High-speed microsimulations of traffic flow

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Kai Nagel
aus Köln

1995

Berichterstatter: Prof. Dr. A. Bachem
Priv.-Doz. Dr. M. Schreckenberg

Tag der letzten mündlichen Prüfung (requirements for Ph.D. fulfilled on): 12. Dez. 1994

(This version compiled on March 1, 1995.)

Contents

Chapter 1

Introduction

Traffic jams are annoying, cost money, and pollute the environment. Meanwhile, traffic jams are only the most visible part of a far bigger problem: In the western part of Germany[1] alone, people travel 748.3 billion km per year (pkm = passenger kilometers), and goods are moved 266 billion tkm (ton kilometers) per year [171]. This is more than 2 000 000 000 pkm and 725 000 000 tkm per day.

Moreover, 13.7% of the Gross National Product of Germany is absorbed by trade, transportation, and telecommunication [57]; in the U.S. it is 14.8% by transportation alone [39]. In Germany, nearly 10% of the federal budget is spent by the department of transportation (1991/92, [57]). Thus, transportation is a huge part of the economies of industrialized countries.

The traditional answer of transportation planners and politicians to bottlenecks has been to increase the transportation supply. Yet, demand has always caught up, and during the last years it became increasingly clear that there are big problems in transportation which cannot be solved by the traditional policies.

For example in Germany, traffic produces 67.9% of all carbon monoxides, 58.4% of all NO_x, and 18.2% of all carbon dioxides [171], and it consumes 28% of all energy [57]. Car traffic is a noisy and dangerous ingredient of urban life. Airports, rail lines for high speed trains, and man-made water ways destroy the last undisturbed natural areas in the industrialized parts of Europe. Airplanes inject water vapor and pollution directly into the stratosphere.

An intuitive reaction is that there should be ways to do this without putting so much stress on the environment. And indeed, there is a large number of propositions, ranging from modern technology to changes in social habits. Examples are among others:

- fuel-saving engines,
- Advanced Traffic Information/Management Systems (ATIS/ATMS) using Advanced Transport Telematics (ATT) or the American counterpart Intelligent Vehicle Highway Systems (IVHS),
- routing optimization,
- freight distribution centers, integrated freight logistics,
- transit systems (e.g. trains),
- reducing freight transportation by making industrial production more local

[1] Statistical numbers in good quality are not yet available for the new parts of Germany.

- tele-commuting (i.e. working from a computer terminal at home)
- long-term traffic-reducing urban planning.

In the United States, the Clean Air Act will force communities to find policies which keep the air quality better than a given standard.

However, because of the high complexity of the problem, singular actions are often not very effective and sometimes even counter-productive. For example, addition of a new road may lead to a *de*crease of the vehicle throughput [24]. Simulations show that the introduction of a transit system may lead to *in*creased pollution [165] (private vehicles remain at home and are used by another household member for a high number of short trips with a cold engine). More than half of all passenger kilometers are not work related: In Germany, 63.2% of all pkm fall into the categories shopping (9.2%), leisure (45.3%), and vacation (8.7%) [171].

These examples should have made clear that there are no simple solutions. It will be difficult to reduce the stress on the environment when the transportation demand keeps on increasing at current rates; and reducing the transportation demand cannot be achieved only by technical solutions in freight transportation but will have to include changes in social habits. Obviously, this will need thorough discussion for finding a widely accepted balance of interests.

One of the best methodologies today to deal with complex systems like the transportation system is simulation. Fast desktop computers allow people to get a "hands-on" experience with different possible scenarios, and (parallel) supercomputers give the necessary power to deal with big problems.

As a first step, one would like to simulate "everything" which is related to transportation—from settlement and infrastructure over social habits over trip planning and execution to air pollution dynamics. It seems necessary to include all these elements since, for example, changes in the settlement structure will affect air pollution in non-foreseeable ways.

This is in fact the approach of the American TRANSIMS project at the Los Alamos National Laboratory [165]; the European project EUROTOPP [153] uses similar ideas. Large interdisciplinary groups are necessary to integrate all the different aspects of such an approach. The research presented here is part of a collaboration between the University of Cologne and the TRANSIMS project. It is planned to work with more than 30 people during the next 5 years on the project of transportation simulation, analysis, and optimization.

In this collaboration, the Cologne group concentrates so far on road traffic, with special focus on theoretical developments, on computational aspects of simulation, and on algorithmical aspects. Road traffic currently contributes more than 81% of all passenger and 52.7% of all freight transportation [171]. And despite widespread efforts, the share of road transportation is still increasing. In this sense, it makes sense to start with road traffic when dealing with transportation systems.

This work presents results of traffic simulations using very fast microscopic models. "Micro" refers to the fact that each individual vehicle is resolved, which is an inevitable requirement for, e.g., individual routing analysis. High computational speed, also on (parallel) supercomputers, is achieved by keeping the model extremely simple. This makes, on one hand, real time simulations of large road networks possible, and it allows, on the other hand, the use of dynamical systems techniques in conjunction with many repeated Monte Carlo runs to characterize the phase space of the transportation system. Last, but not least, the simplicity allows theoretical work: Critical exponents clarify the connection between fluid-dynamical and elaborate car-following models, and analytical methods help classifying the general properties of traffic flow models.

This text has two main parts: traffic flow, and network traffic. A third part contains computational results, an outlook, and the conclusion.

The first part consists of Chapters 2 to 7. Chapter 2 explains some necessary background for the traffic flow investigations. It starts with an overview of theoretical models for one-dimensional transport and for traffic as found in the literature. It is, to my knowledge, the first overview which, by using well-known results from the area of driven diffusive systems, classifies fluid-dynamical theories and particle hopping models for traffic flow into one common framework. The second part of this chapter reviews traffic flow simulation concepts and models, and argues why a fast microscopic approach is necessary. Chapter 3 presents the basic model, and gives first results on its properties. Chapters 4 and 5 discuss two different limits of the model: In the first case, all randomness is set to zero; in the second case, this is only done for the part of the randomness which leads to spontaneous initiation of traffic jams. Chapter 6 then summarizes further results and explains them in the light of the results from Chapters 4 and 5. Chapter 7 summarizes all traffic flow results, and shows how the traffic model investigated here helps closing the gap between fluid-dynamical theories, car following behavior, and other, simpler, particle-hopping models.

Chapters 8 to 10 describe traffic in networks. Chapter 8 discusses basic issues, Chapters 9 and 10 contain network examples, where road network elements such as nodes and edges can be processed and vehicles follow individual travel plans. Chapter 9 demonstrates, by using a simple example, how questions such as efficient adaptive road pricing, can be treated by using this fast microsimulation, and that these questions cannot be separated from the system dynamics. Quite generally, it turns out that high system performance is intrinsically coupled to high fluctuations and therefore low predictability. The second example (Chapter 10) is a (single lane) network implementation of the freeway network of Northrhine-Westfalia, showing that this approach makes algorithmic investigations using realistic individual travel plans possible in much faster than real time.

Chapter 11 tests computational performance by implementing the basic model on many different supercomputers. These results are compared to performance values from other projects, and it is checked how performance decreases when implementations become more realistic, e.g. when programming freeway networks. Chapter 12 is an outlook, describing indications for further work, or extensions of the model which have already been investigated by others. Chapter 13 shortly summarizes and concludes the text.

Several appendices contain, for completeness, technical details, or work which is important for the overall picture but less important for the main arguments pursued in this text.

Except for Chapter 3, it should be possible to read only parts of this text without losing too much understanding. People with a background in driven diffusive systems may want to start with the summary in Chapter 7 and just skip back when they need additional information; and the network Chapters 8 to 10 build on, but are largely independent of, the rest of the text.

Part A

TRAFFIC FLOW

Chapter 2

Background

Especially when demand exceeds supply, it makes sense to use existing transportation facilities as efficiently as possible. Naively, one would want to keep, say, a freeway in the regime of maximum vehicle throughput (= capacity). However, due to a traditional bias as well as due to a lack of tools, transportation science mostly has concentrated on traffic at densities *far below* maximum throughput—as can for example be seen from the fact that the Highway Capacity Manual [63] distinguishes five different regimes (Levels of Service A to E) for traffic at densities below capacity, but only one (Level of Service F) above.

Most of currently used traffic flow theory has been developed in the 1950's, see, e.g., [47] for a review. Steady state solutions for mathematical car-following models correspond to steady state throughput-density curves, and due to the lack of refined technology this standard was sufficient to explain measurements. Jam-waves could be explained by using the analogy between traffic and kinematic waves in, e.g., rivers for a fluid-dynamical treatment [100, 175].

However, recent and refined measurements [2, 56, 69] point out that especially the regime of maximum throughput is far from being completely understood. At the same time, extensions of the fluid-dynamical treatment using new methods from nonlinear dynamics indicate a phase transition (route to chaos) in the equations [88], consistent with theoretical developments based on measurements [1, 132]. Simulations of these equations allow for the first time systematic investigation of large scale traffic flow phenomena near capacity [80].

Nonetheless, the fluid-dynamical description is not really useful to clarify the connection between the macroscopic, emergent phenomena (e.g. jams) and the microscopic dynamics (e.g. driving behavior). In addition, simulations based on fluid-dynamical equations are known to be problematic in complex geometries [145]. In the context of traffic this would mean that each ramp or each change in the road characteristics (speed limit, number of lanes, slope, ...) could trigger spurious disturbances which stem from the discretization of the equations.[1] In fluid-dynamics, simple particle-based models (lattice gas automata, LGA) [44] are used to complement the partial differential equation (PDE) approach. LGA have proven to be extremely useful in complex geometries (e.g. for porous media [145]), where they avoid the instabilities of the discretized PDEs.

[1]The fact that this does not seem to happen in network implementations based on the Lighthill-Whitham theory is most probably a consequence of the fact that, together with numerical dissipation due to the discretization, the Lighthill-Whitham approach is stable for all frequencies. This is no longer true for more realistic fluid-dynamic models, which, though, have not yet been tried for road networks.

For vehicular traffic with its much higher particle granularity, an LGA approach is even more natural than for fluids. Instead of with "First Principles" as in LGA, one starts with behavioral rules, and in the same way as in LGA, one can attempt to find the hydrodynamic limit. In this way, CA traffic models are not only useful for computational applications, but also for theoretical developments (theory of traffic flow), directly placed between extensive microscopic and averaging fluid-dynamical models.

The first section of this chapter will be devoted to fluid-dynamical models for traffic. Then, particle based models for one-dimensional transport and their relations to the fluid-dynamical models will be reviewed. Section 3 treats models for car-following, and Section 4 reviews simulation models for traffic flow.

Section 1 is somewhat more comprehensive because it organizes results found in the traffic science literature and results found in the physics literature into one common framework. To my knowledge, such a review does not exist elsewhere.

2.1 Fluid-dynamical models for traffic flow

2.1.1 General equations

One might use the standard fluid-dynamical conservation equations for mass and momentum as a starting point for a fluid-dynamical description of traffic:

$$\partial_t \rho + \partial_x(\rho\, v) = 0$$

and

$$\frac{\mathrm{d}v}{\mathrm{d}t} \equiv \partial_t v + v \cdot \partial_x v = F/m \ ,$$

where ρ is the density and v the velocity.[2] $\mathrm{d}/\mathrm{d}t$ is the individual derivative, F is the force acting on mass m. The first equation (of continuity) describes mass conservation; the second one (momentum conservation) describes the fact that the momentum of a point of mass may only be changed by a force. Obviously, for traffic, F has to take care of vehicle and driving dynamics.

2.1.2 Fluctuations

A standard first step in fluid-dynamics [25] is to assume that v and ρ fluctuate statistically around average values $\langle\, v \,\rangle$ and $\langle\, \rho \,\rangle$, i.e.

$$v = \langle\, v \,\rangle + v' \ , \qquad \langle\, v' \,\rangle = 0$$

and

$$\rho = \langle\, \rho \,\rangle + \rho' \ , \qquad \langle\, \rho' \,\rangle = 0 \ .$$

In this case, one only assumes that $\langle\, v \,\rangle$ and $\langle\, \rho \,\rangle$ fluctuate *slowly* in space and time; for the general subtleties of hydro-dynamical theory see, e.g., [90].

