Symmetry breaking in optimal timing of traffic signals on an idealized two-way street

Mark J. Panaggio,[1,*] Bertand J. Ottino-Löffler,[2] Peiguang Hu,[3] and Daniel M. Abrams[1,4]

[1]*Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60208, USA*

[2]*Department of Mathematics, Institute of Technology, Pasadena, California 91126, USA*

[3]*Engineering Science Program, National University of Singapore, Singapore*

[4]*Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, USA*

(Received 10 June 2013; published xxxxx)

Simple physical models based on fluid mechanics have long been used to understand the flow of vehicular traffic on freeways; analytically tractable models of flow on an urban grid, however, have not been as extensively explored. In an ideal world, traffic signals would be timed such that consecutive lights turned green just as vehicles arrived, eliminating the need to stop at each block. Unfortunately, this "green-wave" scenario is generally unworkable due to frustration imposed by competing demands of traffic moving in different directions. Until now this has typically been resolved by numerical simulation and optimization. Here, we develop a theory for the flow in an idealized system consisting of a long two-way road with periodic intersections. We show that optimal signal timing can be understood analytically and that there are counterintuitive asymmetric solutions to this signal coordination problem. We further explore how these theoretical solutions degrade as traffic conditions vary and automotive density increases.

DOI: 10.1103/PhysRevE.00.002800 PACS number(s): 89.40.Bb, 89.75.−k, 64.60.Cn

I. INTRODUCTION

The physics of traffic flow has been studied for more than half a century [1–7]. On freeways, traffic has been successfully modeled as a nonlinear fluid, making analytical solution possible [2]. On urban grids, however, the nonlinear effects of timed traffic signals make most models analytically intractable [8–12].

The inefficient timing of traffic signals is responsible for up to 10% of traffic delays [13]. These delays cause commuters to waste dozens of hours in traffic each year, leading to billions of dollars in wasted fuel and a large environmental cost [14]. Coordination between traffic signals has proven to be a cost-effective way to reduce these delays dramatically [13].

Signal timing schemes fall into two categories: real time and pretimed [3,15]. Real-time schemes make adaptive use of information about traffic density and localized conditions to trigger light cycle changes [16]. Unfortunately, this information is not readily available at most intersections, and installing the necessary detectors can be prohibitively expensive [17].

Pretimed schemes employ detailed computer simulation and heuristic optimization tools such as genetic algorithms [18–20] to search for optimal timings [17,21]. Once a scheme is generated, it can be relatively inexpensive to implement, but generating such a scheme requires computational resources that are often beyond the capacity of local government. Where traffic demands fluctuate, these schemes can quickly become outdated, so there is no guarantee that they will be optimal by the time they are implemented [16].

II. MOTIVATION

Many theoretical questions about optimal traffic signal timing remain unanswered. For a single one-way street, the best solution is a so-called *green wave* [4,22], in which vehicles leaving a light at the instant it turns green arrive at all subsequent lights at the instant they turn green [8]. In theory, this means that vehicles traveling at the speed limit will never stop at a red light, although in practice this fails when traffic density exceeds a "jamming threshold" [12].

It is impossible to achieve a bidirectional green wave on an arbitrary two-way street due to the inherent frustration of competing demands in each direction [23]. Past theoretical work has focused on maximizing the "bandwidth" [4,5,16,21,24]—the interval of time in which vehicles can progress through all traffic signals without stopping—of a finite segment of road. There are several drawbacks to this approach. First of all, bandwidth is not a direct measure of efficiency, so the solution that maximizes bandwidth may not minimize total trip time, stops, or delay [8]. Second, for long roads, nonzero bandwidth is often unachievable in one direction, making this approach incapable of improving upon one-way schemes without arbitrarily dividing the road into subsections.

III. OUR MODEL

We consider a highly simplified model of an infinite two-way street with traffic lights along the entire length [25]. We fix the spacing between consecutive lights Δx and set $x_n = n\Delta x$, where x_n denotes the position of light n.

We let $\phi_n(t)$ denote the phase of light n at time t, taking light n to be green when

$$2N\pi \leqslant \phi_n(t) < (2N+1)\pi$$

and red when

$$(2N+1)\pi \leqslant \phi_n(t) < (2N+2)\pi,$$

where N is any integer. Note that we ignore the yellow portion of the cycle and assume that the green time is half of the light cycle.

