simplification that enables the mathematical description of the traffic flow process without use of discrete variables. This is of paramount importance because it opens the way to the application of a number of highly efficient optimization and control methods (such as linear programming, quadratic programming, nonlinear programming, and multivariable regulators) with polynomial complexity, which, on its turn, allows for coordinated control of large-scale networks in real time, even under saturated traffic conditions.

The critical simplification is introduced when modeling the outflow u_i of a stream i. Assuming sufficient demand on the link, the outflow u_i at discrete time k is set

$$u_i(k) = (g_i(k)/c)s_i \qquad (12)$$

where $g_i(k)$ is the green time duration for this stream and s_i is the corresponding saturation flow. If the sampling time T is equal to the cycle time c, Fig. 10 illustrates that $u_i(k)$ in (12) is equal to the *average* flow during the corresponding cycle, rather than equal to s_i during the green phase and equal to zero during the red phase (this corresponds to a pulse width modulation in electrical engineering). In other words, (12) suggests that there is a continuous (uninterrupted) outflow from each network link (as long as there is sufficient demand). The consequences of this simplification are as follows.
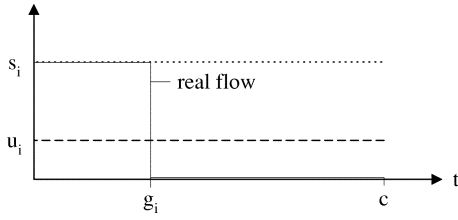
Fig. 10. Simplified modeling of link outflow u_i.

1) The sampling time T of the discrete-time representation cannot be shorter than the cycle time c; hence, real-time decisions cannot be taken more frequently than at every cycle.

2) The oscillations of vehicle queues in the links due to green/red commutations are not described by the model.

3) The effect of offset for consecutive intersections cannot be described by the model.

Despite these consequences, the appropriate use of store-and-forward models may lead to efficient coordinated control strategies for large-scale networks as demonstrated in simulation studies in some of the aforementioned references. In fact, the use of (12) leads to a linear state-space model for road networks of arbitrary size, topology, and characteristics [40], [41] (bold variables indicate vectors and matrices)

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{B}\Delta\mathbf{g}(k) + \mathbf{D}\Delta\mathbf{d}(k) \qquad (13)$$

where the state $\mathbf{x}$ is the vector of the numbers of vehicles x_i in network links i; $\mathbf{B}$ and $\mathbf{D}$ are constant matrices reflecting the network characteristics; $\Delta\mathbf{g} = \mathbf{g}(k) - \mathbf{g}^N$ and $\Delta\mathbf{d} = \mathbf{d}(k) - \mathbf{d}^N$; $\mathbf{g}$ (the control input) is the vector of green times g_i for each stage i in all intersections of the network, while $\mathbf{g}^N$ comprises some corresponding constant nominal green times g_i^N; $\mathbf{d}$ (the disturbance vector) and $\mathbf{d}^N$ comprise the demand flows d_i and the constant nominal flows d_i^N respectively. Suitable bounds for minimum green times and maximum storage capacity of links must also be considered [40].

In order to minimize the risk of oversaturation and the spillback of link queues, one may attempt to minimize and balance the links' relative occupancies $x_i/x_{i,\max}$ where $x_{i,\max}$ (in veh) is the storage capacity of link i. A quadratic criterion that considers this control objective has the general form

$$\Im = \frac{1}{2}\sum_{k=0}^{K}\left(\|\mathbf{x}(k)\|_{\mathbf{Q}}^2 + \|\Delta\mathbf{g}(k)\|_{\mathbf{R}}^2\right) \qquad (14)$$

where $\mathbf{Q}$ and $\mathbf{R}$ are nonnegative definite, diagonal weighting matrices. The first term in (14) accounts for minimization and balancing of the relative occupancies of the network links. To this end, the diagonal elements of $\mathbf{Q}$ are set equal to $1/x_{i,\max}$ [42]. Furthermore, the magnitude of the control reactions can be influenced by the choice of the weighting matrix $\mathbf{R} = r\mathbf{I}$. To this end, the choice of r may be performed via a trial-and-error procedure so as to achieve a satisfactory control behavior for a given application network.

The outlined optimization problem is of the quadratic programming type and may be readily solved by use of broadly available codes even for large-scale networks. For a real-time application, the corresponding algorithm may be embedded in a rolling horizon procedure. Alternatively one may neglect the future demand in (13); set the optimization horizon $K \to \infty$ in (14); and apply the control bounds externally [42]. Under these assumptions one obtains a Linear-Quadratic-Regulator problem, which permits a closed-loop (feedback) solution

$$\mathbf{g}(k) = \mathbf{g}^N - \mathbf{L}\mathbf{x}(k) \qquad (15)$$

where the gain matrix $\mathbf{L}$ results as a straightforward solution of the corresponding discrete-time Riccati equation. This is the multivariable regulator approach taken by the signal control strategy TUC [41], [42] to calculate in real time the network splits, while cycle time and offset are calculated by other parallel algorithms [43]. TUC has been implemented and is currently operational in a part of Glasgow, U.K.'s [44] and Chania, Greece's [45] urban networks with quite satisfactory results, particularly under saturated traffic conditions.

Fig. 11 illustrates instances of some recent simulation results obtained via application of TUC within the microscopic traffic flow simulator AIMSUN [46] for a part of the urban network of Tel Aviv, Israel [43] (notice a light-rail line between the opposite directions of vehicle lanes). The left column of Fig. 11 [i.e., (a), (c), and (e)] displays the traffic situation in a network part at 7:30 A.M., 8:00 A.M., 8:30 A.M., respectively, under fixed-time signal control settings that were optimized according to the utilized morning-peak demand. The right column of Fig. 11 [i.e., (b), (d), and (f)] displays the same network part at the same times when TUC is applied under identical demands. It may be seen that at 7:30 A.M. under fixed-time control, long vehicle queues are created at internal links. The resulting queue spillback phenomena limit the network throughput and lead to the development of queues even at some origin links [see Fig. 11(c) and (e)]. In contrast, TUC manages to keep the link queues within limits, thus protecting upstream intersections from detrimental queue spillback and maintaining a high network throughput. As a result, traffic conditions under TUC are back to normal at 8:30 A.M. (the demand has been served), while long queues persist under fixed control at the same time [Fig. 11(e) and (f), respectively].

For further theoretical developments on control strategies applicable to saturated traffic conditions, see [2], [47]–[49].

E. Integrated Urban-Freeway Traffic Control

Modern metropolitan traffic networks include both urban roads and freeways and employ a variety of control measures such as signal control, ramp metering (see Section III), variable message signs and route guidance (see Section IV). Traditionally, control strategies for each type of control measure are designed and implemented separately, which may result in antagonistic actions and lack of synergy among different control strategies and actions. However, modern traffic networks that include various infrastructure types, are
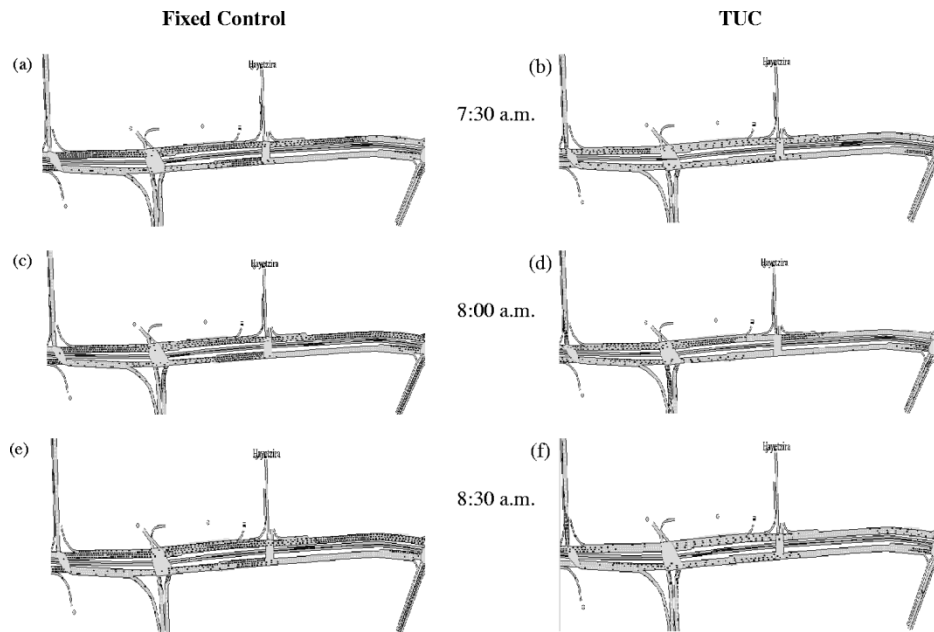
Fig. 11. Comparative microscopic simulation results using fixed-time signal control at (a) 7:30 A.M., (c) 8:00 A.M. (e) 8:30 A.M.; and TUC at (b) 7:30 A.M., (d) 8:00 A.M., (f) 8:30 A.M.

perceived by the users as an entity, and all included control measures, regardless of their type or location, ultimately serve the same goal of higher network efficiency. Integrated control strategies should consider all control measures simultaneously toward a common control objective. Despite some preliminary works on this subject (see [50], for an early status report), the problem of control integration is quite difficult due to its high dimensions that reflect the geographical extension of the traffic network [39], [40], [51]–[55]. For this reason, it appears that store-and-forward modeling (at least for the urban road part) might be the only feasible way to design and operate in real time a unique integrated control strategy. The aforementioned Glasgow implementation covers in fact control measures of various types (signal control, ramp metering, variable message signs) via partial interconnection of three feedback strategies [44], see Fig. 12; the displayed integrated control strategy IN-TUC incorporates the partly cooperating feedback strategies TUC for signal control (see Section II-D), ALINEA for ramp metering (see Section III-C), and a reactive one-shot route guidance strategy for user optimum (see Section IV-C).

III. FREEWAY TRAFFIC CONTROL

A. Motivation

Freeways had been originally conceived so as to provide virtually unlimited mobility to road users, without the annoyance of flow interruptions by traffic lights. The rapid increase of traffic demand, however, led soon to increasingly severe congestions, both *recurrent* (occurring daily during rush hours) and *nonrecurrent* (due to incidents). The increasingly congested freeways within and around metropolitan areas resemble the urban traffic networks before introduction of traffic lights: Chaotic conditions at intersections, long queues, degraded infrastructure utilization, reduced safety. At the present stage, responsible authorities have not fully realized that the expensive freeway or freeway-network infrastructure is strongly underutilized on a daily basis due to the lack of efficient and comprehensive traffic control systems (see Section I-D). In other words, the expensive infrastructure is intended to deliver a nominal capacity that is not available (due to congestion), ironically, exactly at the time it is most urgently needed (during peak hours).

The control measures that are typically employed in freeway networks are the following.

- *Ramp metering*, activated via installation of traffic lights at on-ramps or freeway interchanges.
- *Link control* that comprises a number of possibilities including lane control, variable speed limits, congestion warning, tidal (reversable) flow, keep-lane instructions, etc.
- *Driver information and guidance systems*, either by use of roadside variable message signs or via two-way communication with equipped vehicles (see Section IV).

Ramp metering is the most direct and efficient way to control and upgrade freeway traffic. Various positive effects are achievable if ramp metering is appropriately applied:

- increase in mainline throughput due to avoidance or reduction of congestion;
- increase in the served volume due to avoidance of blocked off-ramps or freeway interchanges;
- utilization of possible reserve capacity on parallel arterials;
- efficient incident response;
- improved traffic safety due to reduced congestion and safer merging.

