Emergent Phenomena in Vehicular Traffic

Geoffrey Poore

5 May 2006

Traffic jams and other features observed in vehicular traffic are examples of emergent phenomena. The current understanding of vehicular traffic is reviewed, with special emphasis on traffic jams. Empirical studies have revealed the presence of multiple phases in traffic, and have shown that traffic jams may form spontaneously. Several methods from physics have been applied to modeling traffic, and the general empirical features have been reproduced.

I. Introduction

Why study traffic? From an individual's perspective, anything about traffic that can be understood and controlled will be of great value. Between 1982 and 2001, the amount of time Americans spent in traffic increased by 236%. During the same time period, the length of rush hour doubled in many major cities, so that many commuters now spend almost a full workweek each year stuck in traffic [1]. This is not only an incredible cost in time, but also an increasingly great cost in actual money, as gas prices continue to climb [2].

Historically, the study of traffic paralleled the adoption of automobiles and the development of a highway system. In the 1930s, Bruce D. Greenshields of the Yale Bureau of Highway Traffic began to study models relating speed and density, and to investigate intersection efficiency. After World War II, increasing automobile use and an expanding highway system prompted increased traffic study [3]. In the 1950s, traffic volumes in excess of carrying capacities were already a problem [4], and this trend has continued, so that congestion is estimated to cost Americans over $78 billion a year in fuel and time [1]. As a result, interest in understanding and controlling traffic has been persistent.

While some physicists contributed to the study of traffic before 1990, the majority of their contributions have been since the early 1990s. Once they began working on traffic, they produced a deluge of experiments, theories, and papers [4]. Vehicular traffic has been known to exhibit phases for some time, and these have been successfully modeled using physics methods [5].

Here, I shall try to present a general overview of traffic, as currently studied and understood by physicists. In Section II, I consider in some detail experimental techniques and what they have revealed about real traffic. In Section III, I summarize various attempts to model traffic and discuss their success.

II. Empirical results

Comparison with empirical results is important in the study of traffic just as it is in other systems investigated by physicists. However, empirical study of traffic involves complications often not present in other systems. Controlled experiments are difficult or perhaps impossible to perform. In addition, there is the problem of human experimental subjects; under ordinary circumstances, experimentalists cannot simply go into the field and attempt to create a traffic jam. As a result, empirical methods are generally limited to passive observation [5].

A variety of observational techniques have been developed; these are considered in Section II.B. The empirical results are considered in Sections II.C and II.D. First, however, it is useful to briefly mention the parameters that should be measured.

A. Parameters

A number of traffic parameters are potentially of interest. These include the following [6]:

- Rates of flow (vehicles per unit time);
- Speeds (distance per unit time);
- Travel time over a known length of road;
- Occupancy (percentage of time that a point on the road is occupied by vehicles);
- Density (vehicles per unit distance);
- Time headway between vehicles (time per vehicle);
- Spacing between vehicles (distance per vehicle); and
- Concentration (measured by density or occupancy).

Though these parameters are most frequently measured, this list is by no means inclusive. For example, in some situations driver reaction times or acceleration/deceleration rates may be of interest [7]. In addition, external parameters such as weather may be relevant [8].

B. Observational techniques

A variety of observational techniques have been developed for studying traffic. These provide a specific context for mathematical definitions of the parameters described above.

Early traffic data collection relied on hand tallies or on pneumatic tubes placed across the road [6]. Today, data is collected with considerably more sophisticated methods. Photography or video recording can provide data for all vehicles along a stretch of road, allowing the tracking of many trajectories [4]. A more limited data set may be obtained by equipping some vehicles with measurement devices, such as differential GPS receivers [7]. However, most data are collected by detectors located along the road, primarily induction loops. Two closely spaced loops provide time and time difference measurements from which flow, velocity, and other quantities may be derived [5].