Since the geometry of the system is invariant under translations of integer multiples of the block length ($x \mapsto x + N\Delta x$),

*markpanaggio2014@u.northwestern.edu

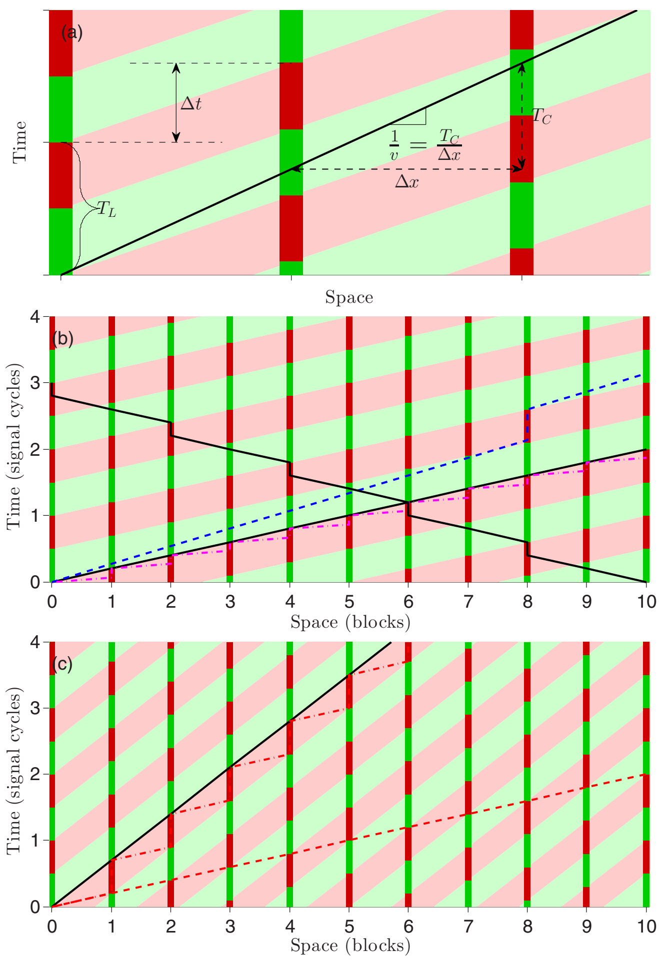

FIG. 1. (Color online) Space-time diagram for traffic flow. The red (dark gray) bars indicate locations and times for red lights and the green (light gray) bars indicate green lights. (a) illustrates the definitions of various variables and parameters in our model. (b) displays various vehicle trajectories. The dashed line (blue) indicates a car traveling slower than the green-wave speed v_g, the dash-dotted line (magenta) indicates a car traveling faster than v_g, and solid lines (black) indicate cars traveling exactly at v_g both eastbound (starting on the left) and westbound (starting on the right). (c) displays vehicle trajectories near the green- and red-wave speeds. The solid line (black) indicates a car traveling at the green-wave speed. The dashed line (red) indicates a vehicle traveling slightly faster than the red-wave speed. The dash-dotted line (red) indicates a vehicle traveling slightly more slowly than the red-wave speed.

we will look for optimal timing schemes that are also invariant under such translations. We therefore set

$$\phi_n(t) = \frac{2\pi}{T_L}(t - n\Delta t), \tag{1}$$

where T_L is the period of the light cycle and Δt is the time offset between consecutive lights (see Fig. 1) with $0 \leqslant \Delta t < T_L$. We set $\phi_0(0) = 0$ without loss of generality.

Consider a single vehicle starting at $x = 0$ and traveling eastbound on this street with constant velocity $v = \Delta x / T_C$, where T_C is the time for a car to travel one block. When the vehicle arrives at a red light, it stops until the light turns green, and then repeats the process (for simplicity we ignore acceleration and assume drivers react instantaneously). From the perspective of the vehicle, at the moment this light turns green, the relative light phases are identical to the initial state. Thus, the speed will be periodic.

The car's effective speed v_{eff}—its average speed as $t \to \infty$—is determined by the fraction of time spent waiting at red lights, suggesting that an appropriate metric for efficiency is $E = v_{\text{eff}}/v$.

