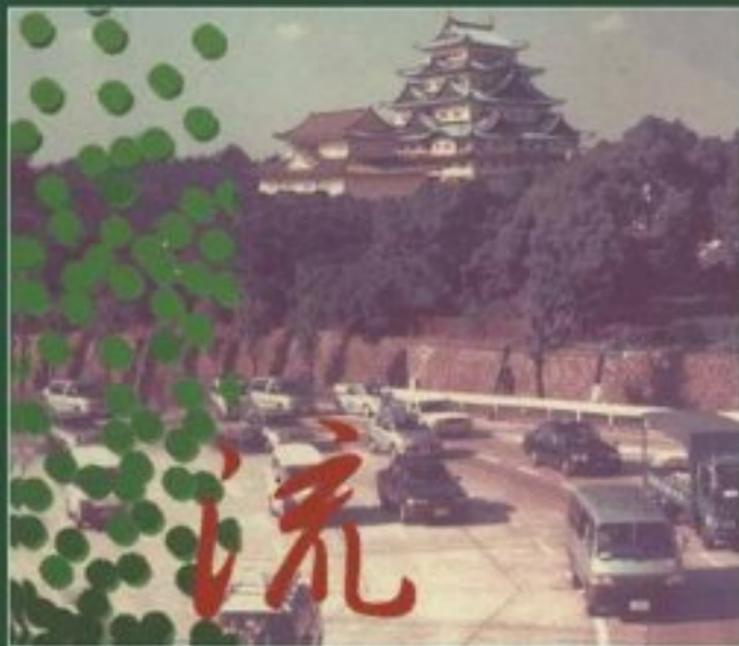


M. Fukui Y. Sugiyama  
M. Schreckenberg D. E. Wolf  
Editors

# TRAFFIC AND GRANULAR FLOW '01



Springer

# TRAFFIC AND GRANULAR FLOW '01



M. Fukui Y. Sugiyama  
M. Schreckenberg D.E. Wolf  
*Editors*

# TRAFFIC AND GRANULAR FLOW '01

With 368 Figures, 32 in Colour



Springer

## Editors

Minoru Fukui  
Nakanihon Automative College  
Fukagaya, Sakahogi-cho, Gifu-ken, 505-0077, Japan  
e-mail: fukui3@cc.nagoya-u.ac.jp

Yuki Sugiyama  
School of Informatics and Sciences  
Nagoya University  
Furouchou, Chikusaku  
Nagoya 464-8601, Japan  
e-mail: genbey@eken.phys.nagoya-u.ac.jp

Michael Schreckenberg  
Universität Duisburg-Essen  
Physik von Transport und Verkehr  
Lotharstraße 1  
47048 Duisburg, Germany  
e-mail: schreckenberg@uni-duisburg.de

Dietrich E. Wolf  
Universität Duisburg-Essen  
Theoretische Physik  
Lotharstraße 1  
47048 Duisburg, Germany  
e-mail: wolf@comphys.uni-duisburg.de

Mathematics Subject Classification (2000):  
90Bxx, 60K30, 76T25, 65Cxx, 68M20, 37M10, 60H35

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek  
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;  
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

ISBN 978-3-642-07304-5 ISBN 978-3-662-10583-2 (eBook)  
DOI 10.1007/978-3-662-10583-2

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag Berlin Heidelberg GmbH. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 2003

Originally published by Springer-Verlag Berlin Heidelberg New York in 2003  
Softcover reprint of the hardcover 1st edition 2003

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: *design & production* GmbH, Heidelberg  
Typeset by the authors using a Springer L<sup>A</sup>T<sub>E</sub>X macro package

Printed on acid-free paper 46/3143LK - 5 4 3 2 1 0

## Welcome Address

Good morning, Ladies and Gentlemen!

Welcome to Nagoya and Nagoya University.

It is my great pleasure to say some words at the opening of this conference on behalf of the host institute of this conference, Center for Integrated Research in Science and Engineering of Nagoya University. The Center is organized as an institute to serve collaborating research and education on the new frontiers of materials science, energy science and global environment science.

I am the coordinator of the project entitled “Traffic simulator based on mathematical models and data base of highway traffic flow” which is one of the interdisciplinary projects of the Center proposed by those outside of the Center but still within Nagoya University. The organizers of this conference include the principal members of the project, and they planned to have an international workshop “Traffic and Granular Flow ‘01” in Nagoya University. I am certainly glad to have the fourth of an international series of conferences for traffic and granular flow dynamics within the period of the project.

Now, I should mention that this time is the hottest time ever of Nagoya University, since as many of you probably know, last week one of our colleagues of Nagoya University, Professor Noyori, got a Nobel prize of chemistry for the first time in the 21st century. This conference is actually the first conference in the campus after the Nobel prize fever at Nagoya University.

I understand that the traffic problem is one of the most urgent problems to be solved in the 21st century. I hope this conference, the first one in the Century, will be looked back in future as that held in the right place and at the right time.

I wish you a great success of the conference and to enjoy Nagoya and Nagoya University.

Thank you.

Yoshito Itoh  
Professor, Center for Integrated Research in Science and Engineering  
Director of University Library  
University Senator  
Nagoya University

## Preface

The biannual international conference *Traffic and Granular Flow* has become an indispensable forum for researchers in this interdisciplinary field, which is rapidly evolving and addresses some of the most urgent scientific and technological challenges. After Jülich, Duisburg and Stuttgart, the fourth conference in this series was the first one to take place outside Germany. Nagoya (Japan) provided a perfect setting to host the 96 participants from 11 nations. Again, many new ideas and developments were discussed.

One highlight of the conference was certainly the dispute between Boris Kerner and Dirk Helbing, which touched upon the very basics of traffic modeling. Boris Kerner passionately promoted his idea that a two dimensional continuum of homogeneous synchronized states in the flow-density plane must be at the core of any realistic traffic model (three phase traffic theory). Obviously the dynamics, in particular the response to perturbations like an on-ramp or to noise, is bound to be fundamentally different in such a model compared to the usual models (propagated by Dirk Helbing) with a fundamental diagram, where the homogeneous states lie on a curve in the flow-density plane.

Unforgettable also the video record of an experiment performed by the project team of Japanese researchers (which is reported in the talk by Makoto Kikuchi), where some 20-30 cars were driving one behind the other in a large round-about until finally a strong enough fluctuation occurred to trigger a jam.

Further topics in the focus of this conference were pedestrian traffic and scale-free networks like the internet. The flow of granular media was not so much at the focus this time, but there were excellent posters discussing new phenomena like the clustering during avalanches or segregation of ellipsoids upon shaking. An impressive account of the complex physics of grains falling through a vertical pipe was given by Mitsugu Matsushita.

We are very grateful for the generous support of several sponsors, first of all the DaimlerChrysler AG, but also the Ford Motor Company, the Center for Integrated Research in Science and Engineering in Nagoya University, the Research Foundation for the Electrotechnology of Chubu, and the Daiko foundation. Without this funding the conference would not have been such a success.

We like to thank the conference secretaries: Atsushi Honda, Takahisa Kajimoto, Namiko Mitarai, Masayo Miyano, Shizuka Nishiura, and Akiko Okumura, as well as the technical and organizational staff: Yoshihiro Ishibashi, Makoto Kikuchi, Katsuhiro Nishinari, Akihiro Nakayama, Shin-ichi Tadaki, Satoshi Yukawa, Tamaki Fukui, and Katsuya Hasebe for the internet service. The manuscripts were prepared for publication by Andreas Keßel, Hubert Klüpfel, Torsten

VIII Preface

Huisinga, Robert Barlovic, Matthias Woltering, Andreas Pottmeier, Roland Chrobok, Guido Bartels, Stephan Dammer, Zeno Farkas, Dirk Kadau, Kwangho Park, Jochen Werth, and Frank Westerhoff. Special thanks go to Birgit Dahm-Courths, who compiled all the manuscripts of the conference proceedings into this book.

The next conference *Traffic and Granular Flow 2003* will take place in Delft (Netherlands), and we hope to see many of you again there.

Duisburg / Nagoya, July 2002

*Yuki Sugiyama*  
*Minoru Fukui*  
*Michael Schreckenberg*  
*Dietrich E. Wolf*

# Contents

<b>Welcome Address</b> .....	V
<b>Preface</b> .....	VII
<b>Contents</b> .....	IX
<b>List of Participants</b> .....	XV

## Traffic Models: Theory

Euler and Lagrange Representation of Traffic Models .....	3
<i>K. Nishinari</i>	
Three-Phase Traffic Theory .....	13
<i>B.S. Kerner</i>	
Spacing-Oriented Analytical Approach to a Middle Traffic Flow CA Model Between FI-Type and NS-Type.....	51
<i>B.-H. Wang, D. Mao, L. Wang, and P.M. Hui</i>	
Continuum Traffic Equations from Microscopic Car-Following Models .....	65
<i>H.-W. Lee, H.K. Lee, H.Y. Lee, and D. Kim</i>	
Metastable Flows in an Extended Burgers Cellular Automaton Traffic Model .....	79
<i>M. Fukui, K. Nishinari, D. Takahashi, and Y. Ishibashi</i>	
Existence and Classification of Travelling Wave Solutions to Second Order Highway Traffic Models .....	85
<i>R.E. Wilson and P. Berg</i>	
Improved Optimal Velocity Model for Traffic .....	91
<i>A. Shibata</i>	
Microscopic Modeling of Synchronized Traffic .....	97
<i>W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg</i>	
Limit Sets and the Rate of Convergence for One-Dimensional Cellular Automata Traffic Models .....	103
<i>T. Namiki</i>	
Localized Defects in a Cellular Automaton Model for Traffic Flow with Phase Separation .....	109
<i>A. Pottmeier, R. Barlovic, W. Knospe, A. Schadschneider, and M. Schreckenberg</i>	
Asymmetric Optimal Velocity Model.....	115
<i>A. Okumura and S. Tadaki</i>	

Chaos and Multifractality in a Time-Delay Car-Following  
Traffic Model ..... 119  
*L.A. Safonov, E. Tomer, V.V. Strygin, Y. Ashkenazy, and S. Havlin*

**Traffic Models: Application**

Optimal Velocity Model and its Applications ..... 127  
*A. Nakayama, K. Hasebe, and Y. Sugiyama*

Breakdown and Recovery in Traffic Flow Models ..... 141  
*K. Nagel, C. Kayatz, and P. Wagner*

Kinetic Theory of Traffic Flows ..... 155  
*E. Ben-Naim and P.L. Krapivsky*

The Generalized Fundamental Diagram of Traffic and Possible  
Applications ..... 169  
*E. Tomer, L.A. Safonov, and S. Havlin*

Stochastic Resonance Towards Traffic Models ..... 187  
*T. Ohira*

An Interpretation of a Traffic Engineer on Vehicular Traffic Flow ..... 199  
*M. Koshi*

Congestion Induced by Bottlenecks in Two-Lane Optimal Velocity  
Traffic Flow Model ..... 211  
*S. Tadaki, M. Kikuchi, K. Nishinari, Y. Sugiyama, and S. Yukawa*

Widely Extended Optimal Velocity Model of Traffic Flow  
and their Linear Stability ..... 221  
*K. Hasebe, A. Nakayama, and Y. Sugiyama*

Stability of Multi-Lane Traffic Flow ..... 227  
*A. Sasoh*

8-Figured Hysteresis Loop of OV Model ..... 235  
*T. Nishi and D. Takahashi*

**Empirical Traffic Data**

Observational Aspects of Japanese Highway Traffic ..... 243  
*S. Yukawa, M. Kikuchi, A. Nakayama, K. Nishinari, Y. Sugiyama,  
and S. Tadaki*

Long-Term Traffic Data from Japanese Expressway ..... 257  
*M. Kikuchi, A. Nakayama, K. Nishinari, Y. Sugiyama,  
S. Tadaki, and S. Yukawa*

The Local Occupation Probability Method for Evaluating Traffic Flows ..... 263  
*Y. Ishibashi and M. Fukui*

**Telematics and Drivers' Behaviour**

Cooperative Driving: Taking Telematics to the Next Level ..... 271  
*R. G. Herrtwich and G. Noecker*

Modeling of Vehicular Behavior from Road Traffic Engineering  
 Perspectives..... 281  
*H. Ozaki*

ITS and the Revolution of Automobiles..... 293  
*Y. Noguchi*

Modelling the Impact of ACC-Systems on the Traffic Flow  
 at Macroscopic Modelling Level..... 305  
*C. Demir*

A Cellular Automaton Model for Dynamic Route Choice Behavior  
 in Urban Roads..... 319  
*M. Fukui, T. Yosida, and H. Oikawa*

Experimental Investigation of Day-to-Day Route Choice Behaviour ..... 325  
*R. Selten, M. Schreckenberg, T. Pitz, T. Chmura, and J. Wahle*

Optimal Traffic States in a Cellular Automaton Model for City Traffic..... 331  
*R. Barlovic, E. Brockfeld, A. Schadschneider, and M. Schreckenberg*

Traffic Flow Analysis Based on Multiagent ..... 337  
*H. Tsunashima, Y. Nishi, T. Honjyo, and T. Sakai*

Traffic Forecast Using a Combination of On-Line Simulation  
 and Traffic Data ..... 345  
*R. Chrobok, A. Pottmeier, J. Wahle, and M. Schreckenberg*

On-Ramp Control..... 351  
*D. Huang*

The Influence of Tollbooths on Highway Traffic ..... 357  
*D. Huang and W. Huang*

**Networks / Internet**

Packet Transport and Load Distribution in Scale-Free Networks ..... 365  
*K.-I. Goh, B. Kahng, and D. Kim*

Phase Transition of Three-Directional Traffic-Flow in 2D Network ..... 377  
*Y. Honda*

Analysis of Minimal Model of Internet Traffic..... 389  
*K. Fukuda, M. Takayasu, and H. Takayasu*

Microscopic Modeling of Packet Transport in the Internet..... 401  
*T. Huisinga, R. Barlovic, W. Knospe, A. Schadschneider,  
 and M. Schreckenberg*