Since induction loops are so widely used, it is worth considering in greater detail how their data is analyzed and how some of the parameters mentioned above may be calculated, closely following the discussion in [4]. Consider first the case of a single induction loop, which can only measure the times at which a vehicle reaches and leaves it. Label vehicles with the index n, and define t_n^0 and t_n^1 as the times when the n^{th} vehicle reaches and leaves the detector, respectively. Then the time between vehicles, or *time headway* (gross or brutto time separation), is

$$\Delta t_n^{00} = t_n^0 - t_{n-1}^0 , \qquad (1)$$

while the time between vehicle *n–1* leaving and *n* arriving, or *time clearance* (netto time separation), is

$$\Delta t_n^{01} = t_n^0 - t_{n-1}^1 . \qquad (2)$$

If ΔN vehicles cross the detector during a time interval ΔT, then the *occupancy* (ratio of time that a point is occupied by vehicles) is

$$O(x,t) = \sum_n (t_n^1 - t_n^0) \big/ \Delta T , \qquad (3)$$

while the vehicle flow during this time is

$$Q(x,t) = \Delta N \big/ \Delta T . \qquad (4)$$

When a second loop is added near the first, vehicle velocities v_n and lengths l_n may be estimated, using the approximation that velocity is constant between the loops. This allows the calculation of the distance from the front of one vehicle to the front of the next, or *headway* (brutto distance),

$$d_n^{00} = v_n \Delta t_n^{00} \qquad (5)$$

as well as the distance between vehicles, or *clearance* (netto distance),

$$d_n^{01} = d_n^{00} - l_{n-1} = v_n \Delta t_n^{01} . \qquad (6)$$

Knowing the v_n also makes possible the calculation of the average velocity

$$V(x,t) = \langle v_n \rangle = \sum_{n=1}^{\Delta N} v_n \big/ \Delta N . \qquad (7)$$

This in turn allows the calculation of the vehicle density, for example,

$$\rho(x,t) = Q(x,t) \big/ V(x,t) , \qquad (8)$$

among other methods. It is important to note that calculating the density from induction loop data may introduce errors unless care is taken. Q averages over time at one loop, but ρ should be an average over space; spatial and temporal averages are being mixed. Thus, some have preferred to calculate ρ with equations that give greater weight to small velocities, rather than using equations such as (8).

C. General empirical results: the fundamental diagram

Functional relations between vehicle flow Q, average velocity V, vehicle density ρ, and occupancy O have been measured since the beginning of traffic study. The relation between vehicle flow and density has been the most important, so that the plot of flow and density is know as the fundamental diagram [4].

A typical fundamental diagram is reproduced in Figure 1, below (note that this figure uses J, rather than Q, for flow). Three phases of traffic flow are generally distinguished in the diagram: (1) *free flow*, the almost linear relation beginning at the origin and labeled F; (2) *synchronized flow*, the leftward part of the area labeled J; and, (3) *wide moving jams*, the rightward part of the area labeled J [5]. However, it is important to note that there is still debate regarding the exact nature of traffic at higher densities in region J [8].

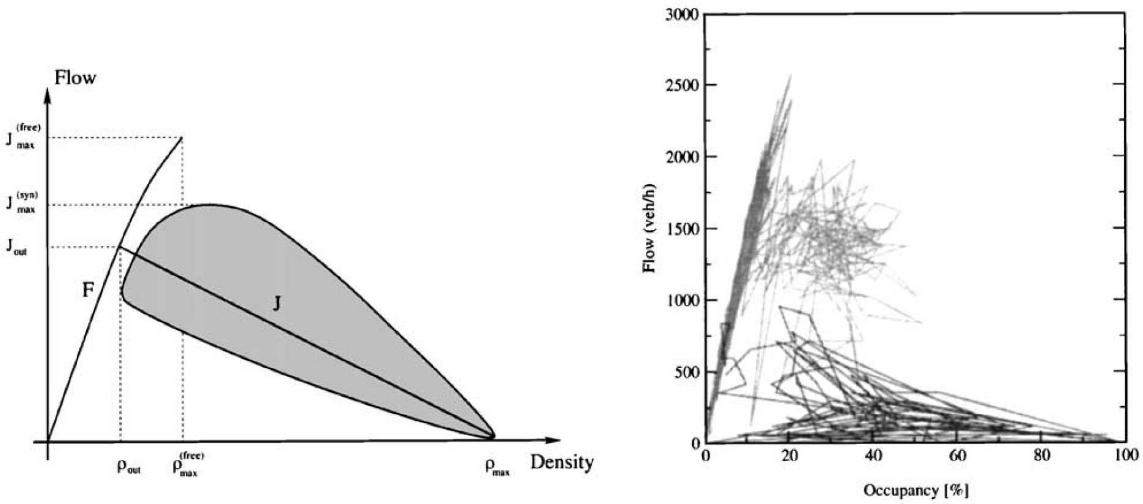

Figure 1. Left: Schematical form of the fundamental diagram. *F* denotes the free flow branch and line *J* is determined by the properties of wide moving jams. Right: Empirical fundamental diagram. One can clearly distinguish the three phases. The differences to the schematical form are mainly due to the use of local measurements for the determination of the empirical fundamental diagram (from [5]).