Inserting these relations and subsequent averaging over the whole equations (e.g. $\langle\, \partial_x[(\langle\, \rho \,\rangle + \rho')\,(\langle\, v \,\rangle + v')] \,\rangle = \partial_x \langle\, \rho \,\rangle \langle\, v \,\rangle + \partial_x \langle\, \rho' v' \,\rangle$) yields

$$\partial_t \langle\, \rho \,\rangle + \partial_x \langle\, \rho \,\rangle \langle\, v \,\rangle + \partial_x \langle\, \rho' v' \,\rangle = 0$$

[2] The traffic literature often uses k instead of ρ.

and

$$\partial_t\langle\, v\,\rangle + \langle\, v\,\rangle\partial_x\langle\, v\,\rangle + \frac{1}{2}\partial_x\langle\, v'v'\,\rangle = \langle\, F/m\,\rangle \;.$$

One can [25] parametrize an averaged fluctuation by the corresponding gradient

$$\langle\, v'A'\,\rangle \approx -\alpha\,\partial_x\langle\, A\,\rangle \;,$$

which leads to the set of equations

$$\boxed{\begin{aligned} \partial_t\rho + \partial_x(\rho\, v) &= D\,\partial_x^2\rho \\ \partial_t v + v\,\partial_x v &= F/m + \nu\,\partial_x^2 v \;, \end{aligned}}$$

where, according to convention, the averaging brackets has been omitted, and the diffusion coefficient D as well as the (kinematic) viscosity ν are assumed to be independent of x and t. It should be noted that similar diffusion terms can also be obtained from other arguments.

2.1.3 Lighthill-Whitham-theory and kinematic waves

If one assumes that the velocity is a function of density only ($v = f(\rho)$), then the momentum equation is no longer necessary. This corresponds to instantaneous adaption; the particles (or cars) carry no memory. Using without loss of generality the current $q(\rho) \equiv \rho\, v(\rho)$, and setting in addition $D = 0$, one obtains

$$\boxed{\partial_t\rho + q'(\rho)\,\partial_x\rho = 0}$$

(*Lighthill-Whitham-equation*) [100, 53], where $q' = \mathrm{d}q/\mathrm{d}\rho$.[3] q' will always mean the derivative of q, even when the prime in connection with other letters often denotes fluctuations.

The equation can be solved by the ansatz $\rho(x,t) = \rho(x-ct)$ with

$$c = q'(\rho) \;.$$

This allows the solution of the *characteristics*: A region with density ρ travels with constant velocity $c = q'(\rho)$, and the resulting straight line in space-time is called characteristic. When $q(\rho)$ is convex, i.e. $q'' < 0$, then for regions of decreasing density ($\rho(x_1) > \rho(x_2)$ for $x_1 < x_2$) the characterstics separate from each other. On the other hand, in regions of *in*creasing density, the characteristics come closer and closer together and ultimately hit each other. When two characteristics hit each other, a density discontinuity appears at this place (a *front*), which moves with velocity

$$c = \frac{q(x_2) - q(x_1)}{\rho(x_2) - \rho(x_1)} = \frac{\Delta q}{\Delta\rho} \;.$$

An illustrative example is a queue, such as at a red light. When the light turns green, the outflow front quickly smoothes out, whereas the inflow front remains steep.

Leibig [97] gives results how a random initial distribution of density steps in a closed system evolves towards two single steps according to the Lighthill-Whitham-theory.

[3] In this text, *current* is the same as *flow* or *throughput*. The physics literature usually uses j or J instead of q.

2.1.4 Lighthill-Whitham with dissipation

Adding dissipation to the Lighthill-Whitham-equation leads to

$$\boxed{\partial_t \rho + q'(\rho)\, \partial_x \rho = D\, \partial_x^2 \rho \ .}$$

The solution of this equation is again a nondispersive wave with phase and group velocity q', but this time with a damping constant D. One way to see this is by linearizing the equation: A stationary and homogenous density $\rho \equiv const =: \rho_0$ is a solution. The ansatz $\rho = \rho_0 + \epsilon\, \rho_1 + \epsilon^2\, \rho_2 + \ldots$ leads in first order in ϵ to

$$\partial_t \rho_1 + q'(\rho_0)\, \partial_x \rho_1 = D\, \partial_x^2 \rho_1 \ .$$

A Fourier-Transform (or using the ansatz $\rho_1 = \hat{\rho}_1 \exp[i(kx - \omega t)]$) yields

$$\omega = q'(\rho_0)\, k - i\, D\, k^2 \ .$$

From the real part of the frequency one obtains the phase velocity c_φ and the group velocity c_g:

$$c_\varphi := \frac{\mathrm{Re}(\omega)}{k} = c_g := \frac{\partial}{\partial k} \mathrm{Re}(\omega) = q'(\rho_0) \ ,$$

which gives are the same kinematic waves as before. The difference is that this time a dissipation (damping) of the wave amplitude is given by $\mathrm{Im}(\omega) = -D\, k^2$: Any disturbance decays exponentially. The reflects the intuitively reasonable effect that traffic jams should tend to dissolve under homogeneous and stationary conditions.

2.1.5 The nonlinear diffusion (Burgers) equation

For a further development, $q(\rho)$ has to be specified. Since we are mostly interested in the behavior of traffic near maximum throughput, we start by choosing the simplest mathematical form which yields a "well-behaved" maximum:

$$q(\rho) = v_{max}\, \rho (1 - \rho) \ ,$$

which, in traffic science, is called the *Greenshields*-model (see [47]). v_{max} is, in principle, a free parameter, but it has an interpretation as the maximum average velocity for $\rho \to 0$.[4] The maximum current q_{max} is reached at $\rho(q_{max}) = 1/2$.

The result is the equation

$$\boxed{\partial_t \rho + v_{max}\, \partial_x \rho - 2 v_{max}\, \rho\, \partial_x \rho = D\, \partial_x^2 \rho \ .}$$

Introducing a linear transformation of variables [115]

$$x = v_{max}\, t' - x' \ , \qquad t = t' \ ,$$

one obtains

$$\partial_{t'} \rho + 2\, v_{max}\, \rho\, \partial_{x'} \rho = D\, \partial_{x'}^2 \rho \ ,$$

[4] Mathematicians would set $v_{max} = 1$; traffic scientists use $1 - \rho/\rho_{jam}$ for the term in parenthesis. ρ_{jam} is the density of vehicles in a jam.

which is the (deterministic) *Burgers equation.*

This equation has been investigated in great detail by Burgers [20] as the simplest non-linear diffusion equation. The stationary solution is a uniform density $\rho(x,t) = const.$ A single disturbance from this state evolves over time into a characteristic triangular structure with height $\sim t^{-1/2}$, width $\sim t^{1/2}$, and bent to the right such that the right side of the disturbance becomes discontinuous. This disturbance moves to the right with velocity $c = q' = 2\,\rho\, v_{max}$.

When interpreting this for traffic jams, one has to re-transform the coordinates. Jams can then move *both* to the left or to the right (with velocities between v_{max} and $-v_{max}$), and the discontinuous front develops at the upstream side of the jam, i.e. where the vehicles enter the jam. One sees that this solution is just the solution of the characteristics, with a dissipating diffusion term added—as should be expected because of $D > 0$.

Some other versions of the Burgers equation are relevant for traffic and have been investigated thoroughly [85, 18, 105]:

- **Noisy Burgers equation:** Adding a Gaussian noise term η to the equation ($\langle\, \eta(x,t)\eta(x',t')\,\rangle = \eta_0\, \delta(x-x')\, \delta(t-t')$) leads to the noisy Burgers equation

$$\partial_t \rho + 2\, v_{max}\, \rho \partial_x \rho = D \partial_x^2 \rho + \eta \; .$$

  This equation does not reach a stationary state any more.

- **Generalized Burgers equation:** The nonlinearity of the Burgers equation can be generalized:

$$\partial_t \rho = \sum_\beta b_\beta \partial_x \rho^\beta + D \partial_x^2 \rho \; .$$

  E.g., a current of

$$q(\rho) = q_{max} - v_{max}\, |\rho - \rho(q_{max})|$$

  leads to a linear Burgers equation

$$\partial_t \rho + \mathrm{sgn}[\rho(q_{max}) - \rho]\; v_{max}\, \partial_x \rho = D\, \partial_x^2 \rho \; .$$

  As we will see later, this linear ansatz is useful for traffic.

  Generalized Burgers equations with arbitrary β have been investigated [18, 85].

The Burgers equation has the same mathematical structure as a well-known equation for surface growth, the Kardar-Parisi-Zhang (KPZ) equation [18, 78, 85]. The connection is most straightforward when one interprets the *fluctuations* of the density, $\rho' = \rho - \langle\, \rho\, \rangle_L$, as a step density for a 1-dimensional surface (i.e. $\rho' = \partial_x h$). Inserting $\rho = \langle\, \rho\, \rangle_L + \rho'$ into the generalized Burgers equation yields

$$\partial_t \rho' = \sum_\beta \tilde{b}_\beta \partial_x \rho'^\beta + D\, \partial_x^2 \rho' \; ,$$

with the same structure as before, but with different coefficients $\tilde{b}_\beta$. The system here is assumed to be closed (periodic boundary conditions), in consequence $\partial_t \langle\, \rho\, \rangle_L \equiv \mathrm{d}\langle\, \rho\, \rangle_L / \mathrm{d}t = 0$. By inserting $\rho' = \partial_x h$ and integrating once, one obtains

$$\partial_t h = const + \sum_\beta \tilde{b}_\beta\, (\partial_x h)^\beta + D\, \partial_x^2 h$$

(generalized *KPZ-equation*).

2.1.6 Including momentum

The equations so far do not explain the spontaneous phase separation into relatively free and rather dense regions, which is observed in real traffic. To obtain this, one can no longer neglect the effect of momentum: One can neither accelerate instantaneously to a desired speed nor slow down without delay. It becomes necessary to include the momentum equation. Here, one has to specify the force term F/m, which describes acceleration and slowing down. At least two properties have to be incorporated: interaction with other cars, and relaxation towards desired speed.

- A first order approximation for the relaxation term is [88, 130]

  $$\frac{1}{\tau}(V(\rho) - v),$$

  where $V(\rho)$ is the desired average speed as a function of density.

  This choice yields exponential relaxation towards the desired speed. The function $V(\rho)$ has to be specified externally, e.g. from measurements.

- A commonly used interaction term [26, 80, 88, 130] is

  $$-\frac{c_0^2}{\rho}\partial_x \rho \ , \qquad\qquad (*)$$

  where c_0 is treated as constant. In traffic, a typical value for c_0 is 15 km/h [87].

  The meaning is that one tends to reduce speed when the density increases, even when the local density is still consistent with the current speed.

  Formally, this term comes from the the pressure term of compressible flow

  $$-\frac{1}{\rho}\partial_x p \ ,$$

  where p is the pressure due to thermal motion of the particles. Assuming an ideal gas ($p = \rho\, RT$) and isothermic behavior $T = const$, one obtains waves similar to sound waves as a solution of the linearized equations.[5] This leads to $(*)$, where c_0 is the speed of the "sound" waves. Note that sound waves move in *both* directions from a disturbance, which means that sound waves alone are not a good explanation for freeway start-stop-waves, contrary to what is written in [14].

A possible momentum equation for traffic therefore is

$$\boxed{\partial_t v + v \partial_x v = -\frac{c_0^2}{\rho}\partial_x \rho + \frac{1}{\tau}\left(V(\rho) - v\right) + \nu \partial_x^2 v \ ,}$$

and the system is closed together with an equation of continuity

$$\boxed{\partial_t \rho + \partial_t(\rho\, v) = D\, \partial_x^2 \rho \ .}$$

Usually, D is set to zero.

[5] True sound waves assume the gas to behave adiabatic, i.e. $p \propto \rho^\kappa$. Then, increasing pressure results in an increased temperature.

For this equation, the homogeneous solution $(v, \rho) \equiv (v_0, \rho_0)$ is unstable for densities near maximum flow for a suitable choice of parameters. Using the methods of nonlinear dynamics [88, 143, 156] (see also [64]), one can go beyond the linear stability analysis. One finds a multitude of stable or unstable fixpoints and limit cycles which suggest that traffic near maximum flow operates on a strange attractor. This can lead to quasi-periodic behavior, exactly as is observed in traffic measurements.

Earlier work [26, 58, 130] has analyzed the same equation without diffusion ($\nu = 0$). The results are reported to be unrealistic for bottlenecks, and the equations do not reproduce the start-stop-wave behavior.

2.1.7 Discussion of fluid-dynamical approaches

Fluid-Dynamical models have been used in traffic science for a long time, with considerable success. But they have shortcomings. Some of the major points are:

- One has to give externally the relation between speed or current and density. This is unsatisfying in terms of the development of a theory. But an even more intricate problem is that there is no agreement on a functional form of the speed-density relation; it is even under discussion if this relation is at all continuous [56, 132].
- Temperature parametrizes the random fluctuations of particles around their mean speed. For gases, fluctuations and therefore temperature increase with density. For granular media, fluctuations *de*crease with density (i.e. inside a jam)—it has been claimed that exactly this inverse temperature effect is responsible for clustering [48]. In this way, assuming isothermic instead of adiabatic behavior as done for the momentum equation seems only half the way one has to go. Helbing [61] discusses this further.
- Helbing [61] also discusses *excluded volume* to take into account the spatial extension of vehicles.