**Granular**

Avalanches and Flow Dynamics of a Collapsing Granular Pile ..... 409  
*L.C. Jia, P.-Y. Lai, and C.K. Chan*

Note on a Micropolar Gas-Kinetic Theory..... 421  
*H. Hayakawa*

Dynamics and Structure of Granular Flow Through a Vertical Pipe ..... 437  
*O. Moriyama, N. Kuroiwa, T. Isoda, T. Arai, S. Tateda,  
 Y. Yamazaki, and M. Matsushita*

Asymmetric Random Average Process: Aggregation and Fragmentation  
 on Continuous State Space..... 449  
*F. Zielen and A. Schadschneider*

Shape Segregation for Bidisperse Mixtures of Ellipses  
 in Two Dimensions ..... 455  
*H.-G. Matuttis, N. Ito, and H. Watanabe*

Bifurcations of a Driven Granular System under Gravity..... 461  
*M. Isobe*

Simulation of the Impact of an Elastic Disk..... 467  
*H. Kuninaka and H. Hayakawa*

Granular Flow in Vertical Pipes: Transition from Dilute to Dense..... 473  
*M. Hou, W. Chen, T. Zhang, and K. Lu*

Spatial Structure of 1/f Noise in Granular Flow Through a Pipe ..... 479  
*A. Nakahara and M. Nakahara*

Collisional Granular Flow on a Rough Slope and its Instability..... 485  
*N. Mitarai, H. Nakanishi, and H. Hayakawa*

The Nature of Occurrence of Queued Flow at Capacity Bottleneck  
 of Ordinary Section..... 491  
*T. Oguchi*

**Pedestrian Dynamics**

Bionics-Inspired Cellular Automaton Model for Pedestrian Dynamics..... 499  
*A. Schadschneider*

Critical Discussion of “Synchronized Flow”, Simulation of Pedestrian  
 Evacuation, and Optimization of Production Processes..... 511  
*D. Helbing, I.J. Farkas, D. Fasold, M. Treiber, and T. Vicsek*

Cellular Automaton Simulations of Pedestrian Dynamics  
 and Evacuation Processes ..... 531  
*A. Kirchner and A. Schadschneider*

Modeling Pedestrians and Granular Flow in 2-Dimensional  
 Optimal Velocity Models..... 537  
*Y. Sugiyama, A. Nakayama, and K. Hasebe*

Evacuation Analysis of Ship by Multi-Agent Simulation Using Model of Group Psychology .....	543
<i>M. Katuhara, H. Matsukura, and S. Ota</i>	
Comparison of an Evacuation Exercise in a Primary School to Simulation Results .....	549
<i>H. Klüpfel, T. Meyer-König, and M. Schreckenberg</i>	
Simulations of Evacuation Using Small World Network.....	555
<i>N. Ohi, M. Ikai, and K. Nishinari</i>	

## **Biology**

Oscillation Patterns in Cytoplasmic Networks of the <i>Physarum</i> Plasmodium.....	563
<i>H. Yamada and T. Nakagaki</i>	
Jamming Bacterial Traffic: Bioconvection .....	569
<i>I.M. Jánosi, A. Czirók, D. Silhavy, and A. Holczinger</i>	
Molecular Dynamics Simulation of Lubricating Films.....	575
<i>T. Murakami, S. Yukawa, and N. Ito</i>	

## List of Participants

1. **Masako Bando**, Aichi University, Miyoshi-cho Kurozasa, Aichi-ken Nishikamo-gun, 470-0796, Japan, bando@vega.aichi-u.ac.jp
2. **Yih-Yuh Chen**, Physics Department, National Taiwan University, N0.1, Section 4, Roosevelt Road, Taipei, 106, Taiwan, Republic of China, yychen@phys.ntu.edu.tw
3. **Stuart Clement**, Transport System Centre, University of South Australia, GPO Box 2471, Adelaide, South Australia, 5001, Australia, stuart.clement@unisa.edu.au
4. **Cesim Demir**, DaimlerChrysler AG, Calwer Straße, 71059 Sindelfingen, Germany, cesim.demir@daimlerchrysler.com
5. **Kensuke Fukuda**, NTT Network Innovation Laboratories, 3-9-11, midori-cho, Musashino, Tokyo, 180-8585, Japan, fukuda@t.onlab.ntt.co.jp
6. **Minoru Fukui**, Nakanihon Automotive College, Fukagaya, Sakahogi-cho, Gifu-ken, 505-0077, Japan, fukui3@cc.nagoya-u.ac.jp
7. **Tamaki Fukui**, Aichi University, Miyoshi-cho Kurozasa, Aichi-ken Nishikamo-gun, 470-0796, Japan
8. **Katsuya Hasebe**, Aichi University, Miyoshi-cho Kurozasa, Aichi-ken Nishikamo-gun, 470-0796, Japan, hasebe@vega.aichi-u.ac.jp
9. **Mami Hashizume**, Waseda University, Ohkubo 3-4-1, Sinjukuku Tokyo, 169-8555, Japan, g9810613@mm.waseda.ac.jp
10. **Shlomo Havlin**, Minerva Center and Department of Physics, Bar-Ilan University, 52900 Ramat-Gan, Israel, havlin@ophir.ph.biu.ac.il
11. **Hisao Hayakawa**, Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, 606-8501, Japan, hisao@yuragi.jinkan.kyoto-u.ac.jp
12. **Dirk Helbing**, Institute for Economics and Traffic, Dresden University of Technology, Andreas-Schubert-Str. 23, 01062 Dresden, Germany, helbing@rcs.urz.tu-dresden.de
13. **Ralf Guido Herrtwich**, Director Telematics Research, DaimlerChrysler AG, FT3/T, Alt Moabit 96a, 10559 Berlin, Germany, ralf.herrtwich@daimlerchrysler.com
14. **Yosuke Hirata**, Toshiba Corporation, 4-1 Ukishima-cho, Kawasaki-ku, Kawasaki, 210-0826, Japan, yosuke.hirata@toshiba.co.jp
15. **Atsushi Honda**, Ryukoku University, Misoratyo 2-13-105, Ootsushi/Shigaken, 520-0223, Japan, atsushi@bird.math.ryukoku.ac.jp
16. **Yasushi Honda**, Departement Computer Science and System Engineering, Muroran Institute of Technology, Muroran, 050-8585, Japan, honda@csse.muroran-it.ac.jp
17. **Takashi Honjyo**, Nihon University, 1-2-1 Izumi-cho, Narashino-shi, Chiba, 275-8575, Japan, c01260@cit.nihon-u.ac.jp
18. **Meiying Hou**, Institute of Physics, Chinese Academy of Sciences, 8, Zhong-Guan-Cun South third Street, Beijing, 100080, China, mayhou@aphy.iphy.ac.cn

19. **Ding-wei Huang**, Department of Physics, Chung Yuan Christian University, Chung-li, 32023, Taiwan, [dwhuang@phys.cycu.edu.tw](mailto:dwhuang@phys.cycu.edu.tw)
20. **Torsten Huisinga**, Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany, [huisinga@traffic.uni-duisburg.de](mailto:huisinga@traffic.uni-duisburg.de)
21. **Miho Ikai**, Ryukoku University, 2-22-17 nomura, Kusatsu Shiga, 525-0027, Japan, [t980006q@ryukoku.seikyoku.ne.jp](mailto:t980006q@ryukoku.seikyoku.ne.jp)
22. **Teruko Ikegami**, Toyota Motor Corporation, 1 Toyota-cho, Toyota, Aichi, 471-8571, Japan, [teru@mail.toyota.co.jp](mailto:teru@mail.toyota.co.jp)
23. **Shuuichi Ikeno**, Japan Highway P.C., 2-18-19, Nishiki, Naka-ku, Nagoya, 460-0003, Japan, [Shuuichi.Ikeno@jhnet.go.jp](mailto:Shuuichi.Ikeno@jhnet.go.jp)
24. **Yoshihiro Ishibasi**, Aichi Shukutoku University, Nagakute-Katahira 9, Nagakute-cho, Aichi Prefecture, 480-1197, Japan, [yishi@asu.aasa.ac.jp](mailto:yishi@asu.aasa.ac.jp)
25. **Masaharu Isobe**, Kanazawa University, Kakuma, Kanazawa, 920-1192, Japan, [isobe@cphys.s.kanazawa-u.ac.jp](mailto:isobe@cphys.s.kanazawa-u.ac.jp)
26. **Yoshito Itoh**, CIRSE, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
27. **Imre M. Janosi**, Department of Physics of Complex Systems, Eotvos University, Pazmany Peter setany 1, Budapest, H-1117, Hungary, [janosi@lecco.elte.hu](mailto:janosi@lecco.elte.hu)
28. **Byungnam Kahng**, School of Physics, Seoul National University, Seoul, 151-742, Republic of Korea, [kahng@phya.snu.ac.kr](mailto:kahng@phya.snu.ac.kr)
29. **Takahisa Kajimoto**, Ryukoku University, 1-16-5 Akanedai, Haibaratyou Udagun Nara, 633-0256, Japan, [t00m002y@ryukoku.seikyoku.ne.jp](mailto:t00m002y@ryukoku.seikyoku.ne.jp)
30. **Makoto Kanagawa**, SANYO Electric Co. Ltd., 180, Ohmori, Anpachi-Cho, Anpachi-Gun, Gifu, 503-195, Japan, [kana@gf.hm.rd.sanyo.co.jp](mailto:kana@gf.hm.rd.sanyo.co.jp)
31. **Mitujirou Katuhara**, National Maritime Research Institute, Shinkawa, 6-38-1, Mitaka City, 181-0004, Japan, [kat@srilot.go.jp](mailto:kat@srilot.go.jp)
32. **Boris S. Kerner**, Research Institute, DaimlerChrysler AG, E224, 70546 Stuttgart, Germany, [boris.kerner@daimlerchrysler.com](mailto:boris.kerner@daimlerchrysler.com)
33. **Macoto Kikuchi**, Osaka University, Machikaneyama-machi 1-1, Toyonaka, 560-0043, Japan, [kikuchi@phys.sci.osaka-u.ac.jp](mailto:kikuchi@phys.sci.osaka-u.ac.jp)
34. **Doochul Kim**, Department of Physics, Seoul National University, Seoul, 151-742, Korea, [dkim@phya.snu.ac.kr](mailto:dkim@phya.snu.ac.kr)
35. **Ansgar Kirchner**, Institute for Theoretical Physics, University of Cologne, Zulpicher Str. 77, 50937 Köln, Germany, [aki@thp.uni-koeln.de](mailto:aki@thp.uni-koeln.de)
36. **Hubert Klüpfel**, Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany, [kluepfel@traffic.uni-duisburg.de](mailto:kluepfel@traffic.uni-duisburg.de)
37. **Masaki Koshi**, Department of Transportation Engineering and Socio-Technology, College of Science and Technology, Nihon University, Narashinodai, Hunabashi, 274-8501, Japan, [koshi@trpt.cst.nihon-u.ac.jp](mailto:koshi@trpt.cst.nihon-u.ac.jp)
38. **Paul L. Krapivsky**, Center for Polymer studies and Department of Physics, Boston University, Boston, Massachusetts, 02215, USA, [paulk@sid.bu.edu](mailto:paulk@sid.bu.edu)
39. **Hiroshi Kumon**, NBC, 1-14-1308, Koyo-cho Naka, Higashinada-ku, Kobe, 658-0032, Japan, [hkumon@nbcom.co.jp](mailto:hkumon@nbcom.co.jp)

40. **Hiroto Kuninaka**, Kyoto University, YoshidaNihonmatsu-cho, Sakyo-ku, Kyoto city, 606-8501, Japan, kuninaka@yuragi.jinkan.kyoto-u.ac.jp
41. **Itigoro Kuroki**, Beingcorp inc, Higasisakura2, Nagoya-si Higasi-ku, 461-0005, Japan, kuroki@beingcorp.co.jp
42. **Pik-Yin Lai**, Department of Physics, National Central University, Chung-Li, 320, Taiwan, R.O.C., pylai@phy.ncu.edu.tw
43. **Hyun Keun Lee**, School of Physics, Seoul National University, San 56-1 Shinrim-dong, Kwanak-gu, Seoul, 151-742, South Korea, hkleee@phya.snu.ac.kr
44. **Hyun-Woo Lee**, Korea Institute for Advanced Study, 207-43 Cheongryangri-dong, Dongdaemun-gu, Seoul, 130-012, Korea, hwlee@kias.re.kr
45. **Masanori Mano**, Japan Highway P.C., 2-18-19, Nishiki, Naka-ku, Nagoya, 460-0003, Japan
46. **Mitsugu Matsushita**, Department of Physics, Chuo University, Bunkyo-ku, Tokyo, 112-8551, Japan, matusita@phys.chuo-u.ac.jp
47. **Hans-Georg Matuttis**, Department of Applied Physics, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8565, Japan, ha@acolyte.t.u-tokyo.ac.jp
48. **Hisashi Mitani**, Fukuoka University of Education, 729 Akama, Munaka-city, Fukuoka, 811-4192, Japan, mitanilh@fukuoka-edu.ac.jp
49. **Namiko Mitarai**, Department of Physics, Kyushu University, 6-10-1, Hako-zaki, Higashi-ku, Fukuoka-shi, Fukuoka, 812-8581, Japan, namiko@stat.phys.kyushu-u.ac.jp
50. **Masayo Miyano**, Ryukoku University, 2-5-5 Nagayoshi Nagahara Hirano-ku, Osaka, 547-0016, Japan, mamiyano@ryukoku.seikyoku.ne.jp
51. **Teruyoshi Murakami**, The University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo, 1138656, Japan, murakami@acolyte.t.u-tokyo.ac.jp
52. **Takashi Nagatani**, College of Engineering, Shizuoka University, Hamamatsu, 432-8561, Japan, tmtnaga@epc.shizuoka.ac.jp
53. **Kai Nagel**, Institute of Scientific Computing, ETH Zentrum / D-INFK, 8092 Zürich, Switzerland, nagel@inf.ethz.ch
54. **Masahide Nakae**, Ryukoku University, 10-1 nishihongoutyohigashi, oumihachimanshi shiga, 523-0818, Japan, t980088i@ryukoku.seikyoku.ne.jp
55. **Midori Nakahara**, Institute of Materials Science, University of Tsukuba, 1-1-1 Ten-noudai, Tsukuba/Ibaraki, 305-8573, Japan, nakahara@kf6.so-net.ne.jp
56. **Akio Nakahara**, Physics, College of Science and Technology, Nihon University, 7-24-1, Narashino-dai, Funabashi, Chiba, 274-8501, Japan, nakahara@phys.ge.cst.nihon-u.ac.jp
57. **Kenichi Nakanishi**, Faculty of Bioresources, Mie University, 1515, Kamihama-cho, Tsu, Mie, 514-8507, Japan, nakanisi@bio.mie-u.ac.jp
58. **Akihiro Nakayama**, Gifu Keizai University, Kitakata-cho 5-50, Ohgaki, Gifu, 503-8550, Japan, g44153g@cc.nagoya-u.ac.jp
59. **Shoichiro Nakayama**, Kanazawa University, 2-40-20 Kodatsuno, Kanazawa, 920-8667, Japan, snakayama@t.kanazawa-u.ac.jp
60. **Takao Namiki**, Division of Mathematics, Hokkaido University, Kita 10 Nishi 8, Kita-ku Sapporo, 060-0810, Japan, nami@math.sci.hokudai.ac.jp

