Novel Freeway Traffic Control with Variable Speed Limit and Coordinated Ramp Metering

Xiao-Yun Lu, Pravin Varaiya, Roberto Horowitz, Dongyan Su, and Steven E. Shladover

Freeway corridor traffic flow is limited by bottleneck flow. If the section upstream of a bottleneck is congested, the bottleneck flow will drop well below its capacity. A logical approach to maximizing recurrent bottleneck flow is to create a discharge section immediately upstream of the bottleneck. This paper proposes a control strategy for combining variable speed limits (VSL) and coordinated ramp metering (CRM) design to achieve this objective when the bottleneck can be represented as a lane (or virtual lane) reduction. At each time step, VSL is designed first, with mainline flow, on-ramp demand and length, and driver acceptance taken into account. VSL values are determined for three locations: (*a*) the critical VSL to regulate the discharge section flow to match bottleneck capacity, (*b*) the VSL for the potentially congested section (mainline storage), and (*c*) the VSL upstream of the congestion tail. With the selected VSL, the first-order mainline flow model is linearized and then used for CRM design with model predictive control to minimize the difference between scaled total travel time (or vehicle hours traveled) and the total travel distance (equivalent to vehicle miles traveled). Microscopic simulation shows that this control strategy should improve bottleneck throughput significantly.

Freeway traffic flow is limited by bottleneck flow. The causes of bottlenecks may vary from case to case. This paper considers traffic control for recurrent (but not nonrecurrent) bottlenecks for simplicity. It is known that congested upstream traffic may drop the bottleneck flow approximately 5% to 20% below its capacity, depending on location and time (*1–7*). The reasons for such drops are as follows: (*a*) the feeding flow into the bottleneck is reduced when the upstream section is congested, and (*b*) flow conservation, that is, bottleneck outflow equals its inflow. A logical way to maximize bottleneck flow is to create a discharging section immediately upstream and to regulate its flow such that the bottleneck's feeding flow is closer to its capacity flow. This paper uses combined variable speed limits (VSL) and coordinated ramp metering (CRM) to achieve this.

X.-Y. Lu, D. Su, and S. E. Shladover, California Partners for Advanced Transportation Technology, University of California, Berkeley, 1357 South 46th Street, Richmond Field Station, Building 452, Richmond, CA 94804-4648. P. Varaiya, Department of Electrical Engineering and Computer Sciences, 271M Cory Hall, and R. Horowitz, Department of Mechanical Engineering, 5138 Etcheverry Hall, University of California, Berkeley, Berkeley, CA 94720-1770. Corresponding author: X.-Y Lu, xylu@path.berkeley.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2229, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 55–65.
DOI: 10.3141/2229-07

Ramp metering (RM) is the most widely practiced strategy to control freeway traffic in the United States, particularly in California. It is recognized that ramp metering can directly control the flow into the freeway (demand) and the average density immediately downstream, which indirectly affects the traffic upstream. After drivers enter the freeway, their collective behaviors are not controlled, which determines the traffic flow pattern. In addition, from the perspective of equity among the on-ramps along a corridor and the ramp queue length limits owing to road geometry, ramp metering has to be switched off if the demand from that on-ramp is too high to avoid traffic spilling back onto arterials. Therefore, from a systems and control viewpoint, using ramp metering alone cannot fully control the freeway traffic in practice. This is the motivation for investigating the combination of RM with VSL. A recent FHWA report (*8*) summarizes the benefits of using VSL, RM, and other strategies in active traffic management (ATM).

The following acronyms are used in this paper: FD, fundamental diagram; MPC, model predictive control; TTD, total traveled distance (or vehicle miles traveled); TTS, total time spent; and TTT, total travel time (or vehicle hours traveled).

This paper focuses on mobility improvements along a stretch of freeway using VSL and RM combined. Several possible ways exist to combine VSL and RM. At each time step:

- Determine RM rate before determining VSL,
- Determine RM and VSL simultaneously using a coupled speed and density dynamics model, and
- Determine VSL first before determining RM rate.

RM was designed before VSL (*9*), which has some practical implications in the sense that many California highways have already implemented RM. Adding VSL is a natural extension. This paper uses the third approach to design a combined traffic-control strategy for maximizing the recurrent bottleneck flow. It determines VSL for maximizing the bottleneck flow, taking into account the following factors: mainline flow, on-ramp demand and length limit (storage capacity), and limits on speed variation over time and space for driver acceptance and safety.

VSL has three parts: (*a*) the critical VSL upstream of the discharge section, which regulates the discharge flow to bottleneck capacity flow (three regulators based on flow, density, and occupancy, respectively, are proposed); (*b*) VSL for the potentially congested (mainline storage) section if the demand is too high for smoothing the traffic; and (*c*) VSL upstream of the congestion tail, to reduce shockwaves by gradually decreasing the VSL. Such a higher-level design leaves

optimization to CRM. It accounts for VSL's inability to change quickly without disturbing drivers, whereas the CRM rate can. In this sense, it is suboptimal but practical. With the designed VSL, the first order mainline flow model is linearized and is then used for CRM design to minimize the difference between scaled TTT and TTD. This is essentially a trade-off between maximizing mainline throughput and accommodating more vehicles. The weight selection depends on the traffic situation. With proper formulation of the constraints, the control design problem is formulated as sequential linear programming, which has efficient numerical solutions.

This paper begins with a literature review, and then presents the overall control strategy for combining VSL and RM, followed by VSL design and CRM design. An integrated traffic simulation is then introduced. It includes microscopic simulation, macroscopic traffic control design, online optimization, and feedback control of microscopic traffic. The paper ends with concluding remarks.

LITERATURE REVIEW

In recent years, several VSL and CRM algorithms have been developed and implemented. Some of these were based on models, but some were not. In the following discussion, both approaches are briefly reviewed for RM, VSL, and their combination.

Ramp Metering

Several RM strategies were reviewed and compared by Bogenberger and May (*10*) and Zhang et al. (*11*). Scariza evaluated four ramp-metering methods: ALINEA (local traffic responsive); ALINEA/Q with on-ramp queue handling; FLOW, a coordinated algorithm that tries to keep the traffic below a predefined bottleneck capacity; and a coordinated linked algorithm to optimize a linear quadratic objective function (*12*). The most significant result was that ramp metering, especially the coordinated algorithms, was only effective when the ramps were spaced closely together. Papamichail and Papageorgiou developed a coordination strategy (HERO) for a stretch of freeway with ALINEA local traffic-responsive ramp metering, which is simple and requires no model (*13*). The implementation of HERO is reported by Papamichail et al. (*14*). A hierarchical structure of CRM strategy for freeway networks is presented in a subsequent study by Papamichail et al. (*15*), which could be considered an extension of the earlier work by Papamichail et al. (*14*) for two layers. On top of that, the third layer includes traffic state and disturbance prediction.

VSL Strategies

Lin et al. present two VSL algorithms for traffic improvement, combined with RM (*16*). These authors believe that VSL can improve not only safety and emissions but also traffic performance by increasing throughput and reducing time delay, primarily for work zones. Two control algorithms were presented. VSL-1 was for reducing time delay by minimizing the queue upstream of the work zone; and VSL-2 was for reducing TTS by maximizing throughput over the entire work zone area. Simulation results showed that VSL-1 may even outperform VSL-2 in speed variance reduction. Alessandri et al. designed VSL using the second order METANET model (*17*). The on-ramp and off-ramp flow are assumed stochastic variables with known probability density function with an optimal control approach and an extended Kalman filter for traffic state estimation. On this basis, a VSL strategy was designed by minimizing an objective function. Several objective functions were proposed including TTT and throughput.