Some recent studies have demonstrated that efficient ramp metering strategies (employing optimal control algorithms)
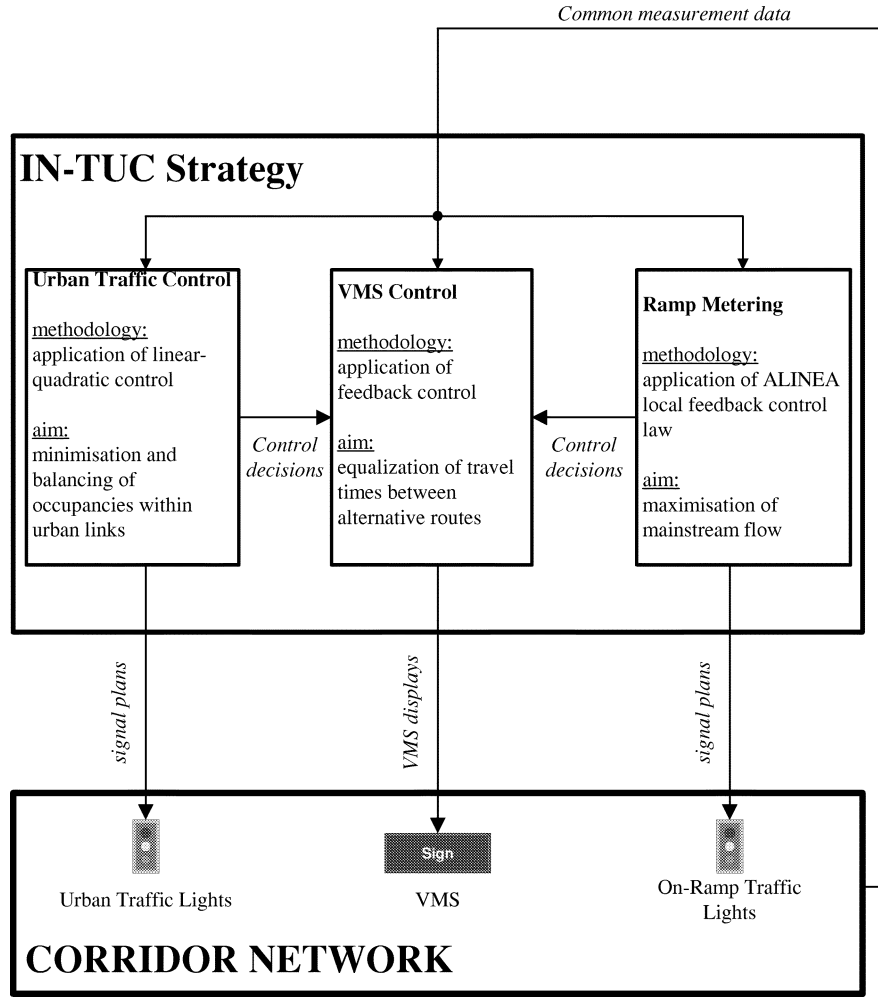
Fig. 12. Functional architecture of IN-TUC strategy as implemented in Glasgow, U.K.

may provide spectacular improvements (50% reduction of total time spent) in large-scale freeway networks [56], [58]. This may also be demonstrated via suitable treatment of real (congested) freeway traffic data; see [59].

B. Fixed-Time Ramp Metering Strategies

Fixed-time ramp metering strategies are derived off-line for particular times of day, based on constant historical demands, without use of real-time measurements. They are based on simple static models. A freeway with several on-ramps and off-ramps is subdivided into sections, each containing at most one on-ramp. We then have

$$q_j = \sum_{i=1}^{j} \alpha_{ij} r_i \qquad (16)$$

where q_j is the mainline flow of section j, r_i is the on-ramp volume of section i, and $\alpha_{ij} \in [0, 1]$ expresses the (known) portion of vehicles that enter the freeway in section i and do not exit the freeway upstream of section j. To avoid congestion

$$q_j \leq q_{\text{cap,j}} \quad \forall j \qquad (17)$$

must hold, where $q_{\text{cap,j}}$ is the capacity of section j. Further constraints are

$$r_{j,\min} \leq r_j \leq \min\{r_{j,\max}, d_j\} \qquad (18)$$

where d_j is the demand at on-ramp j. This approach was first suggested by Wattleworth [60]. Other similar formulations may be found in [61]–[66].

As an objective criterion, one may wish to maximize the number of served vehicles (which is equivalent to minimizing the total time spent [67])

$$\sum_j r_j \rightarrow \text{Max} \qquad (19)$$

or to maximize the total travel distance

$$\sum_j \Delta_j q_j \rightarrow \text{Max} \qquad (20)$$

(where Δ_j is the length of section j), or to balance the ramp queues

$$\sum_j (d_j - r_j)^2 \rightarrow \text{Min.} \qquad (21)$$

These formulations lead to linear programming or quadratic programming problems that may be readily solved

by use of broadly available computer codes. An extension of these methods that renders the static model (16) dynamic by introduction of constant travel times for each section was suggested in [68].

The drawbacks of fixed-time ramp metering strategies are identical to the ones discussed under road traffic control (Section II-C). In addition, fixed-time ramp metering strategies may lead (due to the absence of real-time measurements) either to overload of the mainstream flow (congestion) or to underutilization of the freeway. In fact, ramp metering is an efficient but also delicate control measure. If ramp metering strategies are not accurate enough, then congestion may not be prevented from forming, or the mainstream capacity may be underutilized (e.g., due to groundlessly strong metering).

C. Reactive Ramp Metering Strategies

Reactive ramp metering strategies are employed at a tactical level, i.e., in the aim of keeping the freeway traffic conditions close to prespecified set values, based on real-time measurements.

1) Local Ramp Metering: Local ramp metering strategies make use of traffic measurements in the vicinity of a ramp to calculate suitable ramp metering values. The *demand-capacity strategy* [69], quite popular in North America, reads

$$r(k) = \begin{cases} q_{\mathrm{cap}} - q_{\mathrm{in}}(k-1), & \text{if } o_{\mathrm{out}}(k) \leq o_{\mathrm{cr}} \\ r_{\mathrm{min}}, & \text{else} \end{cases} \quad (22)$$

where (Fig. 13) q_{cap} is the freeway capacity downstream of the ramp, q_{in} is the freeway flow measurement upstream of the ramp, o_{out} is the freeway occupancy (similar to density) measurement downstream of the ramp, o_{cr} is the critical occupancy [at which the freeway flow becomes maximum; see Fig. 13(c)], and r_{min} is a prespecified minimum admissible ramp flow value. The strategy (22) attempts to add to the last measured upstream flow $q_{\mathrm{in}}(k-1)$ as much ramp flow $r(k)$ as necessary to reach the downstream freeway capacity q_{cap}. If, however, for some reason, the downstream measured occupancy $o_{\mathrm{out}}(k)$ becomes overcritical (i.e., a congestion may form), the ramp flow $r(k)$ is reduced to the minimum flow r_{min} to avoid or to dissolve the congestion.

Comparing the control problem in hand with Fig. 1, it becomes clear that the ramp flow r is a control input, the downstream occupancy o_{out} is an output, while the upstream freeway flow q_{in} is a disturbance. Hence, (22) does not really represent a closed-loop strategy but an open-loop disturbance-rejection policy (Fig. 13(a)) which is generally known to be quite sensitive to various further nonmeasurable disturbances.

The *occupancy strategy* [69] is based on the same philosophy as the demand-capacity strategy, but it relies on occupancy-based estimation of q_{in}, which may, under certain conditions, reduce the corresponding implementation cost.

An alternative, closed-loop ramp metering strategy (ALINEA), suggested in [70], reads

$$r(k) = r(k-1) + K_R[\hat{o} - o_{\mathrm{out}}(k)] \quad (23)$$
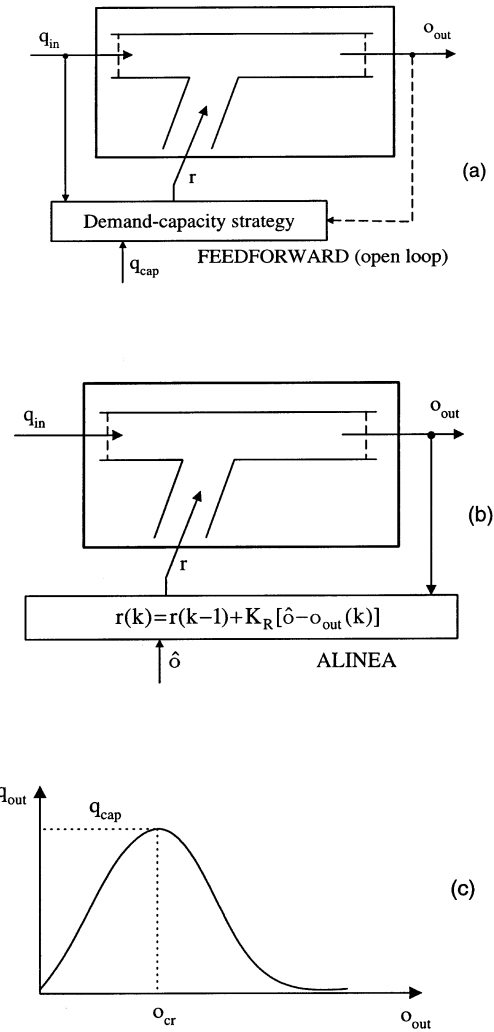
Fig. 13. Local ramp metering strategies: (a) Demand capacity. (b) ALINEA. (c) Fundamental diagram.

where $K_R > 0$ is a regulator parameter and $\hat{o}$ is a set (desired) value for the downstream occupancy [typically, but not necessarily, $\hat{o} = o_{\mathrm{cr}}$ may be set, in which case the downstream freeway flow becomes close to q_{cap}, see Fig. 13(c)]. ALINEA is obviously an integral regulator; hence, it is easily seen that at a stationary state (i.e., if q_{in} is constant), $o_{\mathrm{out}}(k) = \hat{o}$ results from (23), although no measurements of the inflow q_{in} are explicitly used in the strategy. In field experiments, ALINEA has not been very sensitive to the choice of the regulator parameter K_R.

Note that the demand-capacity strategy reacts to excessive occupancies o_{out} only after a threshold value (o_{cr}) is exceeded, and in a rather crude way, while ALINEA reacts smoothly even to slight differences $\hat{o} - o_{\mathrm{out}}(k)$, and thus it may prevent congestion by stabilizing the traffic flow at a high throughput level. Fig. 14 displays a sample from a field implementation of ALINEA in Glasgow, U.K. It may be seen that the measured downstream occupancy o_{out} is maintained close to its set value of 26% except for the final phase where the ramp demand is not sufficient to feed the mainstream. The set value $\hat{o}$ may be changed any time, and thus ALINEA may be embedded into a hierarchical control system with set
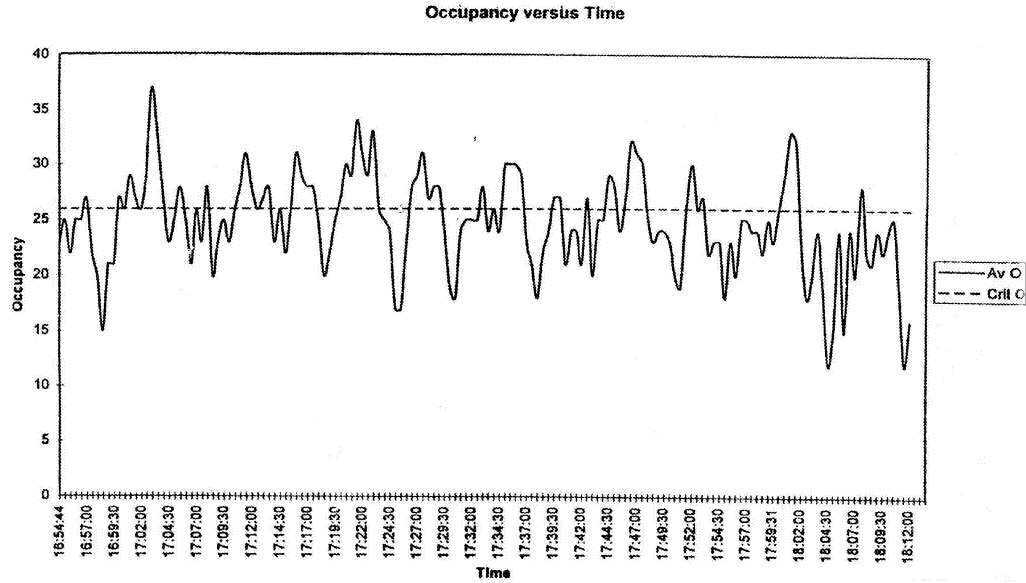
Fig. 14. A sample from the Glasgow implementation of ALINEA.

values of the individual ramps being specified in real time by a superior coordination level or by an operator.

All control strategies calculate suitable ramp volumes r. In the case of traffic-cycle realization of ramp metering, r is converted to a green-phase duration g by use of

$$g = (r/r_{\text{sat}}) \cdot c \qquad (24)$$

where c is the fixed cycle time and r_{sat} is the ramp's saturation flow. The green-phase duration g is constrained by $g \in [g_{\min}, g_{\max}]$, where $g_{\min} > 0$ to avoid ramp closure, and $g_{\max} \le c$. In the case of an one-car-per-green realization, a constant-duration green phase permits exactly one vehicle to pass. Thus, the ramp volume r is controlled by varying the red-phase duration between a minimum and a maximum value. Note that ALINEA is also applicable directly to the green or red-phase duration, by combining (23) and (24)

$$g(k) = g(k-1) + K'_R[\hat{o} - o_{\text{out}}(k)] \qquad (25)$$

where $K'_R = K_R c/r_{\text{sat}}$. Note also that the values $r(k-1)$ or $g(k-1)$ used on the right-hand side of (23) or (25), respectively, should be the *bounded* values of the previous time step (i.e., after application of the $g_{\min}$ and $g_{\max}$ constraints) in order to avoid the windup phenomenon in the I-regulator.