61. **Takayuki Nishi**, Waseda University, Ohkubo 3-4-1, Sinjukuku Tokyo, 169-8555, Japan, takayuki@mpdt.math.waseda.ac.jp
62. **Yasukazu Nishi**, Nihon University, 1-2-1, Izumi-cho, Narashino-shi, 275-8575, Chiba, Japan, tetsu@me.cit.nihon-u.ac.jp
63. **Katsuhiko Nishinari**, Ryukoku University, Seta, Ohtsu, Shiga, 520-2194, Japan, knishi@rins.ryukoku.ac.jp
64. **Shizuka Nishiura**, Ryukoku University, 1-5 setaohetyouyokotani, Ohtsu, 5202194, Japan, t01m003j@ryukoku.seikyoku.ne.jp
65. **Yoshikazu Noguchi**, Toyota Motor Corporation, 1 Toyota-cho, Toyota city, Aichi Pref., 471-8571, Japan, ynoguchi@mail.toyota.co.jp
66. **Takashi Oguchi**, Tokyo Metropolitan University, Department of Civil Engineering, Minamiosawa 1-1, Hachioji, 192-0397, Japan, oguchi-takashi@c.metro-u.ac.jp
67. **Noriko Ohi**, Ryukoku University, 2-324-15 Kunimityou, Narashi Saidaiji Nara, 631-0823, Japan, ohi@pearl.ocn.ne.jp
68. **Toru Ohira**, Sony Computer Science Laboratories Inc., Higashi-Gotanda, Shinagawa-ku, Tokyo, 141-0022, Japan, ohira@csl.sony.co.jp
69. **Hirokazu Oikawa**, Nakanihon Automotive College, Fukagaya, Sakahogi-cho, Gifu-ken, 505-0077, Japan, oikawa@nakanihon.ac.jp
70. **Akiko Okumura**, Saga University, Onimaru 9-5, Saga, 840-0021, Japan, akiko@ai.is.saga-u.ac.jp
71. **Haruo Ozaki**, Civil and Environmental Engineering, Toyo University, Kawagoe, 350-8585, Japan, ozaki@eng.toyo.ac.jp
72. **Andreas Pottmeier**, Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany, pottmeier@traffic.uni-duisburg.de
73. **Akihiro Sasoh**, Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577, Japan, sasoh@ifs.tohoku.ac.jp
74. **Shiro Sawada**, Takuma National College of Technology, 551 Kohda, Mitoyo, Kagawa, 769-1192, Japan, sawada@dt.takuma-ct.ac.jp
75. **Andreas Schadschneider**, Institute for Theoretical Physics, University of Cologne, Zùlpicher Str. 77, 50937 Cologne, Germany, as@thp.uni-koeln.de
76. **Michael Schreckenberg**, Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany, schreckenberg@uni-duisburg.de
77. **Akihiro Shibata**, Computing Research Center, KEK, Oho 1-1, Tsukuba, 305-0801, Japan, akihiro.shibata@kek.jp
78. **Takahiro Shiga**, Toyota Central R&D Labs. Inc., 41-1, Yokomichi, Nagakute, Aichi, 480-1192, Japan, t-shiga@mosk.tytlabs.co.jp
79. **Yuki Sugiyama**, Division of Mathematical Science, City College of Mie, Ishinden Nakano 157, Tsu, Mie, 514-0112, Japan, genbey@eken.phys.nagoya-u.ac.jp
80. **Shin-ichi Tadaki**, Saga University, 1 Honjo, Saga, 840-8502, Japan, tadaki@cc.saga-u.ac.jp
81. **Daisuke Takahashi**, Waseda University, 3-4-1, Ohkubo, Shinjuku-ku, Tokyo, 169-8555, Japan, daisuke@mse.waseda.ac.jp

82. **Misako Takayasu**, Future University - Hakodate, 116-2 Kameda-nakano-cho, Hakodate, Hokkaido, 041-8655, Japan, takayasu@fun.ac.jp
83. **Ki-Wing To**, Institute of Physics, Academia Sinica, Nankang, Taipei, 115, Taiwan, ericto@gate.sinica.edu.tw
84. **Elad Tomer**, Bar-Ilan University, Ramat-Gan, 52900, Israel, tomer@alon.cc.biu.ac.il
85. **Sadayuki Tsugawa**, ITS Research Group, AIST, Namiki 1-2-1, Tsukuba-shi, Ibaraki-ken, 305-8564, Japan, tsugawa.s@aist.go.jp
86. **Hitoshi Tsunashima**, Nihon University, 1-2-1 Izumi-cho, Narashino-shi, Chiba, 275-8575, Japan, tsuna@me.cit.nihon-u.ac.jp
87. **Motohiro Usami**, Waseda University, Ohkubo 3-4-1, Sinjuku Tokyo, 169-8555, Japan, usami@gm.math.waseda.ac.jp
88. **Bing-Hong Wang**, Department of Modern Physics and Nonlinear Science Center, University of Science and Technology of China, Hefei, Anhui, 230026, China, bhwang@ustc.edu.cn
89. **Eddie Wilson**, University of Bristol, Department of Engineering Maths, Queens Building, Bristol, BS8 1TR, United Kingdom, RE.Wilson@bristol.ac.uk
90. **Dietrich E. Wolf**, Department of Physics, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany, wolf@comphys.uni-duisburg.de
91. **Hiroyasu Yamada**, Haruki-dai 5-1-2, Togo-cho, 470-0161, Japan, hyamada@allegro.phys.nagoya-u.ac.jp
92. **Koichi Yamawaki**, Department of Physics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8602, Japan
93. **Taturu Yosida**, Nakanihon Automotive College, Fukagaya, Sakahogi-cho, Kamo-gun, Gifu-ken, 505-0077, Japan, yosida-t@nakanihon.ac.jp
94. **Insuk Yu**, Seoul National University, San 56, Sin Rim, Kwan Ak, Seoul, 151-747, Korea, isyu@snu.ac.kr
95. **Satoshi Yukawa**, Dept. of Appl. Phys., School of Eng., Univ. of Tokyo, 7-3-1, Hongo, Bunkyo-ku, 113-8656, Japan, yukawa@ap.t.u-tokyo.ac.jp
96. **Frank Zielen**, Institute for Theoretical Physics, University of Cologne, Zùlpicher Str. 77, 50937 Cologne, Germany, fz@thp.uni-koeln.de

# **Traffic Models: Theory**

# Euler and Lagrange Representation of Traffic Models

K. Nishinari

Department of Applied Mathematics and Informatics, Ryukoku University, Seta, Ohtsu 520-2194, Japan

**Abstract.** The relations among different traffic models are studied by using the ultra-discrete method and the Euler-Lagrange transformation. It is found that the Burgers CA(BCA) in the Euler form can be transformed into the Lagrange form by using the formulae of the max-algebra. It is also shown that the Lagrange model is related to the optimal velocity model, the slow-to-start model and the Nagel-Schreckenberg model. Moreover, a new hybrid Lagrange model is proposed by extending the BCA, which shows a complex phase transition from free to a jamming state.

## 1 The Euler Form of Traffic Models

There are many different models of traffic flow proposed up to now. In this paper, we consider the relationship among them by using the newly proposed Euler-Lagrange transformation. Also we will propose a new hybrid model combining the good properties of each model.

We focus on the Euler and Lagrange representation of each model. This terminology comes from hydrodynamics. In the Euler form a road is considered as a field and cars are not distinguished. Thus the Euler form has merits in modeling of overtaking the car in front, or more than one lane traffic flow. In the Lagrange form the dependent variable is the position of each car, and we can follow the position of each car individually. We have proposed several Euler models so far in our previous papers. Among them, the most basic and important one is the Burgers CA(BCA), and its Euler form is given by [1]

$$U_j^{t+1} = U_j^t + \min(U_{j-1}^t, L - U_j^t) - \min(U_j^t, L - U_{j+1}^t). \quad (1)$$

It has been shown that (1) is derived from the Burgers equation [2]  $v_t = 2vv_x + v_{xx}$  through the transformation  $U_j^t = L/2 + \varepsilon \Delta x v(j\Delta x, t\Delta t)$ , where  $\Delta x$  and  $\Delta t$  are lattice intervals in  $x$  and  $t$  respectively and  $\varepsilon$  is a parameter used in the ultradiscrete formula [3]. Assuming that  $L > 0$  and  $0 \leq U_j^t \leq L$  for any site  $j$  at a certain time  $t$ ,  $0 \leq U_j^{t+1} \leq L$  holds for any  $j$ . Thus (1) is equivalent to a CA with a value set  $\{0, 1, \dots, L\}$ . Moreover, if we put a restriction  $L = 1$  on the BCA, then the BCA is equivalent to rule 184. There is a parameter  $L$  in this model, and its physical meaning is considered in the following two ways: we consider that a site could either represent a longer segment of the expressway capable of accommodating a maximum of  $L$  cars, or the unit segment of  $L$ -lane expressway.

The maximum velocity of cars in the BCA is 1, and we have extended the BCA to the case of maximum velocity 2 [4,5]. Extension of the BCA to general velocities are, however, found to be difficult as long as we use the Euler form. Euler Models become complex in general when the number of neighboring sites in an evolutionary rule becomes large. It is better to use the Lagrange form when we consider one lane phenomena. It is suitable for representing the case that the order of cars does not change, i.e., every car does not overtake the car in front on one lane.

## 2 Euler-Lagrange Transformation

In this section we derive a Lagrange form of the BCA. For this, we propose the Euler-Lagrange transformation by using new formulae of the max-algebra.

First, we define  $G$  by the equation

$$U_j^t = G_{L(j+1)+\alpha}^t - G_{Lj+\alpha}^t, \quad (2)$$

where  $\alpha$  is an arbitrary constant. The function  $G$  is similar to the function  $F$  in [1], which is the variable in the ultradiscrete diffusion equation. Substituting (2) into (1) we obtain the evolution equation of  $G$  as

$$G_{Lj+\alpha}^{t+1} = \max(G_{L(j-1)+\alpha}^t, G_{L(j+1)+\alpha}^t - L). \quad (3)$$

Thus we can “linearize” the BCA by using  $G$ , and (3) corresponds to the ultradiscrete diffusion equation [1].

From (2),  $G$  increases its value when  $U \neq 0$ , i.e., cars exist. Thus we can put

$$G_j^t = \sum_{i=0}^{N-1} H(j - x_i^t - 1), \quad (4)$$

where  $H(x)$  is the step function defined by  $H(x) = 1$  if  $x > 0$  and  $H(x) = 0$  otherwise, and  $N$  is the total number of cars on the road.  $x_i^t$  is the Lagrange variable that represents the position of the  $i$ -th car at time  $t$ . (4) shows the relation between the Euler variable  $G$  and the Lagrange variable  $x$ . Then replacing  $G$  in (3) into  $H$  by using (4), we have

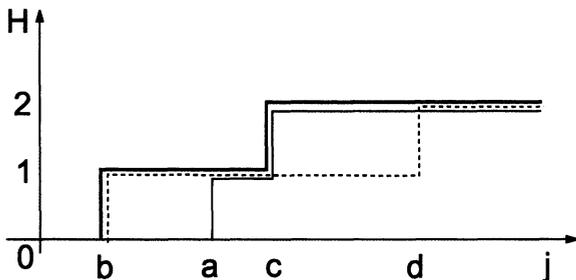
$$\begin{aligned} \sum_{i=0}^{N-1} H(jL + \alpha - x_i^{t+1} - 1) &= \max\left(\sum_{i=0}^{N-1} H(jL + \alpha - x_i^t - L - 1), \right. \\ &\quad \left. \sum_{i=0}^{N-1} H(jL + \alpha - x_i^t + L - 1) - L\right). \end{aligned} \quad (5)$$

The constant  $\alpha$  can be omitted by the Galilei transformation, so we put it as zero in the following. It is noted that the physical meaning of  $\alpha$  is related to the discussion given in [6].