Hegyi et al. identified two functions of VSL: speed homogenization and prevention of traffic breakdown (*18*). Prevention of traffic breakdown avoids high density, which achieves density distribution control through VSL. As an example, a VSL strategy is used to suppress shockwaves considering the whole traffic network as a system.

Wang et al. used an empirical approach to investigate the effectiveness of reducing congestion at a recurrent bottleneck and improving driver safety by using feedback to the driver with advisory variable message signs (VMSs) on an 18-km highway stretch (*19*). The feedback included (*a*) speed limit (piecewise constant with 12-km/h increment); and (*b*) warning information (attention, congestion, and slippery). The VSL strategy was based on the traffic situation upstream and downstream of the bottleneck. Data analysis showed that driver response to the speed limit and messages on the VMS was reasonable, speed was regulated to some extent, and safety was improved by approximately 20% to 30% incident–accident reduction, more significant than mobility improvements.

Papageorgiou et al. evaluated implemented VSL strategies based on data analysis (*20*). Their paper summarizes available information on the VSL impact on FD-aggregate traffic flow behavior as follows:

- Decrease the slope of the flow-occupancy diagram at undercritical conditions,
- Shift the critical occupancy to higher values, and
- Enable higher flows at the same occupancy values in overcritical conditions.

It concluded that there was no clear evidence of improved traffic flow efficiency in operational VSL systems for the implemented VSL strategies.

A simple real-time merging traffic control concept was proposed for efficient toll plaza management in cases where the total flow exiting from the toll booths exceeds the capacity of the downstream highway, bridge, or tunnel, leading to congestion and reduced efficiency due to capacity drop (*21*). The merging control strategy of toll plazas is similar to RM–ALINEA, which is different from a VSL physically since VSLs do not completely stop vehicles. RM using traffic signals decouples platoons into individual vehicles, while VSL intends to keep platoons intact.

Combined VSL and RM

An example of the use of the second-order model for combined VSL and CRM control design is reported by Papageorgiou (*22*). Abdel-Aty and Dhindsa considered the combined effect of VSL and RM in reducing the risk of crashes and improving operational parameters such as speeds and travel times on congested freeways (*23*). Caligaris et al. adopted the METANET model adapted to different vehicle classes for combined VSL and CRM design with MPC (*24*). Alessandri et al. used a second-order model for optimal VSL and RM plus an extended Kalman filter for state estimation (*25*). Optimization was done by minimizing (or maximizing) an

empirical mean cost function according to the Monte Carlo method. Papamichail et al. considered combined VSL and CRM with an optimal control approach (*26*). It claimed an algorithm feasible for large-scale systems and showed by simulation that traffic flow significantly improved with combined VSL and CRM versus using each strategy alone.

Hegyi et al. considered combined RM and VSL based on the FD with MPC approach (*27*). It is believed that RM was effective only when the traffic demand from the combination of the on-ramp and mainline did not significantly exceed downstream mainline capacity flow. Otherwise, flow would break down and RM would have no use. The basic idea in the Hegyi et al. study is that when density is high, the following chain effect would result (*27*): coordinated VSL upstream → reduced density downstream → changing the shape of FD → allowing more vehicles to move in from the on-ramp → preventing or postponing traffic breakdown if there is large demand from the on-ramp → increasing the effective range of RM.

Just because of this, VSL could reduce TTS. In this paper the combined VSL and CRM design takes into account mobility, safety, equity, and driver acceptance instead of just safety, as in most previous VSL practice. However, the results are suboptimal from the overall system viewpoint.

Mainstream traffic flow control using combined VSL and CRM was investigated by Carlson et al. (*28, 29*). These studies used an extended METANET model for tightly coupled VSL and CRM control design for freeway network traffic. A nonlinear optimization process was necessary at each time step.

HIGHER-LEVEL CONTROL STRATEGY

This section presents the main results, that is, design of VSL based on a prespecified RM strategy for a stretch of freeway as shown in Figure 1. The objective is to maximize the recurrent bottleneck flow to approach its capacity flow. The definition of "cell" is from Daganzo (*30*). MPC terminologies from Hegyi et al. (*27*) are used in the discussion below.

In Figure 1*a*, the discharge flow of two lanes will be lower than the bottleneck capacity flow because of conservation if the traffic upstream is congested: $q = 2q_b^{(u)} < Q_b$, where $q_b^{(u)}$ is feeding flow per lane into the bottleneck and Q_b is the total bottleneck capacity flow. Figure 1*b* shows the control strategy, that is, to maximize bottleneck flow by creating a discharge section upstream of the bottleneck.

Recurrent Bottleneck Characteristics

This analysis applies to a recurrent bottleneck that can be represented as a lane reduction. To understand bottleneck flow characteristics, the following concepts are crucial:

- Bottleneck capacity: the physical capacity of the bottleneck or its observed maximum flow,
- Bottleneck discharge flow: exit flow, and
- Bottleneck feeding flow: the flow at the geometric starting point of the bottleneck.

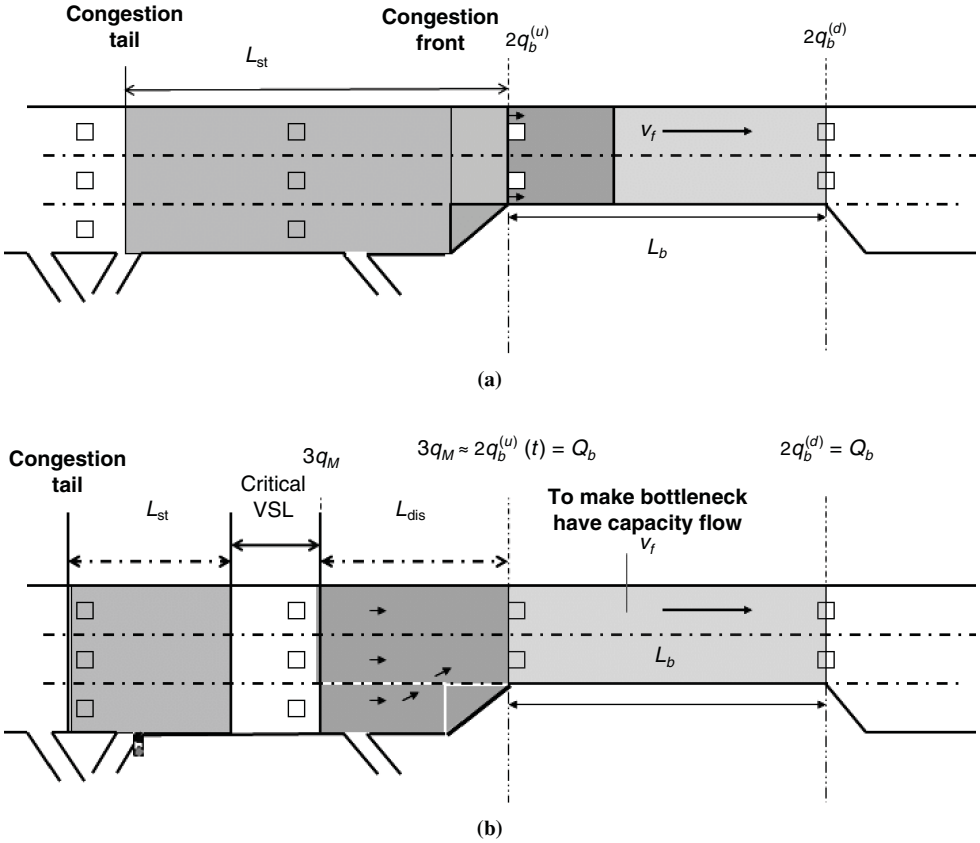


FIGURE 1 Bottleneck characteristics and control strategy [L_{st} = link distance, $q_b^{(d)}$ = lane-wise flow downstream of the bottleneck, L_b = bottleneck length, V_f = free-flow speed, q_m = flow at critical VSL location, t = time (seconds), and L_{dis} = length of discharge section].

