Title
On the Stability of Freeway Traffic

Author
Daganzo, Carlos F.

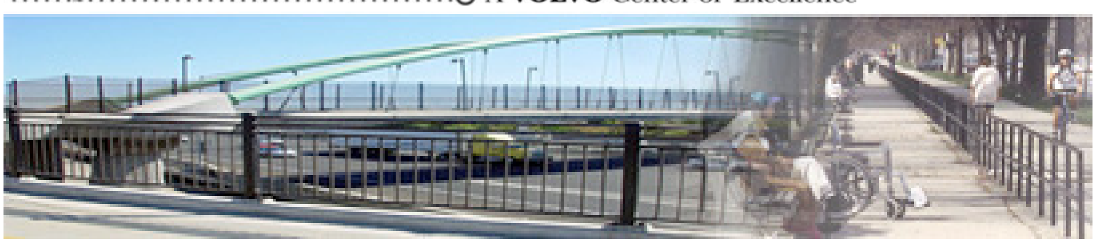
Publication Date
2010-10-01

On the Stability of Freeway Traffic

Carlos F. Daganzo

WORKING PAPER
UCB-ITS-VWP-2010-9

UC Berkeley Center for Future Urban Transport
A VOLVO Center of Excellence



October 2010

On the Stability of Freeway Traffic

WORKING PAPER

Carlos F. Daganzo

UC Berkeley Center for Future Urban Transport,
A Volvo Center of Excellence

October 20, 2010

Abstract

Traffic flow theory is used to analyze the spatio-temporal distribution of flow and density on closed-loop homogeneous freeways with many ramps, which produce inflows and allow outflows. It is shown that if the on-ramp demand is space-independent then this distribution tends toward uniformity in space if the freeway is either: (i) uncongested; or (ii) congested with queues on its on-ramps and enough inflow to cause the average freeway density to increase with time. In all other cases, including any recovery phase of a rush hour where the freeway's average density declines, the distribution of flow and density quickly becomes uneven. The flow-density deviations from the average are shown to grow exponentially in time and propagate backwards in space with a fixed wave speed. A consequence of this type of instability is that, during recovery, gaps of uncongested traffic will quickly appear in the unevenly congested stream, reducing average flow. This extends the duration of recovery and invariably creates clockwise hysteresis loops on scatter-plots of average system flow vs. density. All these effects are quantified with formulas and verified with simulations. Some have been observed in real networks. In a more practical vein, it is also shown that the negative effects of instability diminish (i.e., freeway flows increase) if (a) some drivers choose to exit the freeway prematurely when it is too congested and/or (b) freeway access is regulated in a certain traffic-responsive way. These two findings could be used to improve the algorithms behind VMS displays for driver guidance (finding a), and on-ramp metering rates (finding b).

1 Introduction

It has been proposed (Godfrey, 1969; Herman and Ardekani, 1984; Ardekani and Herman, 1987; Mahmassani and Peeta, 1993; Olszewski et al., 1995; Daganzo, 2007) that plots of average flow vs. average density observed in a city network should follow unimodal curves. It has also been suggested (Daganzo, 2007) that if these curves are reproducible they can be used to build aggregate models of network flow dynamics and control, treating the network as a single reservoir with evenly distributed vehicles. The latter reference also conjectures that curves of this type should be consistently reproduced if the average speed on different parts of the network they describe is evened out by drivers' navigation habits. Geroliminis and Daganzo (2008) shows that indeed, plots of average flow vs. average density for all the streets in Yokohama (Japan) are reproducible and well organized, and called these curves "macroscopic fundamental diagrams" (or MFD's). More recently, Buisson and Ladier (2009) shows that multi-day data from those intersections in Toulouse (France) that have detectors are also well organized .

