Combining Variable Speed Limits with Ramp Metering for Freeway Traffic Control

Xiao-Yun Lu, Tony Z. Qiu, Pravin Varaiya (Fellow IEEE), Roberto Horowitz, and Steven E. Shladover

Abstract—This paper proposes a control strategy for combining Variable Speed Limits (VSL) and Ramp Metering (RM) to maximize the flow of a recurrent bottleneck which can be modeled as a lane reduction. The control strategy can be simply described as: (a) Assuming a known ramp metering rate for each onramp; (b) using Finite Time Horizon Model Predictive Control to design VSL for each link; (3) VSL design is based on a simplified 2nd order METANET model with density (or occupancy) and mean speed as the state variables. Simulations are conducted in Matlab with several performance measures to evaluate the control strategy quantitatively.

I. INTRODUCTION

Freeway traffic management has been developing very rapidly in recent years. There are many strategies to manage the traffic on freeways. Ramp metering is the most widely practiced strategy to control freeway traffic. It has been gradually recognized that ramp metering can directly control the average density immediately downstream of the onramp. Alternatively, ramp metering controls the demand from the onramp into the mainline of the freeway. After entering the freeway, the collective behaviors of the drivers are not controlled. This is why using ramp metering alone to control freeway traffic has limited performance if the demands from onramps and the upstream mainline are high. In addition, from the perspective of equity among the onramps along a corridor and the ramp queue length limit due to road geometry, ramp metering has to be switched off if the demand from that onramp 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 other control strategies such as Variable Speed Limits.

The following acronyms are used throughout the paper: VSL – Variable Speed Limit; RM – Ramp Metering; CTM – Cell Transmission Model; FD – Fundamental Diagram; TOPL (Tool for Operational Planning); SWARM - System Wide Adaptive Ramp Metering; TTT – Total Travel Time; TTS – Total Time Spent; TTD – Total Traveled Distance; MPC – Model Predictive Control; BHL - Berkeley Highway Lab;

VSL attempts to control the collective vehicle speed (or driver behavior) of mainline traffic, which is complementary to RM. Several implementations have been conducted in the UK, France, Germany and Netherlands using VSL to harmonize the traffic mainly for safety instead of for mobility improvement. It is no wonder that evaluations of those strategies did not show significant improvements in mobility, although it is generally accepted that crashes and incidents can be reduced between 25~40 percent [1]. However, one cannot conclude that VSL cannot improve mobility. This paper will focus on mobility improvement along a stretch of freeway using combined VSL and RM.

There are several possible ways to combine VSL and RM depending on what model is adopted and how the control strategy is designed, classified as follows:

- RM rate is determined before determining VSL;
- Determine RM and VSL simultaneously with coupled dynamic model involving speed and density;
- Determine VSL first before determining RM rate.

The paper uses the first approach. Specifically, it assumes that the RM strategy is given. Examples are ALINEA, HERO [2], or those in TOPL [3] based on CTM. At each time step, the known metering rates are used as input to the simplified METANET Model [4, 38] for VSL design. The control objective is to bring the bottleneck flow close to its capacity flow. The main strategy is to create a discharging section right before the bottleneck such that the feeding flow into the bottleneck could be closer its capacity flow. This consideration is based on previous works [5-10], which indicate that congested upstream traffic will reduce the bottleneck flow. A Finite Time Horizon MPC approach is used for control design. The purpose for doing so are three: (a) to avoid infeasibility in optimization encountered in the method using coupled speed-density dynamics since density dynamics is not necessary here; (b) to improve computation efficiency; (c) practical implication: ramp metering has already been implemented in many US highways; to simply add VSL to existing RM could reduce institutional issues.

The paper is organized as follows: Section 2 is for
relevant literature review of RM and VSL strategies; Section 3 presents higher control strategy and an approach for combined VSL and RM; Section 4 is devoted to MPC design based on speed dynamics; Section 5 presents macroscopic simulation results; Section 6 is for concluding remarks.

II. LITERATURE REVIEW

In recent years, model-based traffic control design has been becoming more and more popular. The analysis and control design of ramp metering based on the first order CTM is one example [11]. Another example is to use a second order model for combined Variable Speed Limit and Coordinated Ramp Meter control design in [12, 4, 13, 14]. This paper addresses modeling issues based on practical traffic control design considerations.