The following cases are not distinguished: (*a*) upstream is congested but there is no queue within the bottleneck, and (*b*) both upstream and part of the bottleneck stretch are congested with queue.

Control Objective and Strategy

The control objective is to maximize the bottleneck flow and reduce shockwaves upstream to improve safety and emissions. It can be proved that maximizing the bottleneck flow is equivalent to reducing the TTS under the assumption that all the traffic has to pass the bottleneck. Based on the traffic characteristics, the following control strategy is proposed:

- If the demand upstream is below bottleneck capacity flow, upstream traffic is harmonized by VSL and CRM.
- If the demand is too high from both mainline upstream and the on-ramp and congestion is unavoidable, then create a discharge section with adequate length (approximately 500 m to 700 m) immediately upstream of the bottleneck, and use the critical VSL to control the traffic in the section upstream of the discharge section, as shown in Figure 1, such that the feeding flow to the bottleneck is closer to the bottleneck capacity flow.

This is possible if there is a geometric lane drop (such as a work zone or design) or a virtual lane drop upstream of the bottleneck. A virtual lane drop can be considered as an effective capacity reduction from a weaving section, for example as shown in Figure 2. Because of the lane drop at the bottleneck and excess demand approaching it, the section immediately upstream of the bottleneck is congested. Therefore the flow into the bottleneck drops below its capacity flow. The weaving or lane-changing effect aggravates the situation.

A practical example of virtual lane drop is the freeway diverge at I-80W and I-880S and I-580E for afternoon peak traffic, as shown in Figure 2. Some drivers destined to I-880S or I-580E use I-80W until the last second before changing to the proper lane because the traffic on I-80W is generally light during afternoon peak hours.

Combined VSL and CRM

The flowchart for the overall system including measurement and control design is depicted in Figure 3. The following equity factors need to be taken into account in the design:

- Driver equity to access the freeway from all on-ramps along a corridor, in the sense that the control strategy should not sacrifice the interests of drivers from downstream;
- Minimization of TTS, including the queue time at the on-ramps with or without ramp metering;
- Maximization of TTD (equivalent to saying that the freeway should accommodate more vehicles if demands from on-ramps are too high to avoid spillback to arterials); and
- An overall control design strategy that accounts for the inability of VSL to change quickly, whereas the CRM rate can change quickly: the optimization process could generate different RM rates at each step.

In addition to higher weight for TTD in peak hours to avoid traffic spilling back, once the queue length at the on-ramp reaches a certain level, ramp metering is switched off to allow vehicles to get onto the freeway. In this case, VSL plays a major role because it is the only remaining control mechanism available.

The notations below are grouped according to their functions:

Model parameters:

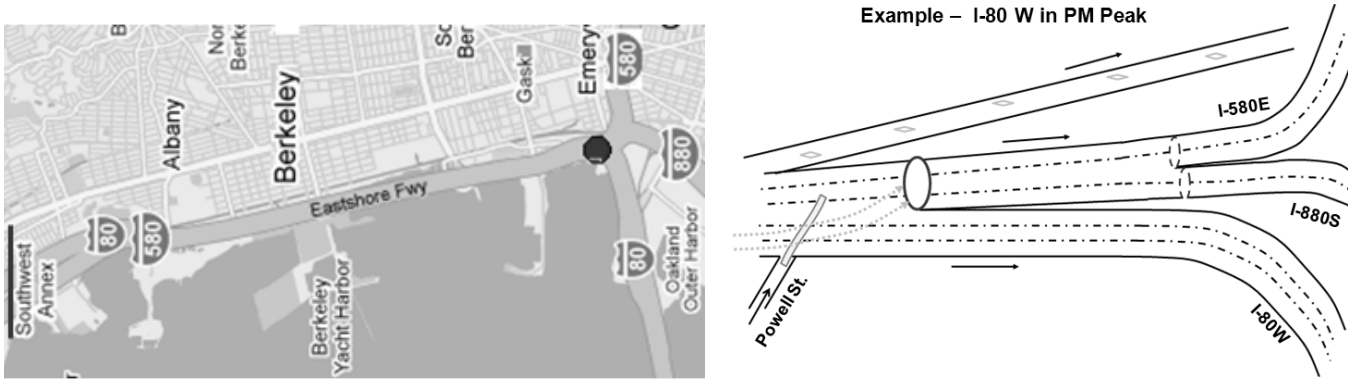
m = link index,
M = critical VSL control link index,
$M+1$ = discharge link index,
k = time index,
L = length of link m,
m_0 = index of the most upstream link affected by the bottleneck (m_0 could be a negative integer),
m_h = link index of the congestion head,
m_t = link index of the congestion tail,
γ = gain parameter to be determined in simulation, and
N_p = prediction steps for each k in MPC.

State and control variables:

q_{M+1}, v_{M+1} = flow and speed at the discharge section;
u_m = desired VSL at link m, to be designed;
$q_m(k)$ = estimated mainline flow at time k;
$\rho_m(k)$ = density of link m at time k;
$u_M(k)$ = critical VSL immediately above the discharge link, control variable; and
$r_m(k)$ = metering flow rate (veh/h), control variable.

Measured or estimated traffic state parameters:

$\overline{q}_m(k-1)$ = flow at time $k-1$, measured;
$\overline{v}_m(k)$ = speed of link m at time k, measured;
$u_0(k)$ = speed in the most upstream link, measured;



(a) (b)

FIGURE 2 I-80 (*a*) map showing west afternoon peak section and (*b*) diagram showing virtual lane drop and weaving at freeway diverge.

