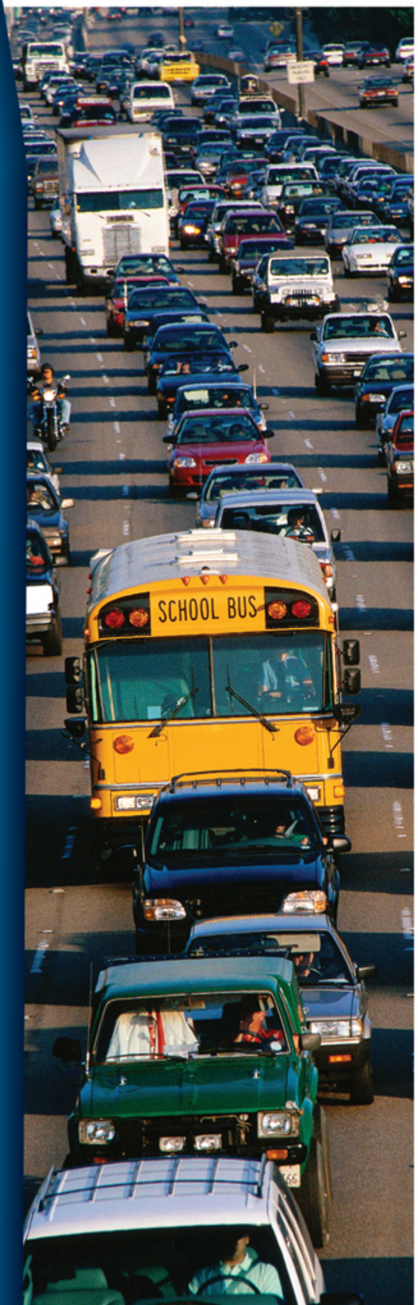
Recurring Traffic Bottlenecks: A Primer

Focus on Low-Cost Operational Improvements

U.S. Department of Transportation
Federal Highway Administration



June 2009

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-HOP-09-037		

4. Title and Subtitle	5. Report Date
Recurring Traffic Bottlenecks: A Primer **Focus on Low-Cost Operational Improvements**	**June 2009**
	6. Performing Organization Code
7. Author(s) **Richard A. Margiotta, CSI. / Neil Spiller, FHWA**	8. Performing Organization Report No.
9. Performing Organization Name and Address **Cambridge Systematics, Inc.** **1265 Kensington Drive** **Knoxville, Tennessee 37922**	10. Work Unit No. (TRAIS)
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address **U.S. Department of Transportation** **Federal Highway Administration / Office of Transportation Management** **1200 New Jersey Avenue, SE** **Washington, D.C. 20590**	13. Type of Report and Period Covered
	14. Sponsoring Agency Code **HOTM**

15. Supplementary Notes

Contact: Neil Spiller, FHWA, HoTM, 202-366-2188 neil.spiller@dot.gov

16. Abstract

This is the updated, 2nd version of this document. (The 1st version is FHWA-HOP-07-130)

While many of the nation's bottlenecks must be addressed through costly major construction projects (i.e., "mega projects") or costly transportation alternative solutions (i.e., high occupancy vehicle or toll lanes, dynamic pricing, investments in transit alternatives, etc.) there is a significant opportunity for the application of operational and low-cost "fixes" at spot-specific locations.

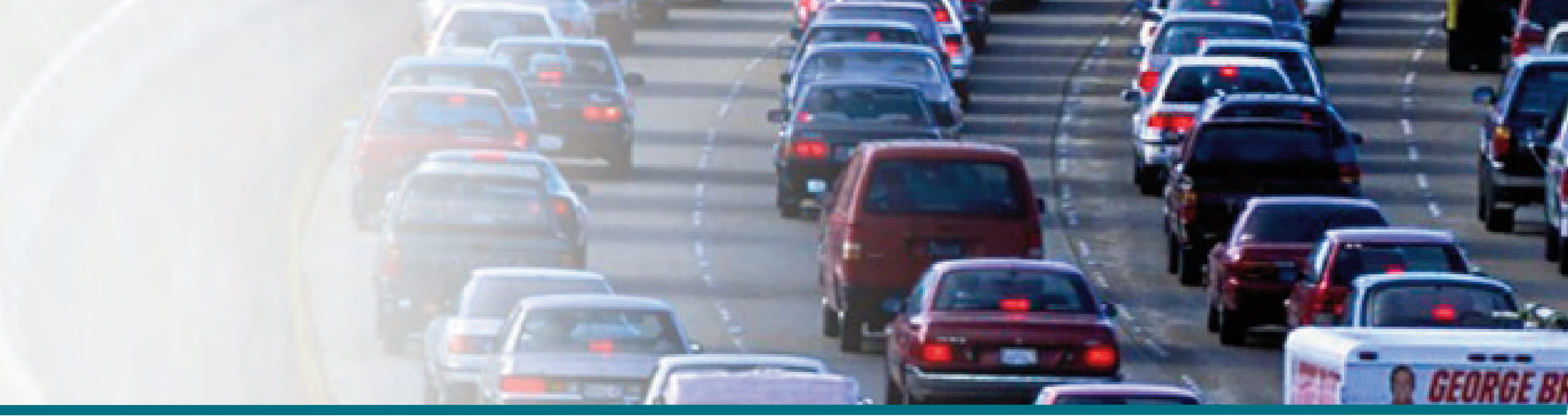
This Primer is the signature product of the Localized Bottleneck Reduction (LBR) Program, which is administered out of the Office of Operations, Office of Transportation Management, at FHWA HQ in Washington, D.C. The LBR program is focused on recurring congestion chokepoints (as opposed to nonrecurring congestion causes) and the operational influences that cause them. Widening, lengthening, retiming, metering, or bypassing these problem areas to unclog them can often be done with lower cost, lesser intensive means than traditionally waiting for a complete facility rebuild or an out-year project.

In much the same way that transportation agencies might have an annualized safety-spot improvement program, so to should they have an annualized congestion-spot program. If the ultimate fix need be a complete facility overhaul (e.g., high cost replacement, upgrade, or defacto new facility) then so be it; but an agency shouldn't limit itself to only "building our way out of congestion."

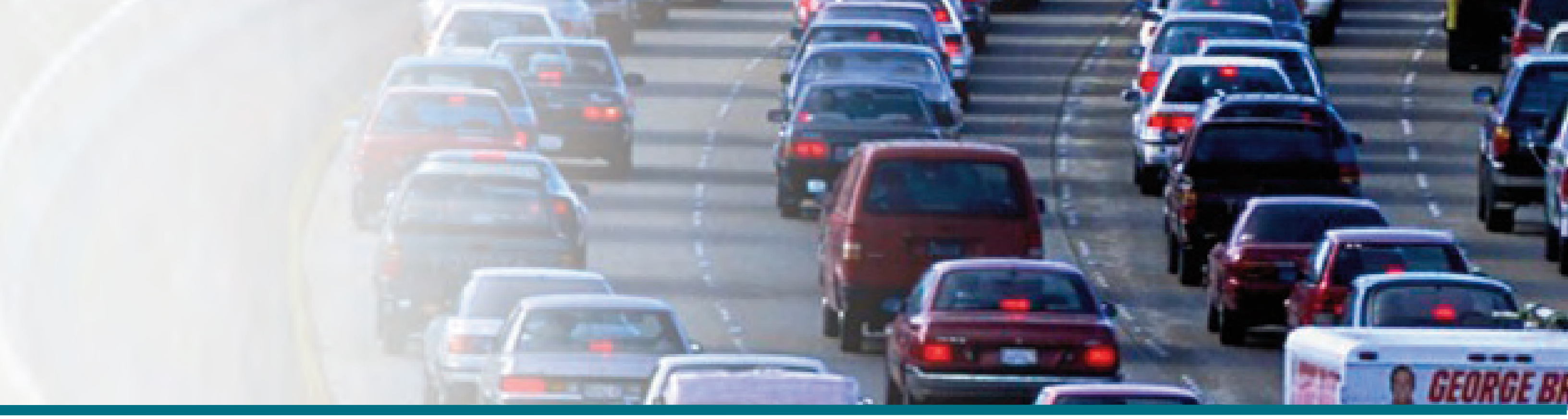
17. Key Words	18. Distribution Statement
bottleneck, chokepoint, recurring congestion, low cost improvements, operational deficiencies, operational influences, lane drops, weaves, merges, metering	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassifed	**Unclassified**	**58**	**na**

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized




TIP
TRANSPORT
INTERNATIONAL
POOL


GEORGE BR

Introduction

The delays arising from traffic congestion are an unavoidably frustrating fact of life. Or are they – unavoidable, that is? Why must we accept to take thirty minutes to make what should be a fifteen minute drive? Speed is distance divided by time; unless of course one has to pointlessly add time sitting in bumper to bumper traffic! In that case, speed drifts into "delay," which in turn drifts into our lives. In bygone days one could expect a vehicle trip to be governed only by the "straight line" distance and the performance of car and driver (with a necessary nod to speed limits too). In today's world, that same trip often factors "expectations" of congestion delay caused by others. Not us, mind you, but others who, if they would only get out of our way, would free that trip back to the bygone era.

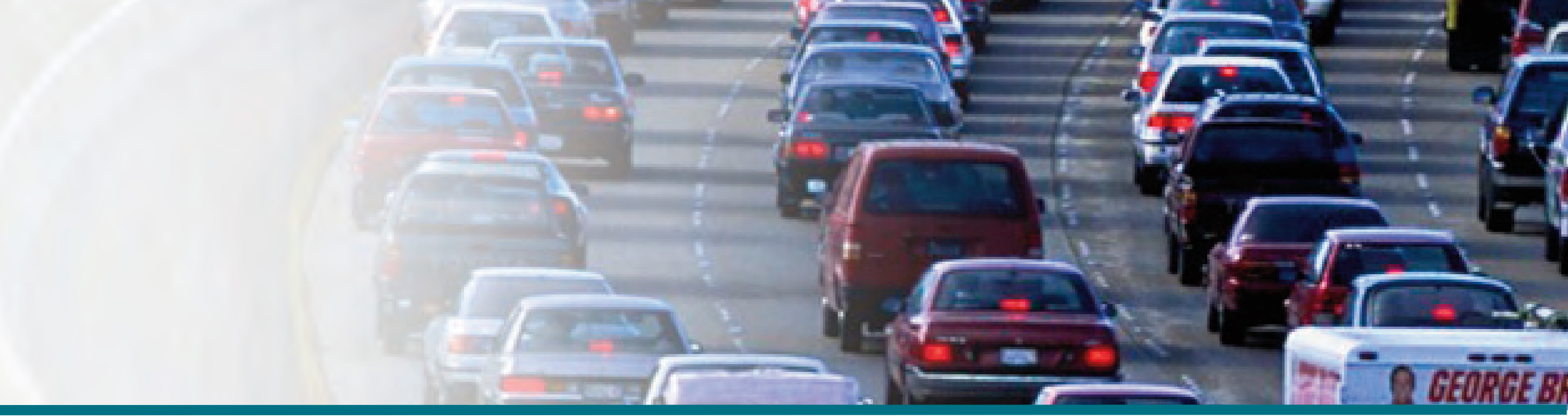
According to a February 2007 Harris Poll, just over one-third (37 percent) of respondents cite traffic congestion as a serious problem in the community, while one-quarter say traffic congestion is a serious problem that is not being addressed. But please don't tell the thousands of practicing traffic engineers, planners, and road workers that congestion is not being addressed. It's just that much like weather forecasting, traffic management is a dynamic activity that makes it an ever evolving profession. And like weather forecasting, we are getting better and better at it, but remain at the whim of unrelenting and ever-evolving "fronts." But there are some conditions we can do something about given the right resolve.

"Close to half of all congestion happens day after day at the same time and location."

Source: http://www.fhwa.dot.gov/congestion/describing_problem.htm.

Much of recurring congestion is due to physical bottlenecks – potentially correctible points on the highway system where traffic flow is restricted. While many of the nation's bottlenecks can only be addressed through costly major construction projects, there is a significant opportunity for the application of operational and low-cost infrastructure solutions to bring about relief at these chokepoints. This document, ***Recurring Traffic Bottlenecks: A Primer – Focus on Low-Cost Operational Improvements***, describes such facility breakdowns and explores the opportunity for near-term operational and low-cost construction opportunities to correct them.

This Primer is intended to be a dynamic work-in-progress. As newer strategies and discourse become available they will be uploaded to the FHWA's "Bottleneck" web site **http://www.ops.fhwa.dot.gov/bn/index.htm** and the Primer will be periodically updated. This Primer constitutes "Version 2." The Primer is a key resource for Federal Highway Administration's Localized Bottleneck Reduction (LBR) Program, providing a virtual forum for peer exchange between members of the transportation community interested in alleviating bottleneck congestion. The LBR program, initiated in 2006, is designed to expand the portfolio of bottleneck reduction tools available to transportation agencies to encompass innovative, readily adopted strategies for reducing congestion at bottleneck locations.



Understanding Bottlenecks

What is a "Traffic Bottleneck"?

Webster's dictionary defines a "bottleneck" as: i) a narrow or obstructed portion of a highway or pipeline, or ii) a hindrance to production or progress. Certainly the elemental characteristics of traffic bottlenecks exist in these descriptions. However, a road does not necessarily have to "narrow" for a bottleneck to exist (e.g., witness bottlenecks caused by a weave condition, sun glare, or a vertical climb).

Traffic bottlenecks (hereafter, bottlenecks) have a myriad of causes and durations. The most egregious ones tend to be freeway-to-freeway interchanges, but we all know that smaller, lesser chokepoints are frustrating too. Many of these chokepoints are "operationally influenced bottlenecks," defined as localized sections of highway where traffic experiences reduced speeds and delays due to recurring operational conditions. The fact that many recurring locations are "facility determinate" (i.e., the design condition contributes to the resulting backup) is both encouraging and discouraging. Facility design is a tangible feature that can always be improved; however the cost or the necessary right-of-way may be prohibitive.

Bottlenecks may be compared to constricted storm pipes that can carry only so much water – during floods the excess water just backs up, much the same as traffic does at bottleneck locations. Like the friction that results from turbulence in a pipe, once traffic flow breaks down to stop-and-go conditions, capacity is reduced – fewer cars can get through the bottleneck because of the extra turbulence. The options available to free up "the pipe" are to carry less water, increase the size of the pipe or system, or remove whatever constriction exists (i.e., remove the kink, reduce cavitations, or reduce head loss due to friction at junctures, et al.). Obviously the pipe comparison is analogous to options available to improve traffic flow; namely, to reduce demand (by effecting driver behavior), build new infrastructure, or improve that which already exists. Bottleneck mitigation is the third of these options.

What is "Congestion"?

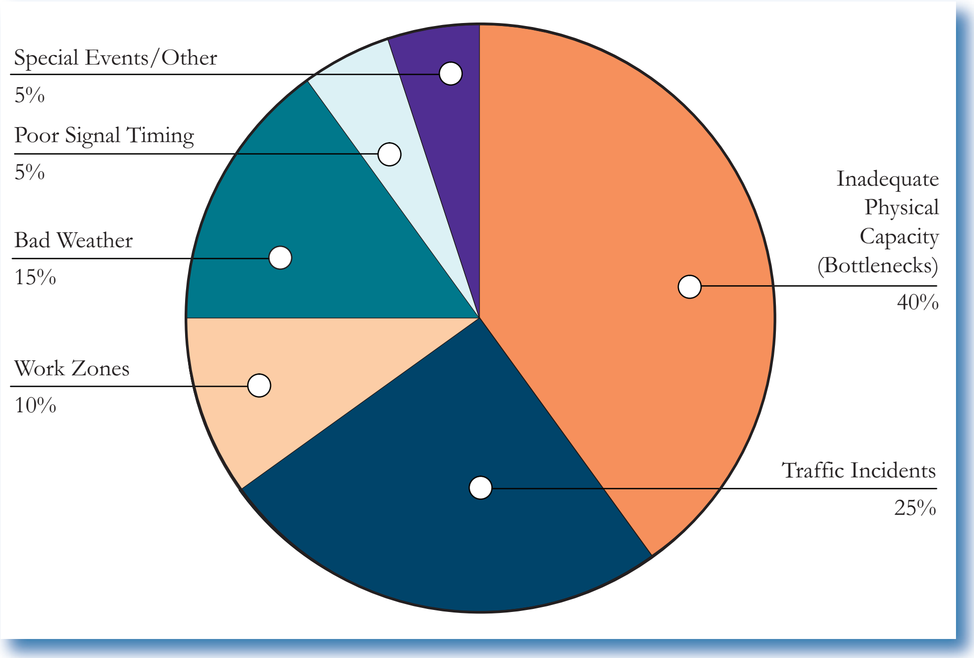
FHWA's Traffic Congestion and Reliability Report defines congestion as *"an excess of vehicles on a roadway at a particular time resulting in speeds that are slower – sometimes much slower – than normal or free flow speeds; [Congestion is] stop-and-go traffic."* The root causes of congestion have long been understood, and there is now broad consensus that congestion generally reflects a fundamental imbalance of supply and demand. That is, during hours of peak usage of the transportation facilities most desirable to motorists, the supply of roadway capacity is insufficient to meet the demand for those facilities.

Traffic congestion can be characterized as either "recurring" or "nonrecurring." This distinction is useful in helping the community of transportation professionals devise strategies that will either mitigate or reduce congestion.

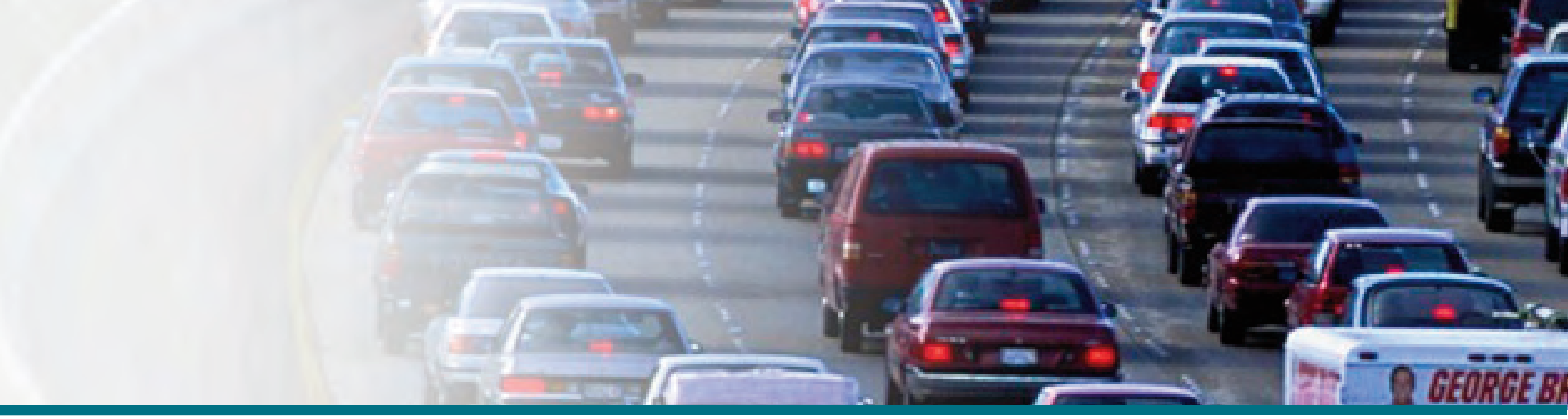
Recurring congestion happens in roughly the same time and place on the same days of the week. It results when physical capacity is simply not adequate to accommodate demand during peak periods. Causes of recurring congestion include inadequate physical capacity and poor signal timing.

Nonrecurring congestion occurs randomly and is caused by events such as work zones, traffic incidents, and bad weather. Even if "planned" in some cases, these occurrences are irregular and are not predictably habitual or recurring in location. Obviously, when these nonrecurring events occur on an already congested facility, the impacts are magnified. Figure 1 presents a pie chart showing the factors that cause traffic congestion.

Figure 1. Sources of Traffic Congestion



Source: http://www.fhwa.dot.gov/congestion/describing_problem.htm.