If the queue of vehicles on the ramp becomes excessive, interference with surface street traffic may occur. This may be detected with suitably placed detectors (on the upstream part of the on-ramp), leading to an override of the regulator's decisions to allow more vehicles to enter the freeway and the ramp queue to diminish.

Comparative field trials have been conducted in various countries to assess and compare the efficiency of local ramp metering strategies; see, e.g., [71]. One of these trials took place at the on-ramp Brançion of the clockwise direction of the Boulevard Périphérique (ringway) in Paris. Several ramp metering strategies were applied over a period of one month each, and 13 typical days (without incidents) per strategy were selected for comparison. The evaluation criteria included total travel time (TTT) on the mainstream; total waiting time (TWT) at the ramp; total time spent (TTS = TTT + TWT); total travel distance (TTD); mean speed (MS = TTD/TTS); and mean congestion duration (MCD), which is the accumulated period of time during the morning peak whereby the measured occupancy is higher than o_{cr}. Table 1 displays an extract of the comparative results for the period 7:00 A.M. to 10:00 A.M. It can be seen that ALINEA outperforms feedforward-based strategies with respect to all evaluation criteria; see [72] and [73] for further field applications.

2) Multivariable Regulator Strategies: Multivariable regulators for ramp metering pursue the same goals as local ramp metering strategies: they attempt to operate the freeway traffic conditions near some prespecified set (desired) values. While local ramp metering is performed independently for each ramp, based on local measurements, multivariable regulators make use of all available mainstream measurements $o_i(k), i = 1, \ldots, n$, on a freeway stretch, to calculate simultaneously the ramp volume values $r_i(k), i = 1, \ldots, m$, for all controllable ramps included in the same stretch [74].

This provides potential improvements over local ramp metering because of more comprehensive information provision and coordinated control actions. Multivariable regulator approaches to ramp metering have been mostly derived by application of the Linear-Quadratic-Regulator (LQR) theory [75]–[86]. The multivariable regulator strategy MET-ALINE may be viewed as a generalization and extension of ALINEA, whereby the metered on-ramp volumes are calculated from

$$\mathbf{r}(k) = \mathbf{r}(k-1) - \mathbf{K}_1[\mathbf{o}(k) - \mathbf{o}(k-1)] + \mathbf{K}_2[\hat{\mathbf{O}} - \mathbf{O}(k)] \qquad (26)$$

Table 1 Comparative Field Results of Local Ramp Metering

CONTROL STRATEGY	TTS veh·h	TTS % change	TTD veh·km	TTD % change	MS km/h	MS % change	MCD min	MCD % change
NO CONTROL	421	–	16463	–	39	–	108	–
ALINEA	354	−15.9	16980	3.1	48	23.1	53	−50.9
DEMAND-CAPACITY	407	−3.3	15143	−8.0	37	−5.1	108	0.0
OCCUPANCY	438	0.4	15673	−4.8	36	−7.7	103	−4.6

where $\mathbf{r} = [r_1 \ldots r_m]^T$ is the vector of m controllable on-ramp volumes, $\mathbf{o} = [o_1 \ldots o_n]^T$ is the vector of n measured occupancies on the freeway stretch, $\mathbf{O} = [O_1 \ldots O_m]^T$ is a subset of $\mathbf{o}$ that includes m occupancy locations for which prespecified set values $\hat{\mathbf{O}} = [\hat{\mathbf{O}}_1 \ldots \hat{\mathbf{O}}_\mathbf{m}]^T$ may be given. Note that for control-theoretic reasons, the number of set-valued occupancies cannot be higher than the number of controlled on-ramps. Typically one bottleneck location downstream of each controlled on-ramp is selected for inclusion in the vector $\mathbf{O}$. Finally, $\mathbf{K}_1$ and $\mathbf{K}_2$ are the regulator's constant gain matrices that must be suitably designed; see [74], [87] for details.

Field trials and simulation results comparing the efficiency of METALINE versus ALINEA lead to the following conclusions.

- While ALINEA requires hardly any design effort, METALINE application calls for a rather sophisticated design procedure that is based on advanced control-theoretic methods (LQR optimal control).
- For urban freeways with a high density of on-ramps, METALINE was found to provide no advantages over ALINEA (the latter implemented independently at each controllable on-ramp) under recurrent congestion.
- In the case of nonrecurrent congestion (e.g., due to an incident), METALINE performs better than ALINEA due to more comprehensive measurement information.

Some system operators hesitate to apply ramp metering because of the concern that congestion may be conveyed from the freeway to the adjacent street network. In fact, a ramp metering application designed to avoid or reduce congestion on freeways may have both positive and negative effects on the adjacent road network traffic. It is easy to see, based on notions and statements made earlier, that, if

an efficient control strategy is applied for ramp metering, the freeway throughput will be generally increased. More precisely, ramp metering at the beginning of the rush hour may lead to on-ramp queues in order to prevent congestion to form on the freeway, which may temporarily lead to diversion toward the urban network. But due to congestion avoidance or reduction, the freeway will be eventually enabled to accommodate a higher throughput, thus attracting drivers from urban paths and leading to an improved overall network performance. This positive impact of ramp metering on both the freeway and the adjacent road network traffic conditions was confirmed in a specially designed field evaluation in the Corridor Périphérique in Paris, see [88].

D. Nonlinear Optimal Ramp Metering Strategies

Prevention or reduction of traffic congestion on freeway networks may dramatically improve the infrastructure efficiency in terms of throughput and total time spent. Congestion on limited-capacity freeways forms because too many vehicles attempt to use them in a noncoordinated (uncontrolled) way. As illustrated in Section I-D, once congestion is built up, the outflow from the congestion area is reduced, and the off-ramps and interchanges covered by the congestion are blocked, which may in some extreme cases even lead to gridlocks. Reactive ramp metering strategies may be helpful to a certain extent, but, first they need appropriate set values, and, second, the scope of their actions is more or less local. What is needed for freeway networks or long freeway stretches is a superior coordination level that calculates in real time optimal set values from a proactive, strategic point of view. Such an optimal control strategy should explicitly take into account:

- the current traffic state both on the freeway and on the on-ramps;
- demand predictions over a sufficiently long time horizon;

- the limited storage capacity of the on-ramps;
- the ramp metering constraints discussed earlier;
- the nonlinear traffic flow dynamics, including the infrastructure's limited capacity;
- any incidents currently present in the freeway network.

Based on this comprehensive information, the control strategy should deliver set values for the overall freeway network over a future time horizon so as to:

- respect all present constraints;
- minimize an objective criterion such as the total time spent in the whole network including the on-ramps;
- consider equity aspects for users of different ramps in the network [56], [85].

Such a comprehensive dynamic optimal control problem may be formulated and solved with moderate computation time by use of suitable numerical algorithms.

The nonlinear traffic dynamics may be expressed by use of suitable macroscopic dynamic models in the state-space form [56], [57]

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k), \mathbf{r}(k), \mathbf{d}(k)] \qquad (27)$$

where the state vector $\mathbf{x}$ comprises all traffic densities and mean speeds of 500-m-long freeway sections, as well as all ramp queues; the control vector $\mathbf{r}$ comprises all controllable ramp volumes; the disturbance vector $\mathbf{d}$ comprises all on-ramp demands and turning rates at bifurcations (including off-ramps). Generally, the evolution of traffic density ρ is described via the conservation-of-vehicles equation, while the mean speed v is calculated via an empirical (static or dynamic) equation in dependence of the traffic density. Finally the traffic flow is by definition $q = \rho v$. The ramp metering constraints are given by (18) while the queue constraints read

$$l_i(k) \leq l_{i,\max} \qquad (28)$$

where l_i are queue lengths. The total time spent in the whole system over a time horizon K may be expressed

$$T_s = T \sum_{k=0}^{K} \left[\sum_{i=1}^{n} \rho_i(k) \cdot \Delta_i + \sum_{i=1}^{m} l_i(k) \right]. \qquad (29)$$

Thus, for given current (initial) state $\mathbf{x}(0)$ from corresponding measurements, and given disturbance predictions $\mathbf{d}(k), k = 0, \ldots, K-1$, the problem consists in specifying the ramp flows $\mathbf{r}(k), k = 0, \ldots, K-1$, so as to minimize the total time spent (29) subject to the nonlinear traffic flow dynamics (27) and the constraints (18) and (28).

This problem or variations thereof was considered and solved in various works [56], [67], [89]–[96]. Although simulation studies indicate substantial savings of travel time and substantial increase of throughput [56], advanced control strategies of this kind have not been implemented in the field as of yet.

Fig. 15 displays an example of (simulated) optimal control application using the generic software tool AMOC (see Sec-

tion III-E) for the Amsterdam ringroad A10 (counter-clockwise direction only) over a typical morning-peak period of 4 h [56]. The freeway A10 in the considered direction has a length of 32 km and includes 21 on-ramps and 20 off-ramps (thereof 4 freeway interchanges with A1, A2, A4, A8); see Fig. 15(a). When no ramp metering is applied, the excessive demand coupled with the uncontrolled entrance of drivers into the mainstream, causes a time-space extended congestion [i.e., very high density values in Fig. 15(b)] that blocks almost half of the freeway off-ramps, thus leading to a strongly reduced throughput as explained in Section I-D. With application of optimal ramp metering, congestion is avoided [Fig. 15(c)], throughput is maximized, and the total time spent by all vehicles (including waiting time at the ramps) is reduced by 43.5% compared to no control. Moreover, in contrast to the no-control case, there is no congestion spillback from A10 into the merging freeway A4 when optimal ramp metering is applied.

### E. Integrated Freeway Network Traffic Control

As mentioned earlier, modern freeway networks may include different types of control measures. The corresponding control strategies are usually designed and implemented independently, thus failing to exploit the synergistic effects that might result from coordination of the respective control actions. An advanced concept for integrated freeway network control may result from suitable extension of the optimal control approach outlined above. More precisely, the dynamic model (27) of freeway traffic flow may be extended to enable the inclusion of further control measures, beyond the ramp metering rates $\mathbf{r}(k)$. Formally $\mathbf{r}(k)$ is then replaced in (27) by a general control input vector $\mathbf{u}(k)$ that comprises all implemented control measures of any type. Such an approach was implemented in the integrated freeway network control tool AMOC [97] where ramp metering and route guidance (see Section IV) are considered simultaneously with promising results; see also [98]–[102].

F. Link Control

Link control may include one or a combination of the following actions:

- variable speed limitation;
- changeable message signs with indications for "keep lane," or congestion warning, or environmental warning (e.g., information about the pavement state);
- lane control measures (e.g., prohibited lane use upstream of heavily used on-ramps or incident locations);
- incident or congestion warning;
- reversable flow lanes (tidal flow).