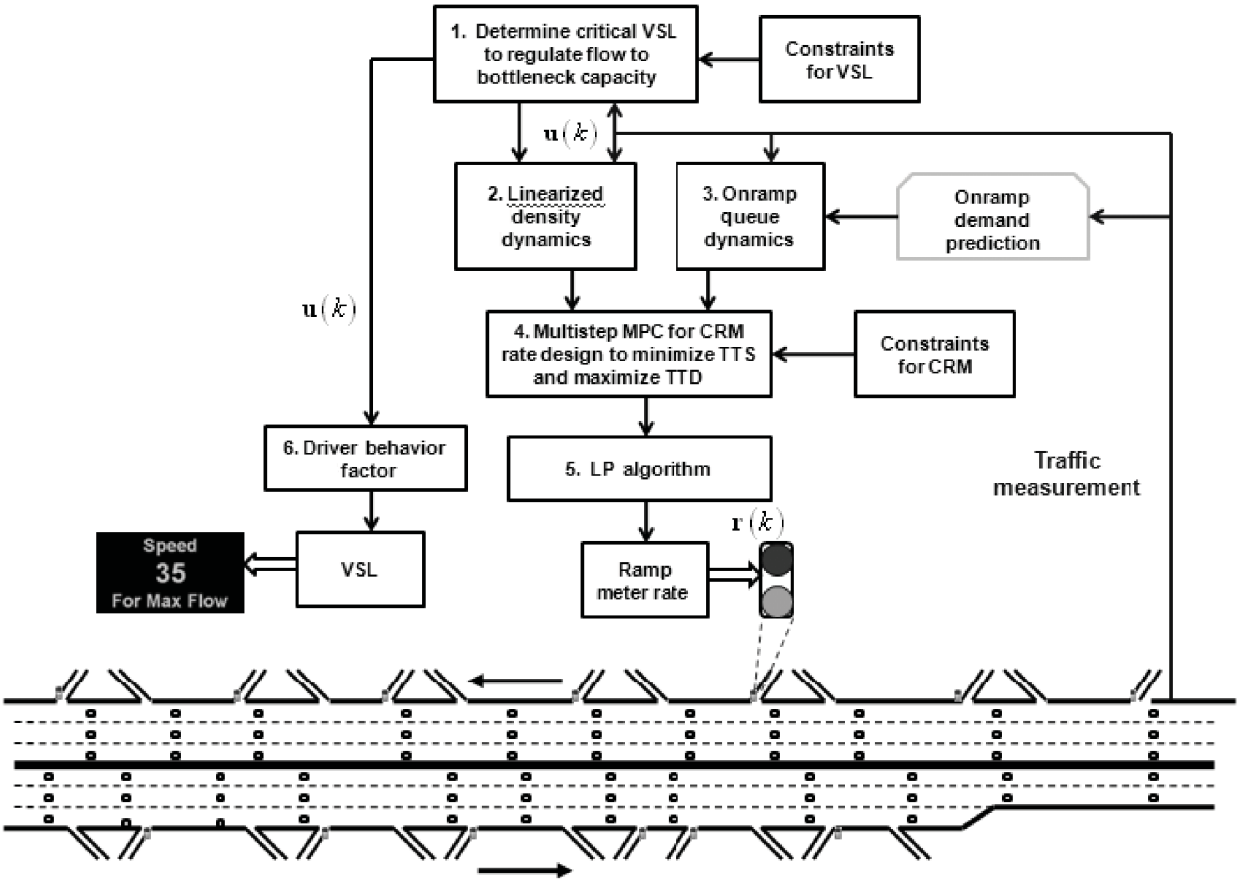


FIGURE 3 Flowchart of overall system: measurement and control design.

$\bar{\rho}_{M+1}$ = discharge link density, measured or estimated;
$s_m(k)$ = total off-ramp flow of a link (veh/h), measured;
d_m = demand from on-ramp m, measured or estimated;
$V_{st}(k)$, $\rho_{st}(k)$ = speed and density of storage section upstream of critical VSL, to be determined;
$L_{st}(k)$ = length of the storage section upstream of critical VSL, to be determined;
Q_m = mainline capacity of link m, known;
Q_b = bottleneck capacity flow, known;
$Q_{m,o}$ = on-ramp m capacity, known;
$L_{m,o}$ = on-ramp m length, known;
V_f = free-flow speed, known;
O_c = critical occupancy, known; and
ρ_c = critical density, known.

It is possible to further divide a link into cells (*18, 26*) in theory. This paper considers each link as one cell for simplicity. It is assumed that each link has exactly one on-ramp but may contain more than one off-ramp.

VSL DESIGN

The designing of VSL is divided into three parts:

1. Design VSL from the most upstream to the congestion tail along the corridor to the critical VSL point to harmonize traffic.
2. Design the critical VSL to maximize bottleneck flow.
3. Determine the V_{st} in the storage section.

Three relevant problems are also addressed:

1. Length of the discharge section,
2. Length of the potential congestion or storage section (or the congestion tail), and
3. Handling the case if the bottleneck is already congested.

Design VSL Upstream of Congestion Tail to Harmonize Traffic

This VSL strategy is designed in two stages according to the traffic. Using m_t and m_h, denote the cell index of the congestion tail and head respectively. Their determination is discussed later. Then,

$$m_0 \le m_t \le m_h \le M$$

Stage 1

The beginning of congestion can be characterized by the measured flow exceeding a threshold. The congestion tail and head are the

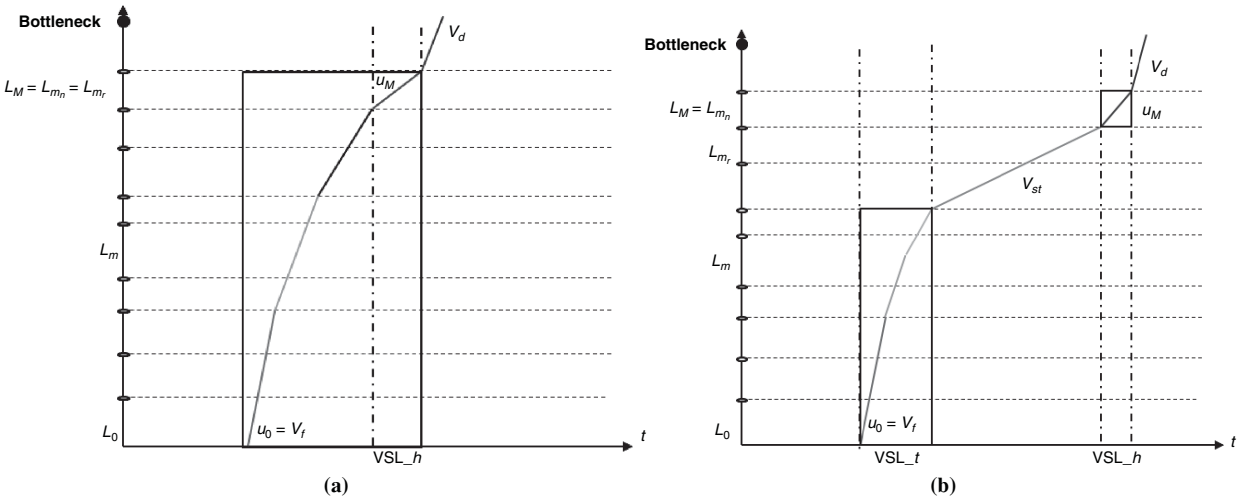


FIGURE 4 Schematic VSL control strategy at (*a*) Stage 1 and (*b*) Stage 2 [L_M = link length of critical VSL section, L_{m_h} = link length of congestion head section, L_{m_t} = link length of congestion tail section, L_m = length of link m, L_0 = length of the most upstream link (free flow), V_d = desired speed at discharge section, u_M = critical VSL, u_m = desired VSL at link m, V_{st} = VSL in saturated (congested) section, VSL_*h* = VSL at congestion head link, and VSL_*t* = VSL at congestion tail link].