So Who's to Blame?

One side of a popular coin says, "There are not enough roads out there!" The other side says, "There are too many cars!" And the edge of the coin reads, "What's out there doesn't work well!" Boy, it's just like humans to blame everything else – except ourselves.

The title of Tom Vanderbilt's 2008 book "Traffic: Why We Drive the Way We Do (And What it Says About Us)" reflects the human traits – and not the physical network – that have conspired to result in the system that we have (and how we use it). Sure, there is the inevitable "state of the system" baseline that discusses the age of the network, some measures of effectiveness (or lack thereof) and a "where do we go from here" look forward. But an equal amount of pages are given to psychologists as to engineers in explaining why the need to overcome our human traits drives the incremental technological advancements; something akin to the old adage that building better and more roads begets buying more vehicles and making more trips. These human factors include driver attention, reaction, and tendencies, and explain that we are creatures of habit or need (witness our penchant to battle the peak hour despite our hatred for it), that we essentially drive "for our personal gain" instead of for the greater good (e.g., reference his comparisons of ants and other animal colonies that seem to work for a common goal), and that we suffer a myriad of transformations (e.g., road rage, selfishness, vehicle envy, distraction-by-gadget, single-driver preference, et al.) that conspire to erode our driving skills and civility, the latter of which we are otherwise forced to address face-to-face when we are outside of our four-wheeled cocoons. Inside our cars we become emboldened; protected by 2,000 pounds of steel and anonymous behind our tinted windows.

George Carlin famously observed that "everyone driving slower than you is an idiot and everyone faster is a maniac."

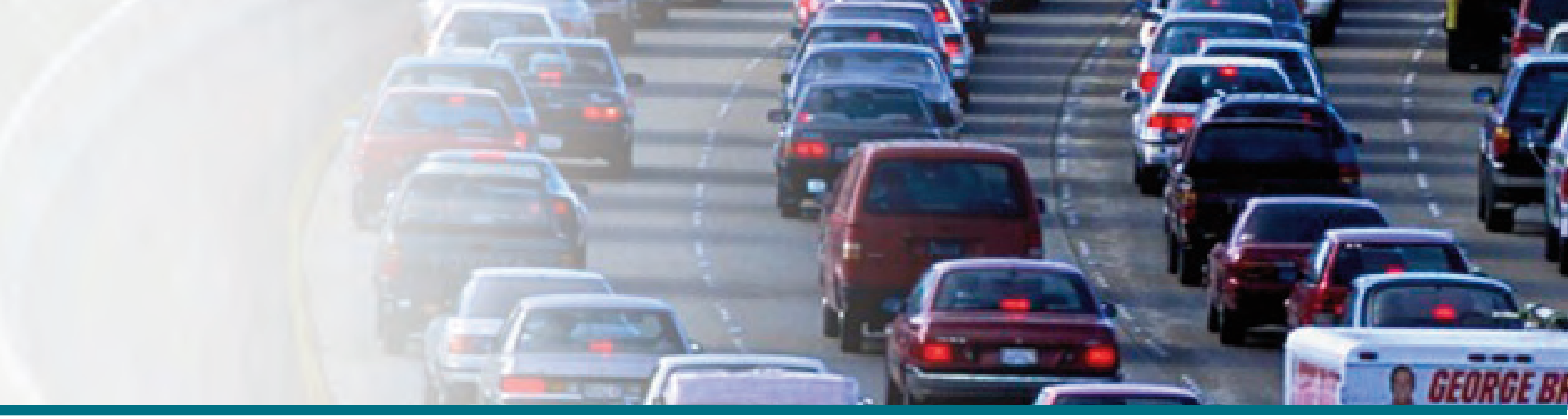
Any slight – in terms of delay, encroachment in our lane, or imposition on our perceived "right" to an unencumbered trip – is a personal affront. A driver passes us; we feel compelled to pass him. We switch to the "fast lane" and invariably so do others, such that we are now in the "slow lane." Vanderbilt cites research that demonstrates 1) how this creates "density friction" that degrades efficiency of traffic for everyone; 2) how you and a neighboring "spotter" car will progress roughly equal; and 3) the fact that all that jockeying gains nary a time advantage when weighed against the nominal case of just staying with one lane. But still we lane jump. Compare to the fact that if one flips a coin enough times, the odds of heads versus tails eventually narrows to 50 percent each. But we continue to switch lanes – or play the odds – thinking that we can beat them, because every once in a while, we do. Further evidence of our human consequence exists in the fact that most all agencies now use the term "crashes" and not "accidents" to record those incidents, to overcome the excuse that "accident" somehow removes

responsibility from our conscience. It's not an "accident" (except perhaps to the wrong place/wrong time victim) if you had been paying due diligence, staying awake, not talking on your cell phone, or not driving while impaired.

Specific to the topic of congestion (and bottlenecks), Vanderbilt's citations confirm what we intrinsically know; that in survey after survey the majority of us consider ourselves superior drivers to practically anyone else despite that in other surveys we are very poor at estimating time spent in delay, distances traveled, and the cost of that delay. A quarter is worth twenty-five cents to everyone, but a minute is overestimated by the guy sitting in a queue, and underestimated by the guy racing to an appointment.

The aforementioned "superiority" manifests itself in our belief that somehow we are more worthy of being on the road and it is everyone else who is clogging it up. We are ridiculously inept at estimating our own speed ("I was going 30 MPH" when in fact he was clocked at 46); estimating time spent circling for a parking space ("I must have spent 5 minutes looking" when in fact she only spent 1 minute 45 seconds from street to space); and estimating others' speed ("everyone travels 50 down my street" when in fact, a radar study found an average of 34 MPH and only two cars doing above 46 MPH in a three-day study). Or maybe in the latter case, it was the same car observed twice; hard to tell, but a moot point in terms of accusing "everyone." We often travel at a speed that the road "allows" and not what the sign mandates. We're not intentionally flouting the law, but hey, we know better than the police and the engineers what it takes to get from here to there! After all, the engineers dutifully allow for a "safety factor" in their design; that's for me, right? Sometimes the prior examples of estimation can be written off as exaggeration but research consistently reproduces like results when people are observed first hand, or when focus groups are asked situational questions.

In the context of "a road never jumped up and bit me" it seems hardly right to "blame" the roads, absent poor maintenance or a malfunctioning signal. So we blame the planners, the engineers, the politicians, and the "idiots" on the road. Never mind that those same "idiots" are the neighbors, customers, and peers we somehow otherwise abide. Clearly our desire to drive more often, and further, and own more vehicles, has strained our major highway arteries. The effort becomes to reduce the *rate* of congestion, if not to eliminate it altogether in as many locations as funding and right-of-way can bear. All the while we will swim upstream against perpetual waves of drivers, mileage, and demand. So the battle against congestion wages on while our human instincts contribute to the malaise.



The Frustration of Congestion

It seems hardly necessary to state that we are frustrated by congestion. We are frustrated because we feel we can't control it. We can control most everything else in our lives. We control what, when, and how much TV we watch because we have the remote and the DVR! We control what, when, and how much we eat. To the best of our ability, we control what we do (our jobs) and where we do it; although we'd like a *little* more control on how much they pay us! We control our pastimes, our hobbies, and our vacations. We somewhat control our spouses and children; or at least we have influence upon them. But unlike the scene in *Bruce Almighty*, we can't part the "sea" of traffic in front of us like Jim Carrey did.

But do we have to *accept* congestion? Our grandparents – and theirs – didn't have to endure the levels of congestion that we see today. But they also didn't enjoy the extent of today's roadway network, the availability of gas stations and services, and the efficiencies and perks of modern cars. Time marches on, not back, so we are left to suffer the ravages of our success.

Why Can't it be Like This All the Time? (Uncongested, That Is!)

Shopping mall planners design parking lots for the twelfth highest demand day of the year so as not to "waste" valuable land that would otherwise be squandered 96 percent of the time. (Similar to insurance actuary tables, the "12^{th} highest" axiom is an accord based on the intrinsic worth of land cost, parking demand, and of course, retail profit.) No doubt, during the Christmas shopping season, we all would love to get those "unnecessary" spaces back! Recurring congestion is sort of like that, except that highway builders don't intentionally under-build; it just seems that way. Highways are built to "design hour" and "design year" tenets, but population and traffic demand are ever growing – and rarely receding – over time. Outside of peak hours, the great majority of highway facilities are more than sufficient for the dynamic demand. Unfortunately, we tend to see highways at their busiest hours. But we occasionally experience the off-hour conditions too, i.e., when traffic flows freely. And like those too-few vacations when we are pampered and coddled, we ask "why can't it be like this all the time?" Recurring congestion is a function of physical constraints, as well as how *and when* drivers interact with a given facility and other vehicles on the facility.

A "Bottleneck" May Be "Congestion," but Congestion is Not Always "Just a Bottleneck"

The word "paper" can mean either a single sheet or a ream; likewise, "congestion" can be a single bottleneck – or much more. A bottleneck is distinguished from congestion in that it occurs at a specific location, and not pervasively along the entire

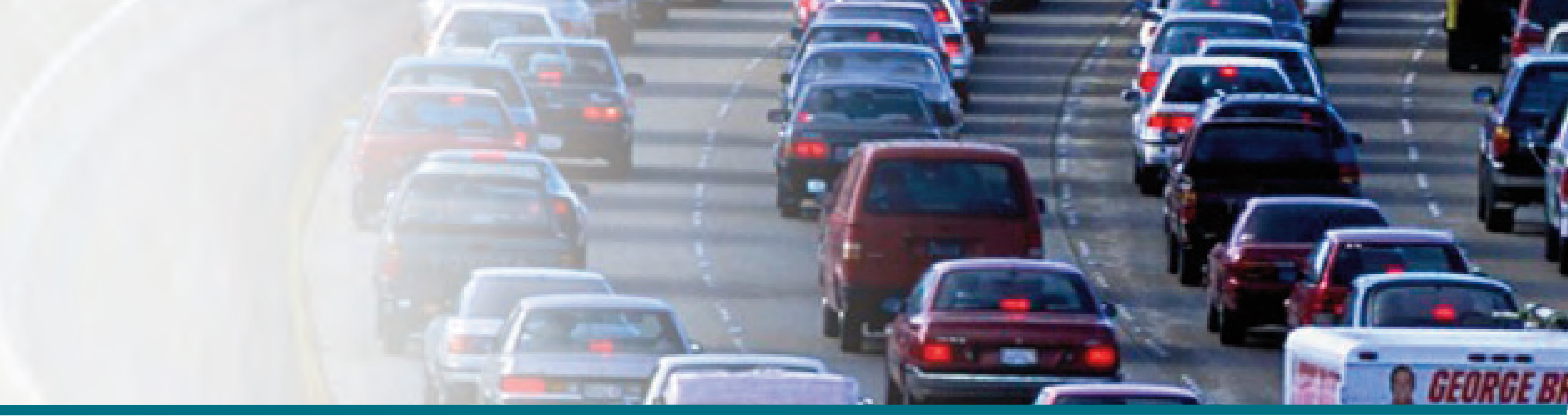
corridor. Recognizing this difference is the first important distinction in developing strategies for one or the other. When too many vehicles compete along all segments of a facility, corridor- or systemwide "congestion" will inevitably result. It is overarching in nature. But when only subordinate segments of that facility are burdened, then operationally recurring bottlenecks are said to exist. In this context then, a bottleneck certainly constitutes "congestion," but congestion cannot be said to be universally analogous to a "bottleneck."

The FHWA estimates that 40 percent of all congestion nationwide can be attributed to recurring bottlenecks (i.e., inadequate physical capacity) and another 5 percent is attributable to inefficient traffic signalization. The good news is that all these things are potentially correctible by remediation. The bad news is that there are many, many candidate locations, and agencies are fiscally constrained on how much they can do. All things being equal, a recurring bottleneck will disappear once traffic demand has decreased to a point where the operational or geometric deficiency is no longer a factor, while a nonrecurring bottleneck will disappear only after the random event has been removed. A nonrecurring bottleneck may be further impacted by traffic volumes, but it is not caused by traffic demand. Conversely, only a physical improvement will relieve a recurring bottleneck.

What Elements Typically Exist to Define a "Bottleneck"?

The Localized Bottleneck Reduction Program focuses on operationally influenced locations; that is to say, those that have a fundamentally design-based cause, resulting in recurring delays of generally predictable times and durations. The root cause of traffic flow degradation at the subject point of a recurrent constriction is almost always a correctable problem. The following conditions either exist or help to identify a recurring bottleneck condition.

- **A traffic queue upstream of the bottleneck,** wherein speeds are below free-flow conditions elsewhere on the facility. (Note: if speeds at all or most-all of the facility are consistently and regularly lower than free-flow speeds, then overarching congestion exists. This is congestion beyond a mere point-specific bottleneck location.)
- **A beginning point for a queue.** There should be a definable point that separates upstream and downstream conditions. The geometry of that point is often coincidently the root cause of the operational deficiency.
- **Free flow traffic conditions downstream** of the bottleneck that have returned to nominal or design conditions.



- As it pertains to an operational deficiency, **a predictable recurring cause.** (Note: this implies that all things being equal, a solution exists that is nevertheless theoretically "correctable" by design, as opposed to, say, an amorphous, random event.)
- **Traffic volumes that exceed the capacity of the confluence to process traffic.** (Note: this applies to recurring events even more-so than nonrecurring.)

What Options Exist to Combat Congestion (and by Extension, Bottlenecks)?

Fixing operationally influenced deficiencies applies to the fourth of the following four strategies available to combat congestion.

- **Bring supply and demand in alignment through congestion pricing.** Congestion pricing or peak-period pricing entails fees or tolls for road use that varies by level of vehicle demand on the facility. As with market pricing in other sectors, road pricing helps allocate limited supply – in this case, that of available road space. With user charges assesses at the point of use, greater efficiency results through improved response to market forces. Charges are typically assessed electronically to eliminate delays associated with manual toll collection facilities. Road-use charges that vary with the level of vehicle demand provide incentives to shift some trips to off-peak times, less congested routes, or alternative modes; or to cause some lower-value trips to be combined with other trips or simply to be eliminated. Congestion pricing has several important objectives. First, it seeks to balance demand with available capacity, i.e., the supply of road space. Second, it seeks to fairly allocate the costs associated with operating, maintaining, and expanding the transportation system to meet growing demand. Third, it seeks to improve operation of the highway system. A fourth objective may include revenue generation.
- **Provide alternatives as to how, when, where, and if to travel.** The goal of this strategy is to reduce the number of vehicles on a given road. This may take the form of promoting alternative commute options such as employee telecommuting options or making transit easier and more attractive to use. Also of interest in managing demand are driver incentive programs that, for example, promote ridesharing and off-peak use.
- **Invest in new highway capacity.** Add new construction on new alignments to preserve or improve system performance.

- **Improve the management and operation of the system.** Improve the day-to-day operation of the system by retiming traffic signals, applying access management techniques, removing operational deficiencies, and improving response time and management of traffic disrupting events like work zones, accidents, and special events. Provide real time information about the system so that travelers can make immediate decisions about when, where, and how to travel, and transportation agencies can make real-time adjustments to improve system operations.

Why Do Recurring Bottlenecks Occur?

Every highway facility has decision points such as on and off ramps, merge areas, weave areas, lane drops, tollbooth areas, and traffic signals; or design constraints such as curves, climbs, underpasses, and narrow or nonexistent shoulders. In many thousands of cases, these operational junctions and characteristics operate sufficiently and anonymously; however, when the design itself becomes the constricting factor in processing traffic demand, then an operationally influenced bottleneck can result.

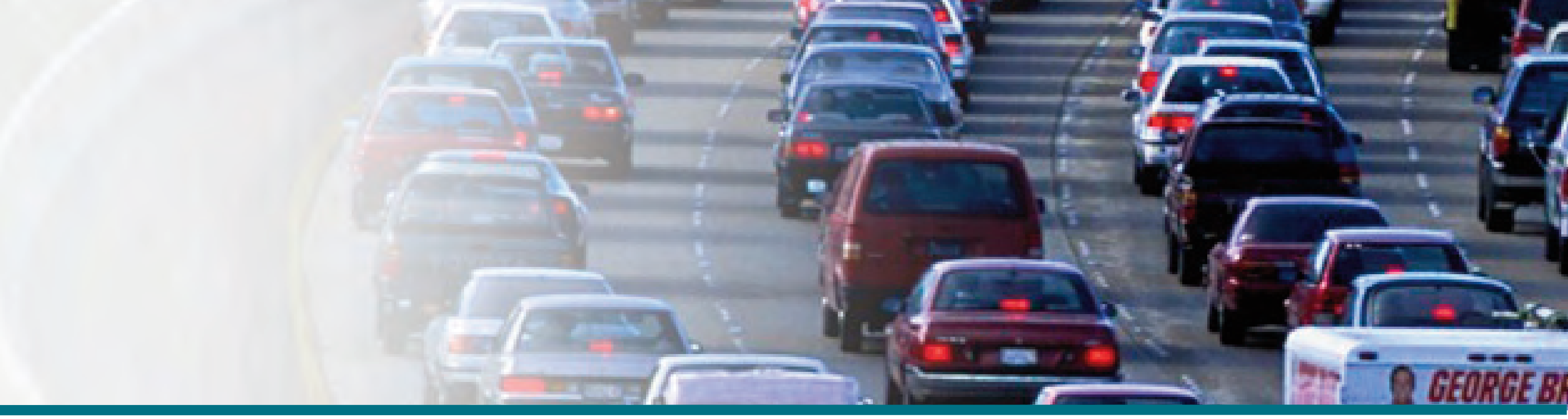
The degree of congestion at a bottleneck location is related to its physical design. Some operational junctions were constructed years ago using design standards now considered to be antiquated, while others were built to sufficiently high design standards but are simply overwhelmed by traffic demand. Whatever the root cause, operationally influenced bottlenecks can occur at:

- A **lane drop** particularly mid-segment where one or more traffic lanes are lost. These typically appear at bridge crossings and in work zones. The latter, however, is a nonrecurring event and is usually remedied when the work zone is removed. Ideally, lane drops should be located at exit ramps where there is a large volume of exiting traffic.

- A **weaving area**, where traffic must merge across one or more lanes to access entry or exit ramps. Bottleneck conditions are worsened when there are confusing or insufficient weaving lengths.

- **Freeway on-ramps,** which are merging areas where traffic from local streets can join a freeway. Bottleneck conditions are worsened on freeway on-ramps without auxiliary lanes, short acceleration ramps, or where there are multiple on-ramps in close proximity.