There are many freeway stretches, particularly in Germany, in The Netherlands, and, more recently, in the United Kingdom, employing a selection of these measures. It is generally thought that control measures of this kind lead to a homogenization of traffic flow (i.e., more homogeneous
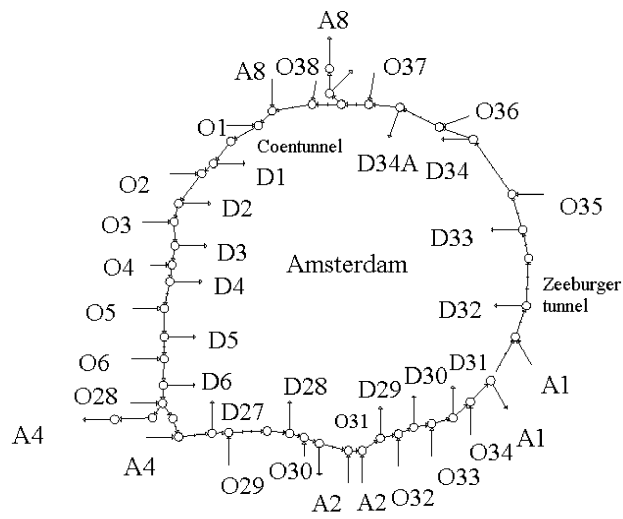
(a)
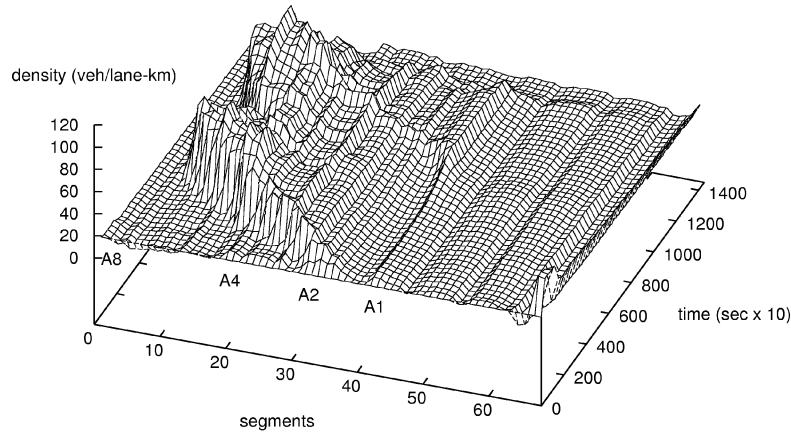
(b)
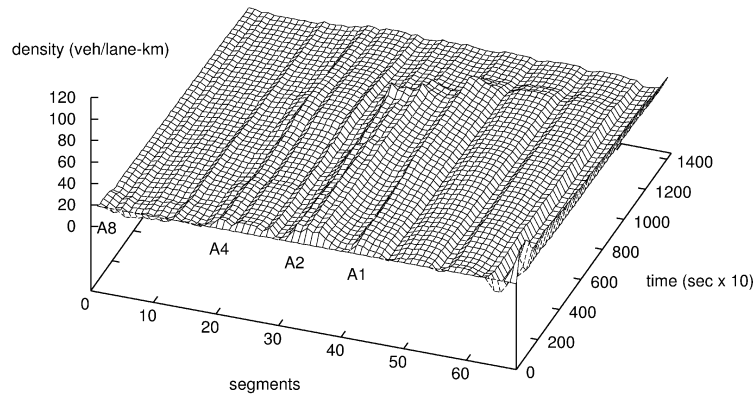
(c)

Fig. 15. Optimal ramp metering for the Amsterdam ringroad (counterclockwise direction). (a) Network sketch. (b) Density profile without and (c) with optimal ramp metering control.

speeds of cars within a lane and of average speeds on different lanes) which is believed to reduce the risk of falling into congestion at high traffic densities and to increase the

freeway's capacity. Very few systematic studies have been conducted to quantify the impact of these control measures (see e.g., [103], [104]) and corresponding validated math-

ematical models are currently lacking. This is one of the reasons why the corresponding control strategies of operating systems are of a heuristic character (e.g., [105]–[108]), while more systematic approaches are based on nonvalidated modeling assumptions [82], [101], [102], [109].

IV. ROUTE GUIDANCE AND DRIVER INFORMATION

A. Introduction

Freeway, urban, or mixed traffic networks include a large number of origins and destinations with multiple paths connecting each origin-destination pair. Fixed direction signs at bifurcation nodes of the network typically indicate the direction that is time-shortest in absence of congestion. However, during rush hours, the travel time on many routes changes substantially due to traffic congestion and alternative routes may become competitive. Drivers who are familiar with the traffic conditions in a network (e.g., commuters) optimize their individual routes based on their past experience, thus leading to the celebrated user-equilibrium conditions, first formulated by Wardrop [110]. But daily varying demands, changing environmental conditions, exceptional events (sport events, fairs, concerts, etc.) and, most importantly, incidents may change the traffic conditions in a nonpredictable way. This may lead to an underutilization of the overall network's capacity, whereby some links are heavily congested while capacity reserves are available on alternative routes. Route guidance and driver information systems (RGDIS) may be employed to improve the network efficiency via direct or indirect recommendation of alternative routes (see [111] for a critical view).

A first classification of RGDIS distinguishes *pretrip* from *en route* advice. Pretrip communication possibilities include the internet, phone services, mobile devices, television, and radio. These communication devices may be consulted by a potential road user to make a rational decision regarding:

- effectuation or postponement of the intended trip;
- choice of transport mode (car, bus, underground, etc.);
- choice of the departure time;
- (initial) path choice.

If the road user has decided to complete the trip by car, she may continue to receive information or advice via appropriate *en-route* devices such as radio services (RDS-TMC), road-side variable message sings (VMS), or special in-car equipment, in order to make sensible routing decisions at bifurcation nodes of the network. While radio broadcasting services and VMS have been in use for more than 25 years (and their number is steadily increasing) (see [112], see also [113] for an early overview), individual route guidance systems employing in-car devices and two-way communication with control centers are in their infancy (some experimental or early operational systems exist in some countries) [114].

At this point, it is appropriate to distinguish among two alternative policies (which in some cases may be combined) of providing en-route information versus explicit route recommendation. Many operators (particularly of VMS-based

systems) prefer the provision of real-time information. Also the majority of drivers (according to some questionnaire results) seem to prefer this option that enables them to make their own decisions, rather than having to follow recommendations by an anonymous system. It should be emphasized, however, that pure information provision has a number of partially significant drawbacks.

- The translation of provided information into routing decisions requires the knowledge of the network which may not be present for all drivers.
- Although the control center disposes over complete information about the traffic conditions in the whole network, only a tiny part of this information can be conveyed to the users due to space limitations on the VMS and other devices. In some cases, only information about the traffic conditions on the downstream links of a bifurcation node are provided. Clearly, this information is not sufficient for a rational route decision for drivers with longer trips through the network, who may be eventually trapped into a severe congestion due to myopic decision making.
- Even if it would be possible to provide more comprehensive information, the drivers would have to make a route decision within a few seconds, i.e., after looking at the VMS and before reaching the bifurcation.
- There is no possibility for the operator or a control strategy to actively influence traffic conditions, as decisions are left with the drivers.

On the other hand, route guidance systems are constrained by the requirement not to suggest routes that would disbenefit complying drivers, else the credibility and eventually the impact of the whole system may be jeopardized. Moreover, route guidance systems call for a genuine control strategy that can optimize the network traffic conditions, e.g., by avoiding traffic congestion on alternative recommended routes due to drivers' overreaction.

B. Travel Time Display

A particular type of driver information system that is gaining increasing momentum due to its relative simplicity and its popularity with drivers is the display (on VMS) of travel times for well-defined stretches downstream of the VMS. This information is readily comprehensible by the drivers, and it may either provide a basis for route choice decisions or simply reduce the drivers' stress, particularly in congested traffic conditions.

For example, some 350 VMS are installed on the Boulevard Périphérique of Paris, France, and on all approaches that lead to this ringway [115], [116]. The displayed message is the current (instantaneous) travel time on the ringway from the particular VMS location to two significant downstream freeway intersections, at distances of approximately 3 km and 6 km downstream from each VMS, respectively. A similar system, providing travel times on the two downstream freeway links of each bifurcation node, is operational

Fig. 16. Example of a Variable Message Sign (VMS) display in the Ile-de-France freeway network around Paris, France; this VMS is located just upstream of a freeway bifurcation whose outgoing links A1 and A3 eventually lead to the Boulevard Périphérique (BP). The sign informs (first line) that the current travel time from the VMS on A1 until BP is 12 min with increasing tendency, while (second line) the travel time from the VMS on A3 until BP is 21 min with decreasing tendency.

in the dense freeway network around Paris; see Fig. 16 for an example.

The calculation of instantaneous travel times for a freeway stretch including several inductive loop detectors is quite simple. The stretch is subdivided into a number N of segments with lengths Δ_i ($\Delta_i = 500$ m in Paris), whereby each segment includes one detector station. A detector station may either directly provide mean speed measurements or it may provide flow and occupancy measurements, from which the mean speed v_i can be deduced with sufficient accuracy (see, e.g., [117], [118]). The travel time τ_i of segment i is then given by

$$\tau_i = \Delta_i/v_i. \tag{30}$$

The *instantaneous travel time* on a freeway stretch is defined to be the travel time of a virtual vehicle that travels the stretch under the assumption that the currently prevailing traffic conditions will not change during the trip. Based on this definition and (30), the instantaneous travel time τ of the whole freeway stretch, consisting of N segments, may be calculated from

$$\tau = \sum_{i=1}^{N} \tau_i = \sum_{i=1}^{N} \Delta_i/v_i. \tag{31}$$

This formula works quite well in practice but it has some drawbacks, notably if the mean speed v_i in a segment becomes temporarily very low, then (31) delivers an unrealistically high travel time for the whole freeway stretch. Alternative formulas have been suggested and evaluated based on real traffic data; see [119], [120].

Clearly, any instantaneous travel time formula based only on current traffic measurements will induce a systematic estimation error if the traffic conditions in the stretch are rapidly changing, e.g., during congestion growth or dissipation. This error grows with the length of the stretch under question and may, under certain conditions, reach unacceptable levels for freeway stretches longer than, say, 10 km. What is needed in this case is a predictive scheme that delivers *predicted travel times* that come closer to the travel times that will be experienced by the drivers during their trip. Predictive travel times may be calculated:

- based on historical information;
- via suitable extrapolation methods (e.g., time series or neural networks);
- by employment of dynamic traffic flow models in real time
- via a combination of the above.

The technical literature addressing travel time estimation and prediction is quite vast; see, e.g., [121]–[130], see also [131] for an empirical comparison of methods.

C. Route Guidance Strategies

1) Basic Notions: A route guidance system may be viewed as a traffic control system in the sense of Fig. 1. Based on real-time measurements, sufficiently interpreted and extended within the surveillance block, a control strategy decides about the routes to be recommended (or the information to be provided) to the road users. This, on its turn, has an impact on the traffic flow conditions in the network, and this impact is reflected in the performance indices. Because of the real-time nature of the operation, requirements of short computation times are relatively strict.

Route guidance strategies may be classified according to various aspects.

- *Reactive strategies* are based only on and react to current measurements without the real-time use of mathematical models or other predictive tools; *predictive strategies* attempt to predict traffic conditions sufficiently far in the future (typically by real-time use of mathematical models) in order to improve the quality of the provided recommendations.
- *Iterative strategies* run several model simulations in real time, each time with suitably modified route guidance, to ensure (at convergence) that the control goal (see below) will be achieved as accurately as possible; iterative strategies are by nature predictive. *One-shot strategies* may either be reactive, in which case they typically perform simple calculations based on real-time data, or they may be predictive, whereby they run one single time a simulation model to increase the relevance of their recommendations.
- Route guidance strategies may aim at either *system optimal* or *user optimal* traffic conditions. In the first case, the control goal is the minimization of a global objective criterion (e.g., the total time spent) even for the price of recommending routes that are sometimes more costly than the regular routes. In the second case, every recommended route should not be more costly than the regular route, even for the price of suboptimality with respect to the global objective criterion. Under a more strict definition, user-optimal conditions imply *equal cost* on all utilized alternative routes connecting any two nodes in the network.

2) One-Shot Strategies: Most one-shot strategies are of the reactive type. Particularly for dense networks, with relatively short links, many bifurcations, and a high number of alternative routes connecting any two nodes, reactive strategies may be highly efficient in establishing user-optimal

conditions on the basis of current traffic measurements. This is because reactive routing recommendations in this kind of networks may be modified at downstream bifurcation nodes if traffic conditions change substantially.

Most reactive strategies are decentralized, i.e., they conduct their calculations at each bifurcation node independently of other nodes. Simple feedback regulators of the P (proportional) or PI (proportional-integral) types have been proposed by Messmer and Papageorgiou [132]. A P-regulator calculates splitting rates β_{nj} as follows:

$$\beta_{nj}(k) = \beta_{nj}^N - K\Delta\tau_{nj}(k) \qquad (32)$$

where $\beta_{nj}(k) \in [0, 1]$ is the portion of the flow arriving at bifurcation node n and destined to node j that is routed through the main direction at time k; β_{nj}^N is the nominal splitting of drivers (in absence of route guidance); $K > 0$ is a regulator parameter; $\Delta\tau_{nj}(k)$ is the instantaneous travel time difference between the main and the alternative direction from node n to node j. Note that $\beta_{nj}(k)$ resulting from (32) is eventually truncated if it exceeds the range $[0, 1]$. The regulator (32) assigns more or less traffic to the alternative direction according to the sign and value of the current travel time difference $\Delta\tau_{nj}(k)$ among both directions thus aiming at equalizing both corresponding travel times, in accordance with the user-optimum requirements. For K sufficiently high, an all-or-nothing (or bang-bang) strategy results from (32) whereby all vehicles are sent to the currently shortest direction. It should be noted that these simple regulators are not very sensitive to varying compliance rates of drivers [133], [134]. An operational system employing decentralized P-regulators in the traffic network of Aalborg, Denmark, was reported in [135], [136]. Multivariable regulators for central route guidance systems have also been suggested [137], [138], as well as heuristic or simplified feedback schemes [139]–[142], and more advanced automatic control concepts [143].