Nonetheless, fluid-dynamical approaches [80, 88, 143, 156] give, for the first time, systematic insight into traffic near maximum flow beyond simple extrapolation of light and dense traffic results. Lee [96] has found the same mechanism for waves in general granular media. Putting all their results together, the overall picture is as follows: Near maximum throughput, the flow becomes unstable due to instabilities in the momentum equation (dynamic mode). These instabilities grow exponentially until the wave is outside the unstable region (i.e. too high density). Then the dynamic mode damps out, but the density discontinuity survives as a kinematic wave.

2.1.8 Kinetic theory

The heuristic introduction of the fluid-dynamical equations for traffic flow remains unsatisfactory. As in standard hydrodynamic theory [140], it should be possible to derive correct hydrodynamic equations for traffic from a kinetic theory. Indeed, such an approach has been followed by Prigogine, Herman, and coworkers [138], and in fact, the hydrodynamic equations of Lighthill and Whitham have been derived from this as a limiting case.

In short, the kinetic approach attempts to give the time development of a distribution function $f(x, v, t)$, where

$$\int_0^\infty \mathrm{d}v\, f(x, v, t) = \rho(x, t)$$

and

$$\frac{1}{\rho(x,t)} \int_0^\infty \mathrm{d}v \, v \, f(x,v,t) = \bar{v}(x,t) \ .$$

Technically similar to the fluid-dynamical theory, the dynamic equation for f is

$$\partial_t f + v \, \partial_x f = \text{force terms} \ .$$

Again similar to the fluid-dynamical theory, mainly two terms are considered for the force:

- The relaxation term
$$-\frac{f - f_d}{T} \ ,$$
where $f_d(x,v,t)$ is the distribution of desired speeds.

- Following standard kinetic theory, the interaction term is at first written as a master equation:
$$\left.\frac{\partial f(x,v_i,t)}{\partial t}\right|_{interaction} = (1-P) \sum_j \left[R(v_i|v_j) - R(v_j|v_i)\right] \ .$$
$R(v_b|v_a)$ is the rate with which cars change from velocity v_a to velocity v_b due to interaction. In gas dynamics, these rates are given by the collision rules. For traffic, it means that an upcoming car slows down to the speed of a slower car. P is the probability of passing: If a car can pass, interaction does not take place. P is itself a function at least of ρ.

  Further approximation of this term leads to the expression
$$\partial_t f|_{interact} = (1-P) \, \rho \, (\overline{v} - v) \, f \ .$$
$\partial_t f(v)|_{interact}$ is positive when $v < \overline{v}$ and negative when $v > \overline{v}$, which means that the interaction term itself has the tendency to slow down the traffic to velocities below $\overline{v}$.

  This approximation is only mean field, meaning that it only takes into account interactions of a car with the *average* field of all the others. Higher correlations are in principle feasible.

The problems with the kinetic approach are:

- The mathematics are relatively laborious *and* one needs a lot of specificity (assumptions about driving behavior etc.) at an early level of the derivation. The approach is therefore not useful for investigating a lot of different assumptions.

- Kinetic theory for traffic is so far only mean field: No two-point correlations have been included into the calculations. Although this is in principle feasible, it will be even more difficult mathematically. Yet, these correlations are most important near maximum flow, where we are interested in.

2.2 Particle models and their relation to fluid-dynamical models

2.2.1 Particle models for one-dimensional transport phenomena

In more recent times, investigations of partial differential equations have been complemented by investigations of discrete models [162].

They are much more straightforwardly to simulate on a computer: Local rules usually only consist of a few integer operations instead of many more floating point operations necessary for discretized partial differential equations. And since in addition no round-off errors and subsequent instabilities occur, much faster algorithms (for a given number of sites) are possible. This is especially useful for the investigation of critical phenomena (see below), where large correlation lengths and times and huge fluctuations call for large systems, long runs, and averages over many runs.

Moreover, these discrete models can be approached by analytical many-body methods from statistical physics (see later).

Out of the multitude of particle hopping models, in the context of this work, two of them are of particular interest. Both use a lattice of size L (open or periodic boundary conditions) which is filled with $N = \rho\, L$ particles.

- **Stochastic asymmetric exclusion** [74, 52, 85, 34]: In each update step, one of the particles is selected randomly (random sequential update). If the site to the right of the particle is free, the particle is moved to this site, otherwise it does not move.
- **Deterministic asymmetric exclusion** [85]: All particles are simultaneously updated according to the same rule as before.

  Biham and coworkers [17] have first used this model for traffic flow considerations. Nagatani [117] has used it for two-lane traffic. Moreover, all CA traffic models on two-dimensional grids (see Chapter 8) are based on this velocity update.

2.2.2 Damage spreading and critical exponents

A helpful concept for understanding critical phase transitions in discrete systems is the notion of "damage spreading" [159]: One simulates two identical copies of the system. At a certain point, a minimal change in one of the copies is made and then the time evolution of the *differences* between the systems is observed.

To be more precise, I will use the picture of traffic jams as an example. In Chapters 5 and 6, this picture will be verified and extended for a subcase of the CA traffic model, and Chapter 7 will explain the meaning and limit of the picture in the full model.

In our traffic system, one way of "damaging" is to change the velocity of one randomly picked car by -1. This car then can, by a chain reaction, cause a jam of a certain life-time; and downstream of this jam, the traffic pattern will be different from the undisturbed model. After this jam has dissolved, the *spatial* amount of damage extends from the disturbed car to the last car involved in the jam, and this length is proportional to the life-time of the jam. For the limit $p_{spont} \to 0$ (but $p_{spont} \neq 0$), i.e. where spontaneous initiation of a jam becomes rare (see Chapter 5), one obtains the following picture:

- For low densities, jams are usually short-lived (i.e. with an exponential cut-off in the life-time distribution). As a result, the average amount of spatial damage is small.
- For high densities, a jam caused by the disturbance will (in the average) survive forever, thus (in the average) causing infinite damage.

  However, dense traffic is characterized by the existence of many jams quasi-randomly distributed over the system. So the additional jam caused by the disturbance will not change the *statistical* properties of the system.

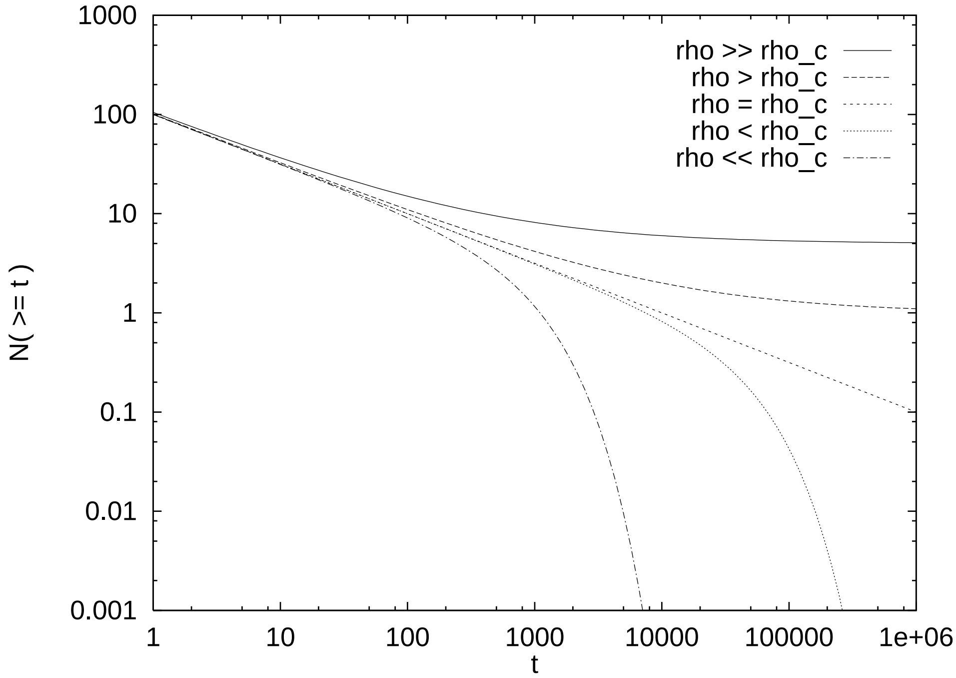

Figure 2.1: Theoretical distribution functions of life-times of traffic jams. The curves show, for different densities, the number of jams with a life-time larger or equal than t as a function of t. (y-axis arbitrary units)

- In between, there should be a density ρ_c, where, when this density is approached from below, jams become increasingly long-lived, with the result that the amount of spatial damage becomes larger and larger. Ultimately, exactly at the critical point, a damage of size infinity (in the thermodynamic limit) is possible.

All these observations are similar to conventional damage spreading observations in cellular automata (CA) [91, 159]: The damage is limited for class I and class II CA, and it *can* be infinite for class VI CA. For class III CA, the damage is practically always infinite, but does not change the statistical properties of the system.

Moreover, at a critical point a system's behavior is characterized by critical exponents. At the critical point, i.e. where jams barely survive, one would expect for the jam survival probability a scaling law:

$$P_{surv}(t) \sim t^{-\delta} ,$$

where $\sim$ means "proportional to in the limit of $t \to \infty$ and $L \to \infty$", and δ is a critical exponent.

But why should that make sense? Why should traffic take place exactly at ρ_c? And if so, should one not use a more realistic model to obtain quantitative numbers (i.e. critical exponents)? The answers to these questions are called in short *self-organized criticality* [8] and *universality* [173].

In our context, the first means that there are microscopic and macroscopic reasons that traffic flow has the tendency to operate near ρ_c. Microscopic, because the outflow from any high density region happens at ρ_c; macroscopic, because optimal use of a road network pushes ρ towards ρ_c. Both results will be explained in detail in later chapters.

Universality means that, in many cases, critical exponents are robust with respect to changes of the model, i.e. grid structure, neighborhood definitions, etc. For the traffic model, we even can explain most of the exponents by a phenomenological theory. In consequence, we expect the results being true for most if not all car following models. The details of this claim will be worked out later in this text.

Traffic practitioners have indeed encountered problems which come out of these fundamental observations. In the TRAF-NETSIM simulation package, a problem was that small control measures could lead to completely different simulation behavior, which is exactly the problem of damage spreading. Rathi and Santiago [139] resolved it by implementing "identical traffic streams", but without a theory it seems doubtful if this is a final solution.[6] Other authors [23] have investigated the variability of NETSIM simulations—an enterprise which also would benefit from more theory such as developed here. For example, it would clarify that variability is density dependent, and where one should expect the highest variability. See also Chapter 9 on variability in network simulations.

2.2.3 Hydrodynamic exponents

Critical exponents have another advantage: They are useful to compare particle hopping models with fluid-dynamical descriptions even when one does not know the hydrodynamical limit of the hopping model. In this way, we will be able to work out the connections of the traffic model proposed in the next chapter to other models in physics and to the fluid-dynamical theories of traffic flow.

For example, dimensional analysis of the KPZ equation (see above)

$$\frac{\partial h}{\partial t} = b_2 \, |\frac{\partial h}{\partial x}|^2 + D \frac{\partial^2 h}{\partial x^2}$$

suggests

$$\frac{H}{T} \sim \frac{H^2}{X^2} , \frac{H}{X^2} ,$$

where H, T, and X are "typical" values for variations in h, t, and x. Since, heuristically, the last term on the RHS is smaller than the first one on the RHS, one finds

$$X^2 \sim H\,T \ .$$

Driving the KPZ equation by noise yields in addition $H \sim X^{1/2}$ [85] and therefore

$$T \sim X^{3/2} ,$$

with the well known "anomalous" dynamic KPZ exponent $z = 3/2$.

More precisely, one considers correlation functions of the type

$$F_h^2(x,t) = \langle \, [h(x_0 + x, t_0 + t) - h(x_0, t_0)]^2 \, \rangle$$

[105] and checks if $F \sim x^{\xi}$ (roughness exponent) or $F \sim t^{\xi/z}$. And by using, e.g., renormalization group arguments, one has shown rigorously that the above exponents for KPZ are exact [18].

[6] Actually, in that case the situation is somewhat more complicated. In the original NETSIM simulation package, one random number sequence was used for all randomized elements, especially also for turning decisions at intersections. And since vehicles in NETSIM do not have trip plans, any small change in the random number sequence (e.g. by the addition of one single car) sends each car to completely different destinations. Using true paths will help a lot in this situation, but a fundamental issue remains: One needs theory to evaluate results of Monte Carlo simulations.

2.2.4 Correspondences between field theories and particle models

What now is left is to identify correspondences between field theories and particle models. For the two hopping models presented above, this is done in the following.