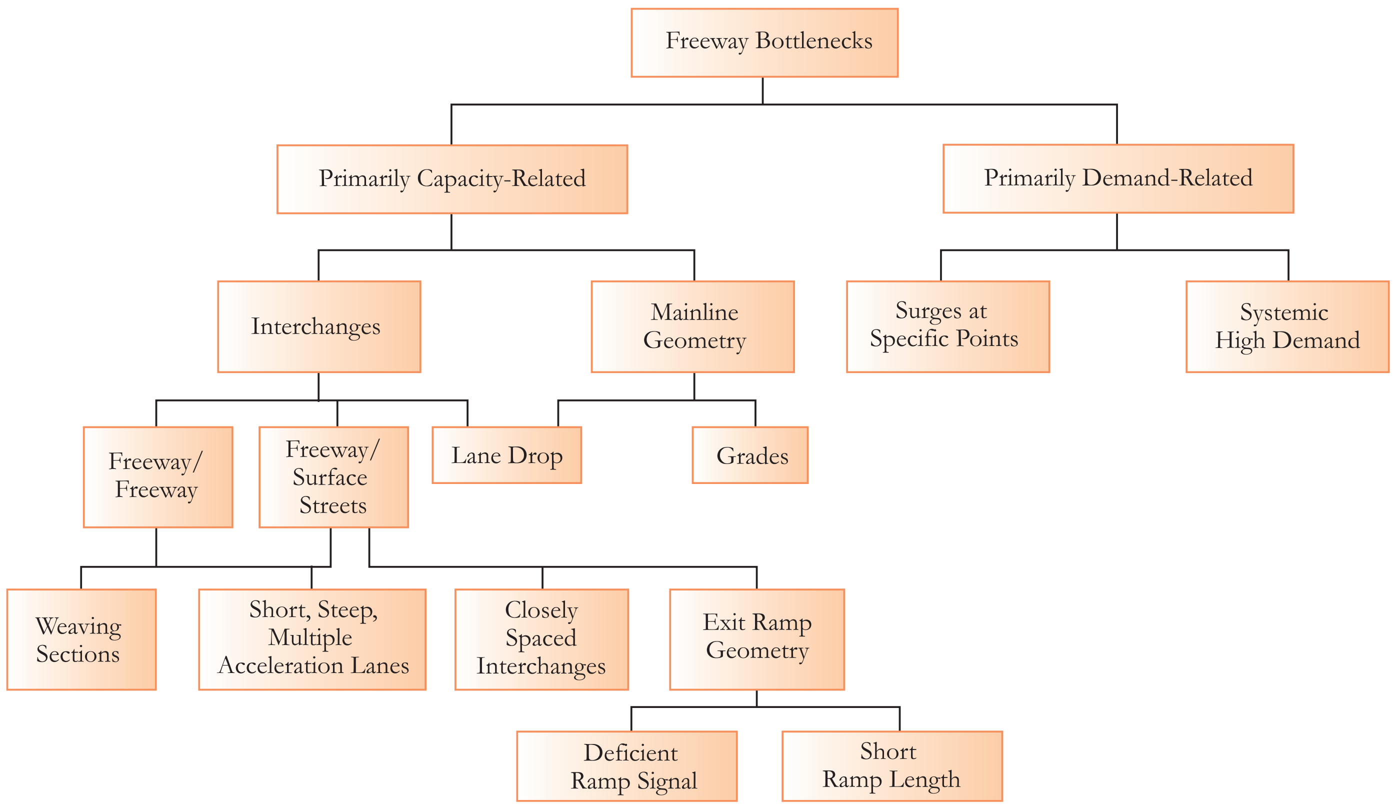
- **Freeway-to-freeway interchanges,** which are special cases of on-ramps where flow from one freeway is directed to another. These are typically the most severe form of physical bottlenecks because of the high traffic volumes involved.



- **Freeway exit ramps,** which are diverging areas where traffic leaves a freeway. Bottleneck conditions are worsened on freeway exit ramps that have a short ramp length, traffic signal deficiencies at the ramp terminal intersection, or other conditions that may cause ramp queues to back up onto freeway mainlanes.
- **Abrupt changes in highway alignment,** which occur at sharp curves and hills and cause drivers to slow down either because of safety concerns or because their vehicles cannot maintain speed on upgrades. Another example of this type of bottleneck is in work zones where lanes may be redirected or "shifted" during construction.
- **Low clearance structures** such as tunnels and underpasses. Drivers slow to use extra caution, or overload bypass routes. Even sufficiently tall clearances may impart problems. In one renowned case, the mere optical illusion of a bridge appearing lower than it really is causes drivers to slow down, resulting in recurring bottleneck conditions.
- **Lane narrowing,** caused by either narrow travel lanes or narrow or non-existent shoulders.
- **Intended interruptions to traffic flow** are literally "traffic disruptions on purpose" that are sometimes necessary in order to manage system flow. Traffic signals, freeway ramp meters, and tollbooths can all contribute to capacity loss.

Figure 2 lists various types of freeway bottleneck causes.

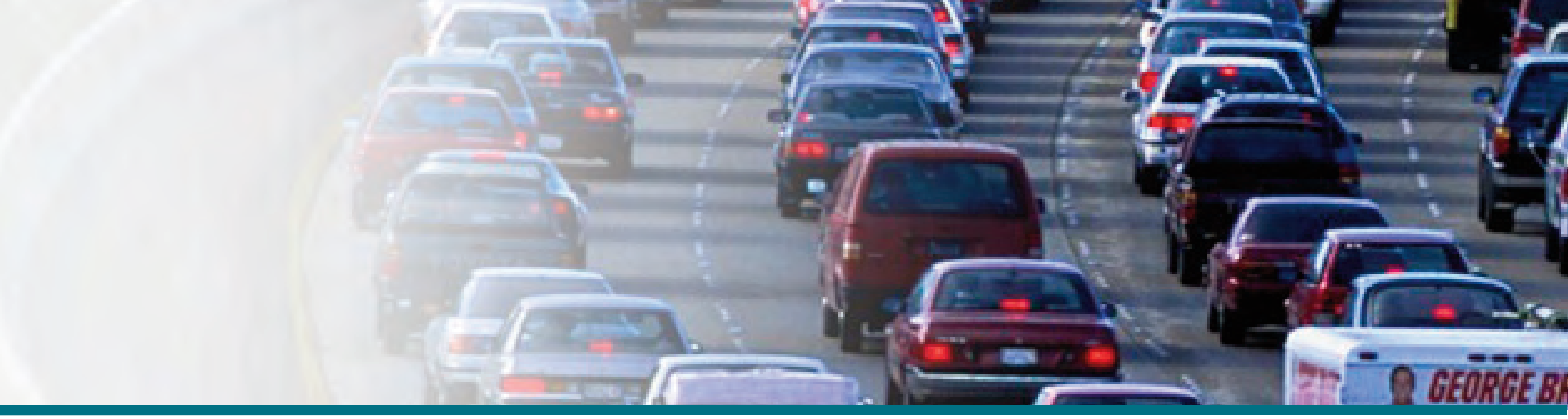
Figure 2. Types of Freeway Bottlenecks



What is Stopping Us from Fixing Bottlenecks?

The knee jerk reaction might be "lack of money." But that's everyone's first complaint about, well, most every problem! In visiting with many states to ascertain if they have a bottleneck-specific program or similar that targets chokepoint congestion, we have found a sampling of reasons.

- **A predisposition to execute major projects.** Certainly, no one is faulting an agency which is attentive to executing major transportation initiatives. There is no shortage of "fronts" on which to fight the congestion battle. A short list would include HOVs, tolling and pricing, transit alternatives, ridesharing programs, and bridge rehabilitation. But the onerous processes involved in many of these initiatives can squeeze out smaller programs.



- **Lack of a champion.** Many successful state or metropolitan planning organization (MPO) programs are the result of one or more persons taking charge to either mandate or adopt a program. High-level administrators form the direction for their agencies. Mid-level managers' production reflects their priorities and skills in executing those initiatives.
- **Lack of a "named" program.** Unless there is an identity, bottleneck remediation is usually relegated to a few projects done "by rote," e.g., as part of an annualized safety program, or as a subordinate part of larger, other-purposed projects.
- **A culture of historical practices.** Example: an agency that dutifully executed an annualized "safety" program looked only at crash rates in determining their annual top ten list of projects. After instituting a congestion mapping process, they identified several significant stand-alone chokepoints that did not correlate with their high crash mapping. Thereafter, high *congestion* hot spots competed with high *accident* hot spots on their unified top ten list of projects.

Common Myths about Bottlenecks

"Bottlenecks are caused only by not enough lanes on an extended highway section."

In the past, recurring congestion was thought to be an overarching or systemic problem (i.e., not enough lanes) that could only be resolved by widening an entire corridor. However, the fact that other subordinate sections *within the system* operate sufficiently demonstrates that the uniform highway segments may not necessarily be under-designed.

Traditional capital solutions often grew from the misconception that a multilane facility should be designed to alleviate recurring congestion during the peak hours each day. The problem is that funding for such large scale projects is limited, and right-of-way is often restricted, such that these projects take many years to complete. As a result, recurring congestion historically goes untreated, or is forced to compete against other worthy projects, until funding becomes available to "catch up" to the problem that has grown from the day the facility opens.

With a shift in focus away from the perception that recurring congestion is systemic (and thus treatable with only large scale projects), it is possible to explore a wider range of improvement strategies that are possible in the short-term. While these will never replace the need for corridor-wide fixes – especially at "mega-bottlenecks" such as major freeway-to-freeway interchanges – low-cost, quick turnaround bottleneck fixes can provide congestion relief at the point of occurrence.

Solving Recurring Bottlenecks is a Win-Win Situation for Nonrecurring Incidents Too!

Strategies to alleviate bottlenecks will lessen delay caused by nonrecurring events such as accidents, weather, work zones, etc. For example, consider an accident that blocks a single lane of traffic. If only two lanes existed prior to the accident occurring, the impact would be greater than if three lanes existed. The capacity loss resulting from the nonrecurring event will be lessened due to the improvements made to the system to benefit the recurring situation.

"Bottlenecks can't be fixed without massive reconstruction."

With the focus of traditional transportation planning and programming efforts on major capital projects, it has been assumed for many years that bottlenecks cannot be fixed without massive reconstruction of an interchange or corridor. However, there are numerous examples where agencies opted to make lower-cost improvements that resulted in significant improvements in traffic flow.

"Improving a bottleneck won't help traffic flow outside of peak periods."

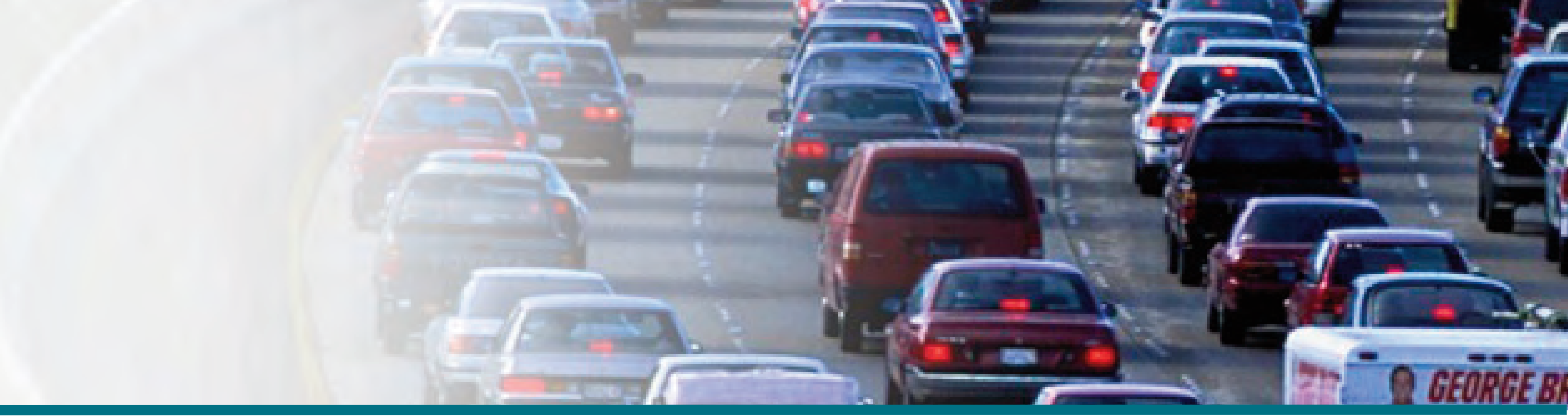
Because traffic-influencing events like incidents, bad weather, work zones, and special events can happen at any time, congestion is not restricted to peak times of the day. Benefits realized from improvements made to address peak-period bottleneck problems will invariably carry over to the times outside of the peak when congestion occurs.

A Rogue's Gallery of Infamous Bottlenecks

On much of the urban highway system, there are specific points that are notorious for causing congestion on a daily basis. These locations – which can be a single interchange (usually freeway-to-freeway), a series of closely spaced interchanges, or lane drops – are focal points for congestion in corridors. Major bottlenecks tend to dominate the congestion problem in corridors where they exist, some even acquiring nicknames from local motorists such as the "Spaghetti Bowl," "Hillside Strangler," "Spaghetti Junction," "Malfunction Junction," and "Mixmaster".

Over the past several years, transportation professionals have come to realize that highway bottlenecks demand special attention. Several national studies have highlighted bottlenecks as a major congestion problem in urban areas. These studies have raised the level of awareness about bottlenecks as a problem, warranting that they be treated as a significant part of the congestion problem.

The American Highway Users Alliance (AHUA) conducted two studies of the nation's urban bottlenecks in 1999 and 2004. The studies produced rankings of the worst bottlenecks in terms of total delay to travelers and discussed what was being done to fix the problems at locations where specific improvements had been scheduled. The studies found that nearly all of the worst bottlenecks are major freeway-to-freeway interchanges in large urban areas. The 2004 study updated the rankings and discussed three bottleneck improvement success stories – bottlenecks identified in 1999 that were subsequently improved or well under construction.



FHWA's first effort related to bottlenecks was in the freight (trucking) arena. Using the AHUA studies as a starting point, the impact of bottlenecks on truck travel was assessed. Bottlenecks outside of urban areas were also considered (e.g., steep grades). A major finding of this study was that in terms of total delay, the urban bottlenecks – typically thought of as a commuter-related problem – are also major sources of truck delay.

States and regions are beginning to recognize the significance of bottlenecks as well. The Ohio Department of Transportation completed a study of freight (trucking) bottlenecks, and the Interstate-95 Corridor Coalition is undertaking a study of all potential bottlenecks in Coalition states. The Atlanta Regional Commission has defined bottlenecks as a specific portion of their Congestion Management Process and is identifying regional and local bottlenecks in their network.

More recently, an effort by a private data provider, Inrix, also identified the nation's worst bottlenecks (Table 1). Whereas previous bottleneck identification efforts were based on analytic procedures using traffic volumes and capacity data, Inrix's approach uses data assembled by them from a variety of sources. As direct travel time measurements become more common and better refined, the science of bottleneck identification and performance will improve.

Improvements Are Possible

"Seven of 18 bottlenecks identified in 1999 – including hotspots in Houston, Albuquerque, Denver, Boston, Los Angeles, and Washington, D.C. – no longer appeared on (subsequent) rankings of the country's worst chokepoints (due to) major reconstruction projects completed or underway."

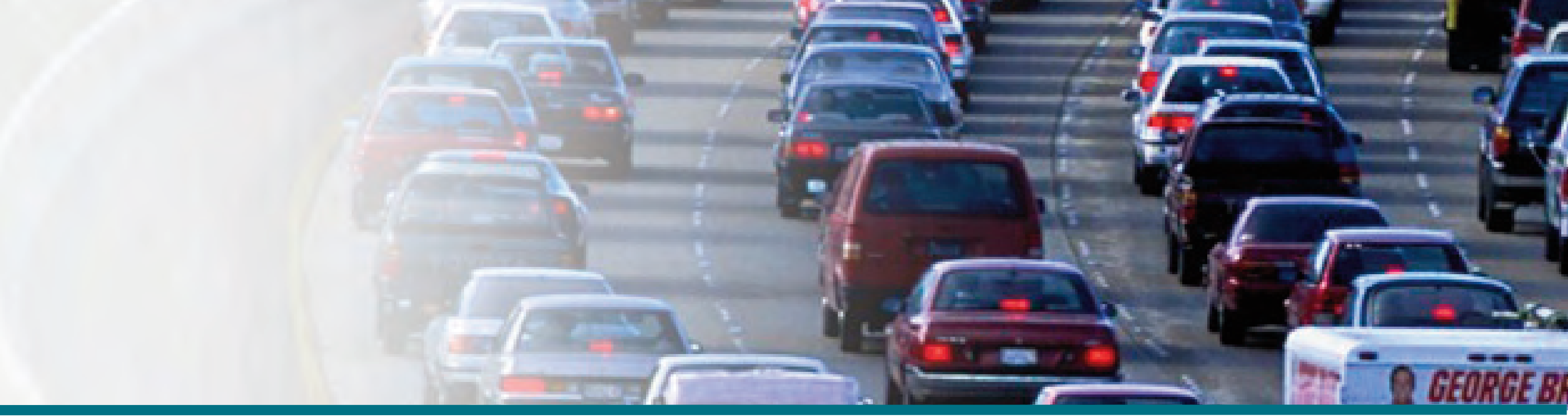
American Highway Users Alliance, 2004

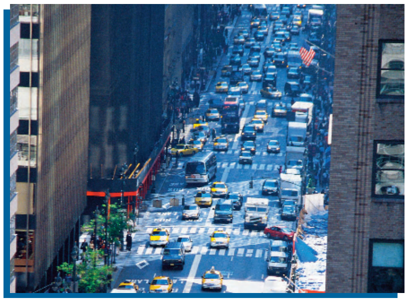
Table 1. The Worst Physical Bottlenecks in the United States

2008

2008 Rank	Area	Road/Direction	Segment/Interchange	State	Length	Hours Congested	Average Speed When Congested
1	New York	Cross Bronx Expressway/I-95 SB	Bronx River Parkway/Exit 48	NY	0.36	94	11.2
2	San Francisco	I-580 WB	Bellam Boulevard	CA	0.38	65	8.1
3	New York	Cross Bronx Expressway/I-95 SB	I-895/Sheridan Expressway/Exit 4A	NY	0.55	93	11.9
4	New York	Cross Bronx Expressway/I-95 SB	White Plains Road/Exit 5	NY	0.27	87	12.3
5	New York	Harlem River Drive SB	3rd Avenue	NY	0.15	81	12.4
6	New York	Van Wyck Expressway/I-678 NB	Liberty Avenue/Exit 4	NY	0.58	77	13.1
7	Los Angeles	Hollywood Freeway/U.S. 101 SB	Vermont Avenue	CA	0.64	77	14.0
8	Chicago	Dan Ryan Expressway/ I-90/I-94 SB	Canalport Avenue/ Cermak Road/Exit 53	IL	0.52	77	13.6
9	New York	Harlem River Drive SB	2nd Avenue/125th Street/Exit 19	NY	0.23	84	12.5
10	Chicago	Eisenhower Expressway/I-290 EB	U.S. 12/U.S. 20/U.S. 45/Exit 17	IL	0.98	57	12.3
11	New York	Cross Bronx Expressway WB/ I-95 SB	Westchester Avenue/Exit 5	NY	1.15	77	14.5
12	Los Angeles	Hollywood Freeway/U.S. 101 NB	Los Angeles Street	CA	0.09	76	11.9
13	Los Angeles	Hollywood Freeway/U.S. 101 NB	Spring Street	CA	0.14	85	14.2
14	Los Angeles	Harbor Freeway/I-110 NB	Adams Boulevard	CA	0.13	73	15.8
15	New York	George Washington Bridge EB/ I-95 NB	Center Avenue	NJ	0.14	68	9.0
16	New York	I-95 NB	U.S. 1/U.S. 9/U.S. 46/Exit 72	NJ	0.42	66	9.7
17	New York	Harlem River Drive NB	Lower Level Washington Bridge	NY	0.09	74	10.3
18	Los Angeles	Hollywood Freeway/U.S. 101 NB	Alameda Street	CA	0.26	73	14.0
19	Chicago	Dan Ryan Expressway/ I-90/I-94 NB	Ruble Street/Exit 52B	IL	0.13	76	16.1
20	Los Angeles	Hollywood Freeway/U.S. 101 SB	Melrose Avenue	CA	0.31	68	15.9
21	Los Angeles	Harbor Freeway/I-110 NB	I-10/I-100/Santa Monica Freeway	CA	1.09	70	16.4
22	New Haven	I-91 SB	I-95	CT	0.47	63	13.4
23	New York	Van Wyck Expressway/I-678 NB	Hillside Avenue/Exit 6	NY	0.27	79	14.4
24	New York	Van Wyck Expressway/I-678 NB	Atlantic Avenue/Exit 5	NY	0.47	75	12.7
25	Chicago	Dan Ryan Expressway/ I-90/I-94 NB	18th Street/Exit 52C	IL	0.34	75	15.7

Source: *"Inrix's National Traffic Scorecard: 2008 Annual Report,"* February 2009, http://scorecard.inrix.com/scorecard/request.asp.





What is FHWA Doing?

With regards to congestion, the Federal Highway Administration (FHWA) promotes a number of efforts to help reduce congestion on the nation's highways. Together with our state partners, who implement these strategies, these efforts can allow for more informed decisions, better coordination, and quicker actions to mitigate the problems.