A. Ramp Metering

The CTM started from the nominal paper by Daganzo [15] based on the first order LWR model [16], under the assumption of a triangular shape of FD. Reference [17] further analyzed the CTM in detail. The model was refined into five modes for each cell according to the traffic situation. The controller determines the maximum flow that an on-ramp can release into the freeway. If no controller is assigned to an on-ramp, its flow is restricted by the ramp capacity and available capacity of the cell to which this on-ramp belongs. CTMSIM [18] is part of the TOPL macroscopic traffic simulation package [3], which provides several ramp metering control options, including ALINEA, and Linear Quadratic Control with Integral action.

A good review of model-based freeway ramp metering approaches is found in [19]. Coordinated ramp metering uses a second order model although speed is not controlled. Other effects including merging and weaving were also considered. The ramp metering methods are mainly two: ALINEA and Coordinated RM, with extension over traffic networks. Several RM strategies were also reviewed and compared in [20].

ALINEA, a simple local traffic responsive RM, is getting more and more popular in practice. Reference [21] evaluated four ramp metering methods: ALINEA-local traffic responsive; ALINEA/Q with onramp queue handling; FLOW - a coordinated algorithm that tries to keep the traffic at a predefined bottleneck below capacity; and the Linked Algorithm, which is a coordinated algorithm that seeks to optimize a linear quadratic objective function. The most significant result was that ramp metering, especially the coordinated algorithms, was only effective when the ramps are spaced closely together.

SWARM is a relatively new ramp meter operating system developed by National Engineering Technology (NET) Corporation. It is totally based on linear regression of measured data for prediction of density instead of model-based. A good review and implementation of SWARM is documented in [22].

B. VSL Strategies

Work in [23] presents two VSL algorithms for traffic improvement, which was combined with RM. The authors of [23] believe that VSL not only can improve safety and emissions, but also can improve traffic performance by increasing throughput and reducing time delay. The methods were primarily proposed for work zones. Two control algorithms were presented. The VSL-1 was for reducing time delay by minimizing the queue upstream of the work zone; the 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 on speed variance reduction.

Work in [14] identified two functions of VSL: speed homogenization and prevention of traffic breakdown. Speed homogenization is the reduction of speed variance. Prevention of traffic breakdown avoids high density, which achieves density distribution control through VSL. As an example, Hegyi et al [14] used a VSL strategy to suppress shock waves. Because traffic at one point in the network would affect other parts of the network, it is necessary to consider traffic control in the network as a whole system.

Work in [24] used an empirical approach to investigate the effectiveness of reducing congestion at a recurrent bottleneck and to improve driver safety by using feedback to the driver with advisory Variable Message Signs (VMS) on a highway stretch (18 km). The feedback includes: (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 20%-30% incident/accident reduction, more significant than mobility.

The Dutch Experiment [1] intended to smooth or homogenize the traffic flow along a stretch of the highway using enforced VSL. Only two speed limits, 70 and 90 km/h, were used, with 1 min update interval. Tests were conducted on multiple stretches totaling 200 km. It showed that speed control was effective to some extent in reducing speed and speed variation and the number of shock waves.

Several empirical studies have been conducted in the U.S. since the 1960’s in several states with varying levels of development for different purposes (improving traffic safety, work-zone safety, or traffic flow) [27]. The outcomes were diverse, with some positive and most negative. The most impressive positive outcome was the work conducted by the state of New Jersey, which was similar to the approach in Germany [25, 26], but with the speed enforced instead of
Papageorgiou [28] evaluated implemented VSL strategies based on data analysis. The 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 under-critical conditions;
- shift the critical occupancy to higher values;
- 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 strategy.

C. Combined VSL and RM

Reference [29] considered both VSL and RM, which are believed to be the two key tools influencing conditions on congested freeways. Their combined effect was also studied in reducing the risk of crash and improvement in operational parameters such as speeds and travel times. References [30, 31] used a second order model for optimal VSL and RM. Reference [32] considered optimal combined VSL and RM based on the METANET model using MPC. Reference [13] considered combined VLS and CRM with an optimal control approach. It claimed an algorithm feasible for large scale systems. It showed by simulation that traffic flow significantly improved with combined VSL and CRM versus using each strategy alone.