There are exceptions, however. Buisson and Ladier (2009) also shows that the average network flows were considerably lower on a day when the network was subject to a large non-recurrent disturbance. On that day the data were more scattered and formed a clockwise hysteresis loop. The lower flows and the sacatter are attributed in Buisson and Ladier (2009) to the disturbance itself, which misdistributed congestion. But that reference does not explain the timing of the disturbance or the direction of the hysteresis loop. Daganzo et al. (2010) shows, using a 2-bin model of network dynamics, that reduced driver adaptation on the day of the disturbance could also contribute to the lower flows. And, using a generalization of the same model, Gayah and Daganzo (2010) explains both, the scatter and the direction of the hysteresis loop. More specifically, Gayah and Daganzo (2010) shows that if drivers are insufficiently adaptive then congestion must become more uneven as the rush progresses and that as a result, flows for any given density are naturally lower when congestion is declining than when it is increasing. This insight strongly suggests that on any network where drivers do not, or cannot, adapt (the latter will happen if the network lacks redundant routes) clockwise hysteresis loops must arise on MFD scatter-plots.

Although most freeways systems lack sufficient route redundancy to guarantee an even distribution of congestion, freeway MFDs have been studied anyway. On the theoretical side, Cassidy et al. (2010) argues that if all the lanes of a freeway system are in the same congested or uncongested regime everywhere along their lengths, then the freeway system must exhibit a roughly triangular MFD. These single-regime conditions are stringent, however, and cannot be expected to hold for large systems. Indeed, well defined freeway MFDs have only been observed so far on systems with up to four ramps (Cassidy et al., 2010). Simulations of larger systems (Ji et al., 2010) show that portions of a freeway system often are congested while others are not, and the MFD is not well defined. Empirical data for even larger systems also reveal similar fragmentation and scatter (Endo et al., 2010), as well as large clockwise hysteresis loops as in the Toulouse street network (Geroliminis and Sun, 2010).

It is therefore valid to ask whether the forces that drive the distribution of congestion in street networks are also at work in freeway networks. The answer is obvious in some cases. For example, if all the traffic on a long homogeneous freeway flows through a single bottleneck, then this freeway's aggregate MFD data will exhibit clockwise hysteresis during any demand-driven episode of bottleneck queuing.[1] However, for freeways with distributed destinations the answer is not as

[1]The reason is that the queue that forms upstream of the bottleneck will be in a fixed congested state with flow equal to the bottleneck's capacity, while the freeway's uncongested upstream portion will have higher flows when the

obvious, especially when these systems are translationally symmetric.

Therefore, to look into this issue, the spatio-temporal distribution of congestion on a freeway of this type is studied below. The freeway is modeled with the kinematic wave theory of traffic flow (Lighthill and Whitham, 1955; Richards, 1956) using the boundary conditions for entrances and exits of the cell transmission model (Daganzo, 1995). This setup describes a freeway system more realistically than the 2-bin model. Furthermore, to isolate the phenomenon at hand the freeway will be a rotationally symmetric ring with even demand all around. This is the same system analyzed in a freeway gridlock study (Daganzo, 1996) which assumed congestion remains evenly distributed along the ring as gridlock develops.

It is shown below that while congestion does remain evenly distributed as it builds, it cannot stay even as it dissipates. Congestion must invariably fragment during this recovery phase. This fragmentation process reduces flow, retards full recovery and produces clockwise hysteresis loops, explaining their persistence in empirical data. More importantly, congestion's natural tendency toward fragmentation means that allowing a freeway beltway to become congested all around is more problematic than previously thought. Section 2 of the paper describes the behavior of the ring when it is in a single regime; sec. 2 the transitions between regimes and the fragmentation process; sec. 3 the effect of adaptive driver diversion to parallel routes; and sec. 4 some practical implications and conclusions.

2 Ring Performance in a Single Regime

Considered is a ring of length L on which traffic follows the kinematic wave (KW) theory with a triangular fundamental diagram as proposed in Newell (1993). The parameters of this model are labeled: v_f (free-flow speed), q_o (capacity), k_o (optimum density), κ (jam density) and $w > 0$ (backward wave speed). Ramps are uniformly distributed along the ring. Their flows are modeled as continuously distributed fluxes measured in units of flow per unit length of freeway (veh/hr-km). This continuum approximation is most accurate when individual ramps carry small flows compared with the freeway. The exit fluxes and influxes at particular locations are determined with the boundary conditions of the CTM (cell transmission model); see Daganzo (1995, 1996).