We refer to a single cycle in which a vehicle passes through N_L lights before stopping and waiting for a time W as a "trip." During a single trip, a vehicle travels a distance of $\Delta x N_L$ in a total time $T_C N_L + W$. The vehicle arrives at light n at time nT_C, and thus N_L will be the smallest positive integer satisfying

$$(N - 1/2)T_L + N_L \Delta t \leqslant N_L T_C < N T_L + N_L \Delta t \tag{2}$$

for some integer N.

Without loss of generality, we can eliminate one of the three free parameters (T_L, T_C, and Δt) above by defining the ratios $r_C = \frac{T_C}{T_L}$ and $r_\Delta = \frac{\Delta t}{T_L}$, so Eq. (2) becomes

$$(N - 1/2) \leqslant N_L (r_C - r_\Delta) < N. \tag{3}$$

It is straightforward to show that $N_L = \left\lceil \frac{1}{2\{r_C - r_\Delta\}} \right\rceil$ and $N = \lceil N_L(r_C - r_\Delta) \rceil$ satisfy Eq. (3) [26], where $\lceil x \rceil$ and $\lfloor x \rfloor$ denote the standard ceiling and floor functions and $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of x modulo 1.

The waiting time W may also be computed: the car stops at time $N_L T_C$ and begins to move again at time $N T_L + N_L \Delta t = \lceil N_L(r_C - r_\Delta) \rceil T_L + N_L \Delta t$, so

$$W = \lceil N_L(r_C - r_\Delta) \rceil T_L + N_L \Delta t - N_L T_C. \tag{4}$$

Thus the efficiency for eastbound traffic can be expressed as

$$\begin{aligned} E_{\text{east}}(r_\Delta, r_C) &= \frac{r_C N_L}{\lceil N_L (r_C - r_\Delta) \rceil + r_\Delta N_L} \\ &= \frac{r_C \left\lceil \frac{1}{2\{r_C - r_\Delta\}} \right\rceil}{\left\lceil \left\lceil \frac{1}{2\{r_C - r_\Delta\}} \right\rceil (r_C - r_\Delta) \right\rceil + r_\Delta \left\lceil \frac{1}{2\{r_C - r_\Delta\}} \right\rceil}. \end{aligned} \tag{5}$$

The efficiency depends on only two parameters, $r_C > 0$ and $0 \leqslant r_\Delta < 1$. It is bounded between 0 and 1, and decreases monotonically with r_Δ except at discontinuities. It reaches a global maximum of 1 at $r_\Delta = r_C$, which represents a green wave, and immediately after a discontinuity at $r_\Delta = r_C + 1/2$, which we refer to as a red wave. A red wave occurs when vehicles arrive at the instant each light turns red. In this worst-case scenario ($r_\Delta = r_C + 1/2$), vehicles travel for T_C seconds and then wait at a red light for the full red time $T_L/2$. This represents the global minimum. A vehicle traveling slightly faster than the red wave ($r_C < r_\Delta - 1/2$), on the other hand, will arrive at the instant before each light changes. As a result, all wait times are infinitesimal, and the efficiency is close to 1. Sample trajectories for vehicles near the green- and red-wave speeds are displayed in Fig. 1. Cross sections of E_{east} for fixed r_C are shown in green in Fig. 2 [26].

For westbound traffic, the time delay between consecutive signals is not Δt but rather $T_L - \Delta t$, and thus the efficiency for westbound traffic is simply $E_{\text{west}}(r_\Delta, r_C) = E_{\text{east}}(1 - r_\Delta, r_C)$, the reflection of the function about $r_\Delta = 0.5$.