In the free flow phase, labeled by *F* in the figure, the density is low enough that vehicle interactions are negligible. Each vehicle can travel with its desired velocity, so flow increases linearly with density. Note that for flow greater than J_{out}, the flow is not uniquely determined by density [5].

All states that are not free flow states are known as *congested* states. Two congested phases have been observed. The wide moving jam phase, the rightward end of the shaded area *J* in the figure, is characterized by a sequence of wide jams (wide jams are those in which the width of the jam is much greater than the widths of the fronts, the zones at the edges of the jam where speeds change rapidly). Within the jams, density is high while velocity and flow are very low. The jams move upstream (against traffic) with a characteristic velocity v_{jam}, which has been measured to be ~15km/h (~9mi/h), and their outflow is independent of their inflow. They can travel through free flow and synchronized flow without being disturbed. Density is difficult to determine for the reasons mentioned previously, so that it is underestimated (this is why, in the right side of the figure, the moving jam phase is almost a flat line) [5].

The synchronized flow phase, the leftward end of the shaded area *J* in the figure, is characterized by congested traffic that is not composed of wide jams. The average velocity is much lower than in free flow, but the flow is larger than in wide jams. There is no functional flow-density relation; as may be seen in Figure 1, the data points are scattered over a wide region. The time series for lanes are highly correlated for synchronized flow [5].

The three phases are not only distinguished by the fundamental diagram. They may also be differentiated using a correlation function. The cross-correlation function

$$cc_{\rho,Q}(\tau) = \frac{\langle \rho(t)Q(t+\tau)\rangle - \langle \rho(t)\rangle\langle Q(t+\tau)\rangle}{\sqrt{\langle \rho^2(t)\rangle - \langle \rho(t)\rangle^2}\sqrt{\langle Q^2(t+\tau)\rangle - \langle Q(t+\tau)\rangle^2}} \quad (9)$$

measures the dependence of the flow Q at time $t+\tau$ on the value of the density ρ at time t. In free flow, $cc_{\rho,Q}(\tau) \approx 1$, meaning that the flow strongly depends on previous densities. In synchronized flow, $cc_{\rho,Q}(\tau) \approx 0$, so flow and density are essentially independent, as described previously [5].

D. The origin and evolution of jams

Several causes of traffic jam formation have been discovered. One interesting result is that jams may form spontaneously due to fluctuations.

Most traffic jams are caused by bottlenecks, such as on- and off-ramps or merging lanes, that locally reduce carrying capacity. This results in the formation of jams upstream, while a free-flow region usually persists downstream. An example of the effects of on- and off-ramps on the speed vs. flow diagram is shown below in Figure 2. In the absence of ramps (location C in the figure), the speed is essentially constant over a range of flows. In the presence of merging traffic (location A), speed decreases dramatically beyond a certain flow, almost in the manner of a step function. Between the on-ramp and the off-ramp (location B), the speed is higher at large flows than it was upstream, but it is still less than it would have been in the absence of ramps. By locally increasing flow, the ramps have effectively reduced capacity (i.e., less of the overall capacity, which is constant, is available to traffic passing from A to C) [8].

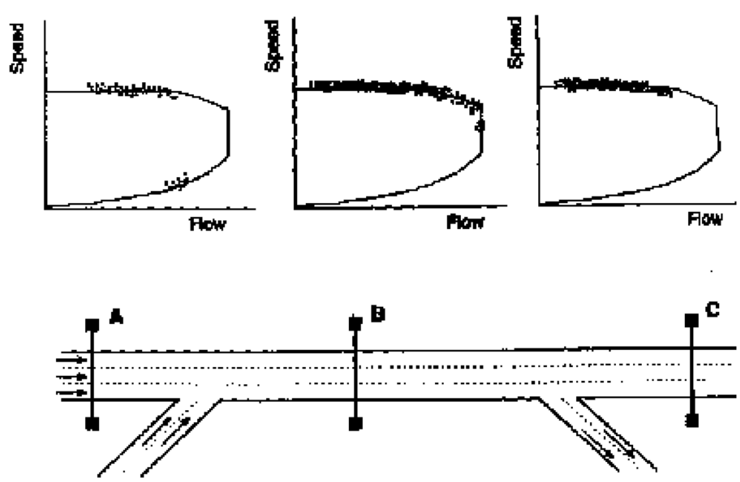

Figure 2. Experimental flow-speed diagrams near on- and off-ramps of a three lane highway. The upper part of the figure shows the empirical results. Each dot represents a five-minute average of the local measurement. The lines serve merely as guide to the eyes. The lower part of the figure shows the location of the detectors (from [9]).