A different kind of one-shot strategies may employ in real time a more or less sophisticated mathematical model of the network traffic flow [144], [145]. Based on the current traffic state, the current control inputs, and predicted future demands, the model is run once, in order to provide information about the future traffic conditions under the current route guidance settings. A regulator is then used to control the predicted future, rather than the current, traffic conditions. Such control schemes are known as IMC (Internal Model Control) strategies in automatic control theory. They are preferable to reactive regulators when the traffic network has long links with a limited number of bifurcation nodes. A control scheme of this kind was applied to the Scottish highway network employing P-regulators [146] or a heuristic expert system [147], [148].

3) Iterative Strategies: Iterative strategies may aim at establishing either system-optimal or user-optimal conditions. For a system optimum, the pursued procedure has already been outlined in Section III. A macroscopic network traffic model may be written in the general form (27), where the control inputs are the splitting rates $\beta(k)$. As already mentioned, macroscopic models describe the traffic flow as a fluid with particular characteristics via the aggregate traffic variables traffic density, flow, and mean speed (see Section I-C). To enable the description of the routing behavior, macroscopic models are extended to incorporate partial densities and partial flows by destination. Partial flows are assigned at bifurcation nodes to downstream links according to the splitting rates $\beta(k)$ which act as control inputs to be optimized. The corresponding optimal control problem, aiming at minimizing the total time spent (29) under the constraints $0 \leq \beta(k) \leq 1$ may be solved by use of the same numerical algorithms as the optimal ramp metering or the integrated control problem [132], [137], [149], see also [150]–[152].

On the other hand, there are also several iterative procedures suggested toward establishing user optimal conditions [153]–[156]. The typical core structure of these iterative strategies is as follows.

a) Set the initial path assignments or splitting rates (control inputs)
b) Run a simulation model over a time horizon H.
c) Evaluate the travel times on alternative utilized paths; if all travel time differences are sufficiently small, stop with the final solution.
d) Modify the path assignments or splitting rates appropriately to reduce travel time differences; go to (b).

The simulation models employed by different algorithms in step (b) may be microscopic, macroscopic, or mesoscopic. *Microscopic models* address and describe the movement of each individual vehicle in the traffic flow in dependence of the movement of the adjacent vehicles, both in the longitudinal (car-following behavior) and in the lateral (lane-changing behavior) sense. Each vehicle has a prespecified destination and its path is decided pretrip or en-route according to the routing decisions of the algorithm. *Macroscopic models* may be expressed in the form (27) as outlined above. *Mesoscopic models* describe the evolution of mean speed macroscopically, but they also consider individual vehicles (or "vehicle packets") which, however, are moved in the network according to the macroscopic mean speed (without employment of microscopic models); the traffic flows at section boundaries are deduced from individual vehicle crossings, while traffic densities may be obtained directly from vehicle counts within a section. The reason why individual vehicles are introduced in mesoscopic models is in order to describe the routing behavior. Thus, as in microscopic models, each vehicle has a prespecified destination, and its path is decided pretrip and en-route according to the routing decisions of step (d), without the need to introduce partial densities and flows as in macroscopic models [153]–[155].

The modification of path assignments or splitting rates in step (d) of the algorithm is typically effectuated in a functionally decentralized way, i.e., each splitting rate or path assignment portion is changed independently of any other. The sign and magnitude of the individual changes depend on the sign and magnitude of the corresponding travel time differences, in a similar way as in (32), with the significant difference that
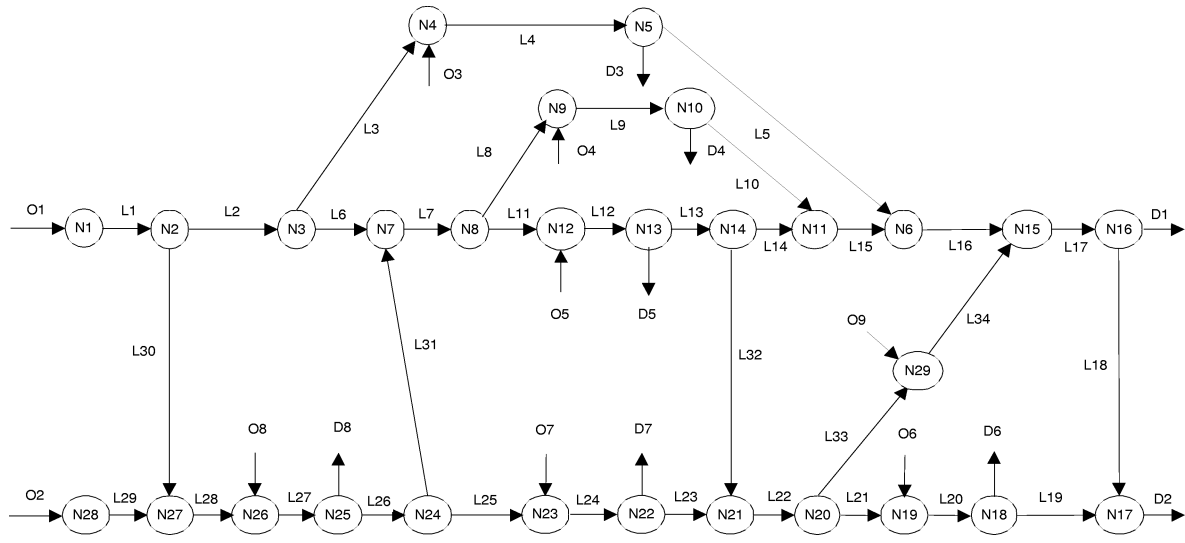
Fig. 17. Example network.

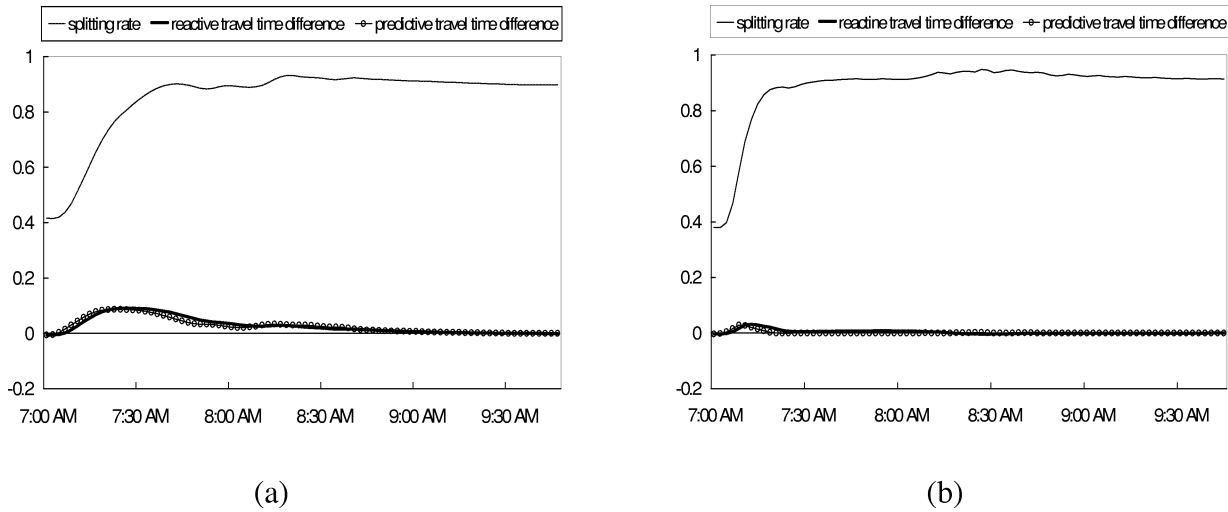(a)   (b)

Fig. 18. Routing results for node N3 to destination D2: PI-regulator, (b) Iterative algorithm.

the travel time differences are here predictive [calculated in step (c)] rather than instantaneous.

The real-time implementation of iterative algorithms for route guidance purposes employs the same rolling horizon procedure outlined in Section II-D in order to reduce the sensitivity with respect to predicted demands and modeling inaccuracies. No field implementation of an iterative route guidance procedure has been reported as yet (although some work is in progress toward this end). Main reasons for this are the relatively recent interest in RGDIS, but certainly also the code complexity of the corresponding algorithms.

4) An Example: Fig. 17 displays a hypothetical freeway network consisting of 29 nodes (N1–N29) and 51 links, thereof nine origin links (O1–O9), eight destination links (D1–D8), and 34 freeway links (L1–L34). Freeway link lengths range from 1 km to 4 km. Splitting for route guidance is considered at nodes N2, N3, N8, N14, N20, N24 toward the destinations D1 and D2, which results in a total number of 11 splitting rates $\beta_{n,j}$ since N20 is connected to D1 via one direction only. Note that the number of possible

paths is much higher. The network traffic flow is simulated macroscopically with appropriate demands at the network origins.

Fig. 18 displays a representative example of routing results. Each diagram displays three histograms relating to a specific (node, destination)-couple (n, j): the splitting rate $\beta_{n,j}$ and two relative travel time differences (instantaneous and predictive). The latter are defined as $\Delta\tau_{n,j}/\tau_{n,j}$, where $\Delta\tau_{n,j}$ was defined in (32) and $\tau_{n,j}$ is the travel time from n to j along the main direction. The control goal (user optimum) is to keep $\Delta\tau_{n,j}(k) \approx 0$ $\forall k$ so long as $\beta_{n,j}(k)$ is not on the bounds 0 or 1, i.e., so long as a splitting actually takes place. Fig. 18(a) displays the results obtained by use of simple PI-regulators while Fig. 18(b) displays the corresponding results from an iterative algorithm (after 100 iterations). These results give rise to the following comments.

1) Instantaneous and predictive travel time differences are not far from each other, the former typically lagging the latter; this may be different in case of net-

works with few bifurcations and very long links, as mentioned earlier.

2) The simple PI-regulator manages to keep travel time differences fairly low; note that, as (32) indicates, such a regulator is based merely on instantaneous travel times at each period k, which can be readily estimated based on real-time measurements.

3) The iterative algorithm leads to virtually full satisfaction of user-optimum conditions (i.e., $\Delta\tau_{\mathrm{n,j}}(k) = 0 \ \forall k$) albeit based on perfect knowledge of future demands, driver compliance, and traffic flow dynamics; in real applications, prediction mistakes are inevitable, leading to clearly less impressive results.

V. FUTURE DIRECTIONS

A. The Theory–Practice Gap

As in many other engineering disciplines, only a small portion of the significant methodological advancements have really been exploited in the field as of yet. It is beyond the scope of this review paper to investigate and discuss in depth the reasons behind this theory–practice gap, but administrative inertia; little competitive pressure in the public sector; industrial interest in (expensive) hardware rather than in (low-cost) methodologies; the complexity of traffic control systems; limited realization of the improvement potential behind advanced methods by the responsible authorities; and limited understanding of practical problems by some researchers may have a role in this. Whatever the reasons, the major challenge in the coming decade is the deployment of advanced and efficient traffic control strategies in the field, with particular focus on addressing traffic saturation phenomena (congestion).

More precisely, the majority of small and big cities even in industrialized countries are still operating old-fashioned fixed-time signal control strategies, often even poorly optimized or maintained. Even when modern traffic-responsive control systems are installed in terms of hardware devices, the employed control strategies are sometimes naïve, poorly tested and fine-tuned, thus failing to exploit the possibilities provided by the relatively expensive hardware infrastructure.

Regarding freeway networks, the situation is even worse. Operational control systems of any kind are the exception rather than the rule. With regard to ramp metering, the main focus is often not on improving efficiency but on secondary objectives of different kinds. Many responsible traffic authorities and decision makers are far from realizing the fact that advanced real-time ramp metering systems (employing optimal control algorithms) have the potential of changing dramatically the traffic conditions on today's heavily congested (hence strongly underutilized) freeways with spectacular improvements that may reach 50% reduction of the total time spent.

With regard to driver information and route guidance systems, there is an increasing interest and an increasing number of operational systems employing variable message signs,

but once more, the relatively expensive hardware infrastructure is not exploited to the degree possible, as implemented control strategies are typically naïve.

On the side of the research community, any effort should be made to enlighten the road authorities, the political decision makers, and the general public about the substantial improvements achievable via implementation of modern traffic control methods and tools. At the same time, it should be emphasized that many methodological works presented at conferences and technical journals address practical problems and concerns only in a limited way. In some cases, proposed traffic control strategies are not even thoroughly and properly tested via simulation, despite the meanwhile high number of available traffic simulators of various kinds. This poses a burden to real implementation of the methods, and perhaps the best way for researchers to familiarize themselves with the practical requirements and constraints is to get occasionally involved in real implementations.