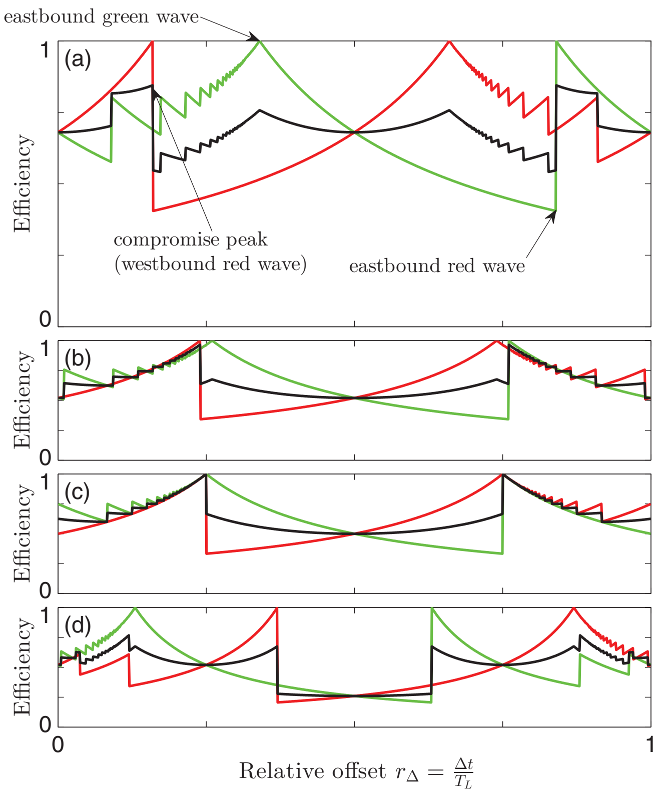

FIG. 2. (Color online) Theoretical efficiency versus r_Δ for (a) $r_C = 0.34$, (b) $r_C = 0.26$, (c) $r_C = 0.25$, and (d) $r_C = 0.13$. Green (light gray) indicates $E_{\rm east}$; red (dark gray) indicates $E_{\rm west}$; and black indicates $E_{\rm tot}$ assuming equal demand in both directions.

On a two-way street, we wish to maximize the weighted average efficiency

$$E_{\rm tot} = w_{\rm e} E_{\rm east} + w_{\rm w} E_{\rm west}, \tag{6}$$

with weights $w_{\rm e}$ and $w_{\rm w}$ dependent on the traffic volumes in each direction. For simplicity we will consider the case of symmetric demand, $w_{\rm e} = w_{\rm w} = 1/2$.

Given T_C as dictated by safety considerations and holding T_L constant, r_C is fixed, and we attempt to choose r_Δ (equivalent to choosing the offset Δt) to maximize efficiency. This simultaneously maximizes the effective velocity $v_{\rm eff}$ and minimizes the total wait time.

With equal demand in both directions, it might seem that the symmetry of the problem suggests a symmetrical optimum, i.e., $r_\Delta = 0$ or $r_\Delta = 1/2$ [22,23]. These are indeed local extrema, but usually not maxima. Another reasonable hypothesis is that the bidirectional optimum will coincide with the optimum in one direction, a green wave [9,22,23]. This is a local maximum but not necessarily the global maximum. Surprisingly, the global optimum instead often occurs when both directions are suboptimal but one direction is favored over the other. This is possible because small perturbations in r_Δ can cause dramatic shifts from local efficiency minima to local maxima near discontinuities in $E_{\rm east}$ or $E_{\rm west}$ (note that the green-wave peak is not discontinuous). When, e.g., a discontinuous peak in $E_{\rm west}$ occurs near the green-wave peak for $E_{\rm east}$, the loss of efficiency by perturbation off the eastbound green wave is offset by gains in the westbound efficiency. As a

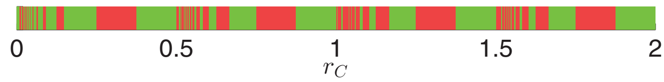

FIG. 3. (Color online) Optimality of the green wave. Intervals of r_C for which the green wave optimizes the bidirectional efficiency are indicated by green (light gray) rectangles. Intervals of r_C for which other timings are optimal are indicated by red (dark gray) rectangles.

result, green-wave timings fail to be optimal for various ranges of r_C (see Fig. 3).

IV. SOME LIMITATIONS

While the efficiency metric in Eq. (6) provides some insight into the ideal signal timing for a two-way street, it has a number of limitations. First of all, it applies only to a single vehicle. In practice, vehicles often travel in groups known as platoons [24,27]. The jagged efficiency peaks described by Eq. (5) and displayed in Fig. 2 may not be achievable by an entire platoon of vehicles. The theory also assumes identical non-interacting cars with constant speeds and perfectly uniform light spacing. In practice, city blocks vary in length even in well-planned urban grids and driver behavior varies. Additionally, the interactions between vehicles can play a significant role in exacerbating congestion [28].