Other obstructions, not related to the form of the road, are an additional cause of jams. For example, traffic accidents may locally reduce carrying capacity. In effect, these create artificial bottlenecks [8].

Jams may also form under high traffic densities for no apparent reason. These jams, known as *phantom jams*, are the result of spontaneous fluctuations in the flow. An example is shown in Figure 3. Each line in the figure represents the trajectory of a single vehicle in the *x*-*t* plane. At the bottom left and the top right, the trajectories are roughly linear and the vehicles are well separated. However, in the central region, there is a curve in the trajectories where a jam spontaneously forms. Since the slope of the jammed region is negative, it is moving upstream. After moving upstream for a distance, the jam spontaneously disappears, just as it formed [8].

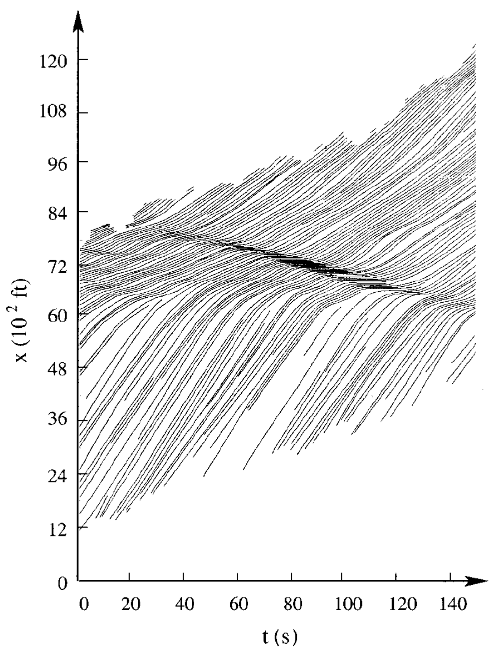

Figure 3. Emergence of a "phantom traffic jam." The depicted vehicle trajectories were obtained by Treiterer and Myers (1974) by aerial photography (reproduction after Leutzbach, 1988). Broken lines are due to lane changes. While the slopes of the trajectories reflect individual vehicle velocities, their density represents the spatial vehicle density. Correspondingly, the figure shows the formation of a "phantom traffic jam," which stops vehicles temporarily. Note that the downstream jam front propagates upstream with constant velocity (from [4]).

III. Theories of traffic

Theoretical models of traffic fall into numerous categories. Here, I shall discuss four: hydrodynamics models, kinetic models, car-following models, and cellular automata models. Each takes methods from physics and applies them to traffic, but with widely varying results.

A. Hydrodynamic models

Hydrodynamic models are macroscopic models that treat traffic as a compressible fluid. Individual vehicles are not present in the description; rather, a continuous density function $\rho(x,t)$ and flow function $Q(x,t)$ are used. In the absence of ramps (i.e., sources and sinks), these functions are related by the conservation equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q}{\partial x} = 0 \,. \tag{10}$$

Since this is a single equation for two unknown functions, additional information is needed. A simple solution is to assume that the flow is a function only of density, $Q(x,t)=Q(\rho(x,t))$. This leads to wave solutions that are stable and cannot describe spontaneous jam formation [5].

In more sophisticated approaches, a Navier-Stokes equation may be used for the velocity, of the form

$$\frac{\partial v(x,t)}{\partial t} + v\frac{\partial v(x,t)}{\partial x} = -\frac{1}{\rho}\frac{\partial P(x,t)}{\partial x} + \nu(\rho)\frac{\partial^2 v(x,t)}{\partial x^2} + \frac{v_e(\rho) - v(x,t)}{\tau(\rho)} \,. \tag{11}$$

$P(x,t)$ is the traffic pressure, related to the velocity variance; $\nu(\rho)$ is a viscosity term which reduces spatial inhomogeneity; $v_{\mathrm{e}}(\rho)$ is an equilibrium velocity, toward which the velocity $v(x,t)$ relaxes; and $\tau(\rho)$ is a relaxation time. In its more sophisticated forms, the hydrodynamic approach succeeds in producing unstable and metastable traffic states [5].