The demand influx from on-ramps, r (veh/hr-km), is assumed to be independent of location, x. The boundary conditions for on-ramps stipulate however that the actual influx is a fixed fraction of the freeway flow if the on-ramps are queued and the freeway is congested. This fixed fraction is denoted a (km^{-1}). Clearly then, although the demand is location-independent, the actual input flow will depend on x if the on-ramps are queued and the freeway is unevenly congested.

Off-ramps are assumed to have enough capacity to discharge everyone who wants to leave. It is also assumed that vehicles take exponentially distributed trips with average length l so that the fraction of traffic exiting per unit length of freeway, e, is the same everywhere: $e = 1/l$ (km^{-1}). The actual exit flux will therefore depend on x if the freeway flow is not the same everywhere.

It is now shown that the freeway's density converges or diverges from evenness over time only because the influxes and exit fluxes depend on x. Consider first the case where the freeway is in the congested regime, with density distribution $k(x,t)$ at time t. Then, if there were no inflows or outflows, and traffic flowed in the direction of increasing x, the kinematic wave relation would be:[2] $k(x, t+dt) = k(x+wdt, t)$. Now note that $[\kappa - k(x+wdt,t)]w$ is the circulating flow, and use this quantity to express the ramp fluxes of the CTM. The exit flux is simply $[\kappa - k(x+wdt,t)]we$; and

queue is growing (demand greater than capacity) than when it is receding (demand less than capacity).

[2] Since $x \in [0, L]$, all sums involving locations should be interpreted from now on as being "modulo L".

the influx is either r (if the ramps are not queued) or $[\kappa - k(x + wdt, t)]wa$ (if they are). Thus, the KW/CTM relation including all inflows and outflows is:

$$k(x, t + dt) = k(x + wdt, t) + rdt + [k(x + wdt, t) - \kappa]wedt, \quad \text{(without ramp queues)}, \tag{1a}$$
$$k(x, t + dt) = k(x + wdt, t) + [k(x + wdt, t) - \kappa]w(e - a)dt, \quad \text{(with ramp queues)}. \tag{1b}$$

Now, inspection of (1) shows that if the density was spatially even at $t = 0$, i.e., $k(x, 0) = \bar{k}(0)$, then it must remain even from then on. This is logical since the problem is rotationally symmetric. This means that $\bar{k}(t)$ satisfies (1) and that the set of equations corresponding to even conditions can be subtracted from the original. The result turns out to depend only on the density's deviations from evenness, $\epsilon(x, t) \doteq k(x, t) - \bar{k}(t)$, as follows:

$$\epsilon(x, t + dt) = \epsilon(x + wdt, t) + \epsilon(x + wdt, t)wedt, \quad \text{(without ramp queues)}, \tag{2a}$$
$$\epsilon(x, t + dt) = \epsilon(x + wdt, t) + \epsilon(x + wdt, t)w(e - a)dt, \quad \text{(with ramp queues)}. \tag{2b}$$

Consider now this system in a rigid and rotating frame of reference attached to observers moving with the backward wave; i.e., with velocity $-w$. Distances in the new frame are related to the old by the relation: $(z, t) = (x + wt, t)$. Now introduce this change of variable in (2a) and (2b). Note that the new distance coordinate is $z = x + wt + wdt$ everywhere in these expressions. Thus, the expressions in the rotating frame of reference can be written as:

$$\epsilon(z, t + dt) = \epsilon(z, t) + \epsilon(z, t)[we]dt, \quad \text{(without ramp queues)}, \tag{3a}$$
$$\epsilon(z, t + dt) = \epsilon(z, t) + \epsilon(z, t)[w(e - a)]dt, \quad \text{(with ramp queues)}, \tag{3b}$$