Stochastic asymmetric exclusion process

The stochastic asymmetric exclusion process corresponds [85] to the noisy Burgers equation with $\beta = 2$, and therefore corresponds to the exponents mentioned above: $X^2 \sim H\,T$ in general, and $H \sim X^{1/2}$, $T \sim X^{3/2}$ in the steady state.

For our needs, however, this statement has to be more precise. The particle process corresponds to a diffusion equation

$$\partial_t \rho + \partial_x q = D\, \partial_x^2 \rho + \eta$$

with a current $q = \rho\,(1-\rho)$ [85, 34]. Inserting the current yields

$$\partial_t \rho + \partial_x \rho - \partial_x \rho^2 = D\, \partial_x^2 \rho + \eta \ .$$

Since we are searching for fluctuations,[7] we insert $\rho = \langle\, \rho \,\rangle_L + \rho'$ and obtain

$$\partial_t \rho' + (1 - 2\langle\, \rho \,\rangle_L)\, \partial_x \rho' - 2\, \rho'\, \partial_x \rho' = D\, \partial_x^2 \rho' + \eta \ .$$

Thus, when transforming into the moving coordinate system $x' = x + (1 - 2\langle\, \rho \,\rangle_L) \cdot t$, one obtains KPZ exponents. In all other coordinate systems, the linear gradient is dominating [18]. A precise treatment of this uses correlations between tagged particles [105].

In terms of kinematic waves, the explanation of this is as follows: The particle process with a given average density $\langle\, \rho \,\rangle_L$ produces kinematic waves of wave velocity $c = q' = 1 - 2\langle\, \rho \,\rangle_L$. Only in the coordinate system of these waves, one sees the anomalous fluctuations. And there is a density, which—not coincidently [86]—corresponds to maximum flow, where these waves do not move in space. When traffic would follow a Burgers equation, then one could detect maximum traffic flow by standing on a bridge: Jam-waves moving in flow direction indicate too low density, jam-waves moving against the flow direction indicate too high density (cf. Figs. 7.1 and 7.2).

For surface physics, these spatial transformations usually do not matter. People are interested into spatially averaged quantities such as the average surface height fluctuations

$$W^2(t) = \langle\, (h(x,t) - \langle\, h \,\rangle_L)^2 \,\rangle_L \ ,$$

which shows anomalous KPZ behavior $W \sim t^{\xi/z}$ independent of spatial transformations.

Something similar is true for traffic flow. According to this theory, a (small)[8] traffic jam of length l would need a time $t \sim l^{3/2}$ until it would be no longer visible in the system. But at the same time, the jam as a whole can move forwards or backwards. Only at maximum flow, the wave velocity $c = q' = 0$, and the jams stays at the same place.

In consequence, when measuring time series of local quantities, for example the density $\rho(x_0, t)$ for fixed x_0, one finds different correlations and a different frequency spectrum at $\rho = \rho(q_{max})$ [105].

[7] This is why the Musha transformation (see above) is not the correct one to obtain critical exponents.

[8] When a jam becomes too big, the system is no longer in the steady state, and a crossing over to other exponents is found.

Deterministic asymmetric exclusion process

The deterministic asymmetric exclusion process corresponds to the deterministic Burgers equation with $\beta = 1$, with a current as described above,

$$q(\rho) = q_{max} - v_{max} \left|\rho - \rho(q_{max})\right| ,$$

with $q_{max} = 1/4$, $v_{max} = 1$, and $\rho(q_{max}) = 1/2$. Further observations for this model are described in Chapter 7, together with summaries of other cases which are relevant for traffic.

2.2.5 Open systems

In an open system, the number of particles (cars) is no longer conserved. We continue to consider a one-dimensional system. Particles move from the left to the right. One imposes open boundary conditions by allowing particles to enter the system at the left end ($x = 0$) at rate ρ_0 and letting them leave the system at the right end ($x = L$) at unit rate. This is equivalent to fixing the density as $\rho(x=0) = \rho_0$, $\rho(x=L) = 0$.

For these systems, one generically has to distinguish two regimes [86]:

- $\rho_0 < \rho(q_{max})$. In this case, the density in the whole system except for the boundary regions is equal to ρ_0. Near the right boundary, ρ decays to zero [86]. The size of the boundary region goes to zero in the hydrodynamic limit [21, 34].
- $\rho_0 > \rho(q_{max})$. In this case, the density in the whole system is equal to $\rho(q_{max}) < \rho_0$, i.e. the system maximizes the flow and adjusts the density [86].

The second regime is interesting: As explained above, the case $\rho = \rho(q_{max})$ shows anomalous fluctuations for a fixed observer. By the automatic selection of this state, this supposedly rare case suddenly happens generically—which may be seen [86] as a case of self-organized criticality [8].

These observations are "generically" true for a large class of systems, but exceptions exist.

For the traffic CA, these observations give the correct ideas, but the details are distinctly different. That will put us into a position to judge where the Burgers equation and the more general Lighthill-Whitham description are incomplete for traffic.

2.3 Car following models

In socio-economic systems such as traffic, one does not have first principles for particle behavior to start from. It is therefore somewhat difficult to discuss the justification of a certain modeling approach—in many cases, one will have to argue from the usefulness of the results. Nonetheless, car following theory has been a field under development for a long time now, and common agreement has established a variety of results.

2.3.1 Mathematical car following theory

Mathematical car following theory assumes a linear relationship between the (re)action and a stimulus (cf. [47] for the following):

$$\text{action} = \text{sensitivity} \cdot \text{stimulus} \ .$$

The stimulus could, e.g., be the distance to the car ahead, or the difference of the velocities. One of the first propositions was

$$v_i(t+\tau) \propto \Delta x_i(t) \ , \qquad (*)$$

where v_i is the velocity of the ith car in a chain, τ is the time which is necessary for the adaption (this is more than the reaction time), and Δx is the distance to the next car ahead. Most of the theory treats vehicles as points. Especially in recent times, results have been obtained starting from this approximation [10, 125, 176].

Deriving $(*)$ once with respect to time leads to

$$a_i(t+\tau) \propto \Delta v_i(t) \ ,$$

where a is the acceleration of i and Δv is the velocity difference to the next car ahead. A generalisation of this ansatz, taking into account that the sensitivity depends on the distance and on the speed, is

$$a_i(t+\tau) = c_1 \cdot \frac{[v_i(t)]^m}{[\Delta x_i(t)]^l} \cdot \Delta v_i(t) \ . \qquad (**)$$

l and m are integer numbers. This equation has been analyzed in great detail in the 1960s, for example with respect to its stability for a single following car (local stability) or for an infinitely long chain of cars (asymptotic stability). On can derive speed-density and therefore flow-density relations from $(**)$. The resulting curves are, however, not extremely realistic when compared to modern measurements [1, 2, 56, 69, 132].

In any case, equation $(**)$ is problematic for numerical modelling. When Δv is zero, the equation allows an arbitrarily small distance to the car ahead, even at very high speed. In the theory, this never happens as long as one starts from realistic initial conditions, but due to numerical inaccuracy for simulations such situations could occur. It is very probable that it will occur in a model with only a few states, as is the model which will be introduced in the next chapter.

As a result, practical implementations either use distance alone as stimulus [51], or they use it to provide some bounds for the velocity-based reactions [16, 177].

2.3.2 Psycho-physiological spacing models

Another approach to car following is based on psychological and physiological observations. Wiedemann [98, 177], before describing a microscopic car-following simulation model, reviews these aspects:

A starting point is that the *angular* resolution of the human eye is limited. Even when a human can detect another car as coming closer, it can estimate its *velocity* only when the car becomes noticeably bigger in the field of vision—i.e. when the angle between left and right parts of the car increases. By using geometric arguments, one quickly sees that this observability threshold is proportional to the speed difference, and inversely proportional to the square of the distance:

$$threshold \propto \frac{\Delta v}{\Delta x^2} \ .$$

Note that this is similar to the mathematical car following model $(**)$ with $l = 2$ and $m = 0$.

Another element of psycho-physiological car following models is that the reactions are, as a first approximation, step functions instead of smooth ones. Thus, after crossing the threshold, slowing down or braking is adjusted with the goal to reach the speed of the car in front together with a certain desired spacing. Measurements, consistent with the physiological argument, indicate that the reaction threshold is largely independent of one's velocity. Thus, sometimes immediate emergency braking may be necessary.

Furthermore, the human's braking reaction usually is not very precise, so that further control action is necessary, leading to a highly stochastic element in car following.

Similar observations are true if the distance to the car in front increases, except that reactions are slower. This can be explained in part again physiologically, and in part by the weaker motivation to react (less dangerous).

More precise results here should be possible using results from ecological psychology. Ecological psychology is concerned with finding general principles as to how organisms, including humans, interact with their environment. Visual perception links the perceiver with his/her environment via an optic flow field, which provides the action relevant visual information. The goal is to find low-dimensional quantities that capture this information as is provided in the flow field.

In the context of car following, the concept of "time to contact" is relevant. Time to contact and its derivative has proven to be a comprehensive descriptor for informing the perceiver about the necessary movement behavior given a goal. This goal could be, for example, to arrive smoothly at a desired car following position [82, 95].

2.4 Simulation and the role of microscopic high speed models

Naively, simulation seems a perfect tool: One simply formulates every aspect of the system as programming code, and then runs the code (i.e. the model) on a computer. When initial conditions and the system dynamics are only realistic enough, this should lead to a good prediction.

There are, though, severe limitations: Simulations are resource-limited; and nonlinearities make predictions even more resource-demanding. The second aspect will be discussed further down; the first leads to a trade-off between resolution, fidelity, system size, simulation speed, and resources [11]:

- **Resolution:** The *level* of detail of a model. All processes of the system which lead, by *upward causation*, to the *emergence* of the phenomena of interest, may be important.
- **Fidelity:** The degree of realism in modeling each process.
- **System size:** The system size which can is covered by the simulation.
- **Simulation speed:** The speed of the simulation, often compared to real time.
- **Resources:** Often computational resources, but also limitations of the time one wants to spend on the programming.

This trade-off has to be balanced in conjunction with the problem and the question under consideration. With respect to resolution, one often distinguishes between macroscopic and microscopic models. In traffic science, macroscopic means that one averages over vehicles, whereas microscopic means that at least vehicles are resolved individually.

Another classification of models is according to their intended use: models for, e.g., signalized intersections, arterial networks, freeway corridors, or rural highways. In the context of this work, models for arterial networks and models for freeway corridors are more important than the others.

For reviews of traffic simulation models see, e.g., [47, 108, 170].

2.4.1 Macroscopic models

For practical network control problems, only macroscopic models are in use, e.g. FREQ, CORQ, NETFLO, or CONTRAM [108, 146, 170]. For CONTRAM and its traffic assignment algorithm, parallelization attempts are reported [50, 65]. TRANSYT [174] is mostly used for signal optimization. In the area of freeway simulation, models often are based on discretizations of fluid-dynamical equations (MACK, FREFLO, FRECON [170, 146], cell-transmission-model [31, 133], [26, 29]). It has been claimed [32] that many of these approximations have not been done systematically, i.e. that the hydrodynamic limit of the finite difference equations would not be consistent with the fluid-dynamical equations. However, it should be kept in mind that this does not necessarily matter because the fluid-dynamical equations themselves are not exact for traffic flow.

2.4.2 Microscopic models

Is is quite obvious that certain questions can only be answered by a microscopic model. To understand the relation between traffic flow phenomena, such as jams, and individual vehicle characteristics (e.g. driving behavior, cruise control, radar-based automatic car following), one has at least to resolve the individual vehicles. Dynamic individual routing just does not make much sense when individual vehicles are not present: For example, to react to incidents in an ATMS/ATT/IVHS environment, destination information is necessary (cf. Fig. 2.2). Other microscopic questions are modal choice or pollution (engine temperature!).

These questions quickly lead to 500 000 or more "intelligent objects" in the system and to the necessity of corresponding computational power.

There are only few well-known simulation models which are at the same time microscopic and operating on a network level. TRAFF-NETSIM [124] was developed for the Federal Highway Administration of the USA more than 20 years ago and has been continuously updated and enhanced since that time. The model is based on a stochastic, time-stepping simulation approach describing the dynamics of traffic operations in urban street networks, which consist of uni-directional links and nodes. A wide variety of options and modules is available, including signal controllers, bus operations, measures of effectiveness, or estimates of emission or fuel consumption.

However, for our needs, NETSIM is too slow, there exists no version which runs efficiently on parallel supercomputers, and it does not allow for trip plans: Vehicles at intersections are distributed according to turn counts (cf. Fig. 2.2).

PARAMICS [181, 110] is a rather new approach, specifically designed for parallel computers (PARAMICS = PARAllel MICroscopic traffic Simulator). It is, so far, the only approach comparable to the work presented here in its possible applications.