To test the predictions of our model and verify that they are relevant when these assumptions are relaxed, we simulated the flow of vehicles on a street with 50 periodically placed traffic lights. We imposed periodic boundary conditions to avoid arbitrarily specifying entrance and exit rates. Instead, cars were randomly placed along the street according to a specified density ρ representing the fraction of the system occupied by vehicles of a finite length (1/25 of the block length for the results displayed in Figs. 4 and 5), and the total number of vehicles in the system was conserved. The trips of these vehicles were simulated during 30 light cycles. In the simulation, vehicles were prevented from passing each other. This caused queues to form at red lights as one might expect. Simulations were repeated with unevenly spaced lights and variable vehicle speed; results can be found in Fig. 4.

V. SIMULATION RESULTS

When the density is less than one vehicle per block, the simulations are indistinguishable from theory. At low to moderate vehicle densities, the computed efficiency remains well approximated by the theory and non-green-wave optima persist [see Fig. 4(a)]. At moderate densities, the efficiency near discontinuous peaks degrades noticeably while the green wave remains highly efficient. Thus a perfect green wave in either direction is optimal for moderate densities. At very high densities, gridlock, the scenario where vehicles at green lights are unable to advance due to the queue ahead of them, becomes a significant issue and the efficiency of all timings degrades. In our model, the only way to avoid gridlock is to set $r_\Delta = 0$ and have all lights change in unison.

The middle and bottom panels of Fig. 4 show the efficiency when vehicle speed varies [Fig. 4(b)] and when the light spacing varies [Fig. 4(c)]. Variation in the light spacing with proportionate variation in the offsets can actually improve

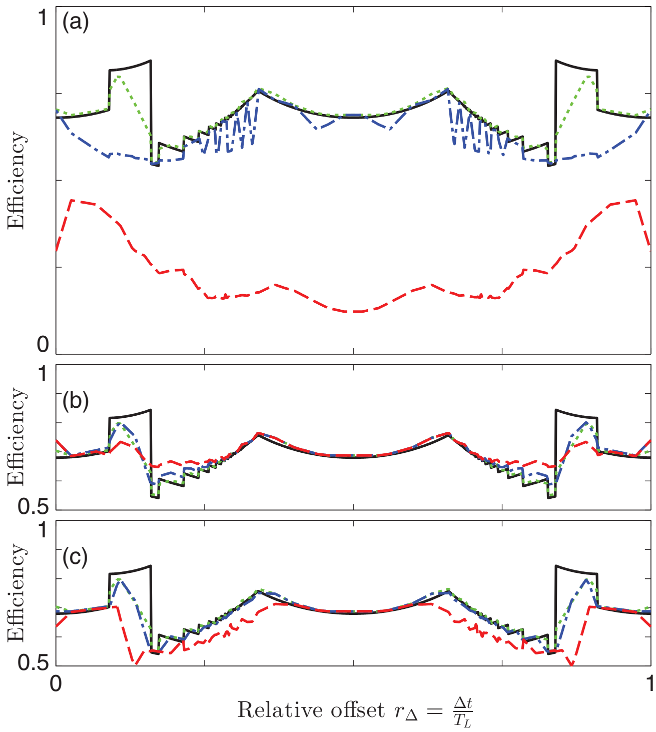

FIG. 4. (Color online) Efficiency from simulation versus r_Δ for $r_C = 0.34$. (a) shows the efficiency for increasing vehicle density: the solid curve (black) indicates the theoretical efficiency ($E_{\rm tot}$); the dotted (green), dash-dotted (blue), and dashed (red) curves represent simulation results for vehicle densities of 10%, 50%, and 90% density, respectively. The bottom panels display the effects of variation in the traffic signal spacing (b) and vehicle speed (c) on the efficiency with 10% traffic density. Solid (black) indicates the theoretical efficiency; the dotted (green), dash-dotted (blue), and dashed (red) curves represent simulation results for 0%, 1%, and 5% standard deviation, respectively.