Recurring Congestion Program Strategies

- **Tolling and Pricing.** Value pricing entails fees or tolls for road use which vary by level of vehicle demand on the facility. Fees are typically assessed electronically to eliminate delays associated with manual toll collection facilities.

- **Public-Private Partnerships.** "Public-private partnerships" (PPP) refer to contractual agreements between a public agency and private sector entity that allow for greater private sector participation in the delivery of transportation projects. FHWA is working with our partners in the public and private sector to further investigate these promising partnerships.

- **Real-Time Traveler Information.** This is "decision-quality" information that travelers can access, understand, and act on to choose the most efficient mode and route to their final destination. Timely and detailed information about traffic incidents, the weather, construction activities, transit and special events, all aid in improving travel time predictability, better choices, and reduced congestion.

- **Traffic Signal Timing.** Signal timing should correspond to the current traffic patterns. Often signals are initially timed, but not re-adjusted when traffic patterns change. This results in inefficiency and unnecessary delays. Goal: work with state and local agencies in congested metropolitan areas and encourage best practices for improved traffic signal timing.

Nonrecurring Congestion Program Strategies

- **Traffic Incident Management.** This utilizes a combination of public safety functions and traffic management functions – it requires cooperation between various public agencies to reduce congestion by clearing traffic crashes and removing stalled vehicles. FHWA is championing laws, policies, and practices that speed up the clearance of major and minor incidents that create congestion.

- **Work Zone Mobility and Highways for LIFE.** The Work Zone Safety and Mobility Rule advocates stronger consideration and management of work zone safety and mobility impacts to reduce congestion during construction projects. The Highways for **Long**-lasting, **Innovative**, **Fast** construction of **Efficient** (also known as HfL) and safe pavements and bridges pilot program as outlined in SAFETEA-LU is meant to accelerate the rate of adoption of innovations and technologies, thereby improving safety and highway quality while reducing congestion caused by construction.

The Localized Bottleneck Reduction Program

The Localized Bottleneck Reduction (LBR) Program promotes operational and low-cost bottleneck mitigation strategies to improve mobility in the short term. Managed by the Office of Operations, the program serves to bring attention to the root causes, impacts, and potential solutions to traffic chokepoints that are recurring events; ones that are wholly the result of operational influences. The goal of the program is to raise awareness of bottlenecks at the state level and promote low-cost, quick-to-implement geometric and operational improvements to address recurring chokepoints. The LBR Program has several activities underway, including:

- This Primer, which overviews the wide range of operational and low-cost strategies available to reduce congestion at bottlenecks.
- A compendium of state best practices in bottleneck identification, assessment, countermeasures, and evaluation, including how bottlenecks are treated in the annual planning and programming processes.
- Regional workshops for state and local agencies to learn and share information on localized bottleneck reduction strategies and how they can be incorporated into state and local planning processes.
- Guidance documents are forthcoming that are aimed at agencies and personnel who have first responsibility to address bottleneck congestion locations.

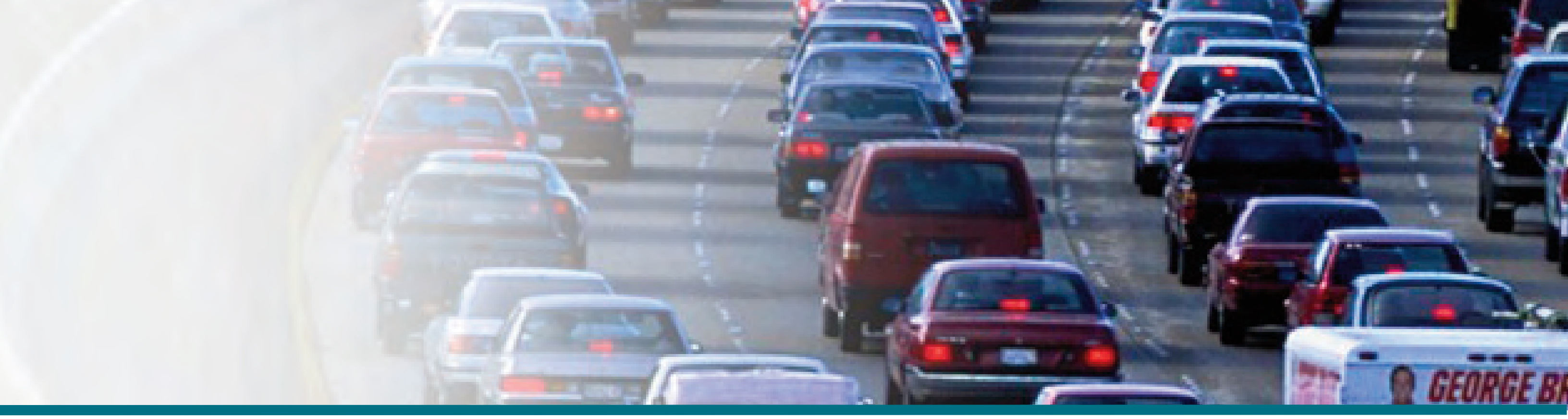
In the longer term, a "toolbox" will be developed. The toolbox will be a one-stop source for guidance on low-cost operational and construction strategies for improving bottlenecks. FHWA will also issue guiding principles and concepts common to low-cost operational improvement programs in order to further state efforts in adopting a comprehensive approach to addressing traffic congestion at bottlenecks.

More Information on the Localized Bottleneck Reduction Program

The FHWA Office of Operations Localized Bottleneck Reduction (LBR) Program brings into focus the root causes, impacts, and potential solutions to traffic chokepoints that can be mitigated through relatively low-cost improvements. The LBR Program has several activities underway, including:

- *A Primer on localized bottleneck reduction;*
- *A compendium of state-initiated solutions to localized bottleneck problems; and*
- *Regional workshops for state and local agencies to learn and share information.*

More information is available at the FHWA Bottlenecks web site: http://ops.fhwa.dot.gov/bn/index.htm.



Benefits of Localized Bottleneck Improvements

The LBR Program focuses on operationally influenced bottlenecks – small, localized "hot spots" where the design of the roadway itself becomes the constricting factor in processing traffic demand, resulting in recurring delays of generally predictable times and durations. Mega-projects required to resolve major bottleneck problems and systemic congestion (e.g., entire corridor rebuilds, multimile lane additions, and systemwide improvements) are far and above the focus of this program area. Unfortunately, when weighed against these larger, more visible projects, localized bottleneck problems often receive lower priority for funding or are put off entirely until they can be implemented as part of the larger, all-encompassing project. However, in this day and age of fiscal constraints, with agencies facing over-escalating costs and increasingly limited right-of-way, it is evident that "business as usual" in resolving congestion problems no longer applies. Low-cost bottleneck mitigations have several advantages that can help agencies deal with these developments:

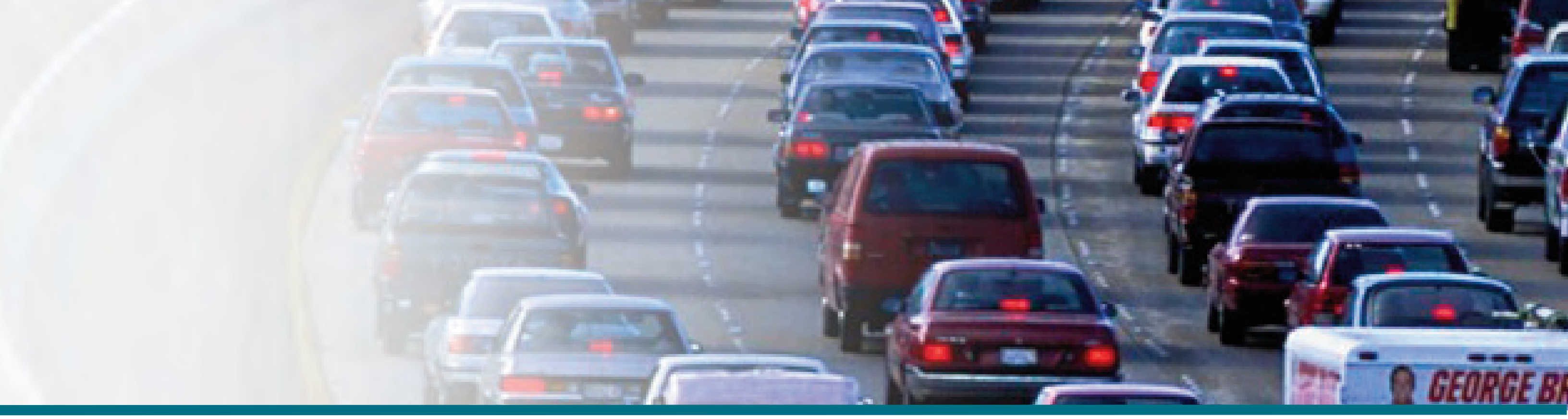
- **They address current problems and therefore have high visibility.** Agencies are under increasing pressure to do something immediately about congestion problems. Because low-cost bottleneck treatments are small in scale, they can be implemented quickly, so benefits start accruing immediately.
- **They are highly cost-effective and usually have positive safety impacts**. Low-cost bottleneck treatments could mitigate or reduce crashes within weaving and merging areas, thereby increasing the cost effectiveness relative to safety merits.
- **They will be required as transportation funding for mega-projects becomes more constrained.** Major reconstruction projects are often justified as the only valid solutions to relieve congestion at the worst bottleneck locations. However, the cost of executing such projects is usually enormous. Low-cost bottleneck improvements provide an effective way to stretch scarce resources.
- **Lower cost means more locations can be addressed.** More spot solutions can be implemented throughout a region, addressing more corridors than just a few large projects.

- **They are less invasive on the physical and human environments.** The environmental footprint of low-cost bottleneck projects is very low, both in terms of disruptions during construction and final design.
- **They are not necessarily just short-term fixes.** For some low-cost treatments, congestion benefits will play out over many years, not just a few. In fact, when combined with other forms of treatment (e.g., demand management and operations), they may be part of a long-term solution for a problem location or corridor.
- **They may be considered part of major reconstruction projects to address current problems.** Some state DOTs have successfully incorporated low-cost bottleneck treatments within the context of larger, multiyear reconstruction projects.

FHWA Survey Suggests Opportunities for Lower Cost Solutions

Respondents to a FHWA Division Office survey reported that 71 percent of bottlenecks occurred on freeways, with 42 percent interchange related bottlenecks. Respondents cited that 36 percent of bottlenecks had no specific plans underway for improvement. Full interchange or freeway reconstruction were the most commonly perceived solutions for resolving bottlenecks. However, at least a portion of these bottlenecks might be candidates for low-cost solutions that can be implemented in the short term to improve traffic flow.

Recent discussions with several state partners have reinforced their position. Specifically, high-ranking state DOT officials and mid-level staffers engaged in day-to-day operations have all opined that there are "tremendous" and "significant" benefits to pursuing low-cost operational improvements. These benefits range from the direct (reductions in delay, increases in throughput) to the indirect (public confidence and agency image-boosting effects).



Identifying and Assessing Bottlenecks

Bottleneck locations can be identified through direct observation by agency personnel, aerial photographs, or video surveillance data. Some metropolitan areas have even conducted public outreach efforts to solicit input from motorists on bottleneck locations. Traffic analysis tools can systematically identify bottleneck locations by analyzing road segments for congestion or poor levels of service.

Once bottleneck locations have been identified, the root cause and severity must be determined. Special travel time runs, aerial photography, or video of suspected bottleneck areas can be used to pinpoint sources of operational deficiencies. On freeways equipped with detection technology, dynamic surveillance can be used to identify where and how often bottlenecks occur, and how severe they are. Archived traffic data can be used to measure whether the problem is growing or receding.

Sometimes, the operational cause of a bottleneck is evident, intuitive, or anecdotal. However, when multimile corridor congestion is prevalent, microsimulation modeling can assist in identifying, separating, and analyzing bottleneck dynamics within the corridor.

Data collection to support bottleneck analysis should be sufficient to capture the duration and extent of congestion. Typically, 15-minute traffic volume counts for all ramps and mainlanes for a four-hour peak period is adequate. Other data can be collected through travel time runs, video, or origin-destination studies.

Bottleneck analysis is necessary to study not only the subject location, but also the impacts of potential bottleneck remediation on upstream and downstream conditions. The analysis will justify action to correct bottlenecks, confirm the benefits of bottleneck remediation, or check for hidden bottlenecks along a corridor. When conducting bottleneck analysis, care should be taken to ensure that:

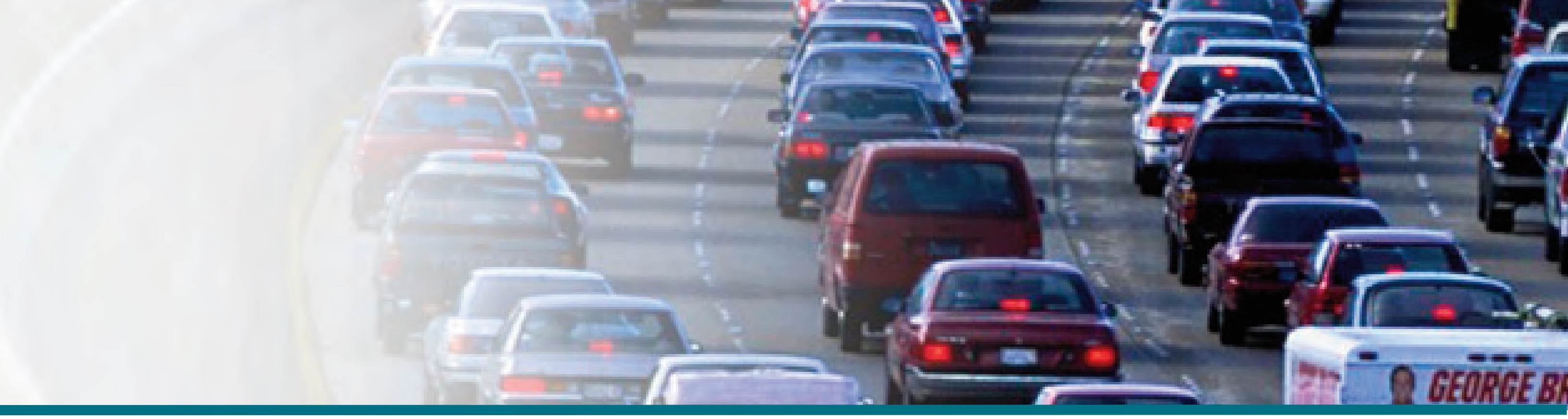
- Improving traffic flow at the bottleneck location doesn't just transfer the problem downstream. The existing bottleneck may be "metering" flow so that a downstream section currently functions acceptably, but the increased flow will cause it to become a new bottleneck.

- Future traffic projections and planned system improvements are inclusive in the analysis. Safety merits also should be strongly considered.

- "Hidden bottlenecks" are considered. Sometimes, the queue formed by a dominant bottleneck masks other problems upstream of it. Improving the dominant bottleneck may reveal these hidden locations. It is important to take into account the possibility of "hidden bottlenecks" during the analysis stage.
- Conditions not traditionally considered by models are accounted for. There are several bottleneck conditions, such as certain types of geometrics and abrupt changes in grade or curvature, that can't be analyzed by current analysis tools. Engineering judgment will need to be exercised to identify those problems and possible solutions.

Operation Bottleneck Solicits Public Feedback on Bottleneck Locations in Arkansas

Metroplan, the MPO for the Little Rock, Arkansas region, conducted Operation Bottleneck in October 2008, a public outreach effort designed to identify traffic bottlenecks as well as auto, bike, and pedestrian safety issues throughout the region. The program received 3,000 responses in 1½ months, with on-line submissions constituting the highest return. Metroplan is currently reviewing, classifying, and analyzing the feedback, and the next step will be to coordinate with local jurisdictions.



How Do Bottlenecks Disperse?

Recurring congestion usually disappears when the crush volume drops back to a level that is manageable by the facility design. *Nonrecurring* congestion usually disappears when the dynamic event is removed. The rate at which merged lanes unite is strongly governed by the rate of dispersal at the constricted point (i.e., the "nozzle") and thus is less easily influenced on its own merits. In layman's terms, it hardly matters how well the two lanes "knit" together when there's no place to go! There is some evidence that in *nonrecurring* events, strong direction in the form of flagmen or specific messages (e.g., "Take Turns," "Merge Here," "Stay in Lane") may improve the rate of dispersal marginally. Such deployments are known as "Dynamic Lane Merges" (explained further below) as they rely on some aspect of proactive instruction. Realistically, the great majority of nonrecurring chokepoints are comparatively "static"; that is, there are signs forewarning the merge, but there is no proactive management of the merge. The information supplied at recurring chokepoints is even worse; barely much beyond a free for all. Finally, and to a lesser degree, the rate of arrival of vehicles upstream joining the back of the queue has some impact on the bottleneck.

"So You Think You Can Merge?"

Are you a "profiteering" lane merger, who seeks only your own personal gain to advance, or are you an "altruistic" driver who yields to others for the benefit of all?

Whether you are an "early" merger (upstream of the point of confluence) or "late" merger (at the last possible moment) does not brand you good or bad, Republican or Democrat, or even right-minded or left-minded. Anecdotally, both sides claim pre-eminence in their preference.

In the end, the prevailing conditions (i.e., whether the merge is executed at free-flow speeds or at a crawl; what guidance is provided) and the generalization that we are each left to our own devices as to consider when it is best to merge, conspire to make this more a study in human behavior, and less a study in efficiency.

Understanding Merging at Recurring Bottlenecks

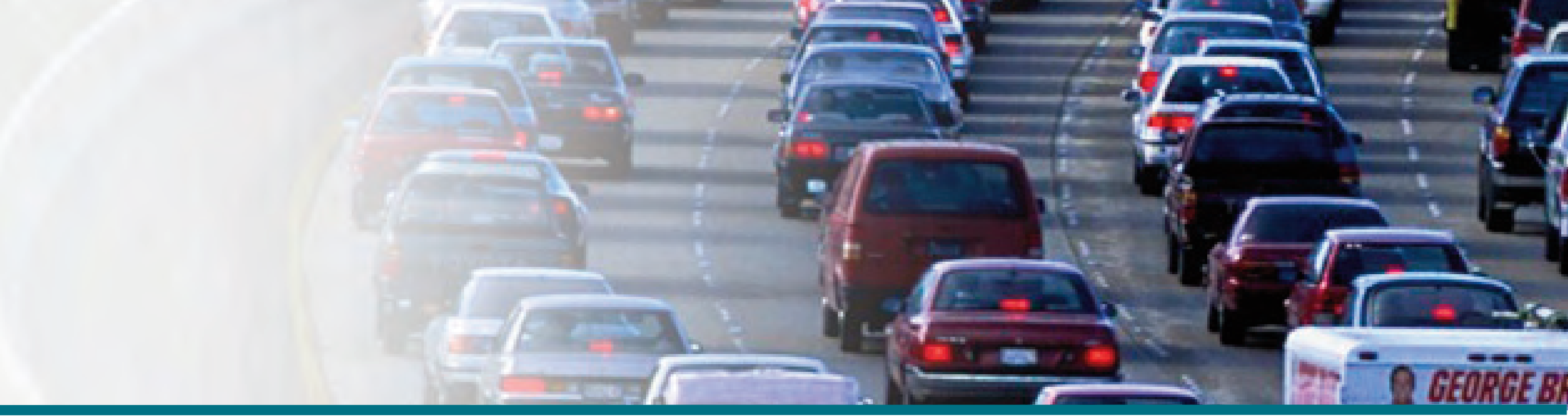
The Difference in Merging for Recurring and Nonrecurring Conditions

Merging in a recurring backup is different than in a nonrecurring one. Outside of ramp metering, the former is essentially "cat herding" with implicit rules (often local in culture or habit) at best. Typically, not much guidance is given. Drivers just "suddenly" encounter taillights ahead, whereas, in a nonrecurring event, there is more apt to be instruction in the form of orange cones, signs, flagmen, or police. There is often direction to motorists *how* ("Take Turns"), *where* ("Merge Here"), and even *what* ("All Lanes Thru") to do/expect, and there can even be enforcement (of lane jumpers) or simply order (traffic cops) from chaos.