same $m_0 = m_t = m_h = M$ (Figure 4). The VSL for each link in the potential influence zone can be determined as follows:

$$u_{m-1}(k) = u_m(k) + \max\left\{-5, \min\left\{(\eta\alpha_m(k) + (1-\eta)\beta_m)\left[u_M(k) - u_{m_0}(k)\right], 0\right\}\right\}$$

$$m_0 < m_h = m_t = M$$

$$u_0(k) = V_f$$

$$q_m(k) = \bar{q}_{m-1}(k-1) + R_m(k) - s_m(k)$$

$$R_m(k) = \min\left\{d_m(k), Q_{m,o}, Q_m - \bar{q}_{m-1}(k-1)\right\}$$

$$\alpha_m(k) = H(Q_m - q_m(k))$$

$$\beta_m = H(1/L_{m,o})$$

$$0 \le \eta \le 1 \tag{1}$$

where $\alpha_m(k)$ reflects the demand from the on-ramp and $\beta_m(k)$ the on-ramp length, and η is to balance the priorities between on-ramp demands and storage capacity along the corridor for equity. The recursive algorithm is the first in Equation 1. Negative 5 in the braces is the limit in miles per hour for VSL changes over time to encourage driver acceptance. The harmonic function $H(\cdot)$ is defined as: Let $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ be a real vector. Then

$$H(x_m) = \frac{\frac{1}{x_m^2}}{\sum_{\mu=1}^{M}\frac{1}{x_\mu^2}} = \frac{\prod_{\mu=1,\mu\neq m}^{M} x_\mu^2}{\sum_{\mu=1}^{M}\prod_{\eta=1,\eta\neq\mu}^{M} x_\eta^2} \tag{2}$$

The following properties are straightforward:

$$\sum_{m=1}^{M}\alpha_m(k) = 1 \qquad \sum_{m=1}^{M}\beta_m = 1$$

$$\sum_{m=1}^{M}(\eta\alpha_m(k) + (1-\eta)\beta_m) = 1 \tag{3}$$

The algorithm determines $u_m(k)$ by interpolating $u_0(k)$ and $u_M(k)$. The coefficients for the interpolation are determined by mainline acceptance capability, on-ramp demand, and physical length. The algorithm can be explained as follows:

- The VSL is monotone decreasing from the most upstream link (cell) of the affected zone with free-flow speed to the congestion tail.
- If the on-ramp demand is higher, the speed reduction at that link will be greater to allow more vehicles to be injected from the on-ramp.
- Short on-ramp length leads to more VSL reduction at that link for a similar reason.

Stage 2

After the VSL is applied for traffic at Stage 1, it is possible that a congestion section upstream of the critical VSL could further backpropagate if the demand from upstream is high, which is the case for I-80W in peak hours. This is called Stage 2 (Figure 4*b*). It is characterized as $m_0 \le m_t \le m_h = M$ and can be identified with criteria $\rho_m(k) > \rho_{st}$. The density threshold ρ_{st} is to be determined later. The VSL algorithm for Stage 2, the VSL for each link downward to the congestion tail ($0 \le m \le m_t$), can be specified as follows:

$$u_{m-1}(k) = u_m(k) + \max\left\{-5.0, \min\left\{(\eta\alpha_m(k) + (1-\eta)\beta_m)\left[V_{st}(k) - u_{m_0}(k)\right], 0\right\}\right\}$$

$$u_{m_0}(k) = V_f$$

$$q_m(k) = q_{m-1}(k-1) + R_m(k) - s_m(k)$$

$$R_m(k) = \min\left\{d_m(k+1), Q_{m,o}, Q_m - q_{m-1}(k-1)\right\}$$

$$\alpha_m(k) = H(Q_m - q_m(k))$$

$$\beta_m = H(1/L_{m,o})$$

$$0 \le \eta \le 1 \tag{4}$$

where $u_{m_0}(k)$ is the VSL at the most upstream link.

The explanation of the algorithm is similar to the above. $V_{st}(k)$ is the VSL in the congested section, which is determined later.

Determine Critical VSL and Discharge Section Length

Two methods are presented based on integral control to determine $u_M(k)$: (*a*) a flow-based regulator to regulate the bottleneck feeding flow to the capacity flow (Equation 5), and (*b*) a density-based regulator (Equation 6).

$$u_M(k) = u_M(k-1) + \gamma \cdot \min\{(Q_b - \bar{q}_{m-1}(k)), \bar{v}_{M+1}(k) \cdot (\rho_c - \bar{\rho}_{M+1}(k))\} \quad (5)$$

where ρ_{m+1} is density at link $m + 1$.

$$u_M(k) = u_M(k-1) + \begin{cases} \varsigma_1 \cdot (\rho_c - \bar{\rho}_{M+1}) & \text{if } \bar{\rho}_{M+1} < \rho_c \\ \varsigma_2 \cdot (\rho_c - \bar{\rho}_{M+1}) & \text{if } \bar{\rho}_{M+1} > \rho_c \end{cases} \quad (6)$$

The two control gains (ς_1 and ς_2) may be different to adapt to differences in the traffic situation. Such flexibility can also be used in antiwindup strategy to avoid control oscillation. In practice, the density can be replaced with occupancy, and the critical density replaced with critical occupancy:

$$u_M(k) = u_M(k-1) + \begin{cases} \varsigma_{o1} \cdot (O_c - o_{M+1}) & \text{if } o_{M+1} < O_c \\ \varsigma_{o2} \cdot (O_c - o_{M+1}) & \text{if } o_{M+1} > O_c \end{cases} \quad (7)$$

where ς_{o1} and ς_{o2} are control gains to be determined and o_{M+1} is occupancy at discharge section.

VSL algorithms Equations 1 and 4 need the length of the discharge section. The discharge section length L_{dis} is determined by the distance required for the vehicle to accelerate from zero speed (the worst case) to the desired speed:

$$L_{dis} = \frac{V_{tgt}^2}{2a_{ave}} + 200 \quad (8)$$

where a_{ave} is average acceleration and V_{tgt} is desired speed at the bottleneck. The added 200 m takes into account other important effects such as weaving and lane changing.

Determine (V_{st}, ρ_{st}, L_{st}) for Stage 2

Implementation of the VSL algorithm in Equations 1 and 4 needs the estimation of congestion tail (Figure 4*b*). For saturated traffic, the shockwave backpropagation speed is nearly constant (*31*). With VSL, it is expected that the shockwave will be diminished or even avoided. Therefore, the speed of the congestion tail propagation is expected to be smaller than the shockwave speed without control. It depends on

- Measured upstream mainline flow,
- Flow from on-ramp or RM rate,
- Roadway storage capacity, and
- VSL for the congested section.

The VSL in this section can be specified as follows:

$$V_{st}(k) \cdot \rho_{st}(k) \geq Q_b$$
$$\rho_c \leq \rho_{st}(k) \leq \rho_J \quad (9)$$

where ρ_{st} is specified first based on historical data or the operator's experience and ρ_J is jammed density. The worst case is $\rho_{st} = \rho_J$. However, one could operate at density levels $\rho_{st} < \rho_J$ depending on traffic. $V_{st}(k)$ is then determined based on Equation 9 and a static FD relationship for saturated traffic (*32, 33*). After the expected density is determined for the storage section, it can be determined whether a link upstream should be added to the storage section. This could be done using real-time data jointly with density dynamics for one-step prediction:

$$\rho_m(k+1) = \rho_m(k) + \frac{T}{L_m \lambda_m}(\lambda_m \rho_{m-1}(k) u_{m-1}(k) - \lambda_m \rho_m(k) u_m(k) + r_m(k) - s_m(k)) \quad (10)$$

where

ρ = density of link m,
L_m = length of link m,
λ_m = number of lanes in link m, and
$r_m(k)$ = ramp meter rate at time k.

At time step k, after implementing VSL $u_m(k-1)$ and ramp metering $r_m(k-1)$, the right-hand side of Equation 10 can be estimated based on measurements. If $\rho_m(k) \geq \rho_{st}(k)$, then link m is considered to be added to the storage section: $L_{st}(k) := L_{st}(k-1) + 1$; otherwise, $L_{st}(k) := L_{st}(k-1)$.