2.4.3 Frameworks

A severe shortcoming of all the models above is that they only deal with one part of the transportation system: road traffic. However, in recent times it became increasingly clear that future solutions will need an integrated traffic management, and modal split (i.e. the use of different modes of transportation) will become a necessity. The TRANSIMS project [165] is a reaction to that. It is a framework planned to include a wider spectrum of transportation activities than ever before, such as:

- Trip generation of individuals from demographic data

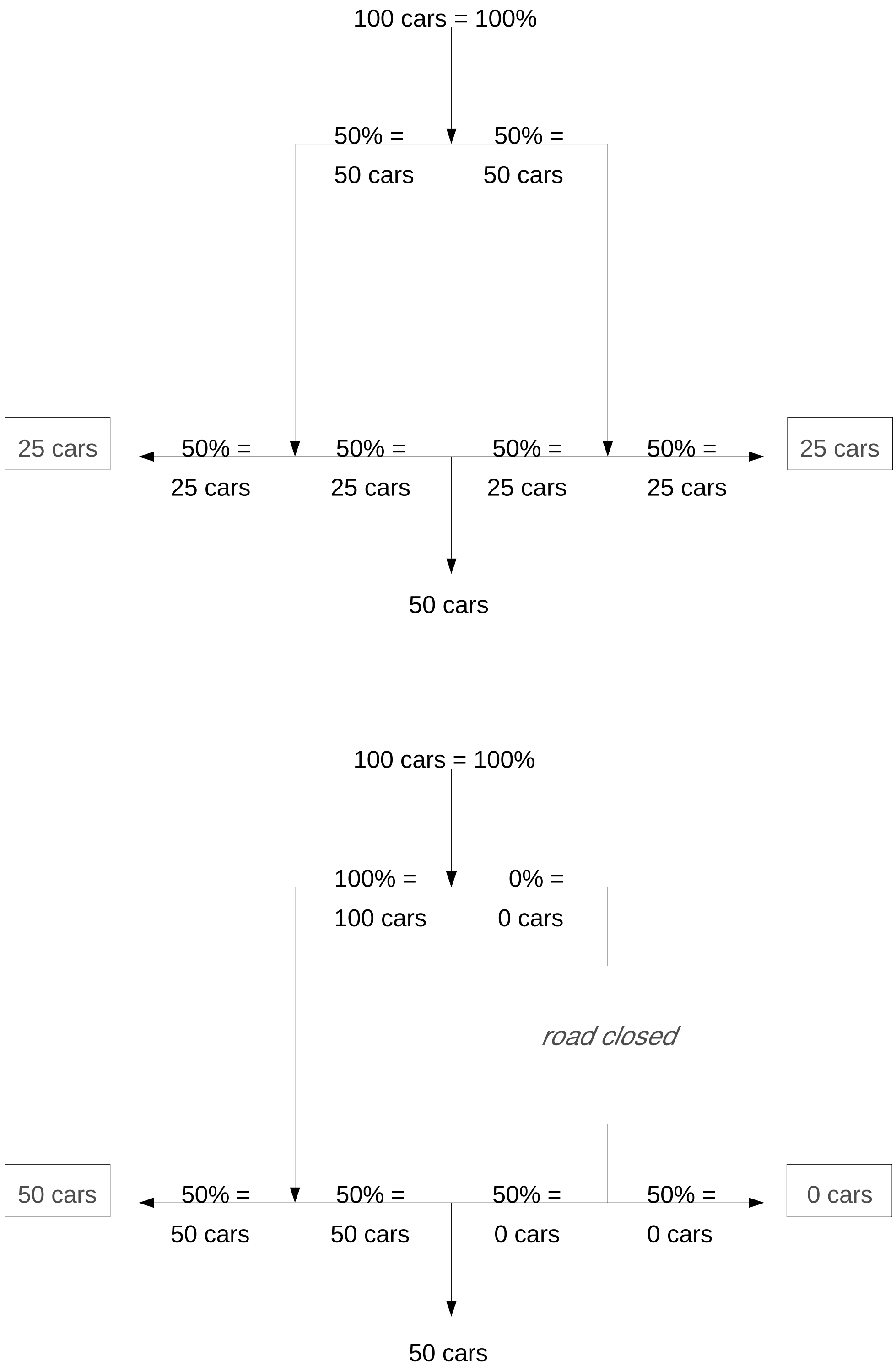

Figure 2.2: Simple example demonstrating the need of microsimulation and travel plans for incident management. *Top:* Undisturbed network. *Bottom:* One road is closed. In both networks, 100 cars are entering from the top, and at junctions they are distributed according to turn-count percentages. From the undisturbed network one sees that 25 cars exit at the right and 25 cars at the left (boxed values). With the closed road, the turn-counts simulation completely misses the correct destination pattern: 50 cars exit to the left, and none to the right (boxed values).

- Planning of trips, routing
- Execution of trips (microsimulation)
- Incident detection
- Meteorological module: air pollution
- Urban development

Current status is that a demo version shows the feasibility of the approach, and a large team of scientist and engineers currently works on its realization. This thesis, in a collaboration between Cologne and Los Alamos, is also part of the TRANSIMS project.

EUROTOPP [153] is a framework similar to TRANSIMS, but is does not have parallel supercomputing plans so far.

PARAMICS, coming from the supercomputing side of traffic microsimulation, is planning to extend towards a framework similar to TRANSIMS.

2.4.4 Nonlinear Dynamics and Statistical Physics: A dynamic systems approach to transportation

So far, together with a general overview over freeway simulation models, it has been shown that microscopic modelling is a necessity for many of the urgent questions. Another desire is the wish for simulation speed. Besides for real time applications, where the simulation has to run many times faster than reality, here another argument will be shown: The necessity of systematic analysis of traffic/transportation as a dynamic system.

Even with a perfect prediction machine, a complete traffic flow forecast would not be possible. Nonlinear Dynamics [152] tells us that tiny changes in the initial conditions can lead to completely changed long-term behavior even in deterministic systems. The best-known example for this is the butterfly effect ("A butterfly in Japan may influence tomorrow's weather in Europe.") [152].

But the situation is not completely lost. Short-term predictions are always possible, and depending on the region in phase space, such a prediction will be better or worse. And in addition it will be possible to identify "average" traffic regimes and to predict their changes under changes of parameters: In the context of the routing problem, one would like to know how different algorithms perform, e.g., in light vs. dense traffic, and if one can generalize these results.

In addition, it will turn out that the optimization process itself affects the system dynamics, in a way that optimization of average quantities (such as overall cost) will lead to high fluctuations of individual quantities: Although the overall connection between two given points may be faster as the result of the optimization, unexpected breakdowns of the whole system may happen more often than before. In short, the optimization procedure becomes part of the system dynamics and can no longer be seen separately as in other optimization problems, such as job scheduling or tour planning [93].

Quite generally, it appears that simulation does not remove the old task of science: The search for structure is as necessary as ever. Clearly, to deal with such questions, a dynamic systems approach to systems such as transportation is necessary. Barrett and Rasmussen [12] describe steps towards formalizing this search for structure; similar questions are currently discussed in the Artificial Life (alife) research community [144].

One of the currently best concepts to characterize the phase space of a spatial stochastic system is "damage spreading" [159] (see above). Investigations like this need a lot of averaging over many

Monte Carlo runs. Since, in addition, things become worse near a critical point ("critical slowing down": Large system sizes and long times are necessary to avoid finite size effects [66]), models useful for this kind of analysis have to run by orders of magnitude faster than traditional forecasting models.

2.4.5 Microscopic high speed models

In consequence, a traffic simulation model with high resolution *and* high speed would be useful. PARAMICS achieves this by plainly setting on parallel supercomputers together with a not too sophisticated car following dynamics. But according to the arguments above, ultimately one has to sacrifice fidelity. The DYNEMO model [154] achieves simulation speed by moving the individual vehicles according to a macroscopic flow, thus indeed reducing car-following fidelity to achieve computational speed.

Another approach is to reduce fidelity directly on the car following level. Similarly to the *lattice gases* in fluid dynamics [160], space, time, and the possible states are discretized. The roadway is separated into boxes, and each box is either occupied by exactly one vehicle, or it is empty. Movement of vehicles is restrained to the boxes. Occupied boxes (i.e. vehicles) can have further properties, such as speed, vehicle type, and so on. In each update step, all these properties have to be moved with the vehicle, which costs computational resources.

Already an early paper on traffic simulation [46] contains an idea which comes close to a lattice gas: Vehicles are restricted to boxes, and the only difference to a lattice gas is that the boxes are not equally spaced. Instead, spacing reflects velocity at certain points of a network: Regions of short spacing allow only small movements in each time-step and thus low velocity.

A modern version of the lattice gas method (including single bit coding) has been introduced for traffic simulation in 1986 [28]. Here, boxes are equally spaced, and vehicles move with different velocities. The model has been developed towards a complete simulation tool allowing different vehicle types and having different modules for multi-lane freeways, two-lane highways with oncoming traffic, and urban traffic [151]. The main problems of this model are:

- The employed bit-coding scheme does not vectorize. As a result, the model did not offer itself for large scale Monte-Carlo simulations on the hardware available at that time. But then, without a large-scale speed advantage, bit-coding is too inflexible for modern traffic science question.
- Vehicle movement is only from one site to the next. The model allowed for velocities up to $v_{max} = 6$, but this had to be achieved by 6 single site movements. In consequence, when the model did not use bit-coding, it was relatively slow.

In recent time and started by the independent publication of [17] and of the first results of this thesis work [120], CA models for traffic are increasingly used, this time by Statistical Physicists, partially for one-dimensional [9, 15, 92, 111, 116, 117, 147, 150, 167, 172] and partially for two-dimensional [30, 112, 119, 118, 166] investigations (i.e. traffic on a square grid). I will mention most of the approaches again in the text.

Chapter 3

A cellular automaton model for traffic flow

In this chapter, a basic model for (single lane) traffic flow is proposed, which is an adaption of the lattice gas method to traffic. Phenomenology and average behavior of the model are studied and compared to reality. The influence of the two free parameters of the model is investigated, followed by some observations concerning calibration.

3.1 This work

According to arguments given in the last chapter, for high simulation speed one ultimately has to sacrifice fidelity, and one acceptedly useful way of doing this is to use the lattice gas methodology. But how much fidelity does one really need?

Due to the lack of "first principles" in traffic science, only an empirical answer can be possible. From all possible elements of driving dynamics, one has to select the most important ones.

This work is exactly about taking this idea to the extreme: Individual cars are preserved, but driving behavior is simplified as much as possible without losing the essentials of traffic jam dynamics. And the simplification matches computer architecture considerations as much as possible.

The limits of simplification are certain properties of vehicular traffic which one wants to preserve, for example spontaneous jam formation. It will come out that the model specified in the next section is "minimal" in the sense that every further simplification leads to losing one of these essential properties.

This approach will turn out to indeed run fast enough to use averaging methods from Statistics / Statistical Physics, for real time applications in large road networks, for algorithm development for dynamical routing, or for long-term urban planning considerations.

Moreover, since a local and simultaneous update will be used, the model will formally be a cellular automaton (CA) [180]. This will simplify further analysis: Many spatial methods of statistical physics have been developed in connection with CA. Analytical treatment is possible (which is, though, laborious, and at the edge of currently available methods, see later), and it allows computational research for finding the essentials of traffic jam dynamics. This is especially useful for traffic near maximum throughput, which is the range which is most interesting for planning purposes but explained worst by conventional methods.

Many models could be used to follow this route. However, I will, by testing state-of-the-art programming techniques such as single-bit coding, argue that a more efficient use of current computer architectures will be difficult to achieve. And I will argue that traffic has measurable "universal" properties which will be reproduced by any microscopic model whatsoever, extensive or not.

3.2 The model

The basic computational model is defined on a one-dimensional array of L sites with open or periodic boundary conditions. This could, for example, be an edge in a road network. Each site is either occupied by one vehicle, or empty. Each vehicle has an integer velocity with values between zero and v_{max}. The number of empty sites in front of a vehicle is denoted by *gap*. For an arbitrary configuration, one update of the system consists of the following four consecutive steps, which are performed simultaneously for all vehicles:[1]

1.) **Acceleration:** If the velocity v of a vehicle is lower than v_{max} and if there is enough room ahead ($v \leq gap - 1$), then the speed is increased by one:

---
if $(v \leq gap - 1)$ **then** $v := \max[v_{max}, v + 1]$.

2.) **Slowing down (due to other cars):** If the next vehicle ahead is too close ($v \geq gap + 1$), speed is reduced to *gap*:

---
if $(v \geq gap + 1)$ **then** $v := gap$.

3.) **Randomization** (*which is applied after rules 1 & 2*): With probability p, the velocity of each vehicle (if greater than zero) is decreased by one:

---
with probab. p **do** $v := \max[v - 1, 0]$.

4.) **Car motion:** Each vehicle is advanced v sites.