B. Kinetic models

Kinetic models are microscopic models that treat traffic as a gas of interacting particles (vehicles). They have their origin in the kinetic theory of gases. In the kinetic theory of gases, a distribution function $f(\mathbf{r},\mathbf{p};t)d^3\mathbf{r}d^3\mathbf{p}$ describes the number of particles at time t in a volume $d^3\mathbf{r}$ about $\mathbf{r}$ with momentum $d^3\mathbf{p}$ about $\mathbf{p}$. The time evolution of f is governed by the Boltzmann equation,

$$\left[\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m}\cdot\nabla_{\mathbf{r}} + \mathbf{F}\cdot\nabla_{\mathbf{p}}\right] f(\mathbf{r},\mathbf{p};t) = \left(\frac{\partial f}{\partial t}\right)_{collision} , \tag{12}$$

where $\mathbf{F}$ is the external force and the partial time derivative on the right side represents the rate of change of f due to particle collisions. In kinetic models of traffic, $f(\mathbf{r},\mathbf{p};t)d^3\mathbf{r}d^3\mathbf{p}$ is replaced by $f(x,v;t)$, describing the number of vehicles at time t in a length dx about x with velocity dv about v. An additional distribution, $f_{\mathrm{des}}(x,v)$, is introduced to describe the distribution that drivers desire to achieve. In the version developed by Prigogine and coworkers, the analogue of the Boltzmann equation was of the form

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{rel}} + \left(\frac{\partial f}{\partial t}\right)_{\mathrm{int}} , \tag{13}$$

in which the first term on the right describes relaxation of f toward f_{des} in the absence of vehicle interactions (analogous to the collision term of the Boltzmann equation) and the second term accounts for changes in f due to interactions (analogous to the force term).

Difficulties arise when an explicit form of the equation is derived. Simplifying assumptions can produce unphysical results, while more complex treatments are difficult to solve. There are several additional difficulties: in dense traffic, vehicle size should be taken into account; all vehicles should not be assumed to be of the same type; and lane-changing is not accounted for. As a result of these and other problems, kinetic models have had limited success [8].

C. Car-following models

Car-following models assign an equation of motion to each vehicle. This equation is analogous to the Newton's equation of an individual particle in a classical system of interacting particles. A vehicle is modeled as responding to the stimuli of the surrounding traffic by accelerations and decelerations. In the earliest and simplest models, the difference in velocities between the n^{th} and $(n+1)^{\text{th}}$ vehicle was assumed to be the only stimulus for the n^{th} vehicle, leading to the equation of motion

$$\ddot{x}_n(t) = \frac{1}{\tau}\left(\dot{x}_{n+1}(t) - \dot{x}_n(t)\right), \tag{14}$$

where τ sets the time scale. Such models are easily adapted to take into account such factors as delayed driver reaction (e.g., $\ddot{x}_n(t)$ goes to $\ddot{x}_n(t+T)$), nonlinear response to stimuli (nonlinear $\dot{x}$ terms), and response to multiple vehicles ahead (e.g., $\dot{x}_{n+2}(t)$ terms). The weakness of car following models is that, in forms complex enough to be realistic, they involve several phenomenological parameters that must be calibrated. In addition, they assume that the vagaries of driver behavior can be approximated, at least on average, by a manageable continuous function [8].

D. Cellular automata models

Cellular automata models are discrete in space, time, and velocity, and are thus ideal for computational work. Space is generally discretized so that a cell is occupied by one vehicle at most. Time evolution follows a set of simple rules containing stochastic elements [5].

One simple CA model that can reproduce many of the observed features of traffic is the Nagel-Schreckenberg (NaSch) model. The state of a vehicle n is characterized by a position x_n and a velocity $v_n \in \{0, 1, 2, \ldots, v_{\max}\}$. The gap between the n^{th} vehicle and the vehicle in front of it is $d_n = x_{n+1} - x_n$. At each timestep, the arrangement of vehicles on the space lattice is updated in parallel according to four steps:

1. Acceleration: If $v_n < v_{\max}$, $v_n \to \min(v_n + 1, v_{\max})$.
2. Deceleration due to other cars: If $d_n \leq v_n$, $v_n \to \min(v_n, d_n - 1)$.
3. Randomization: If $v_n > 0$, $v_n \to \max(v_n - 1, 0)$ with probability p.
4. Vehicle movement: $x_n \to x_n + v_n$.

Step 1 models the desire of drivers to drive fast; Step 2 prevents collisions; Step 3 accounts for natural fluctuations, as well as introducing an asymmetry between acceleration and deceleration; and Step 4 moves the vehicle with the speed determined in the previous three steps [5].