One might argue "What's the difference? I'm in bumper-to-bumper traffic regardless!" The great difference is the greater *potential* (in nonrecurring) for herding those cats. Nonrecurring congestion is better suited for empirical study, i.e., characterized by observation and trial instead of theory. The "Dynamic Lane Merge" (as we will see later) is a prime example of this. Conversely, recurring congestion is best suited for (and pursued by) the academic and scholarly; focusing on the root causes, the queuing theories, and the algorithmic "solutions" that might bend or break the cycle of recurrence. However, in real life recurring situations, drivers are mostly left to their own devices how and when to merge. Their *tendencies* may be studied, but "Exit ½-Mile Ahead" doesn't really help once you are in the backup. And therein lies the fundamental point; a recurring merge relies on a driver's predilection, as well as the aforementioned culture. For comparison, at nonrecurring events, an identical dynamic traffic control plan can be developed and implemented most anywhere and drivers will more or less treat it the same.

Which is Best? "Early" or "Late" Merging?

Can a better recurring merge be developed? Merging takes place at-speed or "at-crawl." The former is most often associated with free flow on-ramp maneuvers, while the latter is most often associated with bumper-to-bumper congestion. In either condition the motorist has the additional choice to merge "early" (upstream) or "late" (at point of confluence). This creates a matrix of four possible merge conditions; 1) at-speed "early;" 2) at-speed "late;" 3) at-crawl "early;" and 4) at-crawl "late." To further complicate things, guidance concerning where, when, and how best to merge can vary from modest to no forewarnings in *recurring* conditions to fully deployed Traffic Control Plans (TCPs) in *nonrecurring* conditions. Given that this primer is focused on the *recurring* bottleneck genre, the purpose of this section was to research if early or late merging was best for these noncontrolled situations; i.e., when no active TCP exists.



What Instruction is Given to Motorists?

Let's look at the difference between recurring and nonrecurring conditions. Work zones and special events are nonrecurring (random or irregular) conditions, wherein, dynamic TCPs are instituted for the expressed purpose of mitigating the planned event, and are almost universally mandated by the oversight agency. Orange cones, signs, flashers, and dynamic messages are typically deployed. The motorist is given forewarning and structured instruction how to negotiate the event zone. However, in recurring situations (chronic and habitual backups) there is minimal guidance. In the best case (e.g., a freeway lane drop) there are a couple of yellow standard warning signs and some pavement markings, which dutifully forewarn for free flow conditions, but are essentially superfluous in jammed conditions. At many other recurring congested locations (e.g., peak hour crowding onto ramps or through intersections, et al.) there is no guidance short of a free for all. This is because engineers build for, and sign for, the unconstrained condition. The result is that in many recurring situations drivers are rarely, if ever, given instruction *where* and *how* to merge, as they would be in a nonrecurring situation. This is important in framing the subject question; namely, in recurring, jammed conditions, is there academic evidence endorsing "early" or "late" merging?

Research could really only uncover data-driven studies in nonrecurring situations (wherein, motorists are more or less directed to follow signs in a work zone or event-driven scenario) and anecdotal discussions (in newspapers and the on-line community) in the recurring genre. Admittedly, in all of these cases, it mattered little in the discussion whether the cause was a recurring or nonrecurring event, but rather the fact that for whatever reason, jammed conditions existed and debate ensued. In the computer simulation community, driver behavior is a random variable used simply to enable the merge algorithms to execute and deduce rate of discharge, queue lengths and duration of the congestion. Again, the point being that in simulation, the queues merge – somehow – but no import is given *where* that merge occurs.

On the whole, drivers are typically left to their own strategies as to *how* to merge together at recurring chokepoints. Personal preference, impatience, frustration, speed-differentials, and other human and vehicular traits conspire to influence safety and reduce efficiency. Altruistic drivers are unselfish and yield – in varying degrees – to proactive drivers, who seek only their own benefit to cut in line. The only real conclusion that can be drawn is suggested by the similarity in methodologies used in the work zone studies. Specifically, in setting up "Dynamic Work

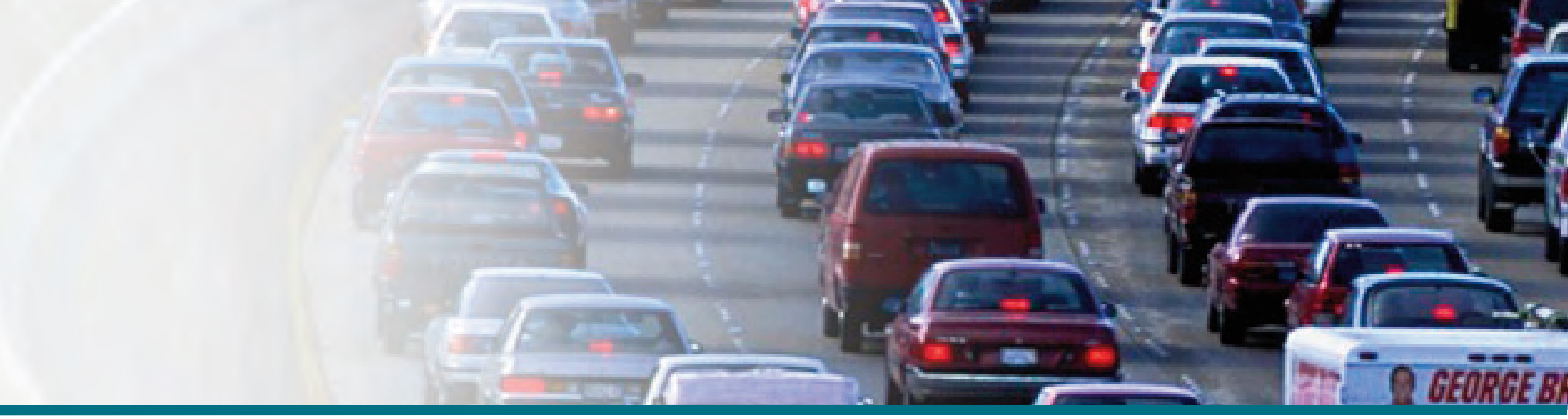
Zones" (see "Studies: Post 2000" below) the mere fact that all of these trials presumed to set up – and study – a "late merge" scenario speaks to the engineering community's penchant towards this method over the "early merge" option for stop-and-go conditions. One theme, however, remained constant. Regardless of the amount of forewarning and direction given to motorists (e.g., "light" guidance in recurring situations and "heavy" guidance in nonrecurring situations) personal preference seemed to win the day. Absent absolute enforcement, motorists were observed to – or opined to – merge when and how they preferred, with less regard for any instruction. Anecdotal conversation and local opinion columns (e.g., "Traffic Doctor," "Mr. Gridlock," etc.) have occasionally debated the merits of merging early (upstream) versus late (at point of constriction). The fact that the debate even lingers gives strength to the notion that no consensus emerged from academic and Internet research. In Tom Vanderbilt's 2008 book "Traffic: Why We Drive the Way We Do (And What it Says About Us)" the author goes so far as to introduce the entire book with the "early" versus "late" debate, and tells why and how he became a convert to the late merge.

Early Attempts to Direct Motorists How to Merge

When the Interstates were built in the 1960s and 70s there was often "instruction" by local engineers and the media of how to engage Interstate ramps, acceleration and deceleration lanes, etc. Of course, at that time, traffic was less congested on the whole, and the merging and diverging were essentially lessons in how to enter and exit Interstates. Academia has touched on queue theory, gap analysis, and related safety-oriented aspects, but none of these studies have focused much on educating motorists how to merge efficiently, unless one considers a "queue" or a "traffic stream" as an entity that can deduce instruction. Nevertheless, the academic community has essentially confirmed, via queuing theory and micro simulation that the discharge rate after the merge governs congestion on the segment. In layman's terms, there is a finite capacity of the single lane downstream of the constriction. Very little of what happens upstream can refute the laws of physics; that only one vehicle can occupy the discharge space at a time; and in a jammed situation, the lead vehicle does so from essentially a crawl speed.

Studies: Post 2000

There have been studies and trials of Dynamic Lane Merge Systems (DLMS) or similar, wherein detectors and variable message boards are designed "to utilize the best aspects of the early and late merge strategies" (Minnesota DOT, 2004).

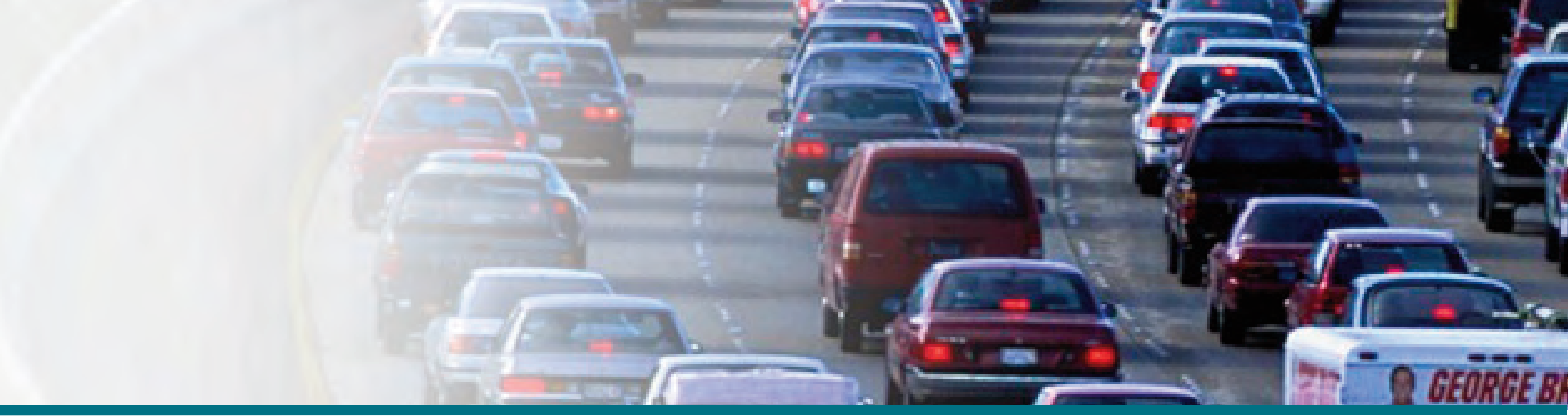


Briefly, the system switches "on" for late merges, and "off" for early merges. The Minnesota study, plus earlier, similar studies by Michigan, Kansas, and Maryland, was conducted in nonrecurring situations; specifically, in work zones. Generally, once queues were detected as growing, a series of dynamic messages would kick in that would direct motorists where and how to merge during the stop-and-go durations. Typically the messages advise motorists to "Stay in Lane," "Merge Here" (at the end of the dropped lane), "Fill All Lanes," and even "Take Turns" or similar. As the queues clear, the messages would turn off, returning the decision of how to merge upstream of the work zone back to the drivers. Arguably, the real test factors driving these trials are safety and increasing throughput in work zones (i.e., in nonrecurring conditions); that is to say that none of the trials was necessarily testing motorists' preferences during the relatively more benign recurring condition. The mere fact that in jammed conditions, all of these trials presumed to set up a "late merge" scenario speaks to the engineering community's penchant towards this method over the "early merge" option for stop-and-go conditions. The converse seems also to provide endorsement; namely, that absent jammed conditions the late merge is "relaxed" and early merging is "allowed." Note that we don't use the word "encouraged" for the latter condition; we merely note that the decision of how and when to merge is abdicated to the driver. The 2004 Minnesota study/comparison of four different deployments (locations, conditions) yielded promising success only in terms of queue reduction and driver behavior. There is anecdotal evidence in the study to support that drivers are less aggressive, and perceive shorter travel times when the overall distance traveled under congested conditions is reduced. However, all things being equal, the rates of dissipation downstream of the constriction don't seem heavily influenced by "early" or "late" merging in stop-and-go conditions, as evidenced by the study's finding that 1,600 vehicles per hour per lane downstream of the constriction seems a maximum flow rate for any deployment – a number that not coincidently is the suggested maximum flow rate listed in the Highway Capacity Manual for "short-term freeway work zones." (The reader is cautioned that the conclusions herein are merely the observations of the Minnesota study authors and should not be considered tantamount to a broad brush conclusion for all bottleneck conditions and factors.) Further, these deployments, with their garish orange barrels, cones, and VMS boards, scream "temporary condition," and are expensive to deploy on an ad-hoc basis – not the kind of thing that one could expect to be adopted as standard practice

everywhere at recurring chokepoints. Excepting for some basic, generic instruction in states' drivers manuals ("wait for a safe gap in traffic" – typ.) little has been done at the national level to educate drivers how to merge safely and efficiently, as compared to other national education efforts promoting seat belt compliance, school zone safety, traveler information, or pedestrian rights and practices. The perceived reason for this may simply be the expectation that there will always be drivers who feel they know best how and when to merge in a queue, irrespective of any instruction to the contrary. The altruistic view is to leave gaps, yield to your neighbor, take your turn but don't force your turn, and generally don't deny him or her entry into your lane. The more proactive view is to take first opportunity to cut in line, perhaps "line jump" to chase whichever line seems to be moving, and scuttle the principles of any orderly manner. Anecdotal evidence from many local traffic blogs and an Internet search finds strong sentiment from both camps as to why they think their method is best.

Merging "At-Speed" versus Merging "At-Crawl"

Under free flow conditions, the Highway Capacity Manual states that 65 mph freeways are capable of enabling up to 2,300 pcphpl – a number that can be compared to the 1,600 pcphpl "work zone" capacity-restricted volume above. (The 2,300 number takes into account safe vehicle separations and optimum conditions.) This roughly 30 percent reduction in capacity is attributable to the fact that in stop-and-go conditions, the inefficiencies of merging are magnified. Whether or not those two adjacent lines merge early, late, or otherwise, they are still metered by the fact that only one vehicle can proceed past the constriction at a time. As upstream volumes dissipate, free flow speeds increase, and vehicles enjoy a "head start" of momentum through the opening, allowing for slightly higher discharge rates that would fall above 1,600 pcphpl but cap at 2,300. In layman's terms, merging at-speed upstream of a constriction slightly improves the discharge capability, but there is still a finite time necessary for all vehicles to clear in one lane downstream.



Merge Principles

Can we increase the efficiency of merging prior to the discharge point? In two words – *be orderly*. Not surprisingly, safety improves too. It is repeatedly shown that traffic is inherently safer when all vehicles are traveling at or near the same speed. Think of an orderly progression on a crowded escalator. Everyone is safely cocooned because they are going the same speed. Now imagine the bumping and chaos that would occur if impatient folks pushed past others.

Principle #1: "Go Slow to Go Fast"

"Go slow to go fast" is an increasingly trendy expression in traffic circles. It speaks to the seemingly paradoxical idea that if we slow down the rate of our "mixing" we can get past a constriction faster. A well known example (actually the winning entry in a 2006 contest to demonstrate the meaning of "throughput maximization") is the "rice" experiment. In the first case, dry rice is poured all at once into a funnel. In the second case, the same amount is poured slowly. Repeated trials generally conclude about a one-third time savings to empty the funnel via the second method. And, it should be noted, there is a tipping point reached as one graduates from a v-e-r-y slow pour, to a medium pace, and so on. What lesson does the rice experiment teach us about traffic?

For What it's Worth

Webster's on-line dictionary defines "merging" (using 'traffic' as the colloquial) as "to blend or come together ***without abrupt change.****"*
(Emphasis added)

The densely packed rice (or traffic) in the first trial creates friction in the literal sense and the practical sense, respectively. The denser the traffic, the smaller the safety cushion around each driver, and the more cautious (i.e., slower) he becomes. A classic "bell curve" diagram also serves to explain how traffic throughput reaches an apex up to the point where traffic friction and conflict conspire to begin a decline in the rate of throughput and speed. There exist some examples of validation of this principle at intersections (e.g., traffic signalization, roundabouts, vehicle detection) that demonstrates that slowing or stopping some traffic benefits the aggregate flow, and is far better than the free-for-all converse. In the bottleneck and corridor genres, we have ramp metering and speed harmonization (see below), respectively, providing examples on freeways.

Principle #2: Keep Sufficient Gaps

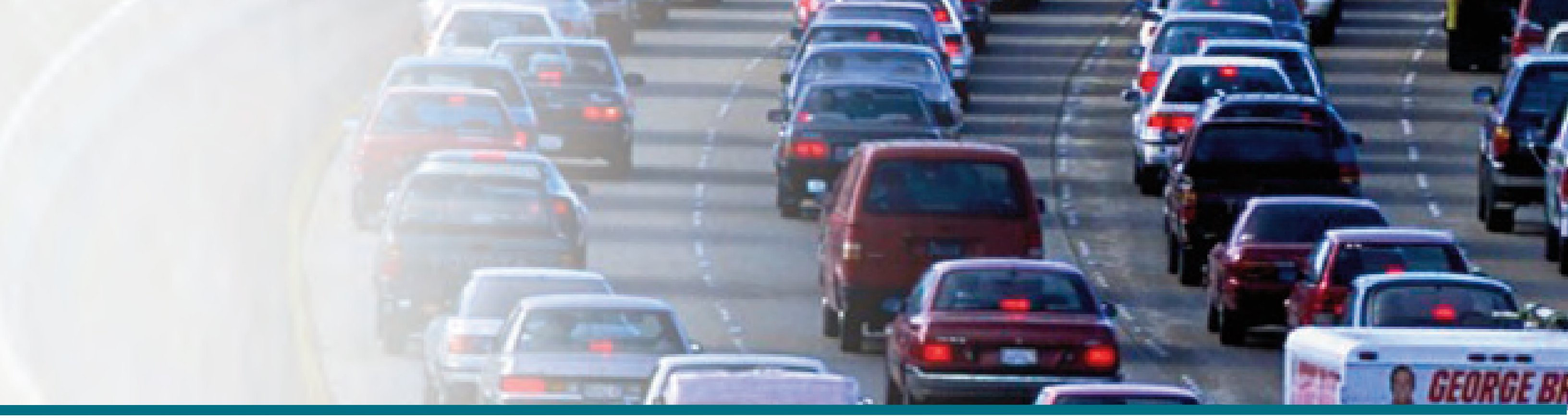
Keeping sufficient (or ideally, the largest possible) gaps leads to uniform and free(er) traffic flow. Gaps allow for small adjustments in braking, accelerating, and drifting. The larger the gap, the lesser the "ripple" affecting adjacent and following vehicles, which otherwise would react by slowing. Gap maintenance is achieved by rote in high-occupancy vehicle (HOV) lanes or high-occupancy toll (HOT) lanes; by selective admittance in the former, and by dynamically shifting the price every

few minutes in the latter. The target benefit is to allow qualifying vehicles the guarantee of a free flow trip, versus the hit-or-miss prospect in the adjacent general purpose (GP) lanes. Both cases have the added (and intended) benefit of removing vehicles and or person-trips from the GP lanes too; so all traffic streams win when these practices are employed. Absent out-and-out violators who can muck up the system, agencies can tweak the lane mandates to keep the systems running at optimum levels.

How does this apply to localized bottlenecks? Theoretically, the same "gapping" principles would hold true in backups; to wit, leaving progressively larger gaps would allow for progressively better progression. (Taken to the extreme, no "bottleneck" would even exist!) The point is that in congested situations the constant brake-tapping in bumper-to-bumper traffic works to self-perpetuate the problem. No one can get much momentum before he or she has to react to the vehicle directly ahead or adjacent. The ripple effects are short, abrupt, and inefficient. The obvious problem with this is that human nature simply won't allow for the patience and orderliness to make this work. The second I create a sufficient gap between me and the car ahead, some "profiteering" lane jumper will fill it. Which is a nice segue into the next principle; zippering.