B. Road Traffic Control Strategies

The number of developed signal control strategies is much higher than what could be included or mentioned in Section III. The need and trend is clearly toward traffic-responsive coordinated strategies. Two avenues may be identified as promising in this respect.

1) Advanced signal control strategies, such as OPAC, PRODYN, CRONOS, and RHODES, have clear limitations regarding the network extent, to which their basic optimization algorithm can be directly applied. The properties of a completely decentralized operation (e.g., independent algorithm application at each intersection) are currently not fully analyzed or understood. This kind of thorough analysis may be useful for enhanced network-wide coordinating layers that reduce the possibly negative impact of decentralization.

2) Store-and-forward based concepts seem, more than three decades after their original conception, to offer a promising background for the development of signal control strategies that are traffic-responsive, coordinated (for large-scale networks), and can cope, under certain conditions, with oversaturation and the imminent inaccuracies of traffic measurements in an urban road environment. In addition, this approach seems ideal for the design of (even more challenging) integrated traffic control strategies involving further traffic systems (freeways) and control measures.

C. Freeway Traffic Control Strategies

With regard to ramp metering, the most important methodological developments are well advanced, although further improvements are desirable and possible both at the reactive and the network-wide levels, as well as at the interconnection of both levels. The most promising and challenging area for control strategies is the design and testing of hierarchical control structures for very large-scale

freeway networks. Control hierarchies should include short-term demand predictions, optimal control algorithms for the coordinated calculation of set values network-wide, and reactive feedback strategies for implementation of the optimal control decisions. Besides efficiency, the equity properties of ramp metering strategies are of particular importance in ubiquitous network-wide ramp metering systems.

The integrated control of freeway networks involving both ramp metering and route guidance measures is currently in a very preliminary phase with some very promising results, but a lot more developments are required to produce integrated control strategies that are efficient, but also applicable in real time to large-scale networks.

Finally link control systems may prove much more useful than at present if their impact is studied more carefully and thoroughly so as to open the way to the design of efficient control strategies. This is perhaps one of the least studied areas within traffic control methodology.

D. Driver Information and Route Guidance Systems

Although the subject of DRGIS is relatively new, a substantial amount of work has been devoted to it and a number of methods and tools have already emerged, but further developments, either completely new or combinations of already suggested methods, are possible and desirable. One-shot methods appear particularly attractive for real-time applications because they are simple, with negligible computational effort. More experience is required regarding their efficiency level and the topological and traffic conditions that are most suitable for their application. On the other hand, the surplus efficiency provided by iterative approaches should be further investigated, and possible combinations (e.g., in order to reduce the computational effort of iterative strategies) should be attempted.

E. A Potential Limitation

Any substantial improvements achieved in the field thanks to beneficial application of control methods, may be countered to some extent due to latent (induced) demand, i.e., due to drivers motivated to use their cars (rather than other transportation means) or to drive through efficiently traffic-controlled areas so as to profit by the reduced travel times. Although the precise nature and extent of this medium-term driver response to beneficial control measures has not been studied sufficiently, we may note the following.

- In the worst case, the originally achieved improvements of travel time due to control measures will be only partially reduced due to latent demand, because full erasure of the improvement would provide no motivation for latent demand to appear.
- A partial (rather than full) improvement of travel times (due to latent demand) addresses a higher benefitting driver population (that includes the latent demand).

REFERENCES

[1] M. Papageorgiou, "Automatic control methods in traffic and transportation," in *Operations Research and Decision Aid Methodologies in Traffic and Transportation Management*, P. Toint, M. Labbe, K. Tanczos, and G. Laporte, Eds. New York: Springer-Verlag, 1998, pp. 46–83.
[2] G. Abu-Lebdeh and R. F. Benekohal, "Development of traffic control and queue management procedures for oversaturated arterials," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1603, 1997.
[3] R. B. Allsop, "SIGSET: A computer program for calculating traffic capacity of signal-controlled road junctions," *Traffic Eng. Control*, vol. 12, pp. 58–60, 1971.
[4] ———, "SIGCAP: A computer program for assessing the traffic capacity of signal-controlled road junctions," *Traffic Eng. Control*, vol. 17, pp. 338–341, 1976.
[5] F. V. Webster, "Traffic signal settings," Road Research Laboratory, London, U.K., Road Res. Tech. Paper no. 39, 1958.
[6] G. Improta and G. E. Cantarella, "Control systems design for an individual signalised junction," *Transp. Res. B*, vol. 18, pp. 147–167, 1984.
[7] A. J. Miller, "A computer control system for traffic networks," in *Proc. 2nd Int. Symp. Traffic Theory*, 1963, pp. 200–220.
[8] R. A. Vincent and C. P. Young, "Self-optimizing traffic signal control using microprocessor: The TRRL "MOVA" strategy for isolated intersections," *Traffic Eng. Control*, vol. 27, pp. 385–387, 1986.
[9] L. De la Bretegue and R. Jezeguel, "Adaptive control at an isolated intersection—a comparative study of some algorithms," *Traffic Eng. Control*, vol. 20, pp. 361–363, 1979.
[10] J. D. C. Little, "The synchronization of traffic signals by mixed-integer-linear-programming," *Oper. Res.*, vol. 14, pp. 568–594, 1966.
[11] J. D. C. Little, M. D. Kelson, and N. H. Gartner, "MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 795, 1981.
[12] N. A. Chaudhary, A. Pinnoi, and C. Messer, "Proposed enhancements to MAXBAND-86 program," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1324, 1991.
[13] N. H. Gartner, "Road traffic control: Progression methods," in *Concise Encyclopedia of Traffic and Transportation Systems*, M. Papageorgiou, Ed, Oxford, U.K.: Pergamon, 1991, pp. 391–396.
[14] N. H. Gartner, S. F. Assmann, F. Lasaga, and D. L. Hom, "A multiband approach to arterial traffic signal optimization," *Transp. Res. B*, vol. 25, pp. 55–74, 1991.
[15] C. Stamatiadis and N. H. Gartner, "MULTIBAND-96: A program for variable bandwidth progression optimization of multiarterial traffic networks," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1554, 1996.
[16] D. I. Robertson, "TRANSYT method for area traffic control," *Traffic Eng. Control*, vol. 10, pp. 276–281, 1969.
[17] M. T. Li and A. C. Gan, "Signal timing optimization for oversaturated networks using TRANSYT-7F," presented at the 78th Annu. Meeting Transportation Research Board, 1999.
[18] T. Van Vuren, "Signal control and traffic assignment," in *Concise Encyclopedia of Traffic and Transportation Systems*, M. Papageorgiou, Ed, Oxford, U.K.: Pergamon, 1991, pp. 468–473.
[19] P. B. Hunt, D. L. Robertson, and R. D. Bretherton, "The SCOOT on-line traffic signal optimization technique," *Traffic Eng. Control*, vol. 23, pp. 190–192, 1982.
[20] J. Y. K. Luk, "Two traffic-responsive area traffic control methods: SCAT and SCOOT," *Traffic Eng. Control*, vol. 25, pp. 14–22, 1984.
[21] N. H. Gartner, "OPAC: A demand-responsive strategy for traffic signal control," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 906, 1983.
[22] J.-L. Farges, J.-J. Henry, and J. Tufal, "The PRODYN real-time traffic algorithm," in *Proc. 4th IFAC Symp. Transportation Systems*, 1983, pp. 307–312.
[23] F. Boillot, J. M. Blosseville, J. B. Lesort, V. Motyka, M. Papageorgiou, and S. Sellam, "Optimal signal control of urban traffic networks," in *Proc. 6th IEE Int. Conf. Road Traffic Monitoring and Control*, 1992, pp. 75–79.
[24] S. Sen and L. Head, "Controlled optimization of phases at an intersection," *Transp. Sci.*, vol. 31, pp. 5–17, 1997.

[25] J. B. Rawlings, "Tutorial overview of model predictive control," *IEEE Control Syst. Mag.*, vol. 20, pp. 38–52, June 2000.

[26] A. Kessaci, J.-L. Farges, and J.-J. Henry, "Upper level for real time urban traffic control systems," in *Preprints 11th IFAC World Congr.*, vol. 10, 1990, pp. 226–229.

[27] N. H. Gartner, F. J. Pooran, and C. M. Andrews, "Implementation of the OPAC adaptive control strategy in a traffic signal network," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 197–202.

[28] D. C. Gazis and R. B. Potts, "The oversaturated intersection," in *Proc. 2nd Int. Symp. Traffic Theory*, 1963, pp. 221–237.

[29] D. C. Gazis, *Traffic Theory*. Boston, MA: Kluwer, 2002.

[30] ——, "Optimum control of a system of oversaturated intersections," *Oper. Res.*, vol. 12, pp. 815–831, 1964.

[31] G. C. D' Ans and D. C. Gazis, "Optimal control of oversaturated store-and-forward transportation networks," *Transp. Sci.*, vol. 10, pp. 1–19, 1976.

[32] M. G. Singh and H. Tamura, "Modeling and hierarchical optimization of oversaturated urban traffic networks," *Int. J. Control*, vol. 20, pp. 269–280, 1974.

[33] P. G. Michalopoulos and G. Stephanopoulos, "Oversaturated signal systems with queue length constraints—I: Single intersection," *Transp. Res.*, vol. 11, pp. 413–421, 1977a.

[34] ——, "Oversaturated signal systems with queue length constraints—II: System of intersections," *Transp. Res.*, vol. 11, pp. 423–428, 1977b.

[35] J. H. Lim, S. H. Hwang, I. H. Suh, and Z. Bien, "Hierarchical optimal control of oversaturated urban networks," *Int. J. Control*, vol. 33, pp. 727–737, 1981.

[36] E. J. Davison and Ü. Özgüner, "Decentralized control of traffic networks," *IEEE Trans. Automat. Contr.*, vol. 28, pp. 677–688, 1983.

[37] E. S. Park, J. H. Lim, I. H. Suh, and Z. Bien, "Hierarchical optimal control of urban traffic networks," *Int. J. Control*, vol. 40, pp. 813–829, 1984.

[38] A. K. Rathi, "A control scheme for high traffic density sectors," *Transp. Res. B*, vol. 22, pp. 81–101, 1988.

[39] K.-J. Kim and M. G. H. Bell, "Development of an integrated traffic control strategy for both urban signalised and motorway networks," presented at the Proc. 1st Meeting EURO Working Group on Urban Traffic and Transportation, Landshut, Germany, 1992.

[40] M. Papageorgiou, "An integrated control approach for traffic corridors," *Transp. Res. C*, vol. 3, pp. 19–30, 1994.

[41] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network-wide signal control," *Control Eng. Practice*, vol. 10, pp. 183–195, 2002.

[42] C. Diakaki, "Integrated Control of Traffic Flow in Corridor Networks," Ph.D. Thesis, Technical University of Crete, Department of Production Engineering and Management, Chania, Greece, 1999.

[43] C. Diakaki, V. Dinopoulou, K. Aboudolas, M. Papageorgiou, E. Ben-Shabat, E. Seider, and A. Leibov, "Extensions and new applications of the signal control strategy TUC," presented at the 82nd Annu. Meeting Transportation Research Board, 2003. Paper no. 03-3040.

[44] C. Diakaki, M. Papageorgiou, and T. McLean, "Integrated traffic-responsive urban corridor control strategy in Glasgow, Scotland—application and evaluation," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1727, 2000.

[45] C. Bielefeldt, H. Condie, C. Diakaki, V. Dinopoulou, and M. Papageorgiou, "Be smart: SMART NETS and TUC," *Traffic Technol. Int.*, pp. 40–43, Apr./May 2002.

[46] J. Barceló, J. Casas, J. L. Ferrer, and D. Garcia, "Modeling advanced transport telematic applications with microscopic simulators: The case of AIMSUN2," in *Traffic and Mobility, Simulation, Economics, Environment*, W. Brilon, F. Huber, M. Schreckenberg, and H. Wallentowitz, Eds, Heidelberg, Germany: Springer-Verlag, 1999, pp. 205–224.

[47] Z. K. Khatib and G. A. Judd, "Control strategy for oversaturated signalized intersections," presented at the 80th Annu. Meeting Transportation Research Board, Washington, DC, 2001.

[48] W. H. Lin, "An enhanced 0–1 mixed integer LP formulation for the traffic signal problem," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 191–196.

[49] H. K. Lo, Y. C. Chan, and H. F. Chow, "A new dynamic traffic control system: Performance of adaptive control strategies for over-saturated traffic," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 406–411.

[50] D. G. Capelle, "Freeway/corridor systems," in *Proc. Int. Symp. Traffic Control Systems*, vol. 1, 1979, pp. 33–35.