In order to simplify (5), we introduce a new formula

$$H(j - \min(a, b)) + H(j - \min(c, d)) = \max(H(j - a) + H(j - c), H(j - b) + H(j - d)), \quad (6)$$

which holds for any constants  $a < c, b < d$  (Fig. 1). From Fig. 1 its proof is



**Fig. 1.** The bold line is the max of the thin line and broken line. The formula (6) holds under the conditions  $a < c$  and  $b < d$ .

trivial. Generalizing (6), we get

$$\sum_{k=1}^n H(j - \min(a_k, b_k)) = \max\left(\sum_{k=1}^n H(j - a_k), \sum_{k=1}^n H(j - b_k)\right) \quad (7)$$

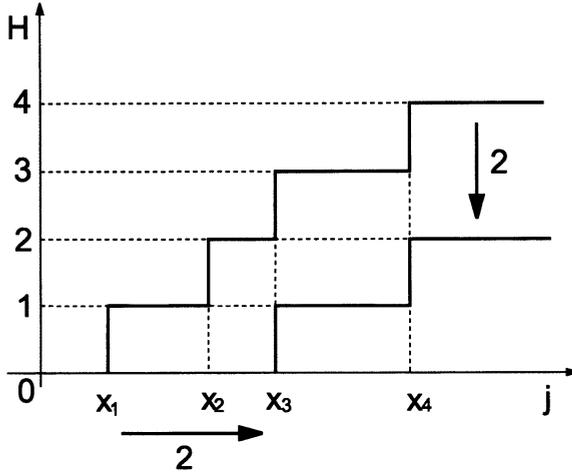
assuming  $a_1 < a_2 < \dots < a_n$  and  $b_1 < b_2 < \dots < b_n$ . This is a fundamental formula for the Euler-Lagrange transformation. Another formula can be obtained by considering the rightmost boundary condition as

$$\max\left(\sum_i H(j - x_i^t) - L, 0\right) = \sum_i H(j - x_{i+L}^t). \quad (8)$$

Here we assume that  $x_j^t = \infty$  if  $j$  is larger than the number of cars (Fig. 2). This formula shows that the subtraction of  $L$  is the same as the shift of  $L$  in the subscript of  $x$ .

By using (7) and (8), since  $x_i^t < x_{i+1}^t$  for any  $i$ , (5) can be transformed into

$$\begin{aligned} \sum_i H(jL - x_i^{t+1} - 1) &= \max\left(\sum_i H(jL - x_i^t - L - 1), \right. \\ &\quad \left. \sum_i H(jL - x_{i+L}^t + L - 1)\right) \\ &= \sum_i H(jL - \min(x_i^t + L + 1, x_{i+L}^t - L + 1)). \quad (9) \end{aligned}$$



**Fig. 2.** The shift of the graph of vertical direction is the same as the shift of horizontal direction. This is illustrated in the case of  $L = 2$ .

Thus comparing the both sides we finally obtain

$$\begin{aligned} x_i^{t+1} &= \min(x_i^t + L + 1, x_{i+L}^t - L + 1) - 1 \\ &= x_i^t + \min(L, x_{i+L}^t - x_i^t - L), \end{aligned} \quad (10)$$

which is the Lagrange form of BCA. If we put  $L = 1$ , then we obtain a Lagrange form of the rule-184 CA as

$$x_i^{t+1} = x_i^t + \min(1, x_{i+1}^t - x_i^t - 1). \quad (11)$$

### 3 Traffic Models in the Lagrange Form

The rule-184 CA is considered as a prototype of all traffic models. Thus in this section we will develop it to a more realistic one by using the Lagrange form.

First, let us extend (11) to the case that the maximum velocity of cars is  $V$ . In this case cars can move forward at most  $V$  sites per unit time if their front sites are empty. The rule-184 CA is the case  $V = 1$ . This is easily modeled by

$$x_i^{t+1} = x_i^t + \min(V, x_{i+1}^t - x_i^t - 1). \quad (12)$$

This corresponds to the Lagrange form of the Fukui-Ishibashi(FI) model [7]. In (12) a car follows only the car in front. We can extend it to the case where the number of cars that a driver can see in front is  $S$ .  $S$  is considered as a “perspective” parameter in this model. Then the model is given by

$$x_i^{t+1} = x_i^t + \min(V, x_{i+S}^t - x_i^t - S). \quad (13)$$

This model is called sometimes as the quick start model or anticipation model. From this extension it is found that BCA (10) is considered as the case  $V = S = L$ . Thus (13) is considered as a generalization of BCA to the case  $V \neq S$ .

Here we comment on a formal relationship between the Optimal-Velocity(OV) model [8] and the FI model. Adding the term  $-(x_i^t - x_i^{t-1})$  in both sides in (12) gives

$$x_i^{t+1} - 2x_i^t + x_{i-1}^t = \min(V, h_i^t) - (x_i^t - x_i^{t-1}), \quad (14)$$

where  $h_i^t = x_{i+1}^t - x_i^t - 1$  is the headway of the  $i$ -th car. It is noted that  $h_i^t$  also represents the number of vacant sites between the  $i$ -th and  $(i+1)$ -th car. Taking a continuous limit  $\Delta t \rightarrow 0$  we obtain (after appropriate scalings)

$$\frac{d^2 x_i}{dt^2} = \min(V, h_i^t) - \frac{dx_i}{dt}. \quad (15)$$

This is nothing but the OV model, and the  $\min(V, h_i^t)$  corresponds to the piecewise linear OV function appeared in [9]. It should be mentioned that, since the uniform flow in FI model is always stable, the sensitivity parameter [8] in (15) becomes large and the uniform flow is always stable in this case.

Next we consider another type of traffic models in which the update rule depends on two preceding steps. So far it depends only on one step. A slow-to-start model has been proposed by M.Takayasu and H.Takayasu [10], which can be written in the Lagrange form as

$$x_i^{t+1} = x_i^t + \min(1, x_{i+1}^t - x_i^t - 1, x_{i+1}^{t-1} - x_i^{t-1} - 1). \quad (16)$$

The inertia effect of cars can be treated by this slow-start rule. It is found that the velocity of cars depends not only on the present headway but also on the past headway in this rule.

The Nagel-Schreckenberg(NS) model [11] is a stochastic CA model which is written in the Lagrange form as

$$x_i^{t+1} = x_i^t + \max(0, \min(V, x_{i+1}^t - x_i^t - 1, x_i^t - x_i^{t-1} + 1) - \eta_i^t), \quad (17)$$

where  $\eta_i^t = 1$  with probability  $p$  and  $\eta_i^t = 0$  with  $1 - p$ . The last term in  $\min$  represents the acceleration of cars, because 1 is added to the present velocity  $v_i^t = x_i^t - x_i^{t-1}$ . From (16) and (17) the last term in  $\min$  is different if we neglect the stochastic term in the NS model. It is known that only the slow-start model shows the metastable state, which is often observed in real traffic. An extension of the NS model to include the slow-to-start rule is treated in [12].

## 4 An Extended Lagrange Model

In this section, let us extend our Lagrange model of BCA with  $V \neq S$  to include the effect of inertia of cars, i.e., we combine (13) and (16). We obtain such hybrid

Lagrange model as

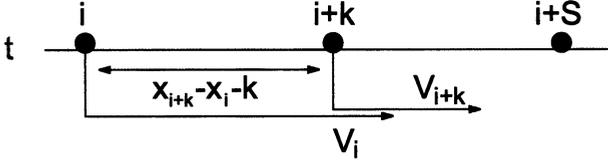
$$x_i^{t+1} = x_i^t + \min \left( V_i^t, \min_{k=1, \dots, S-1} (x_{i+k}^t - x_i^t - k + V_{i+k}^t) \right) \quad (18)$$

$$= x_i^t + \min_{k=1, \dots, S-1} (V_i^t, x_{i+k}^t - x_i^t - k + x_{i+S+k}^{t-1} - x_{i+k}^{t-1} - S) \quad (19)$$

where

$$V_i^t = \min (V, x_{i+S}^t - x_i^t - S, x_{i+S}^{t-1} - x_i^{t-1} - S), \quad (20)$$

and if  $S = 1$  we put  $k = 1$ . The term  $\min_{k=1, \dots, S-1}(\dots)$  in the r.h.s of (18) plays an important role in avoiding the collision of cars. This is because the condition that a car does not overtake all the viewable cars in its front is given by (Fig. 3)



**Fig. 3.** The collision-free condition (21) is illustrated. The sum of the headway  $x_{i+k} - x_i - k$  and velocity of  $(i+k)$ -th car must be larger than the velocity  $V_i$ . This must hold for all the cars between the  $i$ -th and the  $(i+S)$ -th car.

$$x_{i+k}^t - x_i^t - k + V_{i+k}^t \geq V_i^t, \quad k = 1, \dots, S-1. \quad (21)$$

This model includes the slow-to-start rule ( $V = S = 1$ ) and the FI model ( $S = 1$  and neglecting the terms  $x^{t-1}$ ).

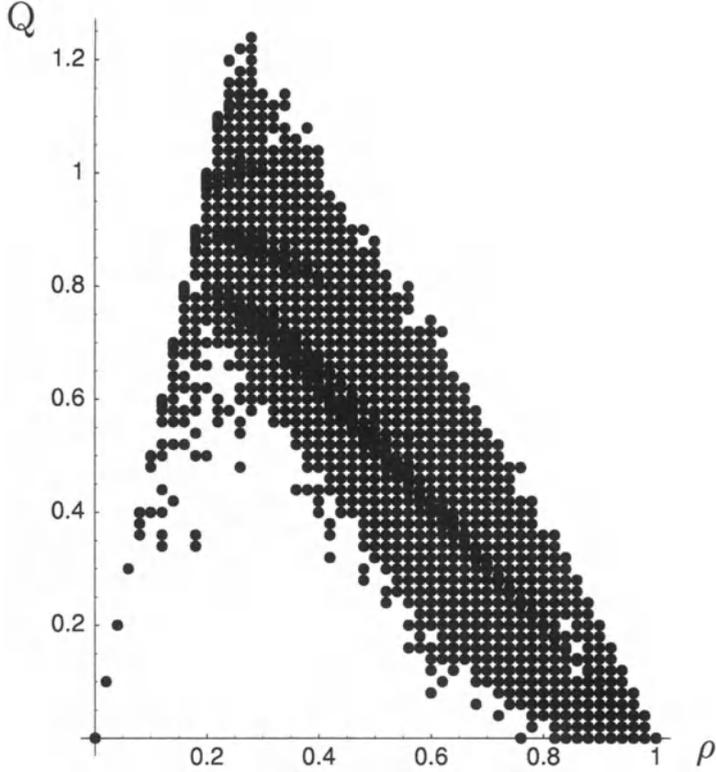
Next, we investigate the behavior of cars in this new Lagrange model. The fundamental diagram, which is the plot of the traffic flow  $Q$  and density  $\rho$ , is given by Fig. 4 for several set of parameters. The flow  $Q$  is defined by the multiplication of the density and the average velocity of cars. From Fig. 4, we observe a complex phase transition from a free to a congested state around the critical density. There are many metastable branches in the diagram as seen in our previous models in the Euler form [5].

The schematic diagram is given by Fig. 5. At the maximum current state  $A$ , the stationary pattern

$$\underbrace{11 \cdots 11}_S \underbrace{000 \cdots 000}_V \quad (22)$$

is observed, then the position of  $A$  in the diagram is given by

$$\left( \frac{S}{V+S}, \frac{SV}{V+S} \right). \quad (23)$$



**Fig. 4.** A fundamental diagram of the new Lagrange model. Parameters are set to  $V = 5$  and  $S = 2$ , and the spatial period is 50 sites. Initial car density is varied from 0 to 1 with step 0.05. At each density, we start calculations from 25 randomly generated initial configurations and superpose all plots in the diagram from time 0 to 100. The average flow over the 25 cases at each time are plotted by thick circles. We see the first order phase transition and many metastable branches in the diagram.

Also at the branching point  $B$ , from the stationary pattern

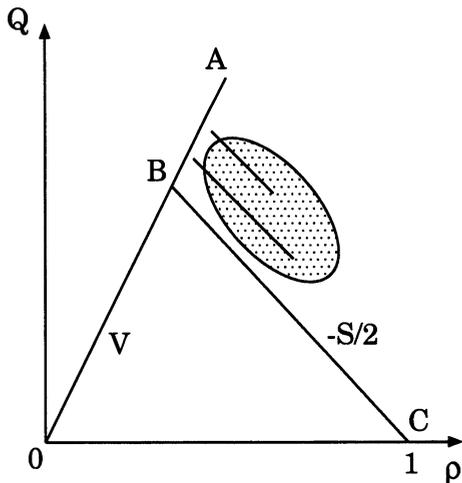
$$\underbrace{11 \dots 11}_{S} \underbrace{000 \dots 000}_{2V} \tag{24}$$

the position is

$$\left( \frac{S}{2V + S}, \frac{SV}{2V + S} \right). \tag{25}$$

The branch  $A - B$  represents overdense free flow and we call it the metastable state which will be a congested state due to perturbations. Let us study the stability of uniform flow in detail in the case  $V = 5$  and  $S = 2$ . The uniform state  $A$  is most unstable to perturbations. From (22) the state is consisted by the block

$$\dots 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0 \dots$$



**Fig. 5.** A schematic fundamental diagram of the new model. The line  $O-A$  represents the free flow and its gradient is  $V$ . The line  $B-C$  shows the strong jam and its gradient is  $-S/2$ . The dotted region represents the weak jam, in which the flow is relatively high in the congested state. We see that there are many metastable branches in it.