where z indexes a particular observer. Note the deviations seen by an observer do not depend on what is seen by other observers. Thus, if we define a case-specific constant C as follows,

$$C = [we] \quad \text{(freeway congested without ramp queues)}, \tag{4a}$$
$$= [w(e - a)] \quad \text{(freeway congested with ramp queues)}, \tag{4b}$$

then (3) reduces to the following ordinary differential equation for any given z:

$$d\epsilon/dt = C\epsilon. \tag{5}$$

This equation describes the evolution of the deviations from evenness seen by any of our observers. It is the equation of exponential growth or decay.

Decay occurs (for all the observers and therefore for the system as a whole) if $C < 0$. In this case, the densities seen by all observers continuously tend toward the same common value $\bar{k}(t)$ and the density distribution becomes increasingly even. We say then that the system is stable. Growth occurs if $C > 0$. In this case, the system is unstable. Disturbances double every $0.69/C$ hrs and the density distribution becomes increasingly uneven with the passage of time.

Note from (4) that instability is the rule rather than the exception. The only stable scenario arises in case (4b) when $a > e$; i.e., when congestion is increasing with on-ramp queues. These queued on-ramps act a a stabilizing servomechanism that constantly drives the system toward evenness when congestion is increasing. But in the late stages of a rush hour when congestion declines, the system is unstable.

If the freeway is uncongested the analysis leading from (1) to (5) can be repeated. The main differences are that the waves and the observers now move in the direction of increasing x at speed v_f and that the on-ramp input flows of the CTM are not affected by the freeway state. The final result of this analysis is still (5) with the following negative value for C:

$$C = [-v_f e] \qquad \text{(freeway uncongested).} \tag{6}$$

This means that an uncongested system naturally tends toward evenness when uncongested.

3 Transitions between Regimes, Fragmentation and Hysteresis

To understand how the ring transitions between regimes, an idealized rush period consisting of two phases (loading and recovery) is now considered. It is assumed that $e \approx 0$ in the loading phase and $(r, a) \approx 0$ in the recovery phase. The loading phase is assumed to last long enough to generate some congestion.

The loading phase occurs entirely in a single regime without any fragmentation because as (6) expresses, the system is stable while in the uncongested regime. This means that all portions of the ring must become congested at the same time when capacity is reached, and that multiple regimes cannot arise on the ring. Since this is the case, the system averages $(\bar{k}, \bar{q})$ must follow the triangular fundamental diagram and reach its capacity q_o.

As (4) shows, however, the loading process is unstable after entering the congested regime if the on-ramps are not queued. In this case, once congestion has started it increases everywhere at different rates and becomes increasingly uneven with the passage of time. This undesirable effect is not visible on the MFD, however. The effect does not arise if the on-ramps are queued; in this case the distribution of congestion is even at the end of the loading phase.

As we have seen, the recovery phase is always unstable and any disturbances grow exponentially. If the on-ramps are not queued during the loading phase, significant disturbances will be present at the start of the recovery phase, so it should not take much time until a small gap with optimum density appears somewhere. As is explained below, the size of this gap will increase with time as its average flow and density decline. Furthermore, since the disturbances in the congested portion of the freeway continue to grow, other gaps can appear within this congested portion, fragmenting it more.

Gaps grow because the flows at the two ends of any gap are different. At the upstream end the flow stays at q_o as the front of the receding freeway queue discharges. But flow is lower at the gap's downstream end because some of the traffic in the gap exits without bridging it. As a result, the gaps' downstream end moves back more slowly than its upstream end, and the gap grows. Clearly, the larger a gap the lower its downstream flow and the more rapid its growth. Thus, the average flow of any gap must decline in time, always remaining below q_o. These lower flows manifest themselves on the MFD plane as the return branch of a clockwise hysteresis loop.