[51] M. Van Aerde and S. Yagar, "Dynamic integrated freeway/traffic signal networks: Problems and proposed solutions," *Transp. Res. A*, vol. 22, pp. 435–443, 1998.

[52] H. Mahmassani and R. Jayakrishnan, "System performance and user response under real-time information in a congested traffic corridor," *Transp. Res. A*, vol. 25, pp. 293–307, 1991.

[53] R. A. Reiss, N. H. Gartner, and S. L. Cohen, "Dynamic control and traffic performance in a freeway corridor: A simulation study," *Transp. Res. A*, vol. 25, pp. 267–276, 1991.

[54] G.-H. Chang, P.-K. Ho, and C.-H. Wei, "A dynamic system-optimum control model for commuting traffic corridors," *Transp. Res. C*, vol. 1, pp. 3–22, 1993.

[55] N. Elloumi, H. Haj-Salem, and M. Papageorgiou, "Integrated control of traffic corridors—application of an LP-methodology," presented at the 4th Meeting EURO Working Group on Transportation Systems, Landshut, Germany, 1996.

[56] A. Kotsialos and M. Papageorgiou, "Efficiency versus fairness in network-wide ramp metering," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 1190–1195.

[57] S. P. Hoogendoorn and P. H. L. Bovy, "State-of-the-art of vehicular traffic modeling," *J. Syst. Control Eng.*, vol. 215, pp. 283–303, 2001.

[58] A. Kotsialos, M. Papageorgiou, M. Mangeas, and H. Haj-Salem, "Coordinated and integrated control of motorway networks via nonlinear optimal control," *Transp. Res. C*, vol. 10, pp. 65–84, 2002.

[59] C. Chen, Z. Jia, and P. Varaiya, "Causes and cures of highway congestion," *IEEE Control Syst. Mag.*, vol. 21, no. 6, pp. 26–32, 2001.

[60] J. A. Wattleworth, "Peak-period analysis and control of a freeway system," Highway Res. Board, Washington, DC, Highway Res. Record 157, 1965.

[61] L. S. Yuan and J. B. Kreer, "Adjustment of freeway ramp metering rates to balance entrance ramp queues," *Transp. Res.*, vol. 5, pp. 127–133, 1971.

[62] D. Tabac, "A linear programming model of highway traffic control," in *Proc. 6th Annu. Princeton Conf. Information Science and Systems*, 1972, pp. 568–570.

[63] C. F. Wang, "On a ramp-flow assignment problem," *Transp. Sci.*, vol. 6, pp. 114–130, 1972.

[64] J. J. Wang and A. D. May, "Computer model for optimal freeway on-ramp control," Highway Res. Board, Washington, DC, Highway Res. Record 469, 1973.

[65] I. C. Cheng, J. B. Cruz, and J. G. Paquet, "Entrance ramp control for travel rate maximization in expressways," *Transp. Res.*, vol. 8, pp. 503–508, 1974.

[66] S. C. Schwartz and H. H. Tan, "Integrated control of freeway entrance ramps by threshold regulation," in *Proc. IEEE Conf. Decision and Control*, 1977, pp. 984–986.

[67] M. Papageorgiou, *Application of Automatic Control Concepts to Traffic Flow Modeling and Control*. New York: Springer-Verlag, 1983.

[68] ——, "A new approach to time-of-day control based on a dynamic freeway traffic model," *Transp. Res. B*, vol. 14, pp. 349–360, 1980.

[69] D. P. Masher, D. W. Ross, P. J. Wong, P. L. Tuan, H. M. Zeidler, and S. Peracek, "Guidelines for design and operating of ramp control systems," SRI, Menid Park, CA, Stanford Res. Inst. Rep. NCHRP 3-22, SRI Project 3340, 1975.

[70] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville, "ALINEA: A local feedback control law for on-ramp metering," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1320, 1991.

[71] M. Papageorgiou, H. Haj-Salem, and F. Middelham, "ALINEA local ramp metering: Summary of field results," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1603, 1998.

[72] H. Haj-Salem, "A motorway-to-motorway control: Overview of simulation and field trial results in France," in *Preprints Triennal Symp. on Transportation Analysis IV*, 2001, pp. 415–420.

[73] H. Haj-Salem, P. Poirier, J.-F. Heylliard, and J.-P. Peynaud, "ALINEA: A local traffic responsive strategy for ramp metering: Field results on A6 motorway in Paris," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 108–113.

[74] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, "Modeling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris—Part II: Coordinated on-ramp metering," *Transp. Res. A*, vol. 24, pp. 361–370, 1990.

[75] L. S. Yuan and J. B. Kreer, "An optimal control algorithm for ramp metering of urban freeways," presented at the 6th IEEE Annu. Allerton Conf. Circuit and System Theory, Urbana, IL, 1968.

[76] A. Kaya, "Computer and optimization techniques for efficient utilization or urban freeway systems," presented at the 5th IFAC World Congress, Paris, France, 1972.

[77] C. H. Knapp, "Traffic estimation and control at bottlenecks," in *Proc. IEEE Int. Conf. Cybernetics and Society*, 1972, pp. 469–472.

[78] L. Isaksen and H. J. Payne, "Suboptimal control of linear systems by augmentation with application to freeway traffic regulation," *IEEE Trans. Automat. Contr.*, vol. 18, pp. 210–219, 1973.

[79] H. J. Payne, W. S. Meisel, and M. D. Teener, "Ramp control to relieve freeway congestion caused by traffic disturbances," Highway Res. Board, Washington, DC, Highway Res. Record 469, 1973.

[80] M. Athans, P. K. Houpt, D. Looze, D. Orlhac, S. B. Gershwin, and J. L. Speyer, "Stochastic control of freeway corridor systems," in *Proc. 1975 IEEE Conf. Decision and Control*, pp. 676–685.

[81] D. P. Looze, P. K. Houpt, N. R. Sandell, and M. Athans, "On decentralized estimation and control with application to freeway ramp metering," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 268–275, Apr. 1978.

[82] M. Cremer, "A state feedback approach to freeway traffic control," in *Preprints 7th IFAC World Congress*, 1978, pp. 1575–1582.

[83] N. B. Goldstein and K. S. P. Kumar, "A decentralized control strategy for freeway regulation," *Transp. Res. B*, vol. 16, pp. 279–290, 1982.

[84] M. Papageorgiou, "Multilayer control system design applied to freeway traffic," *IEEE Trans. Automat. Contr.*, vol. 29, pp. 482–490, June 1984.

[85] L. Benmohamed and S. M. Meerkov, "Feedback control of highway congestion by a fair on-ramp metering," in *Proc. 33rd IEEE Conf. Decision and Control*, vol. 3, 1994, pp. 2437–2442.

[86] H. M. Zhang, S. G. Ritchie, and R. Jayakrishnam, "Coordinated traffic-responsive ramp control via nonlinear state feedback," *Transp. Res. C*, vol. 9, pp. 337–352, 2001.

[87] C. Diakaki and M. Papageorgiou, "Design and simulation test of coordinated ramp metering control (METALINE) for A10-west in Amsterdam," Dynamic Syst. Simulat. Lab., Tech. Univ. Crete, Chania, Greece, Internal Rep. 1994-2, 1994.

[88] H. Haj-Salem and M. Papageorgiou, "Ramp metering impact on urban corridor traffic: Field results," *Transp. Res. A*, vol. 29, pp. 303–319, 1995.

[89] M. Blinkin, "Problem of optimal control of traffic flow on highways," *Automat. Remote Control*, vol. 37, pp. 662–667, 1976.

[90] M. Papageorgiou and R. Mayr, "Optimal decomposition methods applied to motorway traffic control," *Int. J. Control*, vol. 35, pp. 269–280, 1982.

[91] N. Bhouri, M. Papageorgiou, and J. M. Blosseville, "Optimal control of traffic flow on periurban ringways with application to the Boulevard Périphérique in Paris," in *Preprints 11th IFAC World Congress*, vol. 10, 1990, pp. 236–243.

[92] N. Bhouri, "Commande d'un système de trafic autoroutier: Application au Boulevard Périphérique de Paris," Ph.D. dissertation, Université de Paris-Sud Centre d'Orsay, Paris, France, 1991.

[93] Y. Stephanedes and K.-K. Chang, "Optimal control of freeway corridors," *ASCE J. Transp. Eng.*, vol. 119, pp. 504–514, 1993.

[94] H. Zhang, S. Ritchie, and W. Recker, "Some general results on the optimal ramp metering control problem," *Transp. Res. C*, vol. 4, pp. 51–69, 1996.

[95] O. J. Chen, A. F. Hotz, and M. E. Ben-Akiva, "Development and evaluation of a dynamic ramp metering control model," in *Preprints 8th IFAC/IFIP/IFORS Symp. Transportation Systems*, 1997, pp. 1162–1168.

[96] H. M. Zhang, R. Jayakrishnam, and W. W. Recker, "An optimization algorithm for freeway traffic control," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 102–107.

[97] A. Kotsialos, M. Papageorgiou, and A. Messmer, "Optimal coordinated and integrated motorway network traffic control," in *Proc. 14th Int. Symp. Transportation and Traffic Theory*, 1999, pp. 621–644.

[98] J. C. Moreno-Banos, M. Papageorgiou, and C. Schäffner, "Integrated optimal flow control in traffic networks," *Eur. J. Oper. Res.*, vol. 71, pp. 317–323, 1993.

[99] B. Ataslar and A. Iftar, "A decentralised control approach for transportation networks," in *Preprints 8th IFAC Symp. Large Scale Systems*, vol. 2, 1998, pp. 348–353.

[100] J. H. van Schuppen, "Integrated traffic control of motorway networks," in *Preprints 8th IFAC/IFIP/IFORS Symp. Transportation Systems*, 1997, pp. 1156–1162.

[101] A. Alessandri, A. Di Febbraro, A. Ferrara, and E. Punta, "Optimal control of freeways via speed signalling and ramp metering," in *Preprints 8th IFAC/IFIP/IFORS Symp. Transportation Systems*, 1997, pp. 1066–1071.

[102] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed control," in *Proc. Amer. Control Conf.*, 2002, pp. 3600–3605.

[103] H. Zackor, "Beurteilung verkehrsabhängiger Geschwindigkeitsbeschränkungen auf Autobahnen," *Forschung Strassenbau und Strassenverkehrstechnik*, no. 128, pp. 1–61, 1972.

[104] S. Smulders, "Control of freeway traffic flow by variable speed signs," *Transp. Res. B*, vol. 24, pp. 111–132, 1990.

[105] K.-R. Bode and W. Haller, "Geschwindigkeitssteuerung auf der A7 zwischen den Autobahndreiecken Hannover-Nord und Walsrode," *Strassenverkehrstechnik*, vol. 27, pp. 145–151, 1983.

[106] K. Forner and U. Schmoll, "Situationsabhängige Geschwindigkeitsbeeinflussung auf der Bundesautobahn A8 bei Stuttgart," *Strassenverkehrstechnik*, vol. 28, pp. 1–4, 1984.

[107] H. Zackor and W. Balz, "Verkehrstechnische Funktionen des Leitsystems," *Strassenverkehrstechnik*, vol. 28, pp. 5–9, 1984.

[108] R. D. Kühne and P. Schick, "The effects of corridor control systems upon traffic flow," presented at the 80th Annu. Meeting Transportation Research Board, Washington, DC, 2001.

[109] M. Cremer, *Der Verkehrsfluss auf Schnellstrassen*. Berlin, Germany: Springer-Verlag, 1979.

[110] J. G. Wardrop, "Some theoretical aspects of road traffic research," in *Proc. Inst. Civil Engineers*, vol. 1, 1952, pp. 325–362.

[111] R. W. Hall, "Non-recurrent congestion: How big is the problem? Are traveller information systems the solution?," *Transp. Res. C*, vol. 1, pp. 89–103, 1993.

[112] A. Stathopoulos, "Route guidance, collective," in *Concise Encyclopedia of Traffic and Transportation Systems*, M. Papageorgiou, Ed, Oxford, U.K.: Pergamon, 1991, pp. 412–416.

[113] L. Saxton and C. Schenck, "Diversion and corridor control systems in Western Europe," in *World Survey on Current Research and Development on Roads and Road Transport*. Washington, DC: Int. Road Fed., 1977, pp. 692–723.

[114] J.-J. Henry, C. Charbonnier, and J. L. Farges, "Route guidance, individual," in *Concise Encyclopedia of Traffic and Transportation Systems*, M. Papageorgiou, Ed, Oxford, U.K.: Pergamon, 1991, pp. 417–422.