and its flow is  $Q = SV/(V + S) = 1.42857$ . First, we add a small perturbation to the initial condition which corresponds to a breaking of a car. We set it as

$$\dots 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0 \dots$$

then after some time we found that it reaches a stationary state(Fig. 6(a)). The flow in the stationary state becomes  $Q = 1.28571$ . Next, we gradually magnify the perturbation. If we set the initial condition as

$$\dots 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0 \dots$$

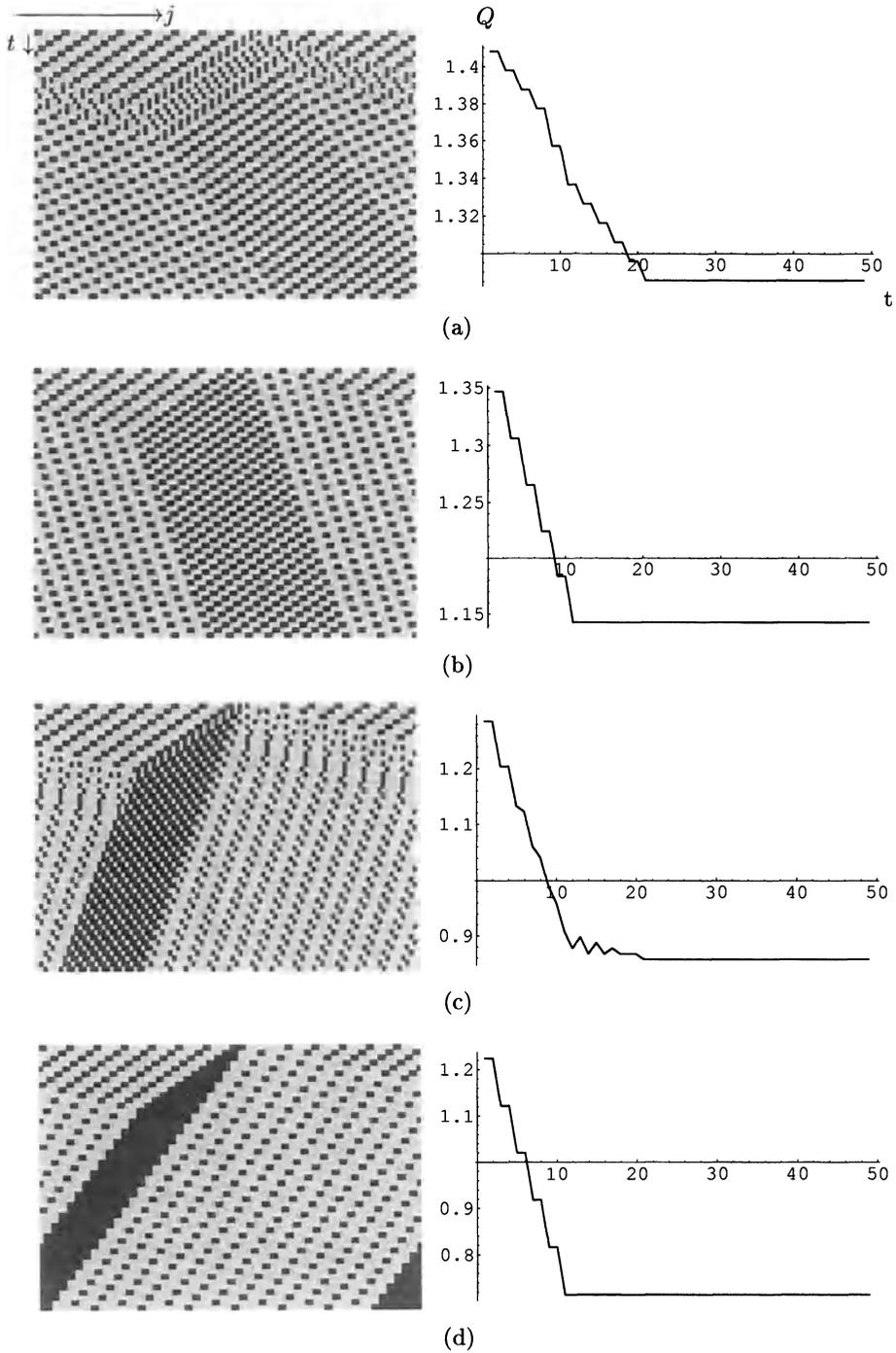
then the final stationary flow becomes  $Q = 1.14286$ (Fig. 6(b)). For the initial condition

$$\dots 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0 \dots$$

the final flow becomes  $Q = 0.857143$ (Fig. 6(c)). We see that in this case the flow oscillates while reaching the stationary value. It is found that the flow is not always monotonously decreasing during relaxation. Finally we add the strongest perturbation as

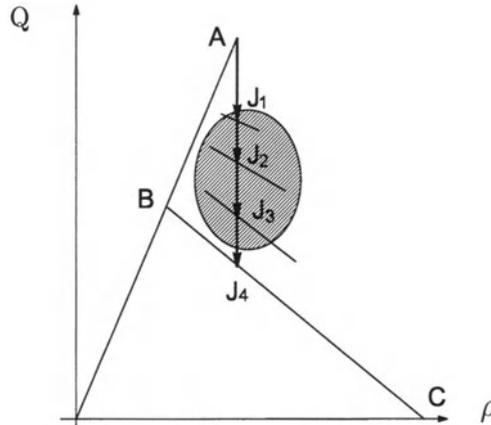
$$\dots 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 \dots$$

and the final flow becomes  $Q = 0.714286$ (Fig. 6(d)), which is the point  $J_4$  on the line  $B - C$  in Fig. 7. The decrease of flow from  $A$  to  $J_4$  under the perturbations is shown schematically in Fig. 7. Thus it is found that perturbation causes a transition from free to a jamming state through several intermediate stages.



**Fig. 6.** The uniform flow in  $A$  is unstable due to perturbations. We magnify the perturbation from (a) to (d). In each perturbation, we draw the space-time behavior and the  $Q$ - $t$  graph. The squares in black and light gray represent 1 and 0, respectively.

Many stationary congested states given above cause data-spreading in the two-dimensional area in the diagram. In the new model therefore it is found that there exist both, a metastable “weak jam” (in the dotted region) and the “strong jam” (line B–C) in the congested state. We expect that these phenomena will be



**Fig. 7.** Flow decreases from  $A$  to  $J_4$  due to perturbations. There are many intermediate states between them, which are all stationary and metastable states.

observed in expressway traffic [13], in which the phase transition from free to congested states shows metastability. So, our new model will be a good tool to analyze complex phase transitions of real traffic flow.

## References

1. K. Nishinari and D. Takahashi, *J. Phys. A* **31**, 5439 (1998).
2. T. Musya and H. Higuchi, *J. Phys. Soc. Jpn.* **17**, 811 (1978).
3. T. Tokihiro, D. Takahashi, J. Matsukidaira, and J. Satsuma, *Phys. Rev. Lett.* **76**, 3247 (1996).
4. K. Nishinari and D. Takahashi, *J. Phys. A* **32**, 93 (1999).
5. K. Nishinari and D. Takahashi, *J. Phys. A* **33**, 7709 (2000).
6. K. Nishinari, *J. Phys. A* to appear.
7. M. Fukui and Y. Ishibashi, *J. Phys. Soc. Jpn.* **65**, 1868 (1996).
8. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995).
9. K. Nakanishi, K. Itoh, Y. Igarashi, and M. Bando, *Phys. Rev. E* **55**, 6519 (1997).
10. M. Takayasu and H. Takayasu, *Fractals* **1**, 860 (1993).
11. K. Nagel and M. Schreckenberg, *J. Phys. I France* **2**, 2221 (1992).
12. A. Schadschneider and M. Schreckenberg, *Ann. Phys.* **6**, 541 (1997).
13. K. Nishinari and M. Hayashi (editors), *Traffic Statistics in Tomei Express Way*, (The Mathematical Society of Traffic Flow, Japan, 1999).

# Three-Phase Traffic Theory

B.S. Kerner

Daimler Chrysler AG, RIC/TS, T729, 70546 Stuttgart, Germany

**Abstract.** Hypotheses and some results of the three-phase traffic theory which has been proposed and developed by the author during the last years are considered. In order to give some insight, types and features of spatial-temporal congested patterns which can occur in the frame of this theory both on a homogeneous road and at a highway bottleneck are briefly discussed. A theory of highway capacity at highway bottlenecks is developed and discussed. Finally, a critical discussion of theoretical results which have been derived within the fundamental diagram approach is made.

## 1 Introduction

Up to now it has been self-evident that *steady* state solutions of a mathematical traffic flow model, i.e., the model solutions where vehicles move at the same distances to one another with the same time-independent vehicle speed (these hypothetical spatially homogeneous and time-independent traffic states are often called also "equilibrium" model solutions; we will use in the article for these traffic states the term "steady" states or "steady speed" states) should belong to a curve in the flow-density plane (see, e.g., [1–40] and the recent reviews [41,42]). This curve which goes through the origin and has at least one maximum is called *the fundamental diagram* for traffic flow. The postulate about the fundamental diagram underlies almost all traffic flow modeling approaches up to now (see reviews [41,42]) in the sense that the models are constructed such that in the unperturbed, noiseless limit they have a fundamental diagram of steady states, i.e. the steady states form a curve in the flow-density plane. An approach to traffic flow modeling which is based on this fundamental hypothesis about steady state model solutions [1,41,42] will be called *the fundamental diagram approach*.

It is well-known that traffic on a multi-lane highway (which will be mostly considered in the paper) can be either free or congested. Congested traffic states can be defined as the traffic states where the average vehicle speed is lower than the minimum possible average speed in free flow (e.g., [43]). It is well known that in contrast to free traffic flow, in congested traffic a collective behavior of vehicles plays an important role (the collective flow by Prigogine and Herman which is related to a part of the fundamental diagram at higher density [5]), and a synchronization of vehicles speeds across different highway lanes usually occurs [43]. In congested traffic, complex spatial-temporal patterns are observed, in particular a sequence of moving traffic jams, the so called "stop-and-go" phenomenon (e.g., the classical works by Treiterer [44] and Koshi *et al.* [43]). Recall

that a moving jam is a moving localized structure. The moving jam is spatially restricted by two upstream moving jam fronts where the vehicle speed and the density change sharply. The vehicle speed is low (up to zero) and the density is high inside the moving jam.

Concerning the important role of highway bottlenecks it should be noted that although congested traffic can occur away from bottlenecks [45], they have an important impact as well as defects in physical systems can play an important role for the phase transitions and for the formation of spatial-temporal patterns. The role of the bottlenecks in traffic flow is as follows: Highway bottlenecks are the most frequent reason for the occurrence of congested traffic (e.g., [46,47]). The bottlenecks can be resulted for example due to road works, on and off ramps, a decrease in the number of highway lanes, road curves and road gradients.

Although the complexity of traffic is linked to the occurrence of spatial-temporal patterns, some of the traffic features can be understood if *average* traffic characteristics are considered. Thus among the important empirical methods in traffic science are empirical flow-density and speed-density relationships which are related to measurements of some *average* traffic variables at a highway location, in particular at highway bottlenecks. The empirical relationship of the *average* vehicle speed on the vehicle density must be related to an obvious result observed in real traffic flow: the higher the vehicle density, the lower the average vehicle speed. When the flow rate, which is the product of the vehicle density and the average vehicle speed, is plotted as a function of the vehicle density one gets what is known as the empirical fundamental diagram.

In 1955 Lighthill and Whitham [1] wrote in their classical work (see page 319 in [1]): "... The fundamental hypothesis of the theory is that at any point of the road the flow (vehicles per hour) is a function of the concentration (vehicles per mile)...". Apparently the empirical fundamental diagram was the reason that the fundamental diagram approach has already been introduced in the first traffic flow models derived by Lighthill and Whitham [1], by Gasis, Herman and Rothery [2], and by Newell [3].

Concerning the fundamental diagram, it must be noted that in congested traffic complex spatial-temporal traffic patterns are observed (e.g., [44,43]). These patterns are extremely spatially non-homogeneous. This spatial behavior of congested patterns is a complex function on time. An averaging of traffic variables related to congested patterns over long enough time intervals gives a relation between different *averaged* vehicle speeds and densities. Thus, the empirical fundamental diagram is related to *averaged* characteristics of spatial-temporal congested patterns measured at a highway location rather than to features of the hypothetical steady states of congested traffic on the theoretical fundamental diagram. For this reason, the existence of the theoretical fundamental diagram is only a *hypothesis*.

It must be noted that on the one hand the empirical fundamental diagram is successfully used for different important applications where some average traffic flow characteristics should be determined (e.g., [47,48]). On the other hand, it is well known in non-linear science that features of steady state solutions of a

mathematical model can make a great influence on the whole model dynamics. Therefore, it is not surprising that the fundamental diagram which defines the steady speed states of a mathematical traffic flow model in most cases determines also the main non-linear spatial-temporal features of the congested patterns which the model can show [17,42]. However, it is obviously that an *average* empirical characteristic of traffic flow - the empirical fundamental diagram - can not explicitly determine empirical *spatial-temporal* features of traffic. Apparently this contradiction can explain the empirical result that the most of empirical spatial-temporal pattern features [49–51] are qualitative different from those which follow from mathematical traffic flow models in the fundamental diagram approach [41,42].

For this reason in 1996-2000 Kerner, based on an empirical traffic flow analysis, has introduced a concept called "synchronized flow" and the related three-phase traffic theory [45,49,52–59]. *The fundamental hypothesis* of the three-phase traffic theory [53] is that steady states of synchronized flow, i.e., the hypothetical states of synchronized flow where vehicles move at the same distances to one another with the same time-independent vehicle speed cover a *two-dimensional region* in the flow-density plane. This means that there is no fundamental diagram for steady states of synchronized flow in this traffic theory.

A discussion of this concept and of main hypotheses to the three-phase traffic theory, a critical comparison of the results of the fundamental diagram approach to traffic flow theory with empirical results, and also some recent empirical results and results of the three-phase traffic theory are covered in the remainder of this article.

## 2 The Concept "Synchronized Flow"

The concept "synchronized flow" introduced by the author consists of the following claims:

1. There are two qualitative different phases, the traffic phase called "synchronized flow" and the traffic phase called "wide moving jam", which should be distinguished in congested traffic [49,52–55]. This distinguishing is based on qualitative different *empirical spatial-temporal* features of these phases.
2. The fundamental basic hypothesis of the three-phase traffic theory [49,53–55]: Hypothetical steady states of synchronized flow cover a two-dimensional region in the flow-density plane. This means that in these hypothetical steady states of synchronized flow, where all vehicles move at the same distances to one another and with the same time-independent speed, a given steady vehicle speed is related to an infinity multitude of different vehicle densities and a given vehicle density is related to an infinity multitude of different steady vehicle speeds. This hypothesis means that there is *no* fundamental diagram for hypothetical steady speed states of synchronized flow.

The fundamental hypothesis of the three-phase traffic theory is therefore in contradiction with the hypothesis about the existence of the fundamental diagram

for hypothetical steady states of mathematical models and theories in the fundamental diagram approach.