Through the steps one to four very general properties of single lane traffic are modelled on the basis of integer valued probabilistic cellular automaton rules [160, 180]. Already this simple model shows nontrivial and realistic behavior.

The Monte Carlo simulations have mainly been carried out with the choice of $v_{max} = 5$ for reasons stated below. It should be noted that the term maximum speed does not reflect the physically possible maximum speed of each car. It reflects, however, a maximum speed as attained in reality, restricted, for example, by speed limits or by preferences of the driver.

Step 3 is essential in simulating realistic traffic flow since otherwise the dynamics is completely deterministic. It condenses three different behavioral patterns into one computational rule:

- **Fluctuations at maximum speed:** Assume a vehicle driving at maximum velocity $v = v_{max}$ and with $gap \gg v$ (free driving). The velocity remains unchanged in rules 1 & 2; in rule 3, it is, with probability p, reduced to $v_{max} - 1$. In the next iteration, according to rule 1, the speed $v_{max} - 1$ is algorithmically reset to v_{max}, before the randomization can again reduce it with probability p. Thus, according to the algorithm, a free vehicle moves with frequency $1 - p$ with $v = v_{max}$, and with $v = v_{max} - 1$ otherwise.

[1] Note that *either* rule 1 *or* rule 2 applies, but never both.

This reflects the fact that people even in light traffic do not keep their speed perfectly constant, for example due to distractions.

- **Retarded acceleration:** Assume a single vehicle with $v = 0$. Taking rules 1 & 3 together, the vehicle accelerates to $v = 1$ with probability $1 - p$. The deterministic acceleration sequence

$$0 \to 1 \to 2 \to \ldots \to v_{max}$$

is thus replaced by a multitude of stochastic sequences, for example

$$0 \to 0 \to 0 \to 1 \to 2 \to 2 \to \ldots \to v_{max} \ .$$

It seems realistic that people often wait longer than necessary when the car ahead already moves.

- **Over-reactions at braking:** When people suddenly see slow traffic ahead, they tend to over-compensate. In the rules, instead of reducing the speed exactly to gap, with frequency p they reduce it further to $gap - 1$ (if ≥ 0).

Without this randomness, every initial configuration of vehicles and corresponding velocities reaches very quickly a stationary pattern which is shifted backwards, i.e. opposite to the vehicle motion, one site per time step (see Chapter 4).

In the framework of car-following models described in Chapter 2, the CA traffic model is difficult to classify. It only uses spacing as a stimulus; the braking rule is similar to a model where the velocity is proportional to the spacing, whereas the acceleration is constant as soon as there is enough room ahead.

When judging the realism of the model beyond these general remarks, one should keep in mind that this model is intended to be a *minimal* microscopic model. As already mentioned above, I will in the end argue that the model goes beyond all of the current fluid-dynamical theory of traffic jam dynamics, and that even for this, most of the exact microscopic properties do not matter. Only *after* having identified the important elements, it will certainly make sense to think about making them more realistic (which will usually mean slower computational speed, though).

For that same reason, this text concentrates on single-lane models. Multi-lane results obtained by other authors will be shortly mentioned in Chapter 12.

Certain choices of the model were originally done for computational reasons, especially for a bitwise implementation:

- Making the model grid-based: (a) Theory for evaluating spatial dynamic systems are much better developed for (few state) grid based models. (b) In view of a planned single-bit implementation, there is no alternative.

- Velocity of the next car ahead is not taken into account: In a bitwise implementation, just checking bit-fields for the presence of a car is much easier to implement (and thus runs much faster) than complicated calculations based on some number different from 0 or 1.

- The choice of the randomization parameter of $p = 1/2$ and the fact that is is equal for acceleration, braking, and free driving, also reflects the fact of an intended bitwise implementation: A bitwise probability of 1/2 is fast to obtain—a randomly selected integer number has its bits randomly occupied with probability 1/2. Other bitwise probabilities are more costly. Random number generation typically consumes about 20% of the computer time for simulations of this traffic flow model.

- Random sequential (cf. Section 2.2) versus simultaneous versus event driven update: Event driven update means that each object keeps on doing the same until some stimulus exceeds a certain threshold—for example that another car comes too close. The most realistic update obviously is the event driven update; and random sequential update is claimed to be a better approximation of it than parallel update because in reality it is more probable that the events occur randomly than in some ordered fashion.

  The exact consequences of the parallel choice for the traffic model will be discussed much later. Roughly, it turns out that for random sequential update to be realistic one would have to make the time scales much shorter. For the CA traffic model, it is certainly much more realistic to assume that all cars move simultaneously. Another reason is computational: For both discrete event simulation and random sequential update an efficient geometric parallelization is hard to achieve:

  For discrete event simulations one needs some more or less optimistic scheduling schemes of running ahead and rolling back [19, 45] to be efficient on parallel machines, which is much harder to implement.

  For random sequential update the problem is that a global determination of the next (randomly chosen) update site obviously is completely inefficient. Localizing this procedure, though, leads to the introduction of a new length scale (the system size between two computational nodes), which might destroy critical properties. Or one would have to implement a scheduler similar to the one for the discrete event simulation.

In an evaluation of the computational speed (Chapter 11), one finds that the bitwise implementation is not much faster on non-vectorizing parallel computers, which are currently the majority of installed supercomputer hardware. Moreover, for a hybrid model together with a more realistic car following logic, which is planned in the TRANSIMS project, the bitwise implementation is useless; the same is already true for the trip plans implementation of Chapter 10. Both arguments together make the use of a bitwise implementation of the model for the future improbable; some of the restrictions above could therefore be relaxed in future work.

3.3 Model vs. reality

Modeling traffic flow as particles which jump from one box to another seems like a quite crude approximation of reality. Can one expect realism from the simulation of such a model? As a first answer to this question, I will make a phenomenological and a quantitative comparison to reality which will show that the model reproduces some aspects of reality rather well. Theoretic arguments supporting this claim will be given later in the text.

3.3.1 Trajectories of cars

A useful way to visualize traffic and traffic jams are space-time diagrams (Fig. 3.1). One can follow the trajectory of each car from top left to bottom right; fast cars have trajectories close to horizontal, slow cars have trajectories close to vertical. An exactly vertical trajectory means that the vehicle does not move at all. In Fig. 3.1 one clearly sees a jam wave, recognizable by the vertical trajectories denoting slow vehicles, and by the high density in that region (trajectories close together). Note that the jam wave moves backwards, i.e. *against* the traffic.

Figure 3.1: Space-time-lines (trajectories) for cars from Aerial Photography (after [168]). Each line represents the movement of one vehicle in the space-time-domain.

Figs. 3.2 and 3.3 show the corresponding plots from the cellular automaton. These plots are standard for cellular automata outputs. Lines are configurations at consecutive time-steps, integer numbers show velocities of cars just before movement. For traffic flow, the plots are at the first view somewhat unusual. Nonetheless, one can follow a car individually from left to right, and if one would connect the positions of one car in space-time, one would get the same trajectories as in Fig. 3.1.

Figs. 3.2 and 3.3 show typical situations at low and high densities. Whereas we find laminar traffic at low densities, there are congestion clusters (small jams) at higher densities, which are formed randomly due to velocity-fluctuations of the cars. If one follows (in Fig. 3.3) one individual car coming from the left, one sees that the car comes in with a speed varying between four and five and then has to stop due to the congestion cluster. There it stays stuck in the queue for a certain time with some small moves, and can re-accelerate to full speed after having left the cluster at its end. So the cluster represents a typical start-stop-wave as found in freeway traffic, and compared to Fig. 3.1, it looks rather realistic.

Figs. 3.4 and 3.5 show larger portions of the system. Vehicles are reduced to dots, i.e. the plot just shows the space-time-position of each vehicle during the evolution.

```
.........................4................................5...........5
.............................4.................................5.......
.5...............................5..................................4..
5.....5...............................5................................
.....4.....4...............................5...........................
.........5.....4................................5......................
.5............3....4.................................4.................
......5..........3.....4.................................4.............
...........5........3......5.................................5.........
................5......4........4.................................4....
.....................4.....5........5.................................4
.........................5......4........4.............................
..............................5.....5........5.........................
```

Figure 3.2: Simulated traffic at a (low) density of 0.03 cars per site. Each new line shows the traffic lane after one further complete velocity-update and just before the car motion. Empty sites are represented by a dot, sites which are occupied by a car are represented by the integer number of its velocity. At low densities, we see undisturbed motion.

```
........4................5......5......5........00.1.......4..........5
..4.........5.................4......5......2...00..2..........5.......
......4..........5................4.......3...0.01....2.............5..
4.........4...........4...............5......01.1.2.....3..............
....5.........5...........5................0.1.0.2..3......4...........
.........5.........4...........4...........1..00...3...3.......5.......
.4............4........4...........5........1.00......3...3.........5..
.....4............5........4............3....000.........2...3.........
.........4.............5.......4...........0.001...........3....3......
.............4..............5......4.......0.00.2.............4....4...
4................4...............5.....2..1.01....2...............3....
....4................4................2..1..01.2....3................4.
........5................5..............0.1.0.2..2.....3...............
.............5................5.........0..00...1..2......3............
..4...............4................3....1..00....2...3.......3.........
......5...............4...............2..0.00......2....3.......3......
..5........4..............4.............01.00........3.....4.......4...
.5.....4.......5..............5.........0.000...........3......4.......
......4....5........5..............4....1.001..............3.......4...
4.........5.....5........4.............1.000.1................4........
....5..........4.....5.......5..........0000..1...................5....
.........4.........5......4.......4.....0000...2.......................
.............5..........4.....5.......0.0001.....2.....................
..................4.........4......1..0.000.2......2...................
.5....................4.........2...0.1.000...3......3.................
......5...................5.......1.0..0001......4......4..............
..5........5...................2...01..001.2.........4......5..........
.......5........4................1.0.0.01.1..3...........5.......4.....
....5.......5.......5.............00.0.0.1.1....3.............4......5.
.........4.......5.......5........01.1.0..0.1......3..............4....
.............4........5.......2...1.0.00..1..2........4...............5
.................5.........4....2..01.01...2...3..........5............
5.....................4........2..01.00.2....3....3............4.......
.....4....................5......00.000...3.....3....4.............5...
```

Figure 3.3: Same picture as Figure 1, but at a higher density of 0.1 cars per site. Note the backward motion of the traffic jam.

3.3.2 Fundamental diagrams

For the next reality check, more aggregated quantities are used. We start with systems with periodic boundary conditions (thus simulating traffic in a closed loop as in car races but only on a single lane).

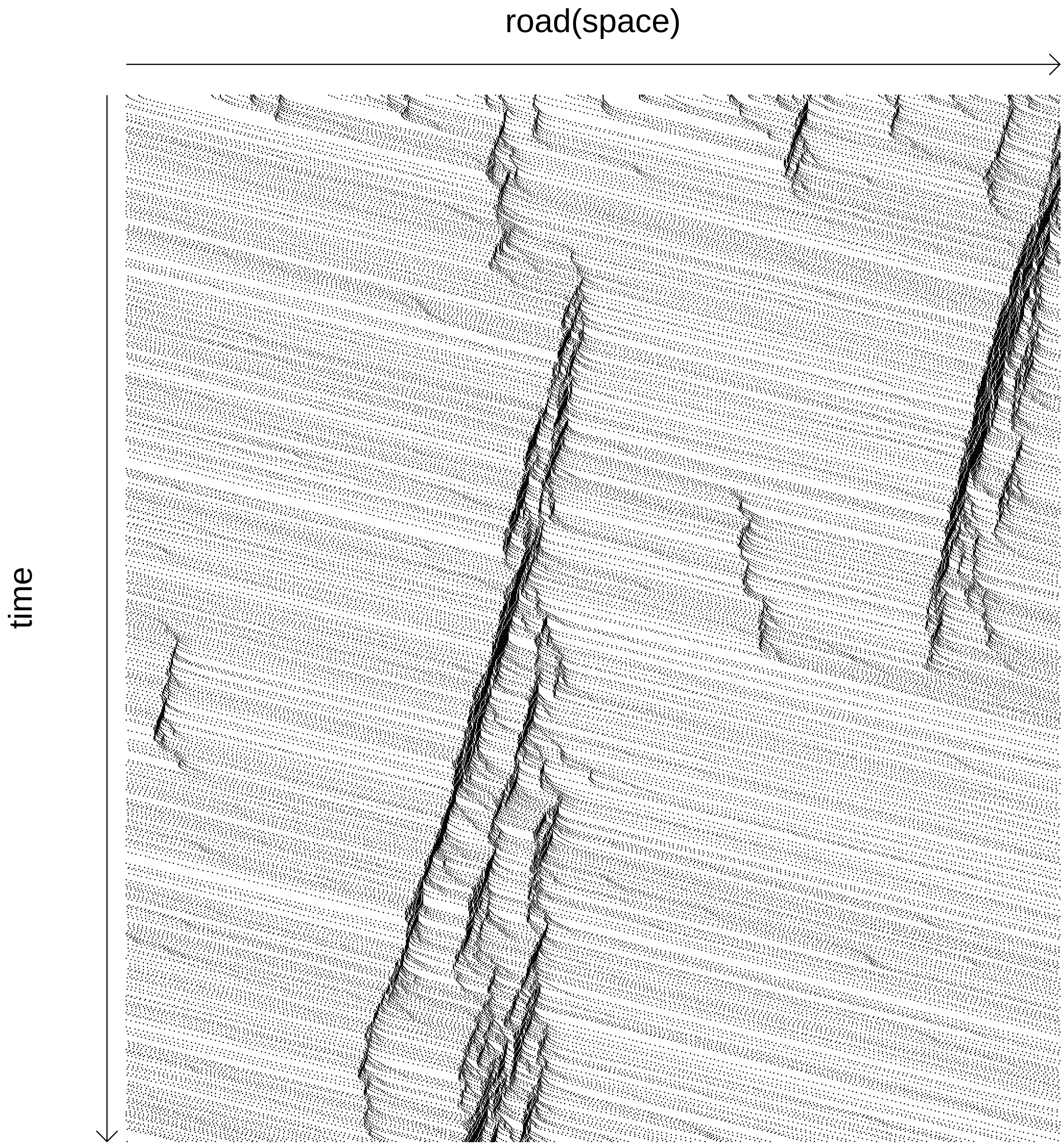

Figure 3.4: Evolution of the model from random initial conditions. Each black pixel represents a vehicle. Space direction is horizontal, time is pointing downwards, vehicles move to the right. The simulation was of a system of size $L = 10000$ with density $\rho \approx \rho(q_{max}) \approx 0.08$; the figure shows the first 1000 iterations in a window of $l = 1000$.