Principle #3: Zippering

Unlike principle #2, which is noted to be fairly impractical to expect, this one could easily be melded into our regular practice; namely, to take turns, or "zipper" merge at the front of the line. The fairness – and simple visualization – of this principle speaks for itself. And there is already precedence that we have been schooled in this; witness the "Yield" condition and many recurring locations where this is the unwritten rule; newcomers quickly adapt! Advocates of zipper merging are proponents of "late" merges; i.e., staying in your lane until the last possible moment and taking turns to get through the chokepoint nozzle. One enterprising fellow in California has gone so far as to adorn his car with a zipper graphic and messages promoting this method. But is this latter day Don Quixote merely tilting at this traffic "windmill"? To be fair, "late" merging seems the slightly favored method in web blog research, although as noted prior, there are unyielding stout proponents of the early merge too. Let's take a closer look.



Imagine two lanes of cars. The left lane (L) is the continuous lane and the right lane (R) is the dropped lane. You are 6^{th} in line in the R lane. If everyone stays put, and the zipper order is L, R, L, R. etc., your neighbor to the left will be 11^{th} and you will be 12^{th} to merge. If, however, you "early merge" and cut in front of him into the L line (and your fellow motorists in the R line all move up in succession) then you will become 11^{th}, the person originally behind you (formerly 14^{th}) now becomes 12^{th}, and your neighbor will have dropped to 13^{th}. You win. Your neighbor loses. But the real winner is the guy originally behind you; he's gained two spots.

Now consider the same scenario except the zipper order is R, L, R, L... etc. In the orderly scenario you would be 11^{th} and your neighbor would be 12^{th} to merge. If you cut in front of him as in the first scenario, the guy behind you (who moves up to fill your space) becomes 11, you are 12, and your neighbor is slotted at 14. You lose. Your neighbor really loses (drops two slots) and the guy behind you (formerly 13) really wins; he gains two spots – again.

Congratulations! In both scenarios, by "early merging" you have definitely improved the slot for the guy behind you! You may or may not have improved your slot. And either way you made your neighbor mad! Of course, in real life, the permutations are mind boggling when one considers multiple jockeying for position in a long queue by you and others, or the dynamic created by multiple lanes. You may think you are making progress one line jump at a time, but your move(s) may have been canceled by someone way up ahead of you; or the woman opposite him. And so on. It effectively becomes akin to playing the lottery – you are at the mercy of others' actions equal to, or more so, than your own. And all for the perceived benefit of getting a few spaces ahead through the nozzle. Of course, there's another way to look at it too; it's all a wash when said and done. It's hard to know. All you may have accomplished by line jumping is to elevate yours or others' stress. Better to sit, wait, take your turn, and put your radio on some calming music.

Principles Put Into Practice: Variable Speed Limits and Speed Harmonization

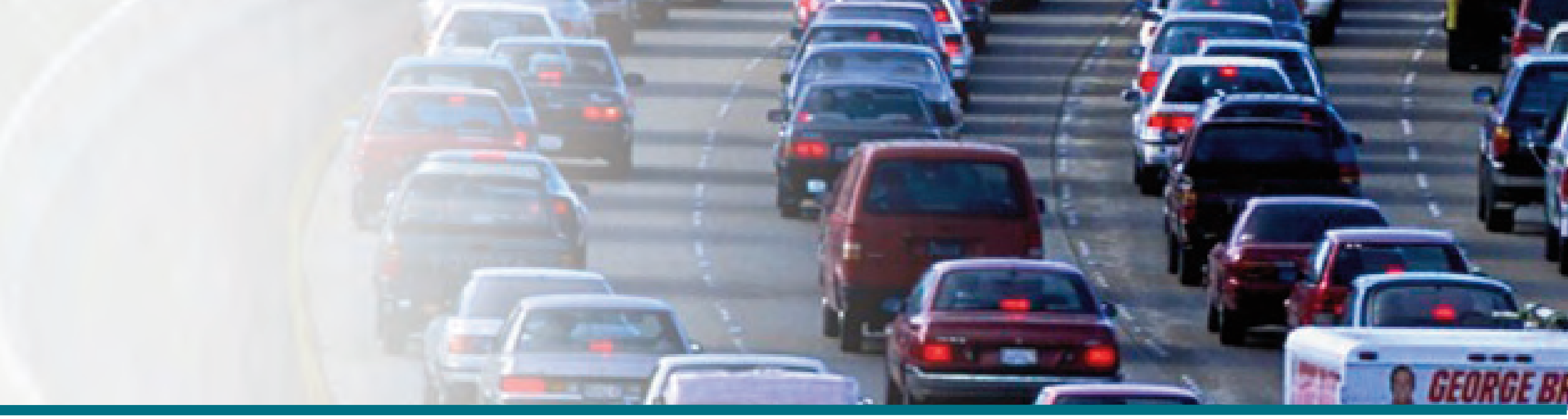
Variable speed limits (tried in work zones; i.e., nonrecurring conditions) and the European concept of "speed harmonization" (nonwork zones) both intend to "harmonize" traffic by regulating speeds. In the latter case, a series of overhead gantries gradually adjust speeds through congested highway segments in order to flatten the sinusoidal effect of traffic speeds bouncing between open sections and interchanges. Speed harmonization is typically effected as the open highway approaches the denser central business district. A great expense is incurred by the

cost of the overhead, spanned gantries, the necessary detectors, the interconnectivity, the necessary operational overhead, and the sheer number of gantries required along the multi-kilometer corridor; a smart highway, as it were. At the end of the day however, such a system is not transferable from facility to facility, so you had better have good justification and need to install it where you did. But that's the bad news too; that such pervasive congestion exists in the first place. What if the congestion disappears (e.g., a parallel facility opens up or a large traffic generator goes away)? The agency is stuck with an expensive boondoggle.

Specific to the question at hand, speed harmonization does not seek to control driver behavior to switch lanes, or mandate how to merge, but rather only to control speeds. "Go slow" (harmonize) as a means to move more traffic than otherwise might have gotten by. As of 2009, this system has not been employed in the United States, although there is intent to try it in the Seattle, Washington area as part of a Federally funded demonstration project.

So How Does One Merge Safely?

What methods work best for motorists to merge traffic efficiently in bottlenecked conditions? The answer, it seems, is largely – perhaps only – dependent on the flow rate of the continuing lane and less-so the dropped lane next to it. The debate between "early" or "late" merge seems more appropriate for human behavioral studies of aggressiveness, impatience, self-indulgence, and simple preference for one method over the other. However, there is general agreement in the engineering community on which method works best for each of two scenarios. Early merges work best under "Scenario 1" (see following page) when traffic is generally free-flow. Late merges seem to work best under "Scenario 2" when traffic is stop-and-go.



Scenario 1:
In at-speed, or nearly free-flow conditions, it is best to merge upstream of the constriction

Motorists in the thru-lane: Allow gaps. By allowing gaps you allow the merging vehicle the opportunity to adjust his/her speed to fill a gap.

Motorists in the dropped lane: Keep up speed and fill gaps at-speed. Use the dropped lane as an acceleration lane; i.e., as an opportunity to match speed with upstream traffic and merge at-speed at earliest and best opportunity.

Benefits: Traffic is safest when all vehicles travel at or near the same speed.

Potential pitfalls: 1) When a vehicle in the dropped lane can't or won't move into the through lane in time, then the dropped lane ends, forcing the vehicle to suddenly slow or stop. This will invariably cause a dangerous ripple as the stopped vehicle now must join moving traffic from nearly a dead stop; at risk of causing a through vehicle to slow to allow it in. 2) Pitfall #1 above also has the potential to cause following vehicles to queue behind the "stopped" vehicle due to an increasingly shortening merge lane. Motorists now forced to slow in the dropped lane become anxious to merge and may cause further ripples. If not allowed to remediate itself, pitfall #2 will morph into "Scenario 2" below. 3) Motorists in the thru lane won't yield (usually by way of speeding up).

Bottom line: Relax. Allow motorists to merge at-speed or fill gaps as opportunity allows.

Scenario 2:
In heavy congestion (stop-and-go conditions) it is more efficient to fill both lanes and zipper-merge at the point of constriction

Benefits: Drivers experience less stress when they understand that each in-turn will have a chance to get past the point of constriction. The zipper merge is orderly and fair.

Potential pitfalls: 1) Queue jumping. This can occur in a number of ways; a) by bypassing cars in line, the "jumper" infuriates others because he is out of turn; or b) by having a second or third vehicle try to fill a single gap, this frustrates altruistic motorists who now feel they are being taken advantage of. 2) Lane blocking. This occurs when a motorist in the continuous lane fails to fairly permit his neighbor to zipper ahead. 3) Line jumping back and forth, as it recalls the old axiom "the other line is always moving fastest." This only serves to disrupt the continuity of the merge ahead.

Bottom line: Remain orderly and fair. Do not line jump.

Low-Cost Bottleneck Improvement Strategies

Here is a sampling of short-term, low-cost operational and geometric improvements. All of these remedies address operational deficiencies, as opposed to other congestion mitigation efforts that address driver choice, travel demand, corridor-wide upgrades, or simply (but expensively) building our way out of congestion.

1. **Shoulder conversions.** This involves using a short section of traffic bearing shoulder as an additional travel lane. Shoulder conversions are appropriate between interchanges or to provide lane congruency with adjacent sections. The shoulder condition be rated for use as a travel lane.

2. **Re-striping** merge or diverge areas to provide additional lanes, provide an acceleration/deceleration lane, extend the merge/diverge area, or improve geometrics to better serve demand.

3. **Lane width reductions.** This involves reducing lane widths and re-striping to add an additional travel and/or auxiliary lane.

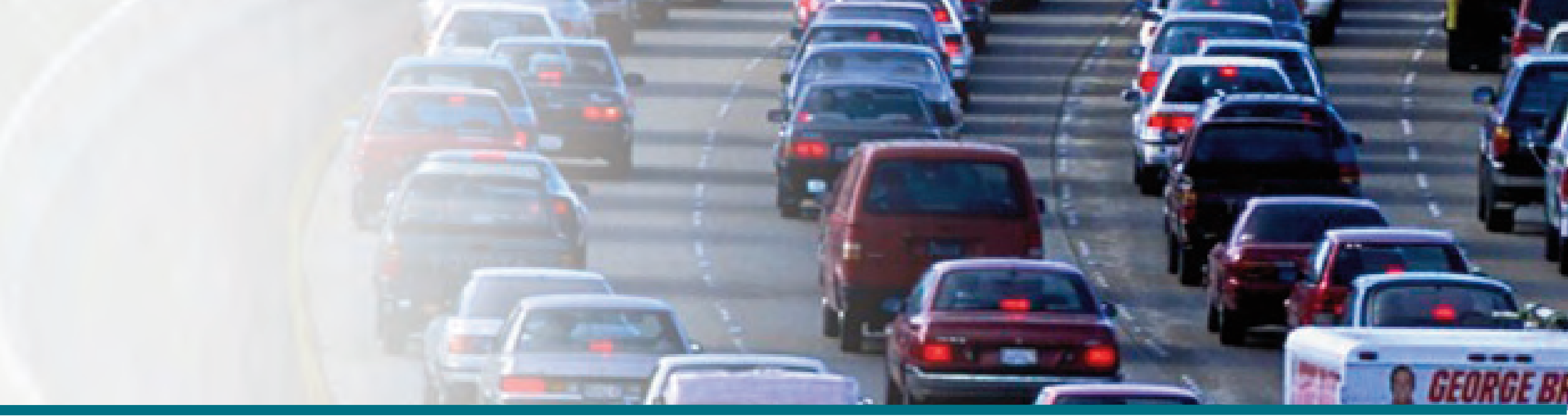
4. **Modify weaving areas** by adding collector/distributor or through lanes.

5. **Ramp modifications.** These could include ramp metering; widening, extending, closing, or consolidating ramps; or reversing entrance and exit ramps to improve operations.

6. **Speed harmonization (variable speed limits).** This is the practice of adjusting speed limits when congestion thresholds have been exceeded and congestion and queue forming is imminent. Speed harmonization can also be used to promote safer driving during inclement weather conditions. This mostly European practice reduces the traffic "shock wave" that results through congested corridors, thereby delaying the onset of a breakdown in traffic conditions. The result is decreased headways and more uniform driver behavior, which indirectly benefit bottlenecks and chokepoints.

7. **"Zippering"** or self-metering that promotes fair and smooth merges. A motorist who is 10th in line knows that he will be 20th to merge into the single lane ahead. This helps to eliminate line jumpers that bull ahead, disrupt the queues, and often block adjacent lanes until they force their way in line. Usually this method of merging requires on-site enforcement, but often is exhibited by regulars who know the process and are willing to abide.

The bottleneck solution, it seems, does not lie with the mechanics of merging, but rather in reducing the underlying impacts in the first place.



8. **Improve traffic signal timing** on arterials. Also, traffic signal timing improvements at ramp terminal intersections will prevent ramp queues from backing up onto freeway mainlanes.

9. **Access management principles** to reduce vehicular conflicts (hence, delays) on arterial corridors

10. **Continuous flow intersections.** These are unconventional at-grade intersections which eliminate one or more left turn conflicts at a main intersection. This is achieved through dedicated left turn bays located several hundred feet prior to the main intersection which allow left turning vehicles to move at the same time as through traffic. The left turn traffic signal phase is eliminated, allowing more vehicles to move through the main intersection and thus reducing traffic congestion and delays. These at-grade intersections achieve traffic flow similar to grade-separated interchanges, but at a considerably lower cost.

11. **High-Occupancy Vehicle (HOV) or reversible lanes.**

12. **Provide traveler information** on traffic diversions.

13. **Implement congestion pricing.** Congestion pricing entails charging fees or tolls for road use that vary by level of vehicle demand on the facility. The objective is to bring supply and demand into alignment. As public acceptance grows and legislative restrictions are relaxed, congestion pricing will increasingly be viewed by transportation practitioners as a powerful and relatively easy to implement strategy to address bottleneck congestion.

In 2006, as part of the research conducted for National Cooperative Highway Research Program Project 3-83 ("Low-Cost Improvements for Recurring Freeway Bottlenecks"), a series of interviews was conducted with state and local transportation agencies to assess the effectiveness of low-cost improvements used at bottleneck locations within their jurisdictions. Table 2 was developed from these responses. The results showed that agencies are using a wide range of strategies to improve bottlenecks, many of them low-cost improvements that can be implemented quickly. The most frequently used operational improvements were ramp metering, auxiliary lanes, and HOV lanes.

Some of the key questions and considerations when selecting improvement alternatives for bottleneck removal include:

- Is there an inside shoulder that would create a usable traffic lane for a short section of freeway?
- If a shoulder is considered for conversion, is there right-of-way (ROW) to allow adding one back for part of the length of the project?
- If there are bridges, are they wide enough to accommodate the extra lane while allowing adequate clearance to barriers (2 feet) and an outside shoulder? If not, are they short enough that a loss of shoulder as a breakdown lane would not be critical (500 feet or less)?
- If changes to an entrance or exit ramp or weaving area are considered, will adjusting the position of ramp gores cause geometric problems which must be resolved?
- Are vertical clearance issues, grade-matching, and sight distance problems created?
- If the bottleneck movement itself cannot be fixed reasonably, can the other traffic which is affected by it be better accommodated?
- Finally, will the improvement invite enough new traffic to cause immediate breakdown again or is this truly the clearing up of a "kink" in the system, without being a capacity addition which will overload some other part of the facility?[1]

[1] These options quoted directly from recent work by the Texas Transportation Institute: *Freeway Bottleneck Analysis Methodology*.

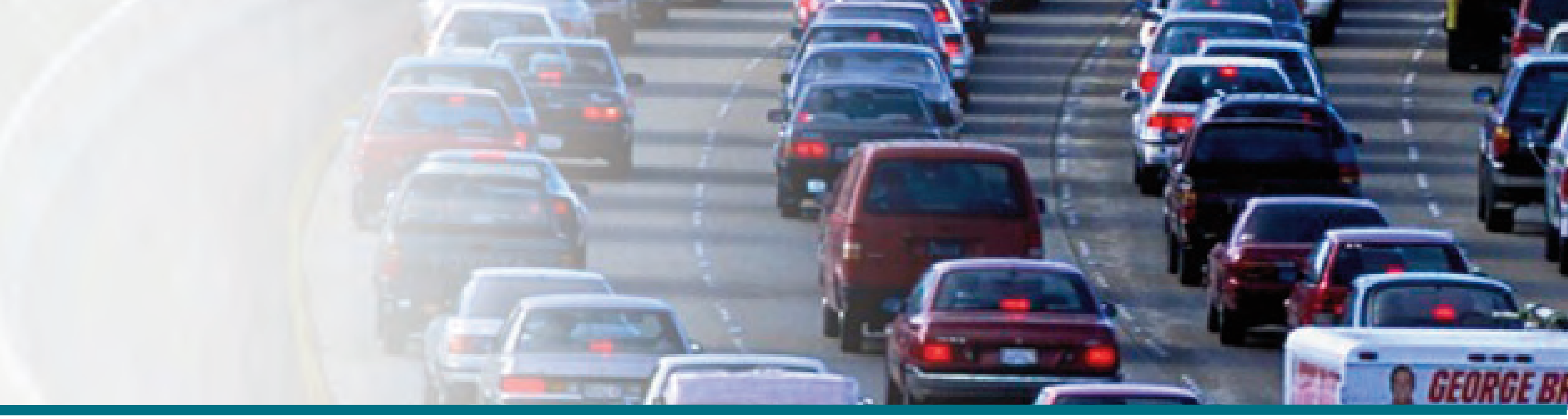


Table 2. Mapping Bottleneck Problems to Mitigation Measures

Bottleneck Types	Mitigation Measures									
	Right Shoulder Conversion	Left Shoulder Conversion	Lane Width Reduction	Auxillary Lanes	Collector-Distributor Road	Re-Stripping to Add More Narrow LanesAll Purpose Lane (Concurrent or Reversible)	HOV Lanes (Concurrent or Reversible)	Ramp Metering	Temporary Ramp Closures	Traveler Information
Heavy On-Ramp Demand	◒	○	◒	●	○	●	○	●	○	◒
Weaving Sections	◒	◒	●	◒	◒	●	●	○	○	◒
Lane Drops	●	○	●	●	○	●	●	◒	◒	◒
Tunnels and Bridges	○	○	●	○	○	○	◒	○	○	◒
Horizontal and Vertical Curves	●	●	○	●	○	●	◒	◒	◒	◒
Narrow Lanes and Lateral Obstruction	◒	◒	○	◒	◒	◒	◒	◒	◒	◒
Inadequate Accelerated and/or Decelerated Lanes	●	○	◒	●	●	◒	◒	◒	●	●

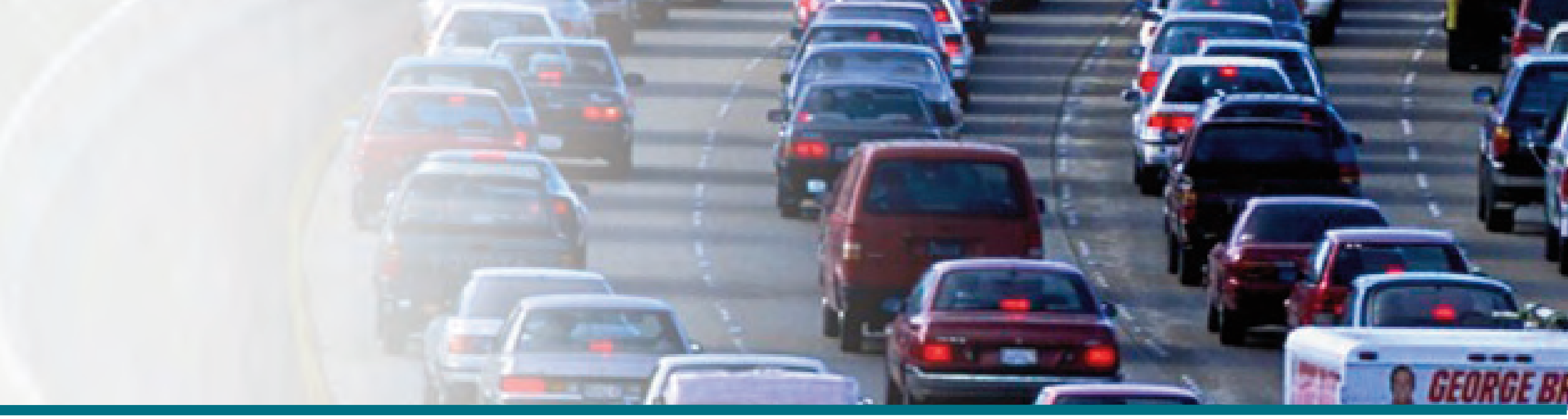
● = good solution ◒ = may be helpful ○ = not applicable

Evaluating Bottleneck Improvement Effectiveness

After implementation, it is often beneficial to conduct an "after" evaluation to gauge the effectiveness of the bottleneck removal project. A conservative approach to evaluating treatment effectiveness is to evaluate operational and safety benefits achieved. Common evaluation methods include microsimulation modeling, cost/benefit analysis, and crash data analysis using data collected before and after project implementation. In addition, the following performance measures are often used to assess the effectiveness of bottleneck improvement strategies: average speed (travel time), lane density, queue lengths, queue discharge rates, vehicle miles of travel (VMT), and vehicle hours of travel (delay). Additional insight could be obtained from before and after opinion surveys of area drivers. These types of evaluations are often not done, yet are important to quantify the benefits achieved through bottleneck mitigation.