### 3 Objective Criteria for Traffic Phases

It should be stressed that the concept "synchronized flow" and the related methodology of the congested pattern study, which has been used for the definition of the three traffic phases below, is based on an analysis of *empirical spatial-temporal* features of congested patterns [49,52,53] rather than on a dynamical analysis of data (e.g., in the flow-density plane) which are measured at only one highway location. First, a spatial-temporal study of traffic must be made. Only after the traffic phases "synchronized flow" and "wide moving jam" have already been distinguished, based on this spatial-temporal data analysis, some of the pattern features can further be studied in the flow-density plane. In particular, this procedure has already been used in [52]: At the first step, a spatial-temporal analysis of empirical data have been made and the phases "synchronized flow" and "wide moving jam" have been identified. At the next step, the traffic phase "synchronized flow" has been plotted in the flow-density plane without any moving jams and some of the features of "synchronized flow" have in addition been studied.

The objective criteria to distinguish the different phases in congested traffic are based on qualitatively different *spatial-temporal features* of these phases [49,52,53]. These objective criteria have been *defined* as the following [54,56,57]:

- A local spatial-temporal upstream moving congested traffic pattern, i.e., a pattern which is spatially restricted by two upstream moving (downstream and upstream) fronts, belongs to the traffic phase "wide moving jam", if at the given "control parameters" of traffic (e.g., the weather and other environment conditions), the pattern possesses the following characteristic, i.e., unique, coherent, predictable and reproducible feature: The pattern as a whole local structure propagates through any states of free and synchronized flows and through any bottlenecks (e.g., at on-ramps and off-ramps) *keeping* the mean velocity of the downstream front of the pattern. This velocity is the same for different wide moving jams.
- The traffic phase "synchronized flow" possesses the characteristic feature to form diverse spatial-temporal patterns upstream of a highway bottleneck. The downstream front of these patterns is usually fixed at the bottleneck. Even if a moving synchronized flow pattern occurs, the velocity of the downstream front of this pattern is not a characteristic parameter: This velocity can change in a wide range during the pattern propagation and it can be different for different patterns.

Examples of the application of these objective criteria for the determination of the phase "synchronized flow" and the phase "wide moving jam" in congested traffic can be found in [51] where a more detailed analysis of these phases has been made.

It should be noted that measured data on sections of highway A5 which have been used in [45,49–59] also comprise the information about vehicle types (number of vehicles and long vehicles) and individual vehicle speeds passing the detector during each one minute interval of averaging. Using the latter information in addition to the spatial-temporal data analysis of one minute averaged data, the determination of the type of synchronized flow in empirical studies of synchronized flow (the type (i), or (ii) or else (iii) [52]) have been made. Besides, the individual vehicle speeds allow us to answer the question whether there are narrow moving jams in synchronized flow or there are not. All these steps of data analysis have been made in [45,49–59]. We will illustrate this in Sec. 9 where an empirical example of a synchronized flow pattern will be studied.

The term "synchronized flow" should reflect the following features of this traffic phase: (i) It is a non-interrupted traffic flow rather than a long enough standstill as it usually occurs inside a wide moving jam. The word "flow" should reflect this feature. (ii) There is a *tendency* to a synchronization of vehicle speeds across different lanes on a multi-lane road in this flow. Besides, there is a *tendency* to a synchronization of vehicle speeds on each of the road lanes (a bunching of the vehicles) in synchronized flow due to a relatively low mean probability of passing in synchronized flow. The word "synchronized" should reflect these speed synchronization effects.

It has already been mentioned, that traffic flow consists of free flow and congested traffic. Congested traffic consists of the phase "synchronized flow" and the phase "wide moving jam". Thus, there are three traffic phases [49,52,53]: 1. Free flow. 2. Synchronized flow. 3. Wide moving jam.

Moving jams do not emerge in free flow, if synchronized flow is not hindered [50]. Instead, the moving jams emerge due to a sequence of two first order phase transitions [49]: First the transition from free flow to synchronized flow occurs (it will be called the F→S-transition) and only later and usually at a different location moving jams emerge in the synchronized flow (the latter transition will be called the S→J-transition and the sequence of both transitions the F→S→J-transitions).

## 4 Short Coming of the Fundamental Diagram Approach for Description of Traffic Congestion

Different explanations of these empirical features of wide moving jams and synchronized flow are up to now in a discussion between different groups (e.g., [30–40,60–70] and the review [42]).

Due to the effort of different scientific groups (see e.g., [6–11,13–24,28–40,71] and the reviews [41,42]) a considerable progress has been made in the understanding of the spatial-temporal congested patterns in different traffic flow models within the fundamental diagram approach. In particular, in this approach two main classes of traffic flow models may be distinguished which claim to show moving jams and other congested patterns upstream from an on-ramp:

- (i) Models where at a sufficiently high initial flow rate, upstream from an on-ramp, moving jams spontaneously occur if the flow rate at the on-ramp beginning from zero gradually is increasing. However, the range of the flow rate to the on-ramp where moving jams spontaneously occur is limited. Beginning at a high enough flow rate to the on-ramp spatial homogeneous states of traffic flow which have been called "homogeneous congested traffic" (HCT) [32] occur upstream of the on-ramp where no moving jams spontaneously emerge (e.g., [29,32–35,42]).
- (ii) Models where as well as in the models of item (i), beginning at a some flow rate to the on-ramp, moving jams spontaneously occur upstream of the on-ramp. However, no HCT occurs in these models.

How do the traffic phase "synchronized flow" and the traffic phase "wide moving jam" emerge in an initially free traffic flow at an isolated bottleneck (i.e., the bottleneck is far away from other effective bottlenecks), e.g., at a bottleneck due to an on-ramp? Empirical observations allow to conclude that the following scenarios are responsible for the phase transitions and for the pattern evolution in traffic flow at the on-ramp [49,51]:

1. Moving jams do *not* emerge in an initially free flow at the on-ramp when the flow rate at the on-ramp is gradually increasing. Rather than moving jams the phase transition from free flow to synchronized flow occurs at the on-ramp.
2. At a low enough flow rate to the on-ramp the vehicle speed in synchronized flow which has occurred upstream of the on-ramp is relatively high. Moving jams do *not* necessary emerge in that synchronized flow. If the flow rate to the on-ramp is high, then the vehicle speed in the synchronized flow is low and moving jams, in particular wide moving jams, emerge in that synchronized flow.
3. The lower the average vehicle speed in synchronized flow upstream of the on-ramp, the higher the frequency of the moving jam emergence in that synchronized flow. This means that the moving jam emergence goes on up to the highest possible values of the flow rate to the on-ramp: Traffic states of high density and low vehicle speed, where moving jams do not emerge, are *not* observed in synchronized flow upstream of the on-ramp.

The empirical results in item 1 and 2 are in qualitative contradiction with the both model classes (i) and (ii) in the fundamental diagram approach. In these models, at high enough initial flow rates upstream of the on-ramp moving jams must emerge in an initial free flow if the flow rate to the on-ramp beginning from zero is gradually increased [42]. The last empirical result in item 3 means that in average the higher the vehicle density is the lower is the stability of traffic flow with respect to the moving jam emergence in that flow. This result of observations which seems to be intuitive obvious for each driver is in a qualitative contradiction with the models of the class (i) in the fundamental diagram approach where HCT, i.e, homogeneous congested traffic of high density and low vehicle speed must occur where moving jams do not emerge [42].

The features of the theoretical diagram [32,33,36,42] given a high enough initial flow rate on a highway upstream of the on-ramp may be illustrated with the following simple *theoretical schema 1*:

- A low flow rate to the on-ramp → different kind of moving jams must emerge.
- A high enough flow rate to the on-ramp → HCT where the density is high and the speed is very low and no moving jams spontaneously emerge must occur.

In contrast to this theoretical result [32,33,36,42], in empirical observations [51] the following *empirical schema 2* is observed:

- A low flow rate to the on-ramp → synchronized flow where the density is relatively low and the speed is relatively high occur where moving jams should not necessary emerge.
- A high enough flow rate to the on-ramp → moving jams must spontaneously emerge in synchronized flow upstream of the on-ramp at any high flow rate to the on-ramp.

Thus, these two schemata (the theoretical schema 1 [32,33,36,42] and the empirical schema 2 [51]) are in contradiction to one another.

It should be noted that in models within the fundamental diagram approach, fluctuations and external perturbations let the system evolve in time through a 2D region in the flow-density plane, as well. However, the dynamics is governed locally by steady state properties. In models within the fundamental diagram approach, these steady states lie on the fundamental diagram, i.e., they are related to a 1D region in the flow-density plane [41,42]. If the steady states form a 2D region, as is the case in the three-phase traffic theory [49,53–55], the dynamics is fundamentally different. This leads also to qualitative differences between the congested patterns obtained in the three-phase traffic theory [51,67,70] or in the fundamental diagram approach [41,42], respectively.

It must be noted that this critical consideration of the application of the fundamental diagram approach for description of congested traffic does not concern some important mathematical ideas which have been introduced and developed in models and theories within the fundamental diagram approach with the aim to describe the traffic flow dynamics, for example the ideas about the modelling of vehicle safety conditions, fluctuations, vehicle acceleration and deceleration, different vehicle time delays and other important effects (e.g., [2,3,6–11,13–24,32,37–39,66–68,71–77]; see also references in the reviews by Chowdhury *et al.* [41], by Helbing [42] and by Wolf [75]). These mathematical ideas are also very important elements of the three-phase traffic theory [45,55,58] and of microscopic models within this theory [67,70]. The main feature of the three-phase traffic theory is that this theory *rejects* the basic hypothesis about the fundamental diagram of earlier traffic flow theories and models. The three-phase traffic theory introduces the new phase of traffic flow, synchronized flow, which steady states cover 2D region in the flow density plane. This allows us to overcome the above problems of fundamental diagram approach and to explain empirical spatial-temporal congested pattern features [49,51].

## 5 Discussion of Hypotheses to the Three-Phase Traffic Theory

### 5.1 Steady Speed States [53,45,55]

To explain the mentioned above and other contradictions of empirical results with the results of traffic flow models and theories in the fundamental diagram approach, the author has introduced the three-phase traffic theory related to the concept "synchronized flow" [49,53,54,56–58,78,79].

The fundamental hypothesis of the three-phase traffic theory has already been formulated in item 2 of the concept "synchronized flow" (Sec. 2): Hypothetical spatially homogeneous and stationary, i.e., time-independent (steady) states of synchronized flow where vehicles move at the same distances to one another with the same time-independent vehicle speed cover a 2D-region in the flow-density plane [49,53,58] (Fig. 1). These steady states are the same for a multi-lane and for a one-lane roads. In other words, in the three-phase traffic theory there is *no* fundamental diagram for synchronized flow.

This is not excluded by the empirical fact mentioned above, that a given vehicle density determines the *average* vehicle speed. Indeed, from empirical observations it may be concluded that at the same distance between vehicles (at the same density) there may be a continuum of different vehicle speeds within some finite range in synchronized flow. Obviously the averaging of all these vehicle speeds leads to one average value at the given density. This is also not ruled out by car following experiments, where a driver has the task to follow a specific leading car and not loose contact with it (e.g., [43]). In such a situation, the gap between the cars will be biased towards the security gap depending on the speed of the leading car. In synchronized flow the situation is different: The gap between cars can be much larger than the security gap.

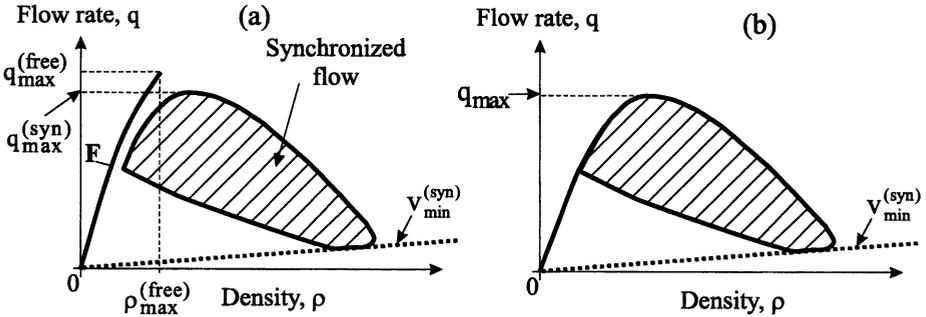
The hypothesis about steady states of synchronized flow which cover a 2D-region in the flow-density plane makes the three-phase traffic theory [49,53,58] almost *incompatible* with all classical traffic flow theories and present models (see, e.g., [1–3,6–9,12,14,13,17,15,18,19,22–33,36,38–42]) which are based on the fundamental diagram approach.

This follows from the diagram of congested patterns at bottlenecks which has recently been found out by Kerner [51,78,82]. This diagram has been postulated on very general grounds within the three-phase traffic theory [51,78] (see Sec. 6.2) and demonstrated for a microscopic traffic model by Kerner and Klenov [67]. In this model, the upper boundary of a 2D-region of steady states of synchronized flow in the flow-density plane is related to the well-known dependence of the safe speed on the gap between vehicles. The low boundary of these steady states is related to the dependence of a synchronization distance  $D$  on the speed. The synchronization distance  $D$  introduced in [67] has been used by Kerner, Klenov and Wolf for the formulation of KKW cellular automata models to the three-phase traffic theory which may show qualitative the same diagram of congested patterns [70] as the continuum model of Kerner and Klenov [67]. The mathematical models within the three-phase traffic theory proposed in [67,70] show

that the fact whether the steady states of a mathematical description of traffic belong to a curve (the fundamental diagram approach) or to a 2D-region in the flow-density plane *qualitatively changes* the basic non-linear spatial-temporal features of the congested patterns which the model allows. This is probably because some of the steady states of the mathematical model can be "fixed points" or attractors for the model. It is well-known in the physics of different non-linear systems that characteristics of these attractors may have a great influence on the whole non-linear dynamics of the mathematical model.

In the section below, some features of steady states of traffic flow on a multi-lane road in the three-phase traffic theory are considered (a more detailed consideration is made in [53,45,55]). These features are further compared with those which follow from traffic theories and models in the fundamental diagram approach.