Discharge Section Clearing

If the bottleneck is already congested, it is necessary to recover the flow in the bottleneck and restore the discharge section to free flow to maximize the bottleneck flow. This can be achieved by first limiting the bottleneck feeding flow at the critical VSL point. The time required to restore the discharge section to free flow can be estimated. It is assumed that the physical capacity of the bottleneck Q_b is a known constant. The potential queue length from the congested section can be modeled as follows:

L_b = bottleneck section length (assumed known),
λ_b = number of lanes at bottleneck,
λ_{dis} = number of lanes in discharge section, and
L_{dis} = discharge section length, calculated as in previous section.

The storage capacities of those two sections are $(\lambda_{dis} L_{dis} + \lambda_b L_b)\rho_J$.

To optimize the traffic flow, it is necessary to reduce the traffic to the desired density:

$$\rho_b = \frac{Q_b}{V_b} \quad (11)$$

where V_b is the desired speed for traffic at capacity flow Q_b in the bottleneck and ρ_b is the corresponding density (as an example, $\rho_b \leq \rho_J$) and both are assumed known.

A constant discharging flow $Q_b \cdot (1 - x\%)$ is further assumed for the bottleneck. The time T_{dis} required to recover density from ρ_J to ρ_b would be

$$T_{\text{dis}} = \frac{T_{\text{dis}}\lambda_{\text{dis}}u_M\rho_M + (\lambda_{\text{dis}}L_{\text{dis}} + \lambda_b L_b)(\rho_J - \rho_b)}{Q_b \cdot (1 - x\%)}$$

$$\Rightarrow T_{\text{dis}} = \frac{(\lambda_{\text{dis}}L_{\text{dis}} + \lambda_b L_b)(\rho_J - \rho_b)}{(Q_b \cdot (1 - x\%) - \lambda_{\text{dis}}u_M\rho_M)} \quad (12)$$

u_M, ρ_M are the desired speed and density at the critical VSL point satisfying

$$u_M\rho_M << Q_b \quad (13)$$

It is clear from Equation 12 that smaller $V_M\rho_M$ will lead to shorter discharging time; and smaller capacity drop in the bottleneck will generate similar results. Suppose that the bottleneck flow maximization control strategy starts (at time t_0) after both the bottleneck and discharge section recover to their capacity.

CRM DESIGN WITH MPC

In MPC design, at time step k, the RM rate is to be determined over the predicted time horizon $k + 1, \ldots, k + N_p$:

$$r = \left[r_1(k+1), \ldots, r_1(k+N_p), \ldots, r_M(k+1), \ldots, r_M(k+N_p)\right]^T \quad (14)$$

where $r(k)$ is the ramp meter rate at time k.

Modeling

The following linearized density and on-ramp queue dynamics models are adopted:

$$\rho_m(k+1) = \rho_{m-1}(k) + \frac{T}{L_m\lambda_m}(\lambda_m\rho_{m-1}(k)u_{m-1}(k) - \lambda_m\rho_m(k)u_m(k) + r_m(k) - s_m(k))$$

$$w_m(k+1) = w_m(k) + T \cdot [d_m(k) - q_{m,o}(k)] \quad (15)$$

where w_m is the on-ramp queue length (number of vehicles).

The first is the conservation of flow (*30*). It is linear because the speed variables $u_{m-1}(k)$ and $u_m(k)$ are already the designed VSL values. This can be justified because (*a*) for strictly enforced VSL, the actual speed will be close to the designed VSL, and (*b*) for advisory VSL, if the density is high enough, even 30% driver compliance will compel speed reductions by the rest of the drivers. Such linearization and decoupling bring great advantages to control design.

Constraints

The following constraints (Equation 16) are adopted for CRM design:

$$0 \le w_m(k) \le L_m^{(r)} \cdot \rho_J$$

$$0 \le r_m(k) \le \min\{d_m(k), Q_{m,o}, \lambda_m(Q_m - \bar{q}_{m-1}(k)), \lambda_m u_m(k) \cdot (\rho_J - \bar{\rho}_m)(k))\}$$

$$0 \le \rho_m(k) \le \min\{\rho_J, \varphi(u_m(k))\} \quad (16)$$

where $\varphi(u_m(k))$ is a function of the control variable.

The first constraint is the on-ramp queue length limit; the second is the direct constraint on RM rate, which is the minimum of the four terms in the braces [i.e., on-ramp demand, on-ramp capacity, and space available in the mainline (last two terms), respectively]. $\lambda_m(Q_m - \bar{q}_{m-1}(k))$ is likely assumed in free-flow cases, and $\lambda_m u_m(k) \cdot (\rho_J - \bar{\rho}_m(k))$ is likely assumed in congestion. This consideration is motivated by the work of Daganzo (*30*).

The third constraint is an indirect constraint on RM rate through the density dynamics. $\varphi(u_m(k))$ is the curve of a specified traffic speed drop probability contour as indicated in Figure 5, with three flow contours for reference (1,000, 1,500 and 2,000 veh/h). For a given acceptable traffic drop probability, the contour gives an upper bound for the feasibility region (*33*).

Objective Function

The following objective function is used at time step k over the predictive time horizon:

$$J = \text{TTS} - \text{TTD}$$

$$\text{TTS} = T\sum_{j=1}^{N_p}\sum_{m=1}^{M} L_m\lambda_m\rho_m(k+j) \quad (\text{TTT})$$

$$+ T\sum_{j=1}^{N_p}\sum_{o} w_o(k+j) \quad \left(\begin{array}{c}\text{time delay due to}\\ \text{on-ramp queue}\end{array}\right)$$

$$\text{TTD} = \alpha_{\text{TTD},0}T\sum_{j=1}^{N_p}\sum_{m=1}^{M-1}\lambda_m L_m q_m(k+j) + \alpha_{\text{TTD},M}T\sum_{j=1}^{N_p}\lambda_M L_M q_M(k+j)$$

$$\alpha_{\text{TTD},M} >> \alpha_{\text{TTD},0} > 0 \quad (17)$$

where J is the objective function and w_0 is the on-ramp queue at Section 0.

Minimizing J minimizes TTS (or density), and maximizes TTD (to maximize mainline flow). Choosing $\alpha_{\text{TTD},M} >> \alpha_{\text{TTD},0}$ emphasizes maximizing the flow on link M.

SIMULATION

An integrated traffic control simulation platform was developed in AIMSUN with application programming interface. The platform includes the following elements:

1. A seven-link network representation of 10 km of I-80W from Carlson to the diverge of I-80 and I-580E and I-880S (with the high-occupancy vehicle lane ignored) (Figure 2);
2. Aggregated field data on traffic speed, flow, and density (occupancy) for feeding into a calibrated macroscopic traffic model;
3. Combined VSL and CRM design using the macroscopic model;
4. Feedback control at the microscopic level with VSL and CRM; and
5. Performance evaluation for comparison of different control scenarios.

The simulation is conducted for 5 h. Demand is high in the first hour and drops afterwards. Figure 6*a* shows the accumulated demand for all the on-ramps and the most upstream mainline. The legend lists the on-ramps in sequence from top to bottom as they occur in the direction of traffic flow.