Potential Issues with Bottleneck Treatments

Because some bottleneck treatments involve innovative solutions that maximize effectiveness with a minimum of new construction, they are occasionally at odds with highway design standards. For example, the addition of slip ramp to a collector/distributor road or the use of a shoulder a through lane at selected locations may not strictly follow allowable design standards. Such deviations have the potential to degrade safety if not properly implemented; the elimination of a shoulder may led to more collisions with roadside features or may impede incident management activities. As it is FHWA's intent to foster creative approaches for low-cost bottleneck improvements, agencies should not see the design standard issue as insurmountable. Rather, they should fully assess the potential safety impacts of strategies and devise ways of addressing them, if necessary. For example, in the case of a shoulder-to-lane conversion, review of crash data, and the specific roadway location (perhaps through a Roadway Safety Audit), it may be determined that a barrier is required to keep vehicles off of the roadside. It may also require a change in incident management policy that would allow emergency vehicles to access incidents from the opposite direction. Finally agencies should be in contact with the FHWA Division offices throughout the process as design review may be required, depending on circumstances.



The second potential issue relates to air quality conformity. Because they are short-term in nature, localized bottleneck improvements may emerge as formal projects that have not been previously identified in Statewide Transportation Improvement Programs or MPO-generated Transportation Improvement Programs. Thus, they may not be part of those projects that have been approved to deal with air quality issues in the region or state. Such occurrences must be dealt with on a case-by-case basis by agencies wishing to undertake bottleneck improvements. One point worth noting: if air quality conformity in a location precludes or discourages major capital expansion (e.g., additional lane-miles), the type of improvements in a localized bottleneck program clearly do not fall in this category.

Examples of How Agencies are Dealing with Bottlenecks

Many transportation agencies have recognized that low-cost treatments can provide effective congestion relief at bottlenecks. A wide variety of improvements have been implemented and many innovative improvements are emerging. This following section provides expanded explanations of how these transportation agencies used strategies to improve congestion at bottlenecks.

- Low cost bottleneck improvements in **Minnesota**;
- **Connecticut DOT's** low-cost efforts to improve a freeway diverge area;
- Expediting bottleneck improvements ahead of major reconstruction project in **New Hampshire**;
- Continuous flow intersection improves intersection operations in **Utah**; and
- Operation Bottleneck Relief in **Los Angeles, California**.

For a more complete listing of representative projects, or for additional background on specific projects, please visit the FHWA "Bottleneck" web site at http://www.ops.fhwa.dot.gov/bn/index.htm or contact the Localized Bottleneck Reduction Program manager.

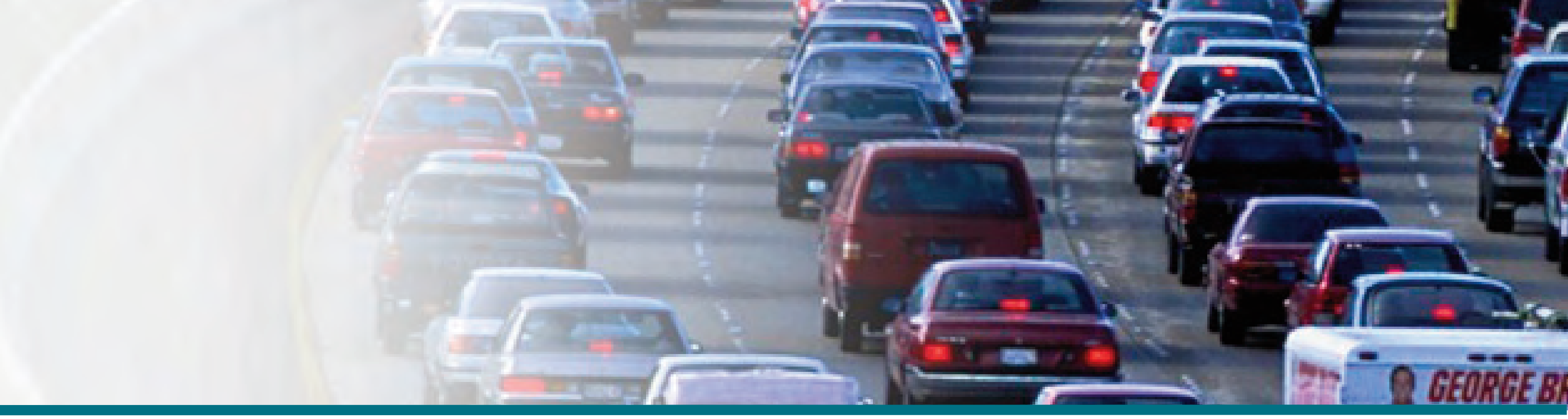
Low Cost improvements in Minnesota Provides Similar Benefits as a Mega Project, but at a Vastly Different Cost

In 2006, Minnesota DOT utilized low-cost strategies to resolve a bottleneck on State Highway 100 between 36th Street and I-394 in St. Louis Park, a suburb west of Minneapolis. This section of highway had only two lanes, while upstream and downstream sections of SH 100 had three lanes. A prior study found that SH 100 in this area was congested for six hours out of every day, the longest duration of congestion than on any other freeway in the Twin Cities metropolitan area.

A full third lane was added in each direction by narrowing the left and right shoulders. In addition, a nearby diamond interchange was connected via collector-distributor roads, reducing access points on SH 100 from seven to four. The improvements eliminated weaving movements that existed along this stretch of on- and off-ramps past the city.

The result yielded a (some say conservative!) 13:1 benefit/cost ratio; an increased throughput of 14,400 vehicles during peak periods daily; and a reduction in backups from 5 to 6 miles previously (depending on direction) to one-quarter mile. Minnesota DOT received an outpouring of positive public reaction, including one local newspaper's "Public Project of the Year" award.

Remarkably, this $7.1M project accomplished the same results as a $138 million project on a parallel freeway facility, IH-494 from Highway 5 to Highway 55. Granted, the IH-494 design/build project was closer to a total facility rehabilitation with noise walls and other costs, but the effective elimination of long bottleneck queues was realized with the SH 100 project for a fraction of the cost.



Re-striping Resolves Bottleneck Conditions at Freeway Diverge Area in Connecticut

Congestion routinely existed on I-84 eastbound in Danbury in advance of the I-84/Route 7 split. The existing lane configuration in the vicinity of the exit ramp consisted of the left lane as an exit-only lane, with the center and right lanes continuing as through lanes. Because significant traffic exited onto Route 7 northbound, this arrangement created weaving problems and bottlenecking, as all traffic bound for Route 7 had to maneuver into the left lane to access the exit.

Connecticut DOT devised a low cost solution which involved re-striping the lanes within the diverge area to better serve demand. The I-84 eastbound lanes were reconfigured to allow the center lane to function as an option lane, where vehicles would have the option to exit or continue on I-84. It was determined that the existing pavement width in the vicinity of the ramp gore would accommodate the option lane and would tie in with the current lane configuration on the ramp, which included two lanes. The project required minor signing and striping revisions and some pavement modification. The bottleneck improvements will be completed as part of a larger interchange improvement project at I-84/Route 37, which will be advertised in June 2009. The cost of the bottleneck improvements is estimated to be $900,000.

I-84 Eastbound at Exit 7 (Route 7 Northbound) – Existing Lane Configuration

I-84 Eastbound at Exit 7 (Route 7 Northbound) – Proposed Lane Configuration

Bottleneck Improvements Expedited Ahead of Major Reconstruction Project in New Hampshire

The existing 19.8-mile four-lane section of I-93 from the Massachusetts state line north to Manchester was originally built in the 1960s and since that time, the facility has not undergone any major system wide upgrades or modernization. The interchanges at Exits 2, 3, and 5 remain essentially as they were originally constructed. Exits 1 and 4 were partially reconstructed in the late 80s and early 90s, but are rapidly becoming incapable of meeting traffic demand. To address these problems, NHDOT identified a major, multiyear reconstruction project. The project will include: widening the mainline from two to four lanes per direction; widening of 44 bridges; reconstruction of all five interchanges on the section; addition of sound barriers; construction of three new park and ride lots; expanded bus service; and the purchase of right-of-way for a future light rail line. However, because this project has an ambitious scale, it will take several years to complete and major congestion now occurs at Exit 5.

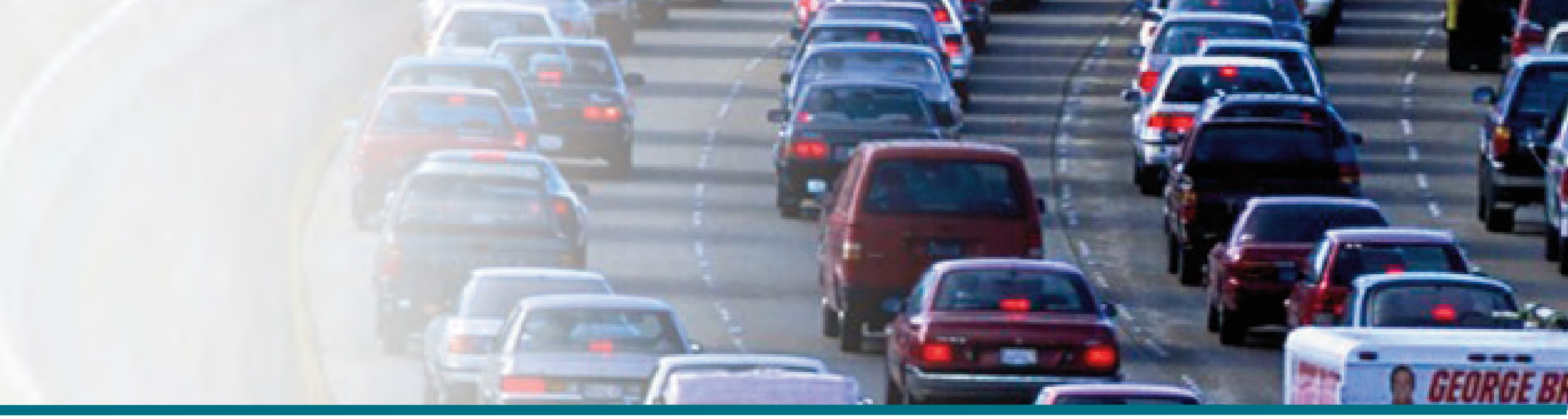
The State of New Hampshire has selected work at Exit 5 (Londonderry) on I-93 to be expedited in order to address the bottlenecking concerns which currently exist. The signalized northbound and southbound ramps back up onto the mainline during peaks. The original project called for much of Exit 5 to be reconstructed in FY 2009. In partial response to the LBR program attention, improvements to two of the existing ramps are being expedited. The southbound off ramp at Exit 5 will be extended. The shoulder area on the mainline will be widened and reinforced to hold some of the queue. The signal at the top of the ramp will be coordinated with loop detection. Reconstruction or widening of two nearby bridges will be accomplished. The estimated cost of this expedited work is $9.13 million. The completed project will meld into the overall 19.8-mile I-93 project

Continuous Flow Design Improves Intersection Operations in Utah

A study of SR-171 (3500 South) at SR-154 (Bangerter Highway) in West Valley City showed that the intersection is one of the busiest in the State, with over 100,000 vehicles passing through the intersection on a typical weekday. Heavy rush hour traffic forces drivers to wait for up to four cycles of the traffic signal to get through the intersection, which contributes to congestion and safety problems that ripple out to affect other intersection on both 3500 South and Bangerter Highway. The study estimates that traffic through the intersection will increase 31 percent by 2030.

Several improvement alternatives for the intersection were evaluated, including a grade separated interchange. However, UDOT decided to implement Utah's first continuous flow intersection (CFI). An unconventional intersection design, CFIs reduce the steps in the traffic signal cycle and place left turning vehicles along a safer path. Compared to a grade separated interchange, the CFI is a low cost alternative.

The CFI improvement project was started in March 2007 and was completed six months later in September 2007, with a total project cost of $7.5 million.



Operation Bottleneck Relief in Los Angeles, California

The City of Los Angeles recently initiated this four-year program to identify and resolve bottlenecks at signalized intersections. The City is severely constrained in expanding mainlines and adding additional lanes at intersections, as right-of-way costs are prohibitive. The program has identified 300 bottleneck locations to date.

Bottlenecks are defined as signalized intersections reaching greater than 40 percent loop occupancy on the upstream system detectors for four 15-minute periods. Once a signal is identified as a bottleneck, the signal timing and phasing is reviewed to see if these can be adjusted to meet demand better. If this is not fruitful, then additional improvement types such as re-striping to provide additional lanes, parking restrictions, or physical widening are evaluated:

The program initially identified 98 intersections that met the occupancy thresholds during peak periods:

- Thirty six percent were successfully treated through signal retiming;
- Fourteen percent were identified as candidates for parking restrictions or lane additions;
- Eight percent were caused by ramp metering overflow from freeways, which will require coordination with Caltrans to adjust metering rates and/or provide extra queue storage; and
- Forty two percent could not be reasonably treated by any of the strategies identified above; demand was simply too high and improvements would be too costly.

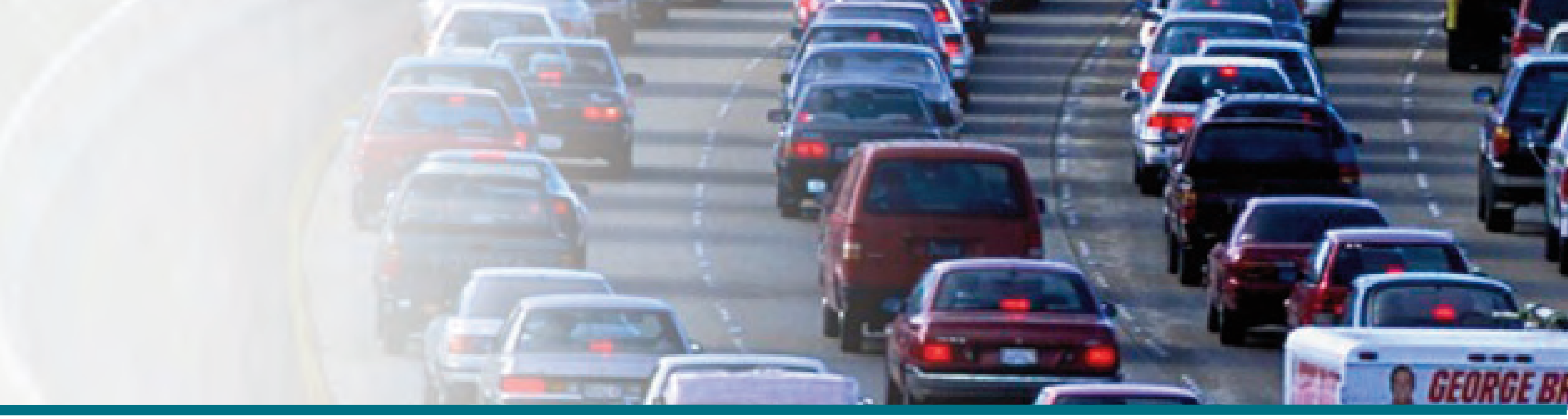
Unfortunately, adaptive signal timing will not resolve the issues for the remaining bottlenecks on the arterial system, as the competing approaches and saturation requires alternatives beyond adaptive timing. However, timing is always the first approach to resolution of the bottlenecks identified through the program.

The City anticipates adding 1,000 more signals over the next 15 years with 300 to 400 in the more immediate future. The OBR program will be used in the long-range planning process to help identify funding sources and solutions. Additionally, it will help building the case for changes both politically and within the community about how bottlenecks are addressed and resolved.

Planning and Programming Bottleneck Improvements

Unless transportation agencies make low-cost bottleneck improvements an explicit presence, it is likely that they will be overlooked or delayed; either deemed part of a "larger problem," or unnecessarily postponed to some indefinite out year. There are many ways to combat this:

- **Create a unique bottleneck program area.** By developing an annual "named" program, agencies can effectively identify, fund, and most importantly, advance low-cost treatments. A stand-alone program also has the added benefit of demonstrating to the public that the agency is actively engaged in fighting congestion.
- **Undertake occasional "special projects" to focus on bottlenecks.** Low-cost bottlenecks can be addressed through occasional "special projects." For example, the Minnesota DOT (MnDOT) conducted a "one-time" special compilation of projects meeting certain candidacy requirements. In much less than one year, MnDOT developed a highly accelerated process for bottleneck identification and prioritization, which led to many effective projects that were implemented in the following two years.
- **Integrate consideration of low-cost bottlenecks into existing programs.** Low-cost bottlenecks can be addressed programmatically even without a special program. By making them part of ongoing planning and processes, the can be part of an agency's congestion arsenal. For example, the Ohio DOT (ODOT) added a congestion-based index ranking to their predominantly safety-based Highway Safety Program (HSP). The result is that congestion weighted problems now have a specific "voice" in the rankings of all candidate ODOT projects. The HSP was used as the mechanism because many of its projects also are low-cost and quickly implementable.



The following provide comparisons of how different state agencies have incorporated low cost bottleneck projects into their planning and programming processes:

- The **California Department of Transportation** (Caltrans) does not have a formal bottleneck planning process; rather, bottleneck issues are addressed at the district level as part of the regional planning process. Much of Caltrans' operational planning is guided by the Transportation Management System Master Plan, which sets forth the types of strategies that should be pursued in improving congestion. In much of California's metropolitan areas, traffic congestion is a 24/7 occurrence, and traffic management is a 24/7 job. Bottlenecks are tweaked "in real time" as part of their Corridor System Management Plans (CSMP), which are developed for some of California's most congested transportation corridors. System monitoring and evaluation is seen as the foundation for the entire process because it can not only identify congestion problems, but also be used to evaluate and prioritize competing investments. Caltrans does not have a direct funding for bottlenecks, although bottleneck projects are routinely programmed through the CSMP process.

- In **Ohio**, bottlenecks are part and parcel of the overarching Ohio Department of Transportation (ODOT) Highway Safety Program (HSP), which ranks all candidate projects and drives the statewide highway project selection and scheduling process. Beginning in 2002, ODOT developed a "congestion mapping" division that uses V/C ratios developed from traffic data recorders and roadway inventory. About the same time, ODOT administration pushed for an annual process of overlaying congestion-index and safety-index "hot spots." As a result, congestion hot spots now have a "voice" in the process regardless of crash indices, and congestion related problems now compete for attention in the HSP listing. Specifically, highway sections with V/C ratios greater than 1.0 are considered "congested" and are added to the listing. Sections with V/C between 0.9 and 1.0, but outside the cities of Columbus, Cincinnati, and Cleveland, are also added. After ODOT headquarters completes their statewide effort of congestion mapping and safety indexing, the respective District engineers are responsible for developing countermeasures for their top-listed candidate projects. District Safety Review Teams sort projects into three scales – low (less than $100K and quickly implementable), medium ($100K to $5M and 1 to 2 years), and high (greater than $5M and necessitating more than years to implement) – and then compete with other projects having the same scale but in other districts.