Note that the three-phase traffic theory is a *behavioral theory* of traffic flow. This means that hypotheses of this theory are based on observed common behavioral fundamental characteristics of drivers on highways. In particular, the fundamental hypothesis of the three-phase traffic theory (Sec. 2) is linked to a driver's ability to recognize whether the distance to the vehicle ahead becomes higher or lower over time [53]. If gaps between vehicles are not very high, this driver's ability is true even if the difference between vehicle speeds is negligible. It has been shown [51,67,70] that the main *qualitative* features of the congested pattern emergence observed in empirical investigations [49,51] can be shown in the three-phase traffic theory where all drivers have the same characteristics and all vehicles have the same parameters. Obviously, in real traffic there are differences in driver's characteristics and in vehicles parameters (e.g., different wish and different safe speeds, aggressive and timid driver's behavior, vehicles



**Fig. 1.** The hypotheses about states of free flow (curve  $F$ ) and spatially homogeneous and time-independent states (steady states) of synchronized flow (dashed region) in the flow density plane for a multi-lane (in one direction) homogeneous (without bottlenecks) road in the three-phase traffic theory (a). For a comparison, in (b) steady states for a homogeneous (without bottlenecks) one-lane road are shown [49,53]. The dotted line is related to a minimum speed in steady states of synchronized flow  $v_{\min}^{(syn)}$  [50,57].

and long vehicles) which may change some spatial-temporal congested pattern parameters and conditions of the pattern emergence. However, these differences in driver's characteristics and in vehicles parameters may first be neglected when *fundamental* traffic pattern features are studied.

## 5.2 Hypothesis about Stability of Steady States [45,53,55]

In three-phase traffic theory, it is suggested that independently of the vehicle density in an initial steady state of traffic flow infinitesimal perturbations of traffic flow variables (i.e., the vehicle speed and/or the density) do not grow in any states of free flow and in any steady states of synchronized flow: In the whole possible density range states of free flow and hypothetical steady states of synchronized flow can exist (Fig. 1 (a)).

To explain this hypothesis from the driver's behavior, let us qualitatively consider a small enough fluctuation in braking of a vehicle in an initial state which is related to one of the states inside a 2D region in the flow-density plane (dashed region in Fig. 1). This small braking may lead to a transition from the initial traffic state to another state with a lower gap. This lower gap is equal to a gap in another state a 2D region in the flow-density plane. Thus, an occurrence of this fluctuation may cause a spatial-temporal transition to another state of synchronized flow. Therefore independent on the vehicle density within the 2D region of states of synchronized flow drivers should not immediately react on this transition. For this reason even after a time delay, which is due to a finite reaction time of drivers, the drivers upstream should not brake stronger than drivers in front of them to avoid an accident. As a result, a local perturbation of traffic variables (density or vehicle speed) of small enough amplitude does not grow.

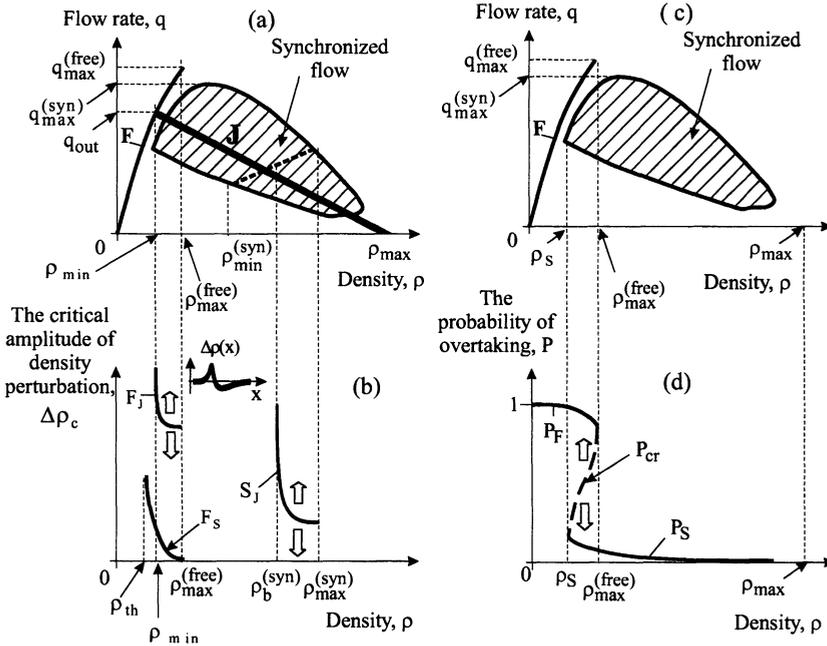
It can be seen that already this hypotheses is in a qualitative contradiction with most of those traffic theories and models in the fundamental diagram approach which are claimed to explain the moving jam emergence. Indeed, in the latter models and theories there is often a range of the vehicle density where steady states of traffic flow which are related to the fundamental diagram have to be unstable (e.g., [13–15,17,22,23,28–33,36,38,41,42]).

## 5.3 Hypothesis about Continuous Spatial-Temporal Transitions between States of Synchronized Flow [55]

The above consideration grounds the following hypothesis of the three-phase traffic theory: Already small local perturbations can cause complex spatial-temporal states of synchronized flow which are related to complex dynamical transitions inside a 2D region of synchronized flow states in the flow-density plane.

This behavior is indeed observed in microscopic models [67,70] where steady speed states of synchronized flow cover a 2D-region in the flow density plane. Complex spatial-temporal states of synchronized flow have also been observed by Fukui *et al.* in their cellular automata model [68].

5.4 Hypothesis about two Different Kinds of Nucleation Effects [45,53,55,59]



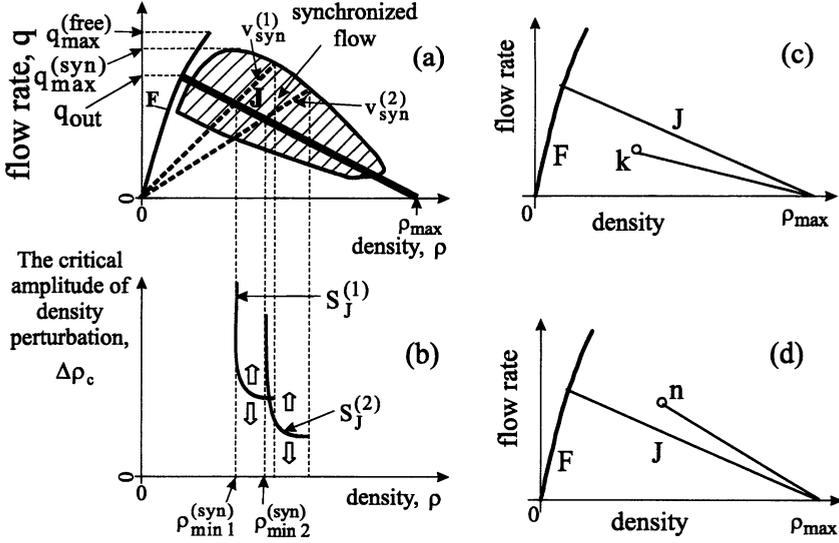
**Fig. 2.** Explanation of hypotheses to the three-phase traffic theory [45,55]. States of free (curve  $F$ ) and synchronized flow (hatched region) in (a, c) are the same as in Fig. 1 (a).

In the three-phase traffic theory, there are two qualitatively different kinds of nucleation effects in states of traffic flow:

1. The nucleation effect which is responsible for the jam's emergence. If an initial traffic state is a state of free flow then this nucleation effect is responsible for the phase transition from free flow to a wide moving jam (the  $F \rightarrow J$ -transition). If an initial state is a steady state of synchronized flow then this nucleation effect is responsible for the phase transition from synchronized flow to a wide moving jam (the  $S \rightarrow J$ -transition) (Fig. 2 (a, b)).
2. The nucleation effect which is responsible for the phase transition from free flow to synchronized flow (the  $F \rightarrow S$ -transition) (Fig. 2 (b-d)).

The existence of two different nucleation effects in free flow means that there are two different metastable regions of the vehicle density in free flow where these nucleation effects are realized:

- (i) The metastable region of the vehicle density in free flow where the  $F \rightarrow J$ -transition occurs if a local perturbation exceeds some critical amplitude and



**Fig. 3.** Explanation of the jam emergence in synchronized flow in the three-phase traffic theory [49]. The densities  $\rho_{min 1}^{(syn)}$  and  $\rho_{min 2}^{(syn)}$  are the threshold densities for the jam emergence for the vehicle speeds  $v_{syn}^{(1)}$  and  $v_{syn}^{(2)}$ , respectively. States of free (curve  $F$ ) and synchronized flow (hatched region) in (a) are the same as in Fig. 1 (a).

- (ii) the metastable region of the vehicle density in free flow where the  $F \rightarrow S$ -transition occurs if a local perturbation exceeds some other critical amplitude.

Although at the same vehicle density in free flow the critical amplitudes for these two phase transitions are very different (see below), the related metastable regions of the vehicle density in free flow can (sometimes partially) overlap one another. This circumstance makes the differentiation between these two qualitative different nucleation effects essentially difficult.

### 5.5 Hypothesis about Critical Amplitude of Local Perturbations and Probability of Phase Transitions in Free Flow [45,55,59]

The following hypothesis of the three-phase traffic theory emphasizes the importance of the  $F \rightarrow S$ -transition in an initially free flow: At each given density in free flow the critical amplitude of a local perturbation of traffic variables which is needed for the  $F \rightarrow S$ -transition (curve  $F_S$  in Fig. 2 (b) is considerably lower than the critical amplitude of a local perturbation which is needed for the  $F \rightarrow J$ -transition (curve  $F_J$  in Fig. 2 (a)). Therefore, at each given density in free flow the probability of an occurrence of the  $F \rightarrow S$ -transition is considerably higher than that of the  $F \rightarrow J$ -transition.

To explain this [59], it can be assumed that the dependence of the mean probability of overtaking  $P$  on the density on a multi-lane road is a Z-shaped

function (Fig. 2 (d)). Synchronized flow occurs, if due to a local perturbation of traffic variables (density or/and vehicle speed) the mean probability of overtaking  $P$  decreases below some critical value  $P_{cr}$ . The critical amplitude of a local perturbation which is able to cause the local avalanche like self-decrease in the mean probability of overtaking can be considerably lower than the critical amplitude of a local perturbation which is needed for the F→J-transition. Indeed, for the occurrence of synchronized flow an initial perturbation itself does not necessarily have to increase avalanche-like: It must only cause an avalanche-like decrease in the mean value of the probability of overtaking. In contrast, for the jam formation the amplitude of the critical perturbation itself must begin to grow avalanche-like in an initial free flow.

The hypothesis about the Z-shaped mean probability of overtaking  $P$  as a function of the vehicle density is confirmed by empirical studies where it have been shown that states of synchronized flow overlap with states of free flow in the density (see Fig. 2 in [55]). This means that at the same density either a state of synchronized flow or a state of free flow is possible. It is obviously that the mean probability of overtaking  $P$  is higher in free flow than in synchronized flow. Thus, the empirical fact that states of free flow and synchronized flow overlap in the vehicle density means that the mean probability of overtaking has a Z-shape. Another example where an overlapping of states of free and synchronized flow in the density in empirical data occurs can be seen below when the widening synchronized flow pattern will be investigated (see Sec. 9.1).

However, it must be noted that this overlapping of states of free flow and states of synchronized flow occurs in a narrow range of the density in the vicinity of the limit density in free flow  $\rho_{max}^{(free)}$ , i.e., when the average speed of synchronized flow is relatively high (see Fig. 2 in [55] and Sec. 9.1). Therefore, if in some empirical data *only* synchronized flows of a relative low vehicle speed are observed then no overlapping of states of free and synchronized flows could be found. As a result, some authors make a conclusion that a dependence of the mean probability of overtaking  $P$  as function of the density is a monotonous decreasing one (e.g., [69]). To find the correct Z-shape of the mean probability of overtaking  $P$  as function of the density [55], however a more precise empirical study is necessary (Sec. 9.1).

At the limit density for free flow  $\rho = \rho_{max}^{(free)}$  the critical amplitude of a local perturbation of traffic variables which is needed for the F→S-transition reaches zero (the curve  $F_S$  in Fig. 2 (b)), respectively the probability of the F→S-transition reaches one. Therefore, the existence of the limit point  $\rho = \rho_{max}^{(free)}$  for free flow in the three-phase-traffic-theory is linked to the occurrence of the F→S-transition rather than the F→J-transition. Indeed, the critical amplitude of traffic variables which is needed for the F→J-transition (curve  $F_J$  in Fig. 2 (b)) is a relatively high finite value even at the limit density  $\rho = \rho_{max}^{(free)}$ .

## 5.6 Hypothesis about Nucleation Effect, Responsible for the Wide Moving Jam Emergence [49,53,45,55]

In a theory of wide moving jams in the fundamental diagram approach which has first been derived by Kerner and Konhäuser in [13] and further developed for a number of different traffic flow models [15,18,19,23,24,32,33,36,39–42], it has been found out that the characteristic line for the downstream front of a wide moving jam (the line J) is a very important characteristic of traffic (line J in Fig. 2 (a)). The slope of the line J is equal to the velocity of the downstream front of the wide moving jam,  $v_g$ . This velocity is the characteristic, i.e., unique, reproducible and predictable parameter which can depend only on the control parameters of traffic (e.g., weather, other road conditions) but the velocity  $v_g$  does not depend on initial conditions and traffic demand. In the flow-density plane, the line J intersects the density-axis at the point related to the density inside the jam  $\rho_{max}$  and the curve  $F$  for free flow at the point  $[\rho_{min}, q_{out}]$ . This latter point is related to the flow rate  $q_{out}$  and the related density  $\rho_{min}$  in the outflow from the wide moving jam when free flow is formed downstream of the jam. The density  $\rho_{min}$  is the threshold one for the jam emergence in free flow [13]: At  $\rho < \rho_{min}$  all states of free flow are stable with respect to the jam emergence. At  $\rho \geq \rho_{min}$  the related states of free flow are metastable with respect to the jam emergence where the related nucleation effect should lead to the jam emergence. These theoretical features of wide moving jams have been found out in empirical observations [81].