Work in [32] considered combined RM and VSL based on the FD. It is believed that RM was effective only when the traffic demand from the combination of onramp and mainline does not significantly exceed downstream mainline capacity flow. Otherwise, flow would break down and RM has no use. The basic idea of the paper is that: when density is high, the following chain effect would result - Coordinated VSL upstream \( \rightarrow \) Reduce density downstream \( \rightarrow \) Changing the shape of the FD \( \rightarrow \) Allowing more vehicles to move in from onramp \( \rightarrow \) Preventing or postponing traffic breakdown if there is large demand from on-ramp \( \rightarrow \) Increase the effective range of RM. Just because of this, the VSL could reduce TTS.

III. Design of Combined VSL with RM

This section presents the main results, i.e., design of VSL based on a pre-specified RM strategy for a stretch of freeway as shown in Figure 1. The objective is to maximize the recurrent bottleneck flow to its capacity flow. The definition of “Cell” is referred to [15]. MPC terminologies are used in the discussion below, which are referred to [32]. The following notations are used in the discussion:

- \( m \) – link index; each link has exactly one onramp; it may have no or more than one off-ramp
- \( n_m \) – the number of cells in link \( m \); each cell contains exactly one traffic detector
- \( M \) – number of links of concern
- \( i \) – cell index (a link may be divided into several cells)
- \( k \) – time index
- \( r_m(k) \) – metering flow rate (veh/hr) - assumed known based on RM strategy
- \( s_m(k) \) – total off-ramp flow rate (veh/hr); assumed to be measured and thus known to time \( k \)
- \( N_p \) – prediction steps for each \( k \) in Model Predictive Control

Reference [29] considered combined RM and VSL based on the FD. It is believed that RM was effective only when the traffic demand from the combination of onramp and mainline does not significantly exceed downstream mainline capacity flow. Otherwise, flow would break down and RM has no use. The basic idea of the paper is that: when density is high, the following chain effect would result - Coordinated VSL upstream \( \rightarrow \) Reduce density downstream \( \rightarrow \) Changing the shape of the FD \( \rightarrow \) Allowing more vehicles to move in from onramp \( \rightarrow \) Preventing or postponing traffic breakdown if there is large demand from on-ramp \( \rightarrow \) Increase the effective range of RM. Just because of this, the VSL could reduce TTS.

Figure 1. Upper: A stretch of freeway with recurrent bottleneck that can be modeled as lane drop; Lower: the discharge flow of 2 lanes will be lower than the bottleneck capacity flow due to conservation if upstream is congested: \( q = 2q^{(a)} < Q_b \); \( q^{(a)} \) – feeding flow per lane into the bottleneck; \( Q_b \) – total bottleneck capacity flow.

A. Recurrent Bottleneck Characteristics

This analysis applies to a recurrent bottleneck that can be modeled as a lane reduction or weaving, such as a work-zone lane closure, geometric design, and freeway split, etc. To understand bottleneck flow characteristics, the following concepts are crucial:

- **Bottleneck Capacity**: Physical capacity of the bottleneck or its observed maximum flow;
- **Bottleneck Discharge flow**: 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.
- **Bottleneck feeding flow**: The flow at the geometric starting point of the bottleneck.

B. Control Strategy

Based on the traffic characteristics, the following control strategy is proposed for the following situation: if the demand is too high from both mainline upstream and onramp and congestion is unavoidable, then create a discharge section with adequate length (500~700m) immediately upstream of the bottleneck by defining a critical VSL as...
shown in Figure 2 such that the feeding flow to the bottleneck is closer to the bottleneck capacity flow. This is possible if there is a geometric or virtual lane drop (due to design of work-zone) upstream of the bottleneck. Virtual lane drop is understood as the situation near a freeway split where some drivers intend to use the lanes with less traffic and to change to the target lane only in the last minute (weaving). 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.