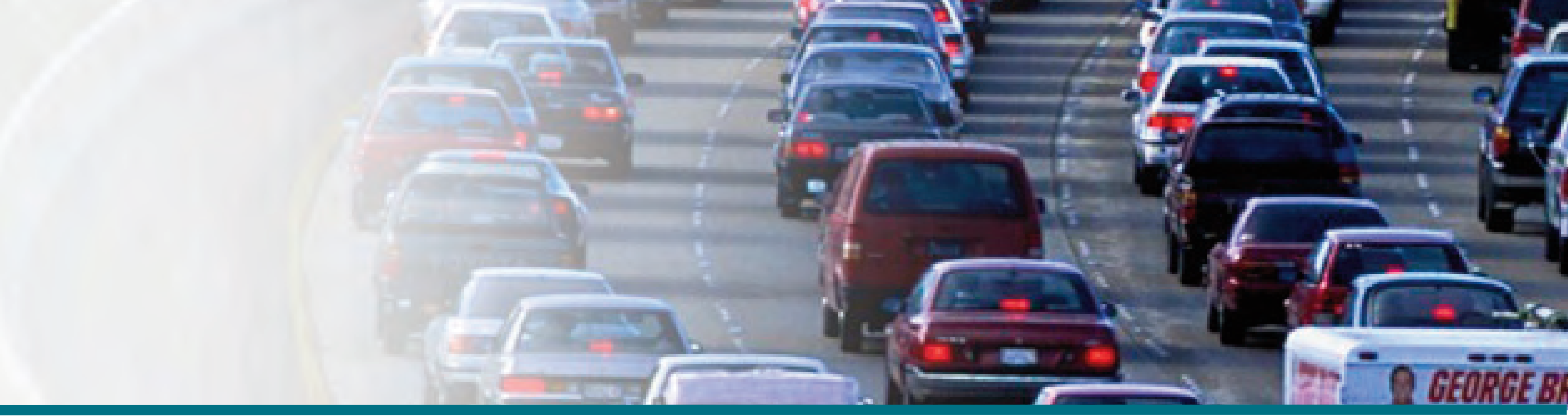
- **Minnesota DOT** was originally driven to explore low-cost congestion relief projects because of budgetary restrictions, but soon realized that these projects could be implemented very quickly and, as a bonus, were highly visible

and popular with the public. In much less than one year, MnDOT developed a highly accelerated process for bottleneck identification and prioritization, which led to many effective projects in the following two years. MnDOT also found that because of lower costs, it could identify multiple locations throughout the region and "spread around" bottleneck reduction projects in a fair and equitable manner. This process consisted of completing a study, which included a five-step process to narrow potential projects into a recommendation list to the state legislature. Evaluation of completed projects produced high benefit/cost ratios, usually greater than 8:1.

- The **Maryland State Highway Administration** (SHA) has a dedicated program of about $5 million per year for the identification and implementation of low cost traffic congestion improvements at intersections. The program's genesis tracks to when SHA asked "what can be done if and when a mega-project's 'no build' alternative is chosen?" The program has been well-received by the public and local governments. Projects typically include low-cost projects that can be implemented quickly, such as signal timing upgrades and adding turn lanes and through lanes at intersections. The Maryland SHA has also has had considerable success with projects to improve freeway ramps and merge areas that have reduced congestion bottlenecks at a low cost.

- In **Florida**, there is not a "bottleneck" planning process, per se; rather, bottleneck-related issues are addressed as part of the Florida Department of Transportation's (FDOT) standard planning process. The planning process, which is managed by the FDOT Systems Planning Office, begins with needs identification conducted at the district level, then projects are developed and proposed for the Cost Feasible Plan. The Cost Feasible Plan is adopted and projects are ranked for inclusion into the 5-year or 10-year programs. Traffic data and the statewide model are used to identify deficiencies, but it is the responsibility of the districts to identify and resolve hot spots.

- **Washington State DOT** (WSDOT) has no direct funding for bottlenecks, but formally recognizes "bottlenecks and chokepoints" in their project planning and development process and devotes a portion of the Washington Transportation Plan (WTP) to them. At the planning stage, WSDOT considers bottlenecks together with traditional corridor improvements in a category called "Congestion Relief" – bottlenecks do not have their own category for assessment or funding. The Congestion Relief projects are ranked (prioritized) using the benefit/cost ratio and other qualitative factors.



Minnesota's Process to Identify and Prioritize Bottleneck Improvements

Step 1: Project Identification

Potential congestion management projects were identified from existing sources:

- Low-cost capacity improvements (e.g., auxiliary lanes)
- Re-striping lane configuration
- Traffic control device improvements (e.g., ramp meters and signal timing)

Step 2: Quantitative Screening

- Project cost < $15 million
- Not in three-year TIP
- Annual hours of delay > 25,000
- Minimum of two hours of congestion

Step 3: Qualitative Screening

- Design readiness
- Cost range
- Congestion benefit
- Construction traffic management
- Future demand changes
- No adverse downstream effects

Step 4: Expert Workshop

Projects were prioritized by an expert group during a half-day workshop.

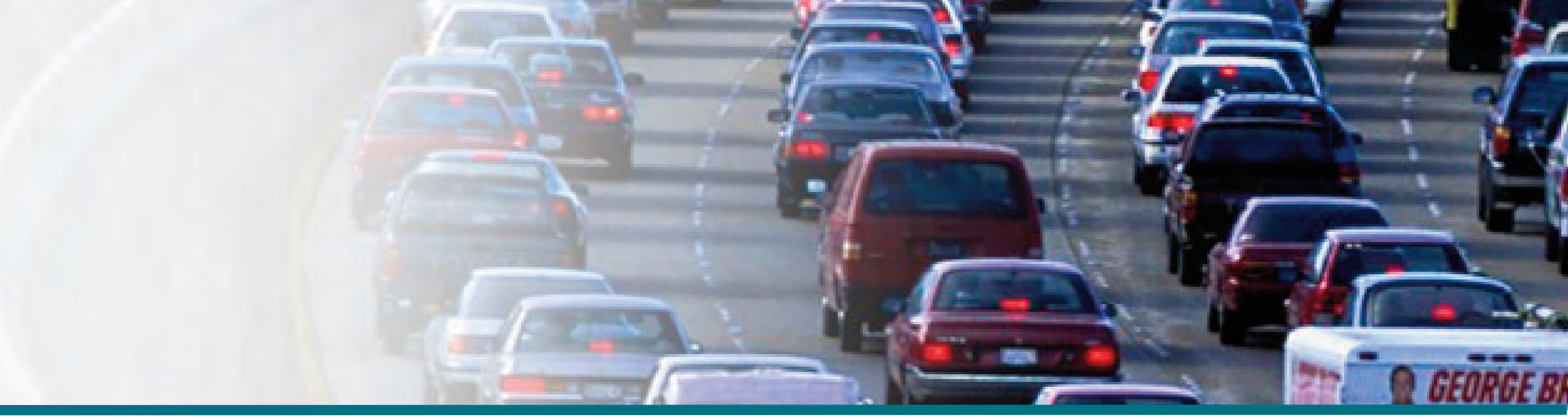
Step 5: Project Planning

The following were prepared for each project:

- Geometric sketches
- Project scope
- Congestion impacts
- Safety impacts
- Benefit-to-cost ratio

Want More Information?

The LBR Program is just one of several program areas dealing with congestion problems. More information may be found at FHWA's "Focus on Congestion" web page at: http://www.fhwa.dot.gov/congestion/links.htm.



Definitions

Auxiliary lanes – Typically, any lane whose primary function is not simply to carry through traffic. This can range from turn lanes, ramps, and other single purpose lanes, or it can be broadened to imply that a traffic bearing shoulder can be opened in peak periods to help alleviate a bottleneck, and then "shut back off" when the peak is over.

Bottleneck – There can be many definitions. Here are a few that are typically used. 1) A critical point of traffic congestion evidenced by queues upstream and free flowing traffic downstream; 2) A location on a highway where there is loss of physical capacity, surges in demand (traffic volumes), or both; 3) A point where traffic demand exceeds the normal capacity; and 4) A location where demand for usage of a highway section periodically exceeds the section's physical ability to handle it, and is independent of traffic-disrupting events that can occur on the roadway.

Capacity – The maximum amount of traffic capable of being handled by a given highway section. Traffic engineers usually speak in terms of "free flow" capacity.

Congestion (specifically, traffic congestion) – FHWA's Traffic Congestion and Reliability Report defines congestion as "an excess of vehicles on a portion of roadway at a particular time resulting in speeds that are slower – sometimes much slower – than normal or free flow speeds. (Congestion is) stop-and-go traffic. Previous work has shown that congestion is the result of seven root causes,[2] often interacting with one another." Since a bottleneck is a cause of congestion, congestion cannot be solely analogous to a bottleneck. Congestion is more. For example, a "congested" corridor may harbor multiple bottlenecks or any combination of the seven root causes of congestion.

Downstream traffic – Traffic that is beyond (past) the subject point on a highway.

Hidden bottleneck – A highway location where some type of physical restriction is present, but traffic flow into this area is metered by an upstream bottleneck so the location does not appear as a bottleneck under prevailing conditions. Removal of the upstream bottleneck will cause the hidden one to emerge as a new bottleneck.

---

[2] The seven root causes are physical bottlenecks (a.k.a. "capacity constraints"), traffic incidents, work zones, weather, poorly timed signals et al., special events, and over-capacity demand (i.e., daily and seasonal peaks superimposed on a system with a fixed capacity). Some sources cite only six root causes because they see over-demand as an inherent sub-element necessary for any of the other causes to exist in the first place. Put another way, absent over-demand, there would just be "volume," but not necessarily "congested" volume.

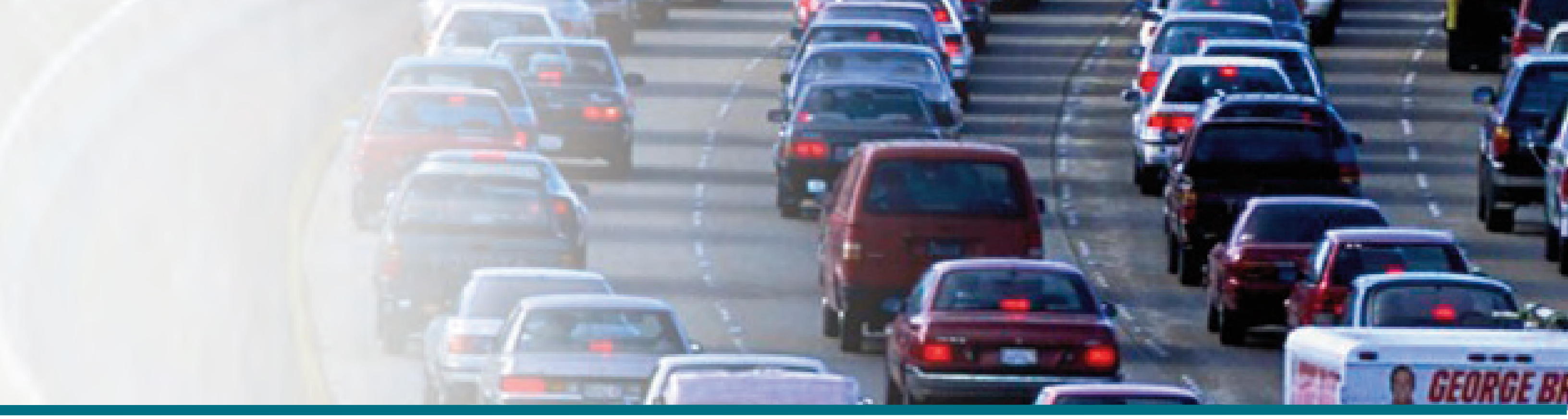
Nonrecurring events – As it pertains to traffic, a delay caused by an unforeseen event; usually a traffic incident, the weather, a vehicle breakdown, a work zone, or other atypical event. Even if planned in many cases, like work zones and special events, they are irregular and not predictably habitual in location and duration.

Ramp metering – The practice of managing access to a highway via use of control devices such as traffic signals, signing, and gates to regulate the number of vehicles entering or leaving the freeway, in order to achieve operational objectives. The intent of ramp metering is to smooth the rate at which entering vehicles will compete with through vehicles. Done properly, ramp metering will calm the "mix" that occurs at these junctions.

Recurring event – As it pertains to traffic, a recurring event is a traffic condition (i.e., a bottleneck or backup) that one can presume to occur in the same location and at the same time daily, albeit for weekday or weekend conditions. Examples would be peak-hour slowdowns at junction points, intersections, and ramps. One can "plan" for these events because one knows by routine that such events will occur time and again in the same manner and place.

Traffic microsimulation tools – Complex microsimulation tools that rely on input of traffic data, intersection "nodes," facility "links," and the associated parameters of each input, in order to output simulated conditions. By changing the inputs, engineers can test different sizes, characteristics, and out-year scenarios of traffic demand.

Upstream traffic – Traffic that has not yet arrived at the subject point on a highway.



Traffic Bottlenecks –

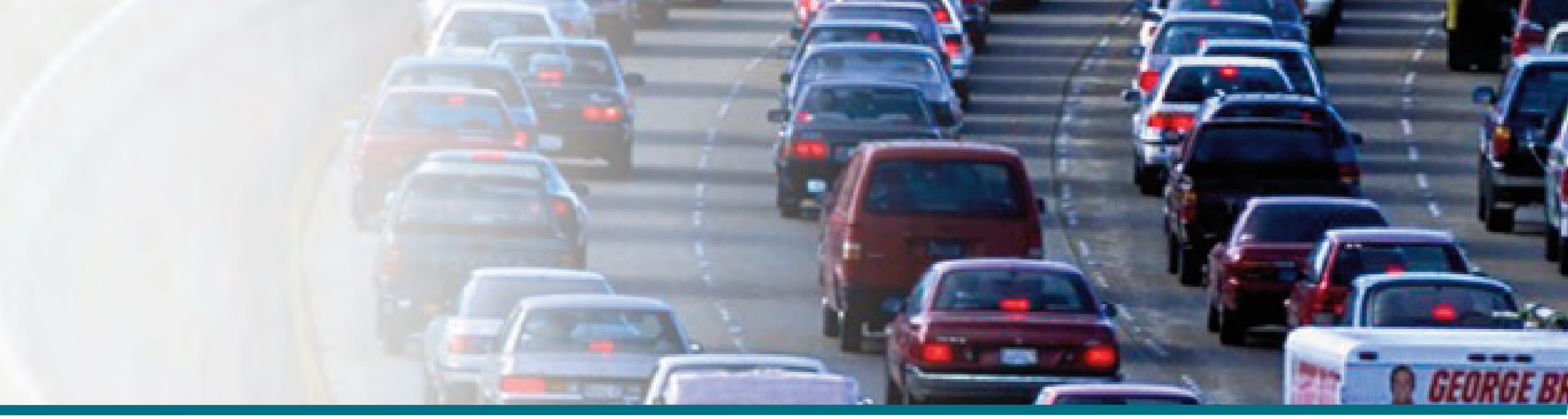
Localized sections of highway where traffic experiences reduced speeds and delays due to recurring operational conditions or nonrecurring traffic-influencing events.

	Recurring – "Predictable" in cause, location, time of day, and approximate duration. **Nonrecurring –** "Random" (in the colloquial sense) as to location and severity. Even if planned in some cases, like work zones or special events, these occurrences are irregular and are not predictably habitual or recurring in location.
	Recurring: Operational Causes – A "facility determinate" condition wherein a fixed condition (the design or function of the facility at that point) allows surging traffic confluence to periodically overwhelm the roadway's physical ability (i.e., capacity) to handle the traffic, resulting in predictable periods of delay. **Nonrecurring: Dynamic Occurrences –** An "event determinate" occurrence, wherein a dynamic situation either reduces available capacity (e.g., loss of lanes due to incident or work zone) or increases demand (e.g., special event).
	Recurring: Ramps, lane drops, weaves, merges, grades, underpasses, tunnels, narrow lanes, lack of shoulders, bridge lane reduction, curves, poorly operating traffic signals. **Nonrecurring:** Work zones, crashes, incidents, special events, weather.
	"Active" bottlenecks – When traffic "released" past the bottleneck is not affected by a downstream restriction (i.e., queue spillback) from another bottleneck. "Hidden" bottlenecks – When traffic demand is metered by another upstream bottleneck(s); i.e., either a lesser or nonexistent bottleneck that would increase or appear, respectively, if only unfettered traffic could reach it.
	Motorists typically refer to bottlenecks in terms of added time delay when compared to the same nondelayed trip, but engineers and agencies also measure performance data: average speed (travel time), lane densities, queue lengths, queue discharge rates, vehicle miles of travel (VMT), and vehicle hours of travel (VHT).
	Data is collected using manual techniques (e.g., floating cars, aerial photography, or manual counts from video recordings) or from dynamic surveillance (e.g., detectors, radar, video, etc.) collected in real time. Modeling, especially microsimulation, can be used to study the impacts of bottleneck remediation on upstream and downstream conditions.
	Recurring: Type I **–** Demand surge, no capacity reduction (typically at freeway on-ramp merges). Type II – Capacity reduction, no demand surge (typically changes in freeway geometry; lane drop, grade, curve). Type III – Combined demand surge and capacity reduction (typically in weaving sections). Nonrecurring: Usually classified by the type of event (e.g., incident, work zone) and severity of impact (e.g., duration of the number of lanes lost, closed, or impassable).
	Recurring: Bottleneck is due to over-demand of volume (i.e., peak-hour conditions). The bottleneck clears from the rear of the queue as volume declines. **Nonrecurring:** Bottleneck is due to loss of capacity due to an incident, or short-term over-demand due to a spot event. The bottleneck clears from the front or rear of the queue, depending on whether the cause is incident-related (former) or volume-related (latter), respectively.
	Recurring: When volume over-demand drops back to manageable levels for available capacity (i.e., when off-peak conditions return). **Nonrecurring:** When dynamic event is removed; queue should dissipate, thereafter.

Traffic Bottlenecks – (continued)

Localized sections of highway where traffic experiences reduced speeds and delays due to recurring operational conditions or nonrecurring traffic-influencing events.

Practical Mitigations:	Recurring: Corridor Congestion	Recurring: Localized Bottlenecks	Nonrecurring
	Dynamic pricing	Use shoulder lane	Improve incident response capabilities; reduce incident impact; reduce on scene time for clearing incidents; reduce facility "downtime" during the event. In work zones, maintain maximum number of open lanes during peak times; shorten durations using innovative methods and contracting practices; minimize number of times a section is an active work zone by combining improvements (e.g., paving and safety) and using highly durable materials; employ least intrusive detour(s). Pre-plan for and coordinate special events to adequately and efficiently handle event traffic, including not only the main event but the subordinate deliveries, VIP access, emergency response, and pre- and post-event activities. Have predetermined detour plans for particular sections of highway in the event of weather- or incident-related events, including available tools (i.e., arrows, sign stands, VMS boards, public information conduits, etc.).
	Transit alternatives	Restripe weave area	
	Ridesharing, telecommuting	Improve merge area	
	High-occupancy lanes	Widen, extend, remove, or consolidate ramps	
	Successive ramp metering	Individual metered or signalized ramp	
	New construction	Improve signalization or intersection design	
	Install frontage roads	Install frontage road	
	Traffic demand management (TDM) techniques	Effect "speed harmonization" as in Europe	
	Build park-and-ride lots	Encourage "zippering"	
	"Downtown" or cordon\ congestion pricing	Use access management techniques	
	Provide traveler information	Provide traveler information	
	Proactive signal timing plans (including adaptive control)		



NOTICE

The publication of this document was sponsored by the U.S. Federal Highway Administration under contract DTFH61-06-00004, awarded to Cambridge Systematics, Inc.

QUALITY ASSURANCE STATEMENT

June 2009

FEDERAL HIGHWAY ADMINISTRATION
Office of Transportation Management, (HoTM)
1200 New Jersey Avenue, SE
Washington, DC 20590
www.fhwa.dot.gov

FHWA-HOP-09-037