The importance of the line J remains also in the three-phase traffic theory. This is apparently *the only one result* of theories and models in the fundamental diagram approach which is compatible with the three-phase traffic theory. However, even at this point there is an important difference between these two approaches: In a lot of models in the fundamental diagram approach, the upper boundary of the metastable density region is related to a point (this point is designated as  $\rho_{max}^{(free)}$  in Fig. 2 (a)) where the critical amplitude for the F→J-transition reaches zero. At higher densities free flow becomes unstable with respect to the jam emergence (e.g., [13–15,17,22,23,28–33,36,38,41,42]). In contrast, in the three-phase traffic theory at this limit point  $\rho_{max}^{(free)}$  the critical amplitude for the F→J-transition is relatively high, i.e., the probability of the wide moving jam emergence is very low (see the curve  $F_J$  in Fig. 2 (b)). In the three-phase traffic theory, as it has already been stressed at this limit point the critical amplitude for the F→S-transition reaches zero, i.e., the F→S-transition occurs rather than the F→J-transition [45,55].

In the three-phase traffic theory, the importance of the line J appears for the moving jam emergence in steady states of synchronized flow [49,53]: The line J (line J in Fig. 2 (a)) determines the threshold of the wide moving jam's existence and excitation in synchronized flow. In other words, all (**an infinite number !**) steady states of synchronized flow which are related to the line J in the flow-density plane are *threshold states* with respect to the wide moving jam emergence. For example, the density  $\rho_{min 1}^{(syn)}$  is the threshold density for the wide moving jam emergence in a steady state of synchronized flow with the vehicle

speed  $v_{syn}^{(1)}$  (Fig. 3 (a)). Correspondingly, the density  $\rho_{min 2}^{(syn)}$  is the threshold density in another steady state with the vehicle speed  $v_{syn}^{(2)}$  shown in Fig. 3 (a).

The line  $J$  separates all steady states of synchronized flow into two qualitatively different classes:

1. In states which are related to points in the flow-density plane lying below (see axes in Fig. 2 (a)) the line  $J$  no wide moving jams either can continue to exist or can be excited.
2. States which are related to points in the flow-density plane lying on and above the line  $J$  are *metastable states* with respect to the wide moving jam emergence where the related nucleation effect can be realized. Only the local perturbations of traffic variables whose amplitude exceeds some critical amplitude grow and can lead to the wide moving jam emergence (up arrows, curves  $F_J$  and  $S_J$  in Fig. 2 (b)), otherwise wide moving jams do not occur (down arrows, curves  $F_J$  and  $S_J$  in Fig. 2 (b)). The former critical local perturbations act as nucleation centers for the wide moving jam emergence.

To explain the hypothesis, note that that wide moving jams cannot be formed in any states of flow situated below the line  $J$ . Indeed, let us consider a state of flow directly upstream of a wide moving jam. This can be related to a point  $k$  in the flow-density plane which is below the line  $J$  (Fig. 3 (c)). Because the value of the velocity of the upstream front of the wide moving jam  $v_g^{(upstream)}$  equals the slope of a line from  $k$  to the point  $[\rho_{max}, 0]$ , the related absolute value  $|v_g^{(upstream)}|$  is always lower than that of the downstream front  $|v_g|$  which is determined by the slope of the line  $J$ . Therefore, the width of the wide moving jam is gradually decreasing. Otherwise, if a flow which is upstream from another wide moving jam is above the line  $J$  (see a point  $n$  in Fig. 3 (d)), then  $|v_g^{(upstream)}| > |v_g|$ , i.e., the width of the wide moving jam is gradually increasing. Therefore, wide moving jams can be formed in states of synchronized flow above the line  $J$ , so that states which are related to points in the flow-density plane lying on and above the line  $J$  are metastable states with respect to the wide moving jam emergence.

The critical amplitude of the local perturbations is maximal at the line  $J$  and depends both on the density and on the flow rate above the line  $J$ . The critical amplitude of the local perturbations is considerably lower in synchronized flow than in free flow (compare the curve  $F_J$  with the curve  $S_J$  in Fig. 2 (b)). The curve  $S_J$  is related to a fixed vehicle speed, dotted line in Fig. 2 (a).

At the same difference between an initial density and a threshold density, the lower the initial speed in synchronized flow is the lower the critical amplitude  $\Delta\rho_c$  of the local perturbation needed for the S→J-transition. This assumption of the three-phase traffic theory is illustrated in Fig. 3 (a, b). In this figure, two dependencies of the critical amplitude of the local perturbation needed for the S→J-transition for two different vehicle speeds in synchronized flow,  $v_{syn}^{(1)}$  and  $v_{syn}^{(2)}$  where  $v_{syn}^{(1)} > v_{syn}^{(2)}$ , are shown (curves  $S_J^{(1)}$  and  $S_J^{(2)}$  for  $v_{syn}^{(1)}$  and  $v_{syn}^{(2)}$ , respectively).

It should be noted that recently the hypotheses about the phase transitions in traffic flow which have been discussed above have been confirmed in a microscopic three-phase traffic theory developed by Kerner and Klenov [67] and in the KKW cellular automata models [70].

## 6 Diagram of Congested Patterns at Highway Bottlenecks

The qualitative difference of traffic flow theories and models in the fundamental diagram approach [41,42] and of the three-phase traffic theory [49,53,45,55,59] can essentially clear be seen if the diagrams of congested patterns at highway bottlenecks which should occur in these different approaches are compared.

Recall that in the diagram of congested patterns at bottlenecks in the fundamental diagram approach which has first been found out by Helbing *et al.* in [32] diverse congested patterns are possible depending on the initial flow rate  $q_{in}$  on a highway upstream of the bottleneck and on the bottleneck strength  $\Delta q$ . If the effective bottleneck is caused by an on-ramp, the bottleneck strength  $\Delta q$  equals the flow rate to the on-ramp  $q_{on-ramp}$ .

The diagram of the congested patterns at on-ramps within the three-phase traffic theory has recently been found out by Kerner [51,78] based on qualitative considerations and derived by Kerner and Klenov based on their microscopic traffic flow model [67]. We will see that this diagram is totally qualitative different from the diagram of congested patterns at on-ramps in the fundamental diagram approach [32,33,36,42].

In particular, in contrast to the fundamental diagram approach [42], in the diagram of congested patterns within the three-phase traffic theory [51,67,78,82] at a high enough flow rate to the on-ramp rather than HCT moving jams always spontaneously emerge in synchronized flow upstream of the bottleneck. At a low enough flow rate to the on-ramp rather than moving jams synchronized flow of higher vehicle speed can occur without an occurrence of moving jams. These theoretical results are in agreement with the related results of empirical observations [51] (see the empirical schema 2 and the related discussion in Sec. 4).

In this section, a qualitative derivation of the diagram of congested patterns at on-ramps made in [51,78] will be shortly considered and compared with the diagram of congested patterns in the fundamental diagram approach. However, first a hypothesis about emergence of synchronized flow at bottlenecks [54] will be briefly considered.

### 6.1 Hypothesis about Emergence of Synchronized Flow at Bottlenecks: Explanation of the Breakdown Phenomenon [54]

The above hypothesis about the priority of the  $F \rightarrow S$  transition in free flow in comparison with the  $F \rightarrow J$  transition is also true at a highway bottleneck. Therefore, in the three-phase traffic theory if a phase transition in an initial free flow at the bottleneck occurs then this transition is the  $F \rightarrow S$ -transition [54]. Thus, the well-known breakdown phenomenon at a highway bottleneck, i.e., a

sharp reduce in the vehicle speed at the bottleneck during the onset of congestion (e.g., [83–85]) is explained in the three-phase traffic theory by the F→S-transition at the bottleneck [54,86].

The existence of the limit point for free flow at the bottleneck (the density and the flow rate in this limit point will be designated  $\rho_{max}^{(free\ B)}$  and  $q_{max}^{(free\ B)}$ , respectively) is also explained in the three-phase traffic theory by the occurrence of the F→S-transition rather than the moving jam emergence (the F→J-transition). However, this limit (maximal) density  $\rho = \rho_{max}^{(free\ B)}$  and the limit (maximal) flow rate  $q_{max}^{(free\ B)}$  at the bottleneck are usually lower than respectively the limit density  $\rho = \rho_{max}^{(free)}$  and the limit flow rate  $q_{max}^{(free)}$  in Fig. 1 (a) which are related to a homogeneous (without bottlenecks) multi-lane road (see for more detail Sec. 7.2). This is an important peculiarity of the F→S-transition at a highway bottleneck in comparison with the case of a homogeneous road considered above. The following hypothesis is related to this case [54]:

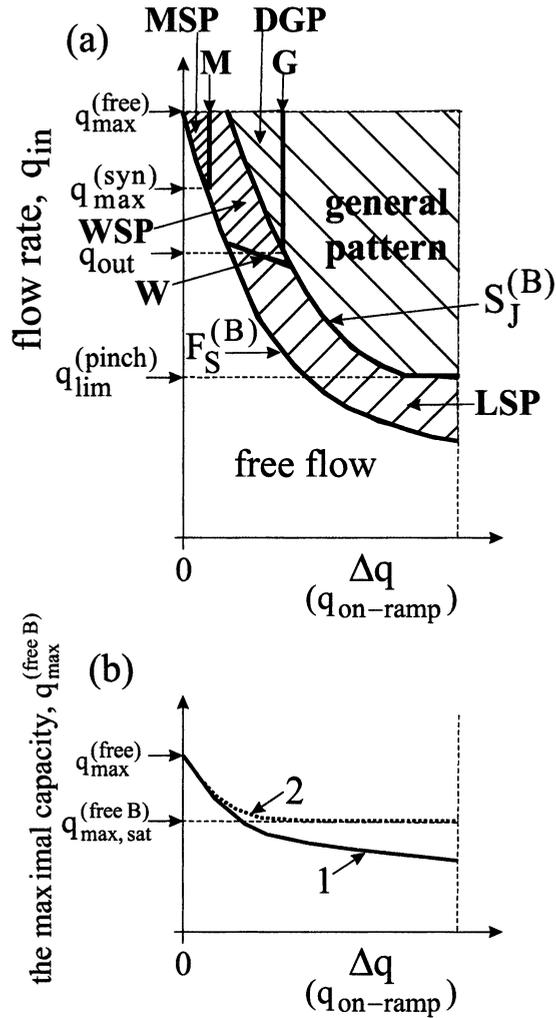
The probability of the F→S-transition per a highway location and a time interval depends on a location of a road. This probability has a maximum at the bottleneck (see Fig. 1 (b) in [54]). Note that the road location where this maximum is reached is called the effective location of the bottleneck (or the effective bottleneck for short).

This hypothesis can also be explained from driver's behavior. Let us consider a road where only one bottleneck exist. Besides upstream and downstream of the bottleneck the road has the same characteristics and it is homogeneous. A feature of the bottleneck is that in the vicinity of the bottleneck each driver should slow down, i.e., decrease the speed. Thus, a local perturbation in the average speed appears at the bottleneck. This perturbation is motionless and permanent because it is localized at the highway bottleneck: The bottleneck forces all drivers permanent to slow down at approximately the same road location.

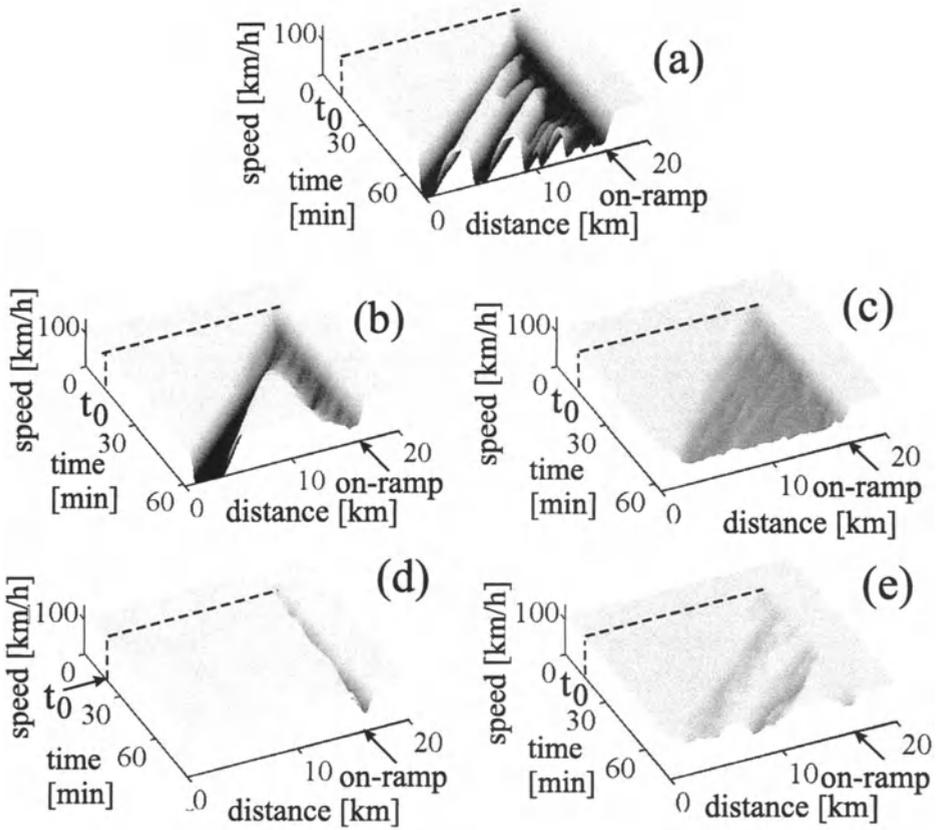
Inside this permanent and time-independent perturbation, where the speed is lower than the speed away from the bottleneck, the vehicle density must be higher than the density away from the bottleneck. Indeed, if the total flow rate across the road upstream of the bottleneck does not depend on time then this total flow rate across the road does not also depend on the highway location, i.e., it remains the same inside the perturbation at the bottleneck and away from the bottleneck. This explains why the probability of the F→J-transition should have a maximum at the bottleneck. This also means that the F→J-transition should mostly occur at highway bottlenecks.

## 6.2 Diagram of Congested Patterns in the Three-Phase Traffic Theory [51,67