As the total number N of cars in the loop cannot change during the dynamics, it is possible to define a constant system density

$$\langle \rho \rangle_L = \frac{N}{L} = \frac{\text{number of cars in the loop}}{\text{number of sites of the loop}} \; .$$

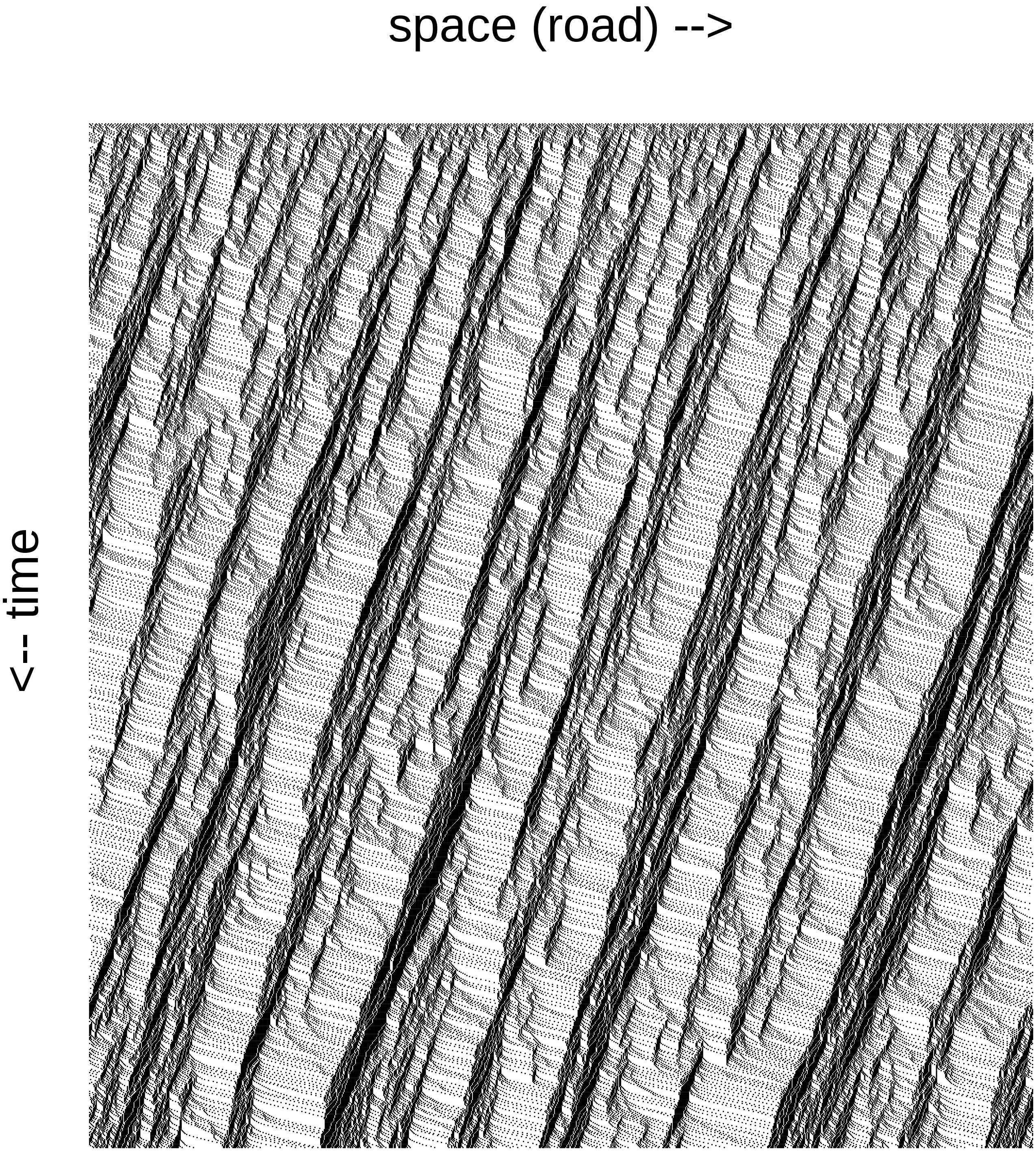

Figure 3.5: Evolution of the model from random initial conditions. Each black pixel represents a vehicle. Space direction is horizontal, time is pointing downwards, vehicles move to the right. The simulation was of a system of size $L = 10000$ with density $\rho \approx \rho(q_{max}) \approx 0.08$; the figure shows the first 1000 iterations in a window of $l = 1000$.

Other system-averaged quantities are the "travel velocity" $\langle\, v^{trav} \,\rangle_N$ and the flow $\langle\, q \,\rangle_L$:

$$\langle\, v^{trav} \,\rangle_N = \frac{1}{N} \sum_{i=1}^{N} v_i \ , \qquad\qquad \langle\, q \,\rangle_L = \frac{1}{L} \sum_{i=1}^{N} v_i \ ,$$

where v_i is the velocity of the ith car and the sum runs over all cars. v^{trav} is in the average identical to the velocity each driver measures for himself when traveling, therefore the name.

For practical reasons, traffic engineers prefer quantities which can be measured at a fixed position. Thus, in order to mimic real conditions, we measure[2] densities (= occupancies) $\langle \rho \rangle_T$ on a fixed site i averaged over a time period T:

$$\langle \rho \rangle_T = \frac{1}{T} \sum_{t=t_0+1}^{t_0+T} \delta_i(t) \ ,$$

where $\delta_i(t) = 0$ (1) if site i is empty (occupied) at time step t.

The time-averaged flow $\langle q \rangle_T$ between sites i and $i+1$ is defined by

$$\langle q \rangle_T = \frac{1}{T} \sum_{t=t_0+1}^{t_0+T} \delta_{i,i+1}(t)$$

where $\delta_{i,i+1}(t) = 1$ if a car motion is detected crossing the boundary between sites i and $i+1$.

For large T, large L, and periodic boundaries one has $\langle \rho \rangle_T = \langle \rho \rangle_L$ and $\langle q \rangle_T = \langle q \rangle_L$.

The *local velocity* is defined as

$$\langle v^{loc} \rangle_T := \frac{1}{n(T)} \sum_{j=1}^{n(T)} v_j$$

where $n(T)$ is the number of vehicles which cross the boundary $i \to i+1$ during the time interval T, and v_j are the vehicles' velocities. It should be noted that v^{trav} generally is lower than v^{loc} due to differences in the statistical sampling. For example, vehicles of velocity $v = 0$ are included for v^{trav} but excluded for v^{loc} (see, e.g., [183]).

With these definitions, it is easy to perform many simulations with different vehicle densities, thus—after relaxation to equilibrium—getting data for the commonly used fundamental diagrams flow $\langle q \rangle_T$ vs. density $\langle \rho \rangle_T$, or $\langle v^{loc} \rangle_T$ vs. $\langle \rho \rangle_T$, etc.

Fig. 3.6 was obtained starting with a random initial configuration of cars with density ρ and velocity 0 and beginning the collection of data after the first t_0 time steps, typically $t_0 = 10^4$. Whereas the broken line indicates the results of averaging over 10^6 time steps, the scattered points represent averages over only 10^2 time steps. The latter should be compared to data from real traffic (Fig. 3.7). For low densities, flow increases linearly with density, reflecting the fact that in this regime adding vehicles does not decrease the velocity and thus just adds to the flow: $q_{low.density} = \rho \, v_{max}$.[3] High densities are similarly easy to explain: Many cars have few holes between them. When a car moves forward, the hole effectively moves backwards. In that sense, flow is due the movement of the holes between jammed cars.

In between, no easy explanation is available. For the long-time averages the maximum flow $q_{max} = 0.318 \pm 0.0005$ is found at the density $\rho = 0.085 \pm 0.004$.

Further simulations show that the position and the form of the maximum of $q(\rho)$ depend on the system size (Fig. 3.6, see [92] for a more detailed analysis). Simulations without randomization, i.e. $p = 0$, do not show a dependence on the system size.

[2] I use the term "measurement" both for reality and simulation, somewhat similar to a "measurement operator" in quantum mechanics or in statistical physics which extracts averaged information from a microscopic system. When it is not clear from the context, I will try to specify if I am referring to reality or to a simulation. E.g., "traffic measurement" refers to measurements of real traffic.

[3] This relation is no longer linear in ρ as soon as one includes passing as well as vehicles of different speed [141].

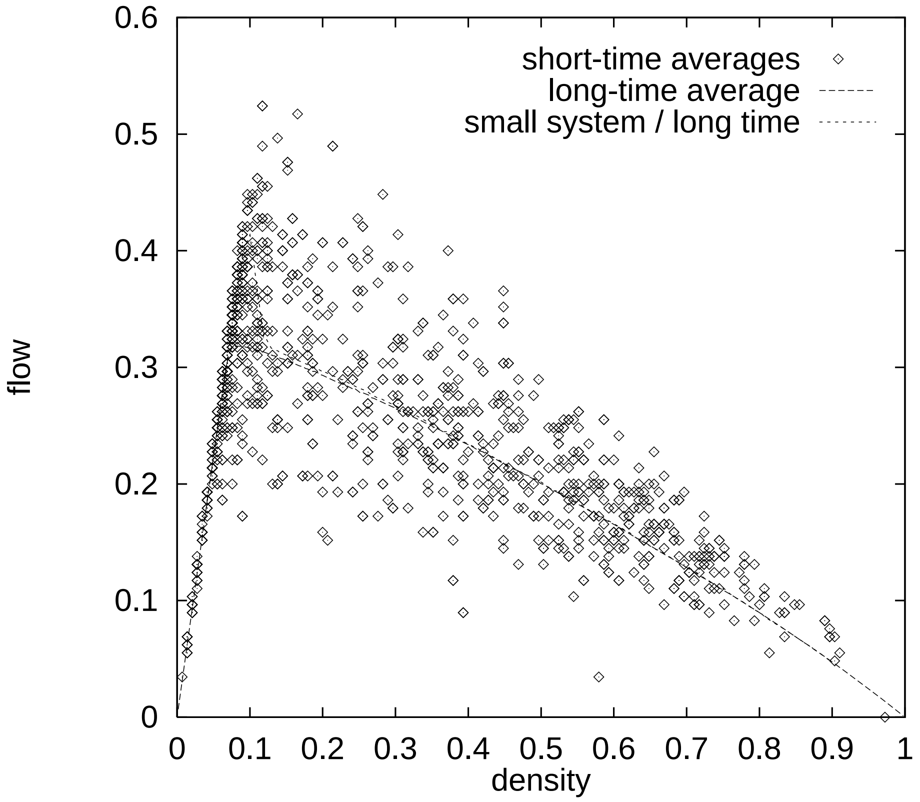

Figure 3.6: Fundamental diagram of the model (throughput versus density). Points: Averages over short times (100 iterations) in a sufficiently large system ($L = 10,000$). Broken line: Long time averages (10^6 iterations) in a large system ($L = 10,000$). Dotted line: Long time averages in a small system ($L = 100$).