It should be intuitive that the lower the freeway flow when fragmentation begins, the lower the maximum flow on this return branch; thus, it is important to postpone the start of fragmentation if at all possible. This effect is quantified below for a limiting case where the recovery involves only one gap. The effect implies that a system that is loaded while its on-ramps have queues will recover more quickly than a system that reaches the same maximum average density without on-ramp queues. This happens, even though both systems would follow the same loading paths on the MFD, because the on-ramp queue servomechanism smooths out the density. The desirability for density smoothness on ring freeways with distributed destinations also suggests that metering schemes that do not completely eliminate freeway congestion are most effective if the metering rates respond to freeway traffic with servomechanisms similar to those of the CTM.

3.1 A Limiting Case: Growth of One Gap

This subsection quantifies both, the flows achieved during recovery and the duration of the recovery for the special case where the recovery involves only one gap.

3.1.1 Recovery flows and hysteresis

Let $(\bar{q}, \bar{k}) = (Q, K)$ be the average flow and average density in the ring at the time when the gap first opens. The average flows in the gap and the ring are now evaluated as a function of the gap size, $G \leq L$ (km). It is assumed that Q and K do not change much while the gap is growing. This is reasonable if $K \approx \kappa$.

The flow in the gap at a distance $y \leq G$ from the upstream end is $q_y = q_o \exp(-ey)$ because of the intervening exits. Thus, the average flow in the gap is $\bar{q}_G = \int_0^G q_o \exp(-ey)dy/G$, which reduces to:

$$\bar{q}_G = q_o[1 - \exp(-eG)]/[eG]. \tag{7}$$

The ring's average flow is therefore $\bar{q} = [q_G G + Q(L - G)]/L$, which reduces to:

$$\bar{q} \;=\; q_o[1 - \exp(-eG)]/[eL] \;+\; Q(1 - G/L). \tag{8a}$$

Similarly, since the average density in the gap is $k_G = q_G/v_f$, the ring's average density, $\bar{k} = [k_G G + K(L - G)]/L$, reduces to:

$$\bar{k} \;=\; [q_o/v_f][1 - \exp(-eG)]/[eL] \;+\; K(1 - G/L). \tag{8b}$$

Equations (8) are the parametric expression of the recovery phase path described by the data on the MFD plane with G as the parameter.

Note that (8a) is concave in G with a well defined maximum in $[0, L]$. Simple manipulations reveal that this maximum is achieved for $G^* = \min\{L, \ln(q_o/Q)/e\}$. In the limiting gridlock case of $Q \approx 0$ the optimum gap is $G^* = L$, and the maximum recovery flow, $\bar{q}^*$, is:

$$\bar{q}^* = q_o[1 - \exp(-eL)]/[eL]. \tag{9}$$

Note this expression equals approximately $1/[eL]$ if $Le >> 1$; i.e., if the ring is several times longer than the typical trip, $l = 1/e$. So if for example the ring is 5 times longer than the typical trip, then the maximum flow reached while recovering from gridlock is only 1/5th of q_o. This illustrates that the maximum flows achieved while recovering from a serious episode of gridlock can be very low despite the complete rotational symmetry of the ring.

3.1.2 Duration of the Recovery Phase

The duration of the recovery phase can be roughly estimated from the growth rate of the single gap dG/dt. To estimate this rate note that the flow at the gap's downstream end is $q_G = q_o \exp(-eG)$, that the corresponding density is $k_G = q_G/v_f$ and that the front of the gap will advance on average at speed $U_G = (Q - q_G)/(K - k_G)$. Since the upstream end of the gap recedes at speed w, the gap grows at an average rate:

$$\frac{dG}{dt} = w + U_G \;=\; \frac{q_o}{\kappa - k_o} \;+\; \frac{Q - q_o \exp(-eG)}{K - k_o \exp(-eG)}. \tag{10}$$

This differential equation for $G(t)$ can be used to estimate the time t_o required to go from some initial gap $G = G_o$ to $G = L$. In the case of very severe congestion, where other gaps may not open wide enough to contribute much to the recovery, this is approximately the time it takes to achieve full recovery.