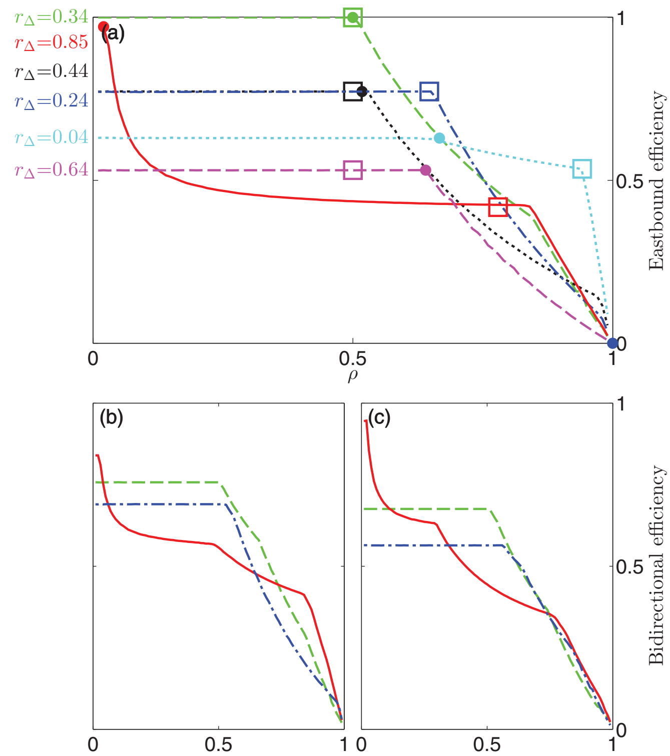

FIG. 5. (Color online) Predictions and simulations of various jamming transitions. (a) displays the eastbound efficiency for various values of r_Δ with $r_C = 0.34$. Of particular interest are the green curve (dashed), which represents an eastbound green wave, and the red curve (solid), which is near an eastbound red wave. The markers correspond to the critical densities for platoon segmentation (circle) and platoon coalescence (square). (b) and (c) show the bidirectional (weighted average of east- and westbound) efficiency for $r_C = 0.34$ and $r_C = 0.26$, respectively, with curves representing green-wave peaks (green dashed curve), discontinuous peaks (red solid curve), and suboptimal timings (blue dash-dotted curve).

efficiency for some ranges of r_Δ. This is reasonable given that lights that are close together behave like a single light and lights that are far apart have smaller wait times relative to the travel times. Variation in the vehicle speed has a smoothing effect on the discontinuities in the efficiency curve. Both of these factors degrade the efficiency in a smooth way, allowing the discontinuous optima to persist when the variation is small. Thus the theoretical predictions are "structurally stable." This feature of the model suggests that the predictions may indeed have value even in real-world systems with nonideal behavior.

VI. DENSITY EFFECTS

To explore the effects of vehicle density in greater detail, we computed the efficiency for fixed (r_Δ,r_C) and increasing ρ. Below a critical density, which we refer to as the "jamming threshold" [29], the predictions of Eq. (6) give a good approximation for the efficiency. Above this threshold, the efficiency degrades, and the theory no longer approximates the observed behavior. For a range of physically relevant parameters the critical density is above 50% of the capacity of the road. Near some discontinuous peaks, however, the critical density is small and few vehicles are able to perform at the level indicated by the theory. This is due to the narrow bandwidth corresponding to these timings [26]. Nonetheless, timings near discontinuous peaks in the bidirectional efficiency can remain optimal for a range of densities beyond the threshold.

The degradation of the quality of the theoretical predictions is due to the assumption that vehicles are noninteracting. Above the jamming threshold, the interactions between vehicles cause delays that the theory ignores. In simulations, vehicles initially clump together forming platoons. These platoons can interact either by coalescing to form even larger platoons or by being segmented at red lights. We can estimate the critical density corresponding to the jamming threshold analytically by deriving the conditions under which this coalescence and segmentation occur at steady state [26]. These predictions are displayed along with the numerical results in Fig. 5.

VII. CONCLUSIONS

In the mid-20th century physicists and engineers began taking an analytical approach to traffic management. Theoretical work has since proceeded along several lines, but we believe that there is still insight to be gained from simple solvable models.

Here we have presented an analysis of traffic flow on an urban arterial road with periodic traffic signals. Our approach allows analytical prediction of optimal signal timing that agrees well with numerical simulations and approximates the behavior of the system even when idealizing assumptions are relaxed. It yields an efficiency metric that can be expressed and computed analytically, yet it reproduces features observed in more complex models—features such as platoon formation [25], discontinuous efficiency curves [23,28], irregular flow patterns [9], and phase transitions due to jamming [1,28,29].