3.3.3 Fluctuations

The variance of v^{loc} is defined as

$$\sigma(v^{loc}) = \sqrt{\frac{1}{N} \sum_i v_i^2 - (v^{loc})^2} \ .$$

This quantity may be used for detecting instabilities in the traffic flow and has in fact been used in Germany to install adaptive speed limits [183]. Fig. 3.8 shows that also in the model $\sigma(v^{loc})$ drastically increases near maximum flow.

3.4 Varying the parameters of the model

In order to get an overview of the influence of the model's two parameters v_{max} (the maximum integer speed) and p (the amount of fluctuation included by the pseudo-random numbers), simulations with different values of these parameters were performed. A relatively large closed system ($L = 10^6$ sites) was used, where it was known from previous tests that this is large enough to prevent visible finite size effects for the measured quantities. Starting from an initial random distribution of vehicles, the system run for 10^4 time-steps in order to let the transients die out. Then, every 10^3 iterations one measurement step was inserted, and the average of 100 such measurements gives one data point. Each

Figure 3.7: Fundamental diagram as found in reality from five-minute averages. Data from a Canadian expressway [56].

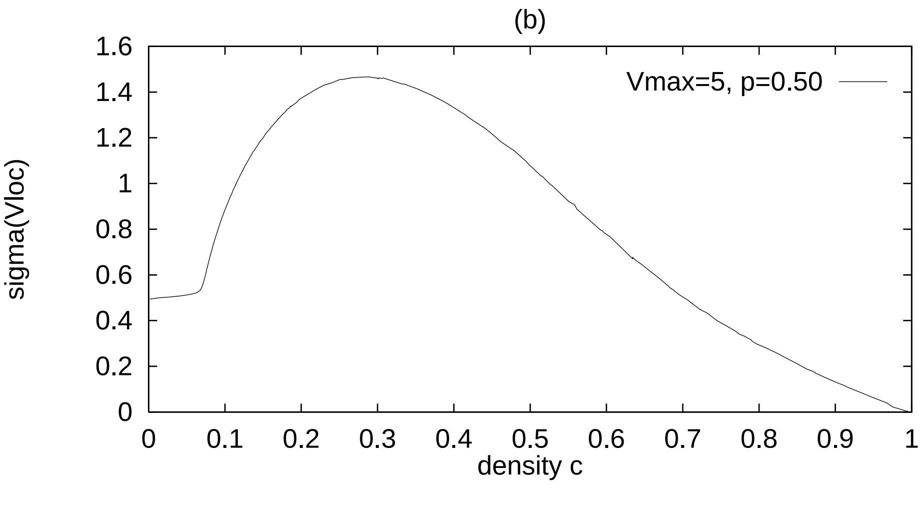

Figure 3.8: Variance of the local velocity.

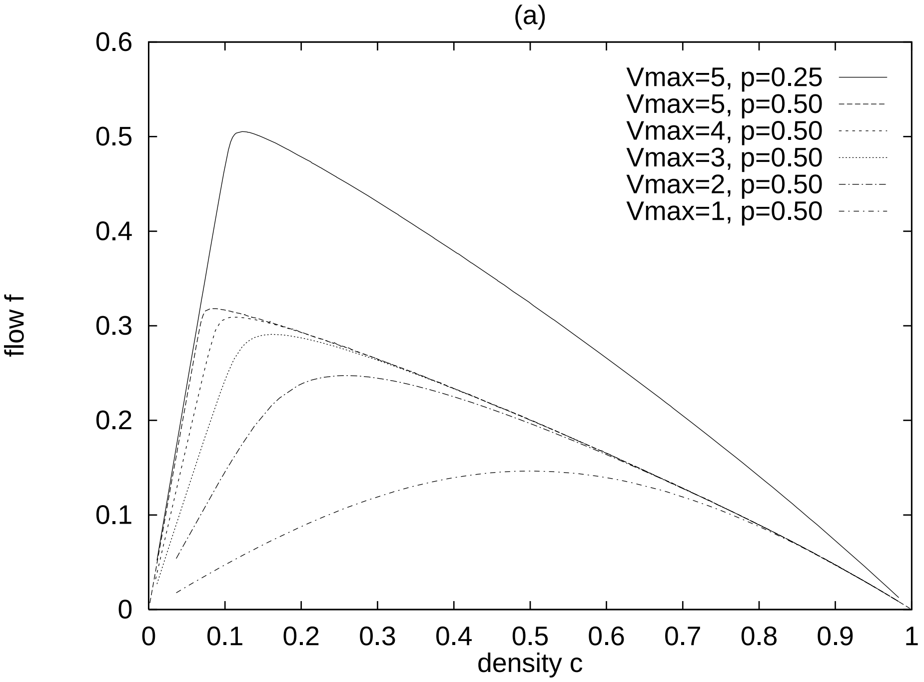

Figure 3.9: Different fundamental diagrams obtained by variations of the two principal parameters of the model p and v_{max} (the latter denoted as V in the legends).

curve for a given pair of parameters (v_{max}, p) contains about 100 such data points, which corresponds to about 6 hours on the NEC or 12 hours on the Parsytec (see Chapter 11 for computational details).

The results are summarized in Figs. 3.9, 3.10, and 3.11, where the three standard fundamental diagrams are given for the model for $v_{max} = 1, ..., 5$ (all with $p = 0.5$) and for $p = 0.25$ and $v_{max} = 5$. Clearly, the model may be adapted, by only varying these two parameters, to a wide range of circumstances (e.g., the influence of bad weather, slopes, ...) [63].

3.5 Quantitative comparison with realistic traffic

In this section, some rough arguments concerning the length scale and time scale of the simulation model are made. The easiest approach to calibrate the model is the claim that in a dense jam each car occupies about $7.5\,m$ of space, which is thus the length of one site, in agreement with [151]. Since the average velocity in free traffic of 4.5 sites per time step should correspond to a velocity of about 120 km/h (in Germany), one arrives naturally at a time for one iteration of

$$7.5\,\frac{\mathrm{m}}{\mathrm{site}} \times 4.5\,\frac{\mathrm{sites}}{\mathrm{time\text{-}step}}/(120/3.6)\,\frac{\mathrm{s}}{\mathrm{m}}$$

$\approx$ (1 second per time-step) which again agrees with [151].

A second possibility is to scale the model by the position of the maximum in the fundamental diagram. From traffic measurements, this maximum is found at about $\rho \simeq$ (30 vehicles per lane and kilometer) = (0.225 vehicles/7.5 m), which is by a factor of about 2 higher than the position of the maximum in the scatter-plot for our model. However, this changes when two-lane traffic with slower vehicles is simulated [141].

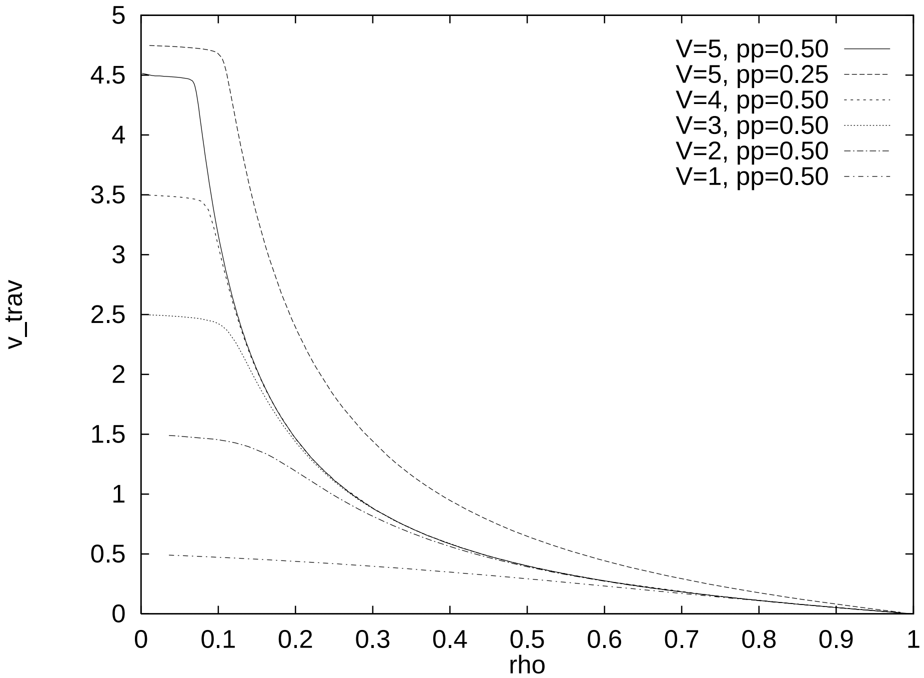

Figure 3.10: Travel velocity v^{trav} as a function of the density. The model very well describes the nearly constant regime for $\rho < \rho(q_{max})$ and the strong decrease of the velocity near maximum flow.

Similarly, freeways have a maximum capacity of about (2000 vehicles per hour and lane) = (0.56 vehicles per second). As our maximum of the flow is only 0.32 vehicles per time step, our model time step should correspond to $0.32/0.56 \approx 0.5$ seconds, thus being by a factor of two lower from the value presented above. However, the definition of "maximum flow" is somewhat unclear. One would think that averaging the reality measurements in Fig. 3.7 in the same way as the simulations would also lead to a much lower flow than 2000 veh/h. In other words: In the simulation, we have an obvious definition for an average maximum flow. It seems however that this definition is not consistent with the more heuristic definition of traffic practitioners.

A fourth possibility for a calibration uses the value of the velocity of the back-travelling start-stop-waves, where a value of about $15\,km/h \approx 4.2\,m/s$ has been measured on freeways. As the maximum capacity of our simulated system is about 0.3 cars per time step, the maximum speed of the back-propagating wave, assuming $\rho = 1$ inside jams, is 0.3 sites ($\approx 2.25\ m$) per time step (i.e., about every third time step a new car arrives at the back of the traffic jam). This would fix one model time step at $2.25/4.2 \approx 0.7$ seconds, thus yielding a value between those of the first and of the third method. However, due to the inexact braking, the density inside jams is smaller than one, thus yielding a higher speed of the backpropagating wave in the model.

In any case, all estimates agree in order of magnitude: One iteration roughly correspond to one second, and one box roughly corresponds to 7.5 m. Since calibration of fundamental diagrams is, as mentioned, different for multi-lane simulations [141], and since the more theoretical results do not depend on the calibration, no further effort for calibration will be done in this work. Calibration for practical applications is the topic of future research.

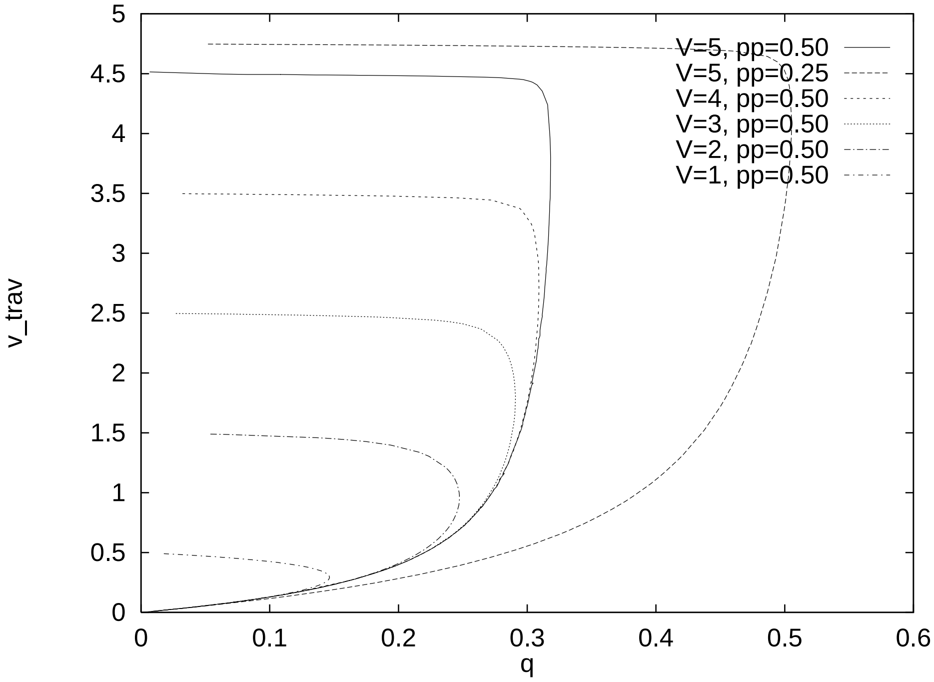

Figure 3.11: Relation between travel velocity v^{trav} and throughput q.

One should note that, contrarily to intuition, the parameters for a discrete traffic flow model are rather fixed once one accepts "one car per lattice site in a jam". The parameters used for the simulations seem relatively reasonable.