This time can be compared with the time t'_o the system would take to recover if the density would stay evenly distributed during recovery; i.e., if $k(x,t) = \bar{k}(t)$. Consideration shows that this idealized recovery would obey (5) with $C = -we$, and that therefore: $t'_o = -[\ln(Q/q_o)]/[we]$. On solving (10) one finds after a few manipulations that the ratio $r = t_o/t'_o$ is given by:

$$r = \frac{b}{-\ln(a)} \int_{\epsilon}^{1} [1 + \frac{a - E(z)}{1 - a + (w/v_f)(1 - E(z))}]^{-1} dz, \tag{11}$$

where $\epsilon = G_o/L, a = Q/q_o, b = eL$ and $E(z) = 1 - \exp(-bz)$.

Calculations show that r is typically several times greater than 1; i.e., that real recoveries are several times longer than those one would naïvely predict assuming spatial uniformity. This estimate assumes that other gaps do not open up, or that if they do they contribute little to the recovery. This is approximately what happens when the ring is very congested as illustrated in the next subsection. Otherwise, the congested region fragments and the congested episode ends before $G(t) = L$. Although in this case r is lower than predicted by (11), it is still greater than 1. These results illustrate that real recoveries should take considerably longer than one might naïvely assume and that, as a rule, eliminating congestion takes more time than creating it.

3.2 Simulation Visualization and Verification

The reader can verify that the effects described in this section also arise in a less idealized case with discrete vehicles and ramps, by accessing the interactive on-line gridlock simulation currently hosted at "*http://www.its.berkeley.edu/volvocenter/gridlock/*". As explained on that web page and the references cited therein, this simulation was built with a cellular automaton model that matches the KW/CTM model of this paper with the resolution of a single jam spacing. To see the effects of interest clearly, run the simulation at low speed by setting the "Sim Speed" slider at 10-20. This slider can be adjusted while the simulation is in progress.

To start, set the "Trip Length" slider at any intermediate value (say 65) and the metering "Interval" slider at 100. This will mimic the loading phase of a rush hour. Press "Play". Now note how the 4 segments of the ring remain evenly loaded during this loading phase, and that queues begin to form in all 4 segments nearly simultaneously. This confirms that the loading phase is stable and that the maximum average flow during loading is close to capacity. Next, pause the simulation when the whole ring is gridlocked, except for a single gap. This terminates the loading phase.

Now click on the "'inf" box to close the on-ramps. This marks the beginning of the recovery phase, as if the demand had vanished. Press "Play" again and then observe how the gap grows, and how the flow across it is initially at capacity. Note too how once the gap spans 1 or 2 ramps, the number of vehicles and the flow at its downstream end is noticeably reduced. In fact, toward the end of the recovery the downstream portions of the gap will often have no flow and no vehicles for extended periods. Clearly, the average flow on the ring never gets close to capacity, as occurred during the loading phase. Like the analytic model, the simulation exhibits strong clockwise hysteresis. Note as well how the ring is unevenly loaded throughout the recovery, which is the hallmark of instability. Finally, note that the recovery phase is much longer than the loading phase, and that this happens every time the simulation is run.

4 Adaptive Driver Navigation

Daganzo et al. (2010) and Gayah and Daganzo (2010) report that if drivers navigate urban areas adaptively, bypassing congested areas when the required detours are reasonable, then MFD data form a well defined curve into the congested regime up to a bifurcation point. Scatter and hysteresis only arise beyond this bifurcation point. It is shown here that driver adaptation also has some beneficial effect on our idealized ring.

To model driver adaptation it will be assumed that in addition to the normal exit flux, qe (veh/hr-km), there is an extra exit flux, $qP(k)$ (veh/hr-km), consisting of drivers who bail out when the density is $k > k_o$. The rate $P(k)$ (1/km) is assumed to be zero when $k = k_o$ and to increase linearly with density up to a value p (1/km) when $k = \kappa$; i.e., $P(k) = p((k - k_o)/(\kappa - k_o)$. Experience suggests that the coefficient p could be a few times greater than e in many cases.