![Figure 2](image)

**Figure 2.** Control strategy: to maximize bottleneck flow; the created discharge flow of 3 lanes is closer to bottleneck capacity flow: \( 3q^{(d)} = 2q^{(e)} = Q_b \)

**C. Modeling**

The model necessary for VSL control design needs to involve speed dynamics. Based on our analysis of the second order model, we select the model for this purpose:

\[
\rho_{m_{k+1}}(k) = \rho_{m_k}(k) + \frac{T}{L_m} (\rho_{m_{k+1}}(k) - \rho_{m_k}(k) + T \nu_{m_k}(k) - \nu_{m_{k+1}}(k))
\]

\[
\nu_{m_k}(k) = \nu_{m_k}(k) + \frac{T}{L_m} (\rho_{m_{k+1}}(k) - \rho_{m_k}(k) + T \nu_{m_k}(k) - \nu_{m_{k+1}}(k)) - \frac{1}{\rho_{m_k}(k) - \rho_{m_{k+1}}(k)} \left[ \frac{\nu_{m_k}(k) - \nu_{m_{k+1}}(k)}{T} \right] \nu_{m_{k+1}}(k) - \nu_{m_k}(k), \quad i = 0, \ldots, M
\]

where: \( v \) – speed; \( \rho \) – density; T - time step; \( L_{m,i} \) - length of cell \( i \) of link \( m \); \( \tau, \nu, \kappa \) - model parameters.

This is a simplified METANET model with two major modifications: (a) there is no further parameterization in the speed control variable \( u_{m,i}(k) \); (b) there is no assumption of the FD. The advantages of doing so include:

- Speed control variable appears linearly;
- 2-DOF for control design: both VSL and RM rate;
- Effectively avoiding model mismatch caused by discrepancies between field data and the FD curve;
- Proper constraints will be added to the optimization problem from an empirical traffic flow drop probability analysis with respect to both speed and density (occupancy) [32].

Here it is assumed that at each time step \( k \), the RM rate

\[
\hat{r} = [\tau_1(k+1), \ldots, \tau_1(k+N_p), \ldots, \tau_1(k+1), \ldots, \tau_m(k+N_p), \ldots, \tau_{m}(k+1), \ldots, \tau_{m}(k+N_p)]
\]

is independently determined by an RM strategy over the time horizon, which is necessary for the prediction over the same time horizon using the model above.

**D. Constraints**

For any given RM rate, the Critical VSL is determined by:

\[
\hat{\lambda}_m \rho_{m_k}(k+j+1) v_{m}(k+j+1) = Q_b
\]

\[
j = 1, \ldots, N_p
\]

which is further relaxed as an inequality constraint:

\[
Q_b - \epsilon \leq \hat{\lambda}_m \rho_{m_k}(k+j+1) v_{m}(k+j+1) \leq Q_b + \epsilon
\]

\[
0 < \epsilon \quad \text{is a small number.}
\]

It is an implicit constraint on the control variable through the density dynamics in (3.2). Therefore it needs converting to direct constraints to the speed control variable by recursively using (3.2) starting from the initial condition at each time step \( k \) and over the predicted time horizon. Denote

\[
u_m = [\bar{u}_m(k+1), \ldots, \bar{u}_m(k+N_p)]
\]

which are the critical VSL and can be calculated based on the RM rate. Based on considerations of safety, driver acceptance and traffic flow characteristics, the following constraints on the VSL control variable are adopted:

\[
0 \leq u_m(k) \leq \bar{v}_m
\]

\[
-5 \leq u_m(k-1) - u_m(k) \leq 5
\]

\[
5 \leq u_{m-1}(k) - u_m(k) \leq 5
\]

The first one is the bounds for the VSL, and the second and the third are the speed increment/decrement limit over time and distance for driver acceptance and enforceability in mile per hour. The following inequality limits the feasible region in the speed and density plane

\[
v_m + a \rho_m^2(k) + b \rho_m(k) \leq \eta_m
\]

where \( \eta_m \) is a design parameter to be tuned off-line. This constraint is from the following consideration: density and speed are upper bounded by a contour on the speed-density plane [33] where \( a \) and \( b \) are determined with filed dat.

**E. Objective Function**

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

\[
J = T \sum_{j=1}^{N_p} \sum_{i=1}^{M} L_{m,i} \hat{\lambda}_m \left[ \alpha_{TTT} \rho_{m}(k+j) - \alpha_{TTD} v_{m}(k+j) \rho_{m}(k+j) \right]
\]

The first term minimizes TTT (to maximize mainline flow); the second term maximizes the TTD (to accommodate more vehicles in mainline). \( \alpha_{TTT}, \alpha_{TTD} \) = (55,1) are selected to match their units for trade-off.