This work provides a theoretical framework for understanding the effects of signal timing on the efficiency of traffic flow. The insight gained from our simple model could help motivate the design of more efficient coordinated traffic signal timing programs on long arterials, particularly during periods of low to moderate traffic demand. Our analysis suggests that timing schemes other than the traditional green-wave approach may be optimal under certain circumstances, and merit further exploration with realistic simulations of complex driver behavior. Our methods could also be used to generate "smart" initial guesses for numerical optimization schemes with more complex efficiency metrics, or, alternatively, our efficiency metric (6) could be modified to apply to arbitrary networks of one- and two-way streets, perhaps allowing exact rather than approximate optimization and yielding more intuitive understanding of results.

ACKNOWLEDGMENTS

The authors thank H. Mahmassani and A. Motter for useful discussions. Research was supported in part by a NICO seed grant.

[1] M. J. Lighthill and G. B. Whitham, Proc. R. Soc. London, Ser. A **229**, 317 (1955).
[2] P. I. Richards, Oper. Res. **4**, 42 (1956).
[3] F. V. Webster, Road Res. Lab. Tech. Papers Report No. 39 (1958).
[4] J. T. Morgan and J. D. C. Little, Oper. Res. **12**, 896 (1964).
[5] G. F. Newell, ITTE Research Report 1967 (unpublished).
[6] D. Chowdhury, L. Santen, and A. Schadschneider, Phys. Rep. **329**, 199 (2000).
[7] B. S. Kerner, *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*, Springer Complexity (Springer, Berlin, 2004).
[8] E. Bavarez and G. F. Newell, Transport. Sci. **1**, 55 (1967).
[9] B. A. Toledo, V. Muñoz, J. Rogan, C. Tenreiro, and J. A. Valdivia, Phys. Rev. E **70**, 016107 (2004).
[10] C. Gershenson, Complex Syst. **16**, 02 (2005).
[11] S. Lämmer and D. Helbing, J. Stat. Mech.: Theory Exp. (2008) P04019.
[12] B. S. Kerner, Europhys. Lett. **102**, 28010 (2013).
[13] National Transportation Operations Coalition (NTOC) Technical Report, 2012 (unpublished), http://www.ite.org/reportcard/TechnicalReport.pdf.
[14] D. Schrank, B. Eisele, and T. Lomax, Texas A&M Transportation Institute (TTI) Technical Report, 2012 (unpublished), http://tti.tamu.edu/documents/mobility-report-2012.pdf.
[15] P. Koonce *et al.*, *Traffic Signal Timing Manual Publication*, Report No. FHWA-HOP-08-024 (U.S. Department of Transportation Federal Highway Administration, Washingtom, DC, 2008).
[16] D. I. Robertson and R. D. Bretherton, IEEE Trans. Veh. Technol. **40**, 11 (1991).
[17] Y. Yin, Transp. Res. B: Methodol. **42**, 911 (2008).
[18] A. S. Abdelfatah and H. S. Mahmassani, Transp. Res. Rec. **1645**, 185 (1998).
[19] G. Abu-Lebdeh and R. F. Benekohal, Transp. Res. Rec. **1727**, 61 (2000).
[20] H. Ceylan and M. G. Bell, Transp. Res. B: Methodol. **38**, 329 (2004).
[21] D. Robertson, Road Res. Lab. Report No. LR253, 1969 (unpublished).
[22] E. Brockfeld, R. Barlovic, A. Schadschneider, and M. Schreckenberg, Phys. Rev. E **64**, 056132 (2001).
[23] B. A. Toledo, E. Cerda, J. Rogan, V. Muñoz, C. Tenreiro, R. Zarama, and J. A. Valdivia, Phys. Rev. E **75**, 026108 (2007).
[24] N. H. Gartner, S. F. Assman, F. Lasaga, and D. L. Hou, Transp. Res. B: Methodol. **25**, 55 (1991).
[25] T. Nagatani, Physica A **368**, 560 (2006).
[26] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.xx.xxxxxx for the details on solving Eq. (3), for details about the discontinuities of E_{east}, for the derivation of the bandwidth, for the equation and derivation of the jamming threshold, and for videos of the platoon dynamics.
[27] G. F. Newell, *Traffic Flow on Transportation Networks (Transportation Studies)* (The MIT Press, Cambridge, MA, 1980).
[28] D.-w. Huang and W.-n. Huang, Phys. Rev. E **67**, 056124 (2003).
[29] M. Sasaki and T. Nagatani, Physica A **325**, 531 (2003).