Multiple strategies have been simulated: status quo, CRM only, VSL only, and a combination of VSL and CRM. Application of

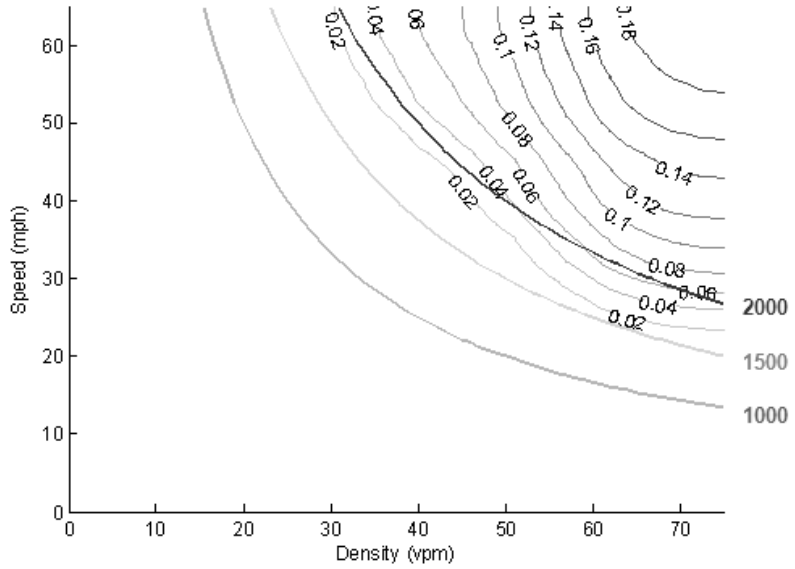


FIGURE 5 Empirical traffic speed drop probability contour versus flow contour (1,000, 1,500, and 2,000 veh/h).

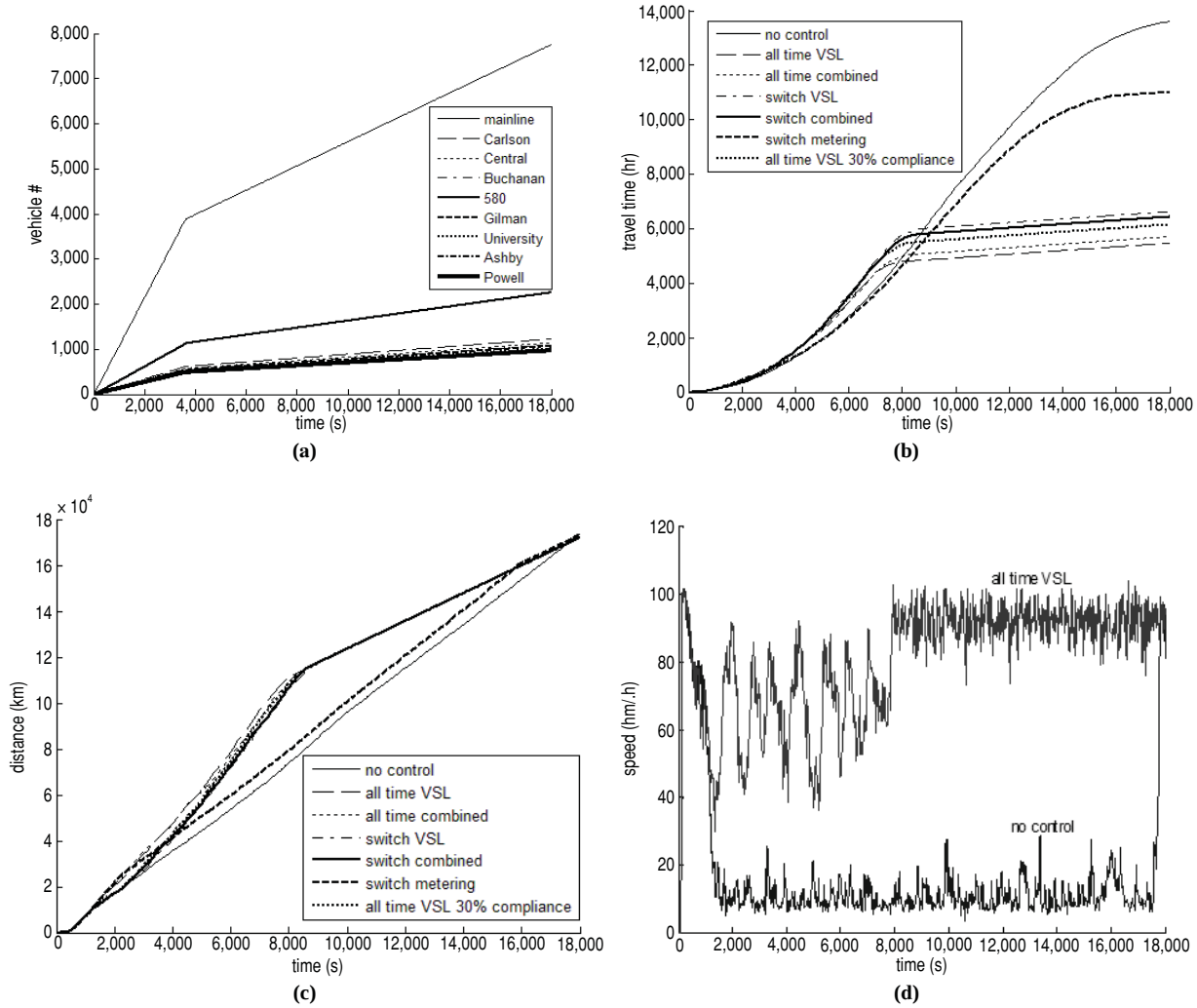


FIGURE 6 Results of 5-h simulation of I-80W corridor under different control scenarios: (*a*) accumulated demand for each on-ramp, (*b*) TTT, (*c*) TTD, and (*d*) speed at diverge.

control has been tried from the very beginning as well as switched on only when traffic flow reaches a certain level. For those control strategies, driver compliance is assumed to be 30% (advisory) and 100% (strictly enforced). Because of the high traffic density, other drivers are forced to follow the posted speed, so the performance for those two cases is very similar. Figure 6*b* and 6*c* show the performance parameters: total delay and TTT for those scenarios with all the vehicles accounted for, including those stored in the virtual on-ramp (an artificial extension of the on-ramp to accommodate all the vehicles). It can be observed that RM control alone improves system performance somewhat but not significantly, while VSL alone and the combination of VSL and CRM improve system performance significantly. This occurs because these on-ramps are short and RM control is switched off if the queue reaches 90% of the storage capacity.

As indicated in Figure 6*c,* TTD is improved significantly in peak hours. Detailed development of the integrated traffic control simulation platform is presented in a separate paper (*34*).

CONCLUDING REMARKS

Freeway corridor traffic flow is limited by bottleneck flows, so maximizing traffic flow requires maximizing bottleneck flow. This paper uses a combination of VSL and CRM to achieve this purpose. The VSL design strategy is to create a discharge section immediately upstream of the bottleneck and regulate the feeding flow into the bottleneck close to its capacity flow.

At each time step, the VSL design is a higher-level scheme that takes into account on-ramp demand and length, mainline flow, and driver acceptance considerations such as limiting VSL variations over time and space. It also linearizes the density model, based on which a refined optimal CRM design is conducted with an MPC approach. The control design problem can be formulated as a linear sequential optimization with proper constraints. The solution searching at each time step is simple, and calculation is very efficient. Simulation has been conducted over the I-80W section from Carlson to the MacArthur Maze, showing that combined VSL and CRM could potentially improve traffic performance measures such as TTS and TTD significantly.