Even this very simple model produces a fundamental diagram with a free flow branch and a congested branch. Though the NaSch diagram lacks some details of the empirical diagram, refinements can reproduce the entire diagram. These include making the probability in Step 3 velocity dependent and adding anticipation of velocity, time-delayed acceleration, and more refined braking. When these refinements are made, CA models can reproduce all three traffic phases, as well as agreeing with empirical single-vehicle data in all phases. In addition, they can reproduce the observed coexistence of phases [5].

The NaSch model also exhibits spontaneous jam formation, as shown in Figure 4. Spontaneous jam formation may be understood as an avalanche of overreactions by the drivers; this result is made possible by Step 3 [5].

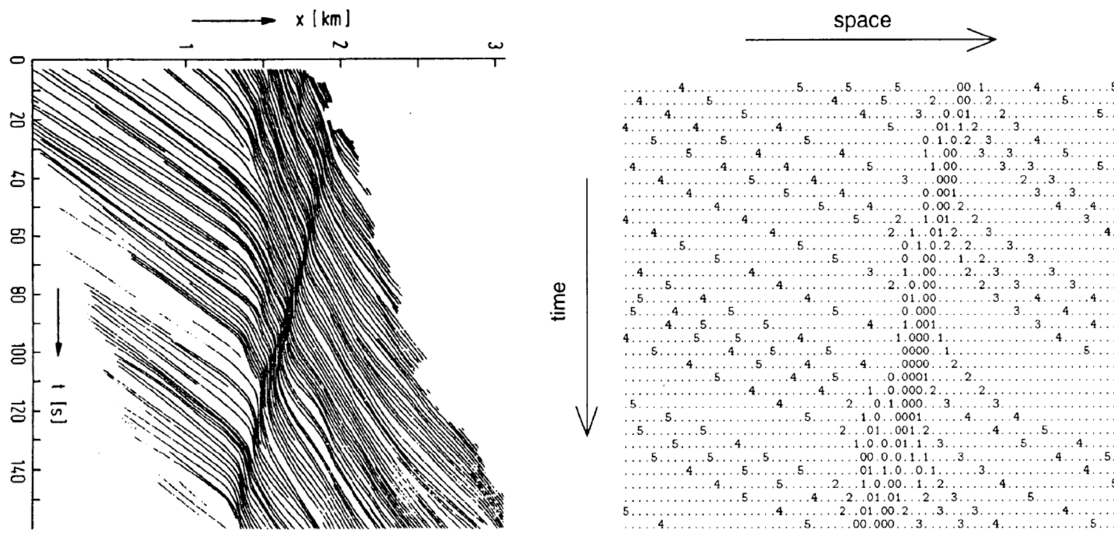

Figure 4. Trajectories of single cars showing spontaneous jam formation. Left: Empirical data; Right: Computer simulation using the NaSch model. Each number 0, 1, . . . , v_{max} = 5 gives the velocity of the corresponding car (from [5]).

IV. Conclusion

Empirical studies of traffic have revealed the presence of multiple phases, and have shown that traffic jams may form spontaneously. Several methods from physics have been applied to modeling traffic, with varied success. The most successful approach seems to be cellular automata, which has reproduced all general empirical features.

While models have successfully reproduced the general empirical features, much remains to be done as additional refinements are made which more accurately reflect the behavior of real drivers and real vehicles.

[1] Longman, Phillip J., “American Gridlock,” *U. S. News & World Report*, 5/28/01.
[2] Lavelle, Marianne, “Feeling the Pain at the Old Pump,” *U. S. News & World Report*, 4/25/06.
[3] Lieu, Henry, “Traffic Flow Theory,” Public Roads, January/February 1999 Vol. 62 No. 4.
[4] Helbing, D., Rev. Mod. Phys. **73**, 1067 (2001).
[5] Schadschneider, A., Physica A **313**, 153 (2002).
[6] Hall, Fred L., “Traffic Stream Characteristics,” Monograph on Traffic Flow Theory (U.S. Department of Transportation, Federal Highway Administration, 1997).
[7] Wagner, P., and I. Lubashevsky, arXiv:physics.cond-mat/0311192 (2003).
[8] Chowdhury, D., L. Santen and A. Schadschneider, Phys. Rep. **329**, 199 (2000).
[9] Hall, F.L., D.E. Wolf, M. Schreckenberg and A. Bachem (eds.), Traffic and Granular Flow (World Scientific, Singapore, 1996)