**IV. MPC CONTROL DESIGN FOR VSL**

MPC design is used here. For any given time starting from \( k \), the control parameters are to be determined in the MPC procedure as the decision parameters for time \( k+1 \):

\[
u = [u_m(k+1), \ldots, u_m(k+N_p), \ldots, u_m(k+1), \ldots, u_m(k+N_p), \ldots, u_m(k+1), \ldots, u_m(k+N_p)]
\]
The MPC mechanism works in logical order for each time step $k$ as depicted in Figure 3. The numerical algorithm [35] and Matlab package for Nonlinear Sequential Programming [36] are used in simulation.

To validate the proposed method, the above control algorithm has been implemented in simulation with the BHL [37] field data, which is a test site that covers 2.7 miles of I-80 eastbound immediately east of the San Francisco-Oakland Bay Bridge in California (Figure 4). Dual loop detector stations provide 60 Hz event data on individual vehicle actuations. Aggregated flow and speed information are extracted from the raw event data. In calibration, the model parameters have been chosen to minimize the quadratic errors between the model computed and the measured values of speed and flows. After the calibration procedure, the following parameter values are adopted for (3.2): $\tau = 0.02$, $\nu = 8.5$, and $\kappa = 32$.

The suggested two locations for VSL signs are between Station 1 and 2, and between 5 and 6 in Figure 4. The simulation starting time is 2:00 PM on December 1 2005, corresponding to the time index 0 on the X axis, and the ending time of simulation is 12:00 AM on December 2 2005, associated with the time index 600 minutes. From Figure 5 it should be noted that VSL control sometimes slows down the traffic flow, for example around the time index 60 (3:00 PM), but on average its effects improve the performance with all the cost functions. The initial conditions of simulation are the same for both controlled and uncontrolled cases, coming directly from the measured BHL data. The deployment of the proposed control strategy is particularly effective against congestion. The VSL improved traffic stability, with more constant flow and a higher average speed as can be seen from Figure 5 comparing the traffic with and without VSL for the peak hours 3:00 PM - 7:00 PM.

The following are the accumulated performance parameters (except the average flow) over the 10 hour simulation period and 5 lanes, showing the improvements with VSL (Table 1).

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Without VSL</th>
<th>With VSL</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT (hours)</td>
<td>5,150</td>
<td>3,510</td>
<td>-31.8%</td>
</tr>
<tr>
<td>TTD (vehicle miles)</td>
<td>157,385</td>
<td>177,645</td>
<td>+12.8%</td>
</tr>
<tr>
<td>Average Flow (veh/hr/lane)</td>
<td>1259</td>
<td>1421</td>
<td>+12.87%</td>
</tr>
<tr>
<td>Objective Function</td>
<td>125,865</td>
<td>15,405</td>
<td>-87.8%</td>
</tr>
</tbody>
</table>

VI. CONCLUDING REMARKS

The METANET model has been simplified by dropping the assumption of FD and the re-parameterization of the speed control variable. With the simplified METANET model involving speed and density dynamics and under the assumption that the RM rate is pre-determined by a separate approach at each time step $k$, VSL control has been designed using Finite Time Horizon MPC. Simulation has been
conducted over the I-80 BHL section, showing that VSL alone improves traffic noticeably.

Since the density directly affects the traffic flow, even a local high density could cause a moving jam [7]. It is necessary to investigate the spatiotemporal characteristics of density for optimal ramp metering.

ACKNOWLEDGEMENTS

This work was supported by the Federal Highway Administration (FHWA) Exploratory Advanced Research Project (Cooperative Agreement DTFH61-07-H-00038) with matching funding from the California Department of Transportation (Caltrans). The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of FHWA or the State of California. This paper does not constitute a standard, specification, or regulation.

REFERENCES

[9] B. Persaud, S. Yagar, D. Tsui, and H. Look, Breakdown-Related Capacity for Freeway with Ramp Metering, Transportation Research Record 1748, Paper No. 01-2636
[38] X. Y. Lu, P. Varaiya, and R. Horowitz, An equivalent second order model with application to traffic control, CD ROM of 12th IFAC Symposium on Control in Transportation Systems, Redondo Beach, CA, USA, September 2 - 4, 2009

2271