ACKNOWLEDGMENTS

This work was supported by the Federal Highway Administration Exploratory Advanced Research Project with cost share funding from the California Department of Transportation. Constructive comments from anonymous reviewers are gratefully acknowledged.

REFERENCES

1. Banks, J. H. The Two-Capacity Phenomenon: Some Theoretical Issues. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 234–241.
2. Bertini, R. L., and M. T. Leal. Empirical Study of Traffic Features at a Freeway Lane Drop. *Journal of Transportation Engineering,* Vol. 131, No. 6, 2005, pp. 397–407.
3. Cassidy, M., and R. Bertini. Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B,* Vol. 33, 1999, pp. 25–42.
4. Hall, F. L., and K. Agyemang-Duah. Freeway Capacity Drop and the Definition of Capacity. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 91–98.
5. Persaud, B., S. Yagar, and R. Brownlee. Exploration of the Breakdown Phenomenon in Freeway Traffic. In *Transportation Research Record 1634,* TRB, National Research Council, Washington, D.C., 1998, pp. 64–69.
6. Zhang, L., and D. Levinson. Ramp Metering and Capacity of Active Freeway Bottlenecks. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
7. Winick, R. M., D. Matherly, and D. Ismart. Examining the Speed-Flow-Delay Paradox in the Washington, D.C. Region: Potential Impacts of Reduced Traffic on Congestion Delay and Potential for Reductions in Discretionary Travel During Peak Periods. Final Report FHWA-09-017. FHWA, U.S. Department of Transportation, Dec. 2008.
8. Brinckerhoff, P. Synthesis of Active Traffic Management Experiences in Europe and the United States. Final Report FHWA-HOP-10-031. FHWA, U.S. Department of Transportation, Mar. 2010.
9. Lu, X.-Y., T. Z. Qiu, P. Varaiya, R. Horowitz, and S. E. Shladover. Combining Variable Speed Limits with Ramp Metering for Freeway Traffic Control. Presented at American Control Conference, Baltimore, Md., June 30–July 2, 2010.
10. Bogenberger, K., and A. May. Advanced Coordinated Traffic Responsive Ramp Metering Strategies. California PATH Working Paper UCB-ITS-PWP-99-19. Berkeley, Calif., 1999.
11. Zhang, M., T. Kim, X. Nie, W. Jin, L. Chu, and W. Recker. Evaluation of On-Ramp Control Algorithms. California PATH Research Report UCB-ITS-PRR-2001-36. Berkeley, Calif., 2001.
12. Scariza, J. R. Evaluation of Coordinated and Local Ramp Metering Algorithms Using Microscopic Traffic Simulation. Master's thesis. Massachusetts Institute of Technology, Cambridge, Mass., 2003.
13. Papamichail, I., and M. Papageorgiou. Traffic-Responsive Linked Ramp-Metering Control. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 9, 2008, pp. 111–121.
14. Papamichail, I., M. Papageorgiou, V. Vong, and J. Gaffney. Heuristic Ramp-Metering Coordination Strategy Implemented at Monash Freeway, Australia. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2178,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 10–20.
15. Papamichail, I., A. Kotsialos, I. Margonis, and M. Papageorgiou. Coordinated Ramp Metering for Freeway Networks: A Model-Predictive Hierarchical Control Approach. *Transportation Research Part C,* Vol. 18, 2010, pp. 311–331.
16. Lin, P.-W., K.-P. Kang, and G.-L. Chang. Exploring the Effectiveness of Variable Speed Limit Controls on Highway Work-Zone Operations. *International Journal of Intelligent Transportation Systems,* Vol. 8, 2004, pp. 155–168.
17. Alessandri, A., A. D. Febbraro, A. Ferrara, and E. Punta. Nonlinear Optimization for Freeway Control Using Variable-Speed Signaling. *IEEE Transactions on Vehicle Technology,* Vol. 48, No. 6, 1999, pp. 2042–2052.
18. Hegyi, A., B. D. Schutter, and H. Hellendoorn. Optimal Coordination of Variable Speed Limits to Suppress Shock Waves. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 6, 2005, pp. 102–112.
19. Wang, H., W. Wang, X. Chen, J. Chen, and J. Li. Experimental Features and Characteristics of Speed Dispersion in Urban Freeway Traffic. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1999,* Transportation Research Board of the National Academies, Washington D.C., 2007, pp. 150–160.
20. Papageorgiou, M., E. Kosmatopoulos, and I. Papamichail. Effects of Variable Speed Limits on Motorway Traffic. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2047,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 37–48.
21. Spiliopoulou, A. D., I. Papamichail, and M. Papageorgiou. Toll Plaza Merging Traffic Control for Throughput Maximization. *Journal of Transportation Engineering,* 2010, pp. 67–76.
22. Papageorgiou, M. *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control: Lecture Notes in Control and Information Sciences.* Springer-Verlag, Berlin, Germany, 1983.
23. Abdel-Aty, M. A., and A. Dhindsa. Coordinated Use of Variable Speed Limits and Ramp Metering for Improving Safety on Congested Freeways. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
24. Caligaris, C., S. Sacone, and S. Siri. Optimal Ramp Metering and Variable Speed Signs for Multiclass Freeway Traffic. *Proc., European Control Conference,* Kos, Greece, 2007.

25. Alessandri, A., A. Di Febbraro, A. Ferrara, and E. Punta. Optimal Control of Freeways Via Speed Signaling and Ramp Metering. *Control Engineering Practice,* Vol. 6, 1998, pp. 771–780.
26. Papamichail, I., K. Kampitaki, M. Papageorgiou, and A. Messmer. Integrated Ramp Metering and Variable Speed Limit Control of Motorway Traffic Flow. *Proc., 17th IFAC World Congress,* Seoul, Korea, 2008.
27. Hegyi, A., B. D. Schutter, and H. Hellendoorn. Model Predictive Control for Optimal Coordination of Ramp Metering and Variable Speed Limits. *Transportation Research Part C,* Vol. 13, 2005, pp. 185–209.
28. Carlson, R. C., I. Papamichail, M. Papageorgiou, and A. Messmer. Optimal Mainstream Traffic Flow Control of Large Scale Motorway Networks. *Transportation Research Part C,* Vol. 18, 2010, pp. 193–212.
29. Carlson, R. C., I. Papamichail, M. Papageorgiou, and A. Messmer. Optimal Motorway Traffic Flow Control Involving Variable Speed Limits and Ramp Metering. *Transportation Science,* Vol. 44, 2010, pp. 238–253.
30. Daganzo, C. F. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research Part B,* Vol. 28, No. 4, 1994, pp. 269–287.
31. Lu, X.-Y., and A. Skabardonis. Freeway Traffic Shockwave Analysis: Exploring the NGSIM Trajectory Data. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
32. Lu, X.-Y., P. Varaiya, and R. Horowitz. Fundamental Diagram Modeling and Analysis Based NGSIM Data. *Proc., 12th IFAC Symposium on Control in Transportation System,* Redondo Beach, Calif., 2009.
33. Chow, A. H. F., X.-Y. Lu, Z.-J. Qiu, S. Shladover, and H. Yeo. An Empirical Study of Traffic Speed Drop for Freeway Management. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
34. Su, D., X.-Y. Lu, P. Varaiya, R. Horowitz, and S. E. Shladover. Variable Speed Limit and Ramp Metering Design for Congestion Caused by Weaving. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.

The Freeway Operations Committee peer-reviewed this paper.