The derivation of (5) is now repeated recognizing adaptation. It should be clear that (1) continues to hold when adaptation occurs if the exit rate e in (1) is replaced by the new exit rate including adaptation: $e_a = e + P(k)$. The only difference in the derivation of (2) is that because e_a depends on k, now terms of order ϵ^2 appear in the new version of (2). For small ϵ these terms can be neglected and a linear expression is obtained. This expression is identical to (2), with e replaced by the following "effective exit rate", e':

$$e' = e + p\frac{2\bar{k} - \kappa - k_o}{\kappa - k_o}. \tag{12}$$

The manipulations leading from (2) to (5) can now be repeated step by step. The final result is structurally similar to (4) and (5):

$$d\epsilon/dt = C'\epsilon, \text{ where } C' = [w(e' - a)] \qquad \text{(freeway congested with adaptation).} \tag{13}$$

It should be understood in (13) that $a = 0$ if there are no ramp queues.

The derivation of (6) does not have to be repeated because drivers do not adapt when the freeway is freely flowing. Thus, (6) continues to hold without modification. This result implies that if the system is uncongested it will behave as without adaptation and therefore will be stable.

If the system is congested, however, its behavior is different from the case without adaptation as (12) indicates. This expression shows that $e' < e$ if $\bar{k} < \frac{1}{2}(\kappa + k_o)$. This means that adaptation reduces instability if congestion is moderate, but increases it if it is heavy. Adaptation can also completely eliminate instability for a range of densities slightly exceeding k_o. To determine this range recall that (13) describes a stable system if $C' < 0$. For a congested system this condition is satisfied if $e' < a$. Consideration of (12) shows that this cannot happen for a congested system (i.e., where $\bar{k} > k_o$) if $p < e - a$. Otherwise, if $p \geq e - a$, there is a bifurcation density in the congested regime, k_B, for which $e' = a$, marking the difference between stability and instability:

$$k_B = k_o + \frac{1}{2}(\kappa - k_o)[1 + (a - e)/p], \quad \text{if } p \geq e - a. \tag{14}$$

For example, if $p = 2$ and $a = 0$, the bifurcation point is one quarter of the way down along the congested branch of the fundamental diagram.

5 Conclusion

This paper has shown with an idealized KW/CTM model that traffic on freeways without on-ramp queues is stable if traffic is uncongested, and unstable if it is congested. Stability means that all portions of the freeway reach saturation, k_o, at the same time while congestion grows. Instability means that flow/density disturbances grow exponentially and propagate with speed $-w$. Instability is most marked when congestion is dissipating, and in this case causes the following additional effects: (i) fragmented congestion; (ii) reduced flow; (iii) clockwise hysteresis loops on the MFD plane; and (iv) extended duration for a full recovery from an episode of severe congestion. These effects were quantified and verified with simulations. Driver adaptation was found to moderate instability upto a bifurcation point.

The analysis in this paper assumed a rotationally symmetric ring with even demand all around. Although in the real world demand is inhomogeneous, the same stability/instability phenomena are at work. The main difference is that inhomogeneous demand will cause different links of a freeway system to reach saturation, k_o, at different times, reducing the average flow as the system is loading. Aside from this, the real system should behave qualitatively as the idealized ring. In particular, if adaptation is not significant and there are no on-ramp queues, density should become increasingly uneven in the congested portions, and these should fragment during recovery. Fragmentation should further increase the variance of the density distribution, causing hysteresis and reducing flow. This phenomenon is clearly seen in Figure 3b of Geroliminis and Sun (2010), which shows with freeway network data from Minneapolis/St.Paul (Minnesota) that the variance of the density distribution is higher during recovery than during loading for the same average density.