[115] C. Lamboley and G. Baudez, "Information based on travel time: A service for managing urban traffic and informing road-users," in *Proc. 1st World Congr. on Applications of Transport Telematics and IVHS*, vol. 1, 1994, pp. 181–190.

[116] H. Haj-Salem, S. Cohen, E. Sididki, and M. Papageorgiou, "Field trial results of VMS travel time display on the Corridor Périphérique in Paris," in *Proc. 4th Int. ASCE Conf. Applications Advanced Technologies Transportation Engineering*, 1995, pp. 368–372.

[117] Z. Jia, C. Chen, B. Coifman, and P. Varaiya, "The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 538–543.

[118] Y. Wang and N. L. Nihan, "Freeway traffic speed estimation with single-loop outputs," U.S. Dept. Transp., Washington, DC, Transp. Res. Record no. 1727, 2000.

[119] H. J. M. van Grol, S. Manfredi, and M. Danech-Pajouh, "On-line network state estimation and short term prediction," *Deliverable 5.2 of DACCORD, Transport Telematics Programme*, 1991.

[120] M. Papageorgiou, "Comments on Annex B of D10.3 and a suggestion for a new travel time estimator," unpublished, 1999.

[121] D. Park, L. R. Rillet, and G. Han, "Spectral basis neural networks for real-time travel forecasting," *ASCE J. Transp. Eng.*, vol. 125, pp. 515–523, 1999.

[122] K. E. Wunderlich, D. E. Kaufman, and R. L. Smith, "Link travel time prediction for decentralized route guidance architectures," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, pp. 4–14, Mar. 2000.

[123] Z. Zhou and W. Wang, "A travel time estimation model for route guidance systems," in *Proc. Int. Conf. Traffic and Transportation Studies*, 2000, pp. 547–552.

[124] N.-E. El Faouzi, "Fusion de données pour l'estimation des temps de parcours via la théorie d'evidence,", Recherche Transports Sécurité, 2000.

[125] J. Rice and E. van Zwet, "A simple and effective method for predicting travel times on freeways," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 229–234.

[126] B. Coifman and M. Cassidy, "Vehicle reidentification and travel time measurement, part I: Congested freeways," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 140–145.

[127] B. Coifman, "Vehicle reidentification and travel time measurement, part II: Uncongested freeways and the onset of congestion," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 146–151.

[128] P. Kachroo, K. Ozbay, and A. G. Hobeika, "Real-time travel time estimation using macroscopic traffic flow models," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 134–139.

[129] S. Kachani and G. Perakis, "Second-order fluid dynamics models for travel times in dynamic transportation networks," in *Proc. 4th IEEE Conf. Intelligent Transportation Systems*, 2001, pp. 253–258.

[130] H. Chen and S. G. Muller, "Use of sequential learning for short-term traffic flow forecasting," *Transp. Res. C*, vol. 9, pp. 319–336, 2001.

[131] X. Zhang, J. Rice, and P. Bickel, "Empirical comparison of travel time estimation methods," Univ. California, Berkeley, CA, California PATH Res. Rep. UCB-ITS-PRR-99-43, 1999.

[132] A. Messmer and M. Papageorgiou, "Automatic control methods applied to freeway network traffic," *Automatica*, vol. 30, pp. 691–702, 1994.

[133] Y. Pavlis and M. Papageorgiou, "Simple decentralized feedback strategies for route guidance in traffic networks," *Transp. Sci.*, vol. 33, pp. 264–278, 1999.

[134] Y. Wang and M. Papageorgiou, "Feedback routing control strategies for freeway networks: A comparative study," in *Proc. 2nd Int. Conf. Traffic and Transportation Studies*, 2000, pp. 642–649.

[135] S. Mammar, A. Messmer, P. Jensen, M. Papageorgiou, H. Haj-Salem, and L. Jensen, "Automatic control of variable message signs in Aalborg," *Transp. Res. C*, vol. 4, pp. 131–150, 1996.

[136] L. Dörge, C. Vithen, and P. Lund-Srensen, "Results and effects of VMS control in Aalborg," in *Proc. 8th Int. Conf. Road Traffic Monitoring and Control*, 1996, pp. 150–152.

[137] M. Papageorgiou, "Dynamic modeling, assignment, and route guidance in traffic networks," *Transp. Res. B*, vol. 24, pp. 471–495, 1990.

[138] M. Papageorgiou and A. Messmer, "Dynamic network traffic assignment and route guidance via feedback regulation," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1306, 1991.

[139] C. R. Berger and L. Shaw, "Diversion control of freeway traffic," presented at the 6th IFAC World Congr., Boston, MA, 1975.

[140] P. E. Sarachik and Ü. Özgüner, "On decentralized dynamic routing for congested traffic networks," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 1233–1238, Dec. 1982.

[141] A. Bolleli, V. Mauro, and E. Perono, "Models and strategies for dynamic route guidance—Part B: A decentralized, fully dynamic, infrastructure supported route guidance," in *Proc. DRIVE Conference*, 1991, pp. 99–105.

[142] Y. Hawas and H. S. Mahmassani, "A decentralized scheme for real-time route guidance in vehicular traffic networks," in *Proc. 2nd World Congr. Intelligent Transport Systems*, 1995, pp. 1956–1963.

[143] P. Kachroo and K. Özbay, *Feedback Control Theory for Dynamic Traffic Assignment*. New York: Springer-Verlag, 1999.

[144] Y. Wang, M. Papageorgiou, and A. Messmer, "A predictive feedback routing control strategy for freeway network traffic," in *Proc. Amer. Control Conf.*, 2002, pp. 3606–3611.

[145] ——, "Performance evaluation of predictive feedback routing for freeway networks," in *CD-ROM Preprints 15th IFAC World Congress*, 2002.

[146] A. Messmer, M. Papageorgiou, and N. Mackenzie, "Automatic control of variable message signs in the interurban Scottish highway network," *Transp. Res. C*, vol. 6, pp. 173–187, 1998.

[147] M. Papageorgiou, P. Gower, A. Messmer, and J. M. Morin, "Control strategies for variable message signs," in *Proc. 1st World Congr. Applications Transport Telematics and IVHS*, 1994, pp. 1229–1236.

[148] J.-M. Morin, "Aid-to-decision for variable message sign control in motorway networks during incident condition," in *Proc. 4th ASCE Int. Conf. Applications of Advanced Technologies in Transportation Engineering*, 1995, pp. 378–382.

[149] A. Messmer and M. Papageorgiou, "Route diversion control in motorway networks via nonlinear optimization," *IEEE Trans. Contr. Syst. Technol.*, vol. 3, pp. 144–154, Mar. 1995.

[150] C. Charbonnier, J. L. Farges, and J.-J. Henry, "Models and strategies for dynamic route guidance—Part C: Optimal control approach," in *Proc. DRIVE Conference*, 1991, pp. 106–112.

[151] S. Lafortune, R. Sengupta, D. E. Kaufman, and R. L. Smith, "Dynamic system-optimal traffic assignment using a state space model," *Transp. Res. B*, vol. 27, pp. 451–473, 1993.

[152] B. Wie, R. Tobin, D. Bernstein, and T. Friesz, "Comparison of system optimum and user equilibrium dynamic traffic assignment with schedule delays," *Transp. Res. C*, vol. 3, pp. 389–411, 1995.

[153] H. S. Mahmassani and S. Peeta, "Network performance under system optimal and user equilibrium dynamic assignments: implications for advanced traveller information systems," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1408, 1994.

[154] M. Ben-Akiva, M. Bierlaire, J. Bottom, H. Koutsopoulos, and R. Mishalani, "Development of a route guidance generation system for real-time application," in *Preprints 8th IFAC/IFIP/IFORS Symp. Transportation Systems*, 1997, pp. 433–439.

[155] M. B. Wisten and M. J. Smith, "Distributed computation of dynamic traffic equilibria," *Transp. Res. C*, vol. 5, pp. 77–93, 1997.

[156] Y. Wang, A. Messmer, and M. Papageorgiou, "Freeway network simulation and dynamic traffic assignment using METANET tools," U.S. Dept. Transp., Washington, DC, Transp. Res. Record 1776, 2001.

Markos Papageorgiou (Fellow, IEEE) was born in Thessaloniki, Greece, in 1953. He received the Diplom-Ingenieur and Doktor-Ingenieur (honors) degrees in electrical engineering from the Technical University of Munich, Munich, Germany, in 1976 and 1981, respectively.

From 1976 to 1982, he was a Research and Teaching Assistant at the Control Engineering Chair, Technical University of Munich. From 1982 to 1988, he was a Free Associate with Dorsch Consult, Munich, and with Institute National de Recherche sur les Transports et leur Sécurité (INRETS), Arcueil, France, from 1986 to 1988. From 1988 to 1994, he was a Professor of Automation at the Technical University of Munich. Since 1994 he has been a Professor at the Technical University of Crete, Chania, Greece. He was also a Visiting Professor at the Politecnico di Milano, Italy, in 1982, at the Ecole Nationale des Ponts et Chaussées, Paris, France, from 1985 to 1987; and at the Massachusetts Institute of Technology, Cambridge, in 1997 and 2000; and a Visiting Scholar at the University of Minnesota, Minneapolis, in 1991 and 1993, the University of Southern California, Los Angeles, in 1993, and the University of California, Berkeley, in 1993, 1997, and 2000. He is an Associate Editor of *Transportation Research—Part C*. He is the author of the books *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control* (Berlin, Germany: Springer-Verlag, 1983) and *Optimierung* (Munich, Germany: Oldenbourg, 1991; 1996), the editor of the *Concise Encyclopedia of Traffic and Transportation Systems* (Oxford, U.K.: Pergamon, 1991), and the author or coauthor of over 200 technical papers. His research interests include automatic control and optimization theory and applications to traffic and transportation systems, water systems, and further areas.

Dr. Papageorgiou is Chairman of the IFAC Technical Committee on Transportation Systems. He is a member of the Technical Chamber of Greece (TEE). He received a DAAD scholarship from 1971 to 1976. He received the 1983 Eugen-Hartmann Award from the Union of German Engineers (VDI), and a Fulbright Lecturing/Research Award in 1997. He is an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Christina Diakaki was born in Chania, Greece, in 1968. She received the Dipl.-Eng. degree in production and management engineering from the Technical University of Crete, Chania, Greece, in 1991, the M.Sc. degree in operations management from the School of Management, Institute of Science and Technology (U.M.I.S.T.), Victoria University, Manchester, U.K., in 1993, and the Ph.D. degree in operations research from the Department of Production Engineering and Management, Technical University of Crete, Greece, in 2000.

Since 1994, she has been Free Associate and a Research Rssociate of the Dynamic Systems and Simulation Laboratory of the Technical University of Crete. She has participated in numerous European and national projects and she is the author and coauthor of several research reports and papers in scientific and technical journals and scientific conferences. Her research interests include operations research and optimization, and applications to traffic and transportation systems.

Vaya Dinopoulou was born in Kozani, Greece, in 1969. She received her Dipl.-Eng. degree and the M.Sc. and Ph.D. degrees in operations control from the Department of Production Engineering and Management, Technical University of Crete, Chania, Greece, in 1992, 1994 and 1999 respectively.

Since 1992, she has been a Research Associate at the Dynamic Systems and Simulation Laboratory (DSSL) of the Technical University of Crete and, since 1999, a Contract Lecturer in the Department of Sciences of the Technical University of Crete. She has participated in European and national research projects and is the author and coauthor of several research reports and papers in scientific and technical journals and scientific conferences. Her main research interests include operations research, reliability systems, queueing theory, optimization and automatic control theories and their application to traffic and transportation systems.

Apostolos Kotsialos was born in Larissa, Greece, in 1972. He received the Dipl.-Eng. degree in production and management engineering and the M.Sc. degree in operations research degrees from the Technical University of Crete, Chania, Greece, in 1995 and 1998 respectively. He is currently working toward the Ph.D. degree in the Dynamic Systems and Simulation Laboratory (DSSL) of the Technical University of Crete, working on integrated freeway control.

Since 1995, he has been with DSSL as a Research Associate, where he has been involved in numerous research projects. His main research interests include traffic control of large scale freeway networks, numerical optimization and its applications.

Yibing Wang received the B.Sc. degree from the Department of Electronics at Sichuan University, Chengdu, China, in 1990, the M.Sc. degree from the Department of Automation, Chongqing University, Chongqing, China, in 1993, and the Ph.D. degree from the Department of Automation, Tsinghua University, Beijing, China, in 1998.

Since 1999, he has been a Postdoctoral Researcher at the Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania, Greece, His main research interests include freeway traffic surveillance and control, route guidance, urban traffic control, automated highway systems and nonlinear systems.