If in other cities adaptive behavior is significant, hysteresis and scatter may be reduced. However, data from the Hanjin freeway in Osaka (Japan) (Endo et al., 2010) is definitely scattered, suggesting that adaptation is not a significant effect in that city. Thus, both the theoretical and real-world evidence so far suggest that well defined freeway MFD's should be the exception rather than the rule if the freeways are long. Driver adaptation, however, can probably be artificially encouraged with VMS signs that warn freeway drivers of congestion ahead. Perhaps, experiments should be conducted on congested freeways to see if VMS signs can increase their average flows.

Finally, we have seen that on-ramp queues on freeways with distributed destinations can act as a servomechanism that smooths out traffic when congestion is growing. We have also seen that for the same average density, smoother congestion dissipates faster with less vehicle delay. Thus, adaptive ramp metering strategies that mimic the effect of on-ramp queues (e.g., using the CTM merge rule) should smooth out congestion, speed up recovery and reduce delay on these types of freeways vis a vis ordinary ramp metering. They should probably be put to the test.

Acknowledgments

This research was supported by NSF Grant CMMI-0856193 and the UC Berkeley Center of Excellence for Future Urban Transport.

References

Ardekani, S. and Herman, R. (1987). Urban network-wide traffic variables and their relations. *Transportation Science*, 21(1):1–16.

Buisson, C. and Ladier, C. (2009). Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board*, 2124:127–136.

Cassidy, M. J., Jang, K., and Daganzo, C. F. (2010). Macroscopic fundamental diagrams for freeway networks: Theory and observation. In *Traffic Flow Theory Committee Summer Meeting, Annency, France*.

Daganzo, C. F. (1995). The cell transmission model. Part II: Network traffic. *Transportation Research Part B*, 29(2):79–93.

Daganzo, C. F. (1996). The nature of freeway gridlock and how to prevent it. In *Transportation and Traffic Theory (J.B. Lesort editor). Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pages 629–646. Elsevier.

Daganzo, C. F. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B*, 41(1):49–62.

Daganzo, C. F., Gayah, V. V., and Gonzales, E. J. (2010). Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. *Transportation Research Part B*, doi:10.1016/j.trb.2010.06.006.

Endo, K., Yoshii, T., and Fujii, S. (2010). Ramp metering strategies well fit to the characteristics of the network topology and traffic demands (in Japanese). In *Proceeding of Infrastructure Planning Conference, Japan Society of Civil Engineering, CD-ROM*, volume 41.

Gayah, V. V. and Daganzo, C. F. (2010). Exploring the effect of turning maneuvers and route choice on a simple network. *UC Berkeley, Volvo Working Paper UCB-ITS-VWP-2010-05.*

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B*, 42(9):759–770.

Geroliminis, N. and Sun, J. (2010). Hysteresis Phenomena of a Macroscopic Fundamental Diagram in Freeway Networks. *(Draft, submitted for publication).*

Godfrey, J. W. (1969). The mechanism of a road network. *Traffic Engineering and Control*, 11(7):323–327.

Herman, R. and Ardekani, S. (1984). Characterizing traffic conditions in urban areas. *Transportation Science*, 18(2):101–140.

Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S., and Quian, X. (2010). Macroscopic fundamental diagram: Investigating its shape using simulation data. In *Compendium of papers CD-ROM, Transportation Research Boarch 89th Annual Meeting, Washington DC*, number 10-3098.

Lighthill, M. J. and Whitham, G. B. (1955). On kinematic waves. I: Flow movement in long rivers. II: A Theory of traffic flow on long crowded roads. *Proceedings of the Royal Society*, A229(1178):281–345.

Mahmassani, H. S. and Peeta, S. (1993). Network performance under system optimal and user equilibrium dynamic assignments: Implications for advanced traveler information systems. *Transportation Research Record*, 1408:83–93.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic I: General theory. II: Queuing at freeway bottlenecks. III. Multi-destination flows. *Transportation Research Part B*, 27(4):281–314.

Olszewski, P., Fan, H., and Tan, Y. (1995). Area-wide traffic speed-flow model for the Singapore CBD. *Transportation Research Part A*, 29(4):273–281.

Richards, P. I. (1956). Shock waves on the highway. *Operations Research*, 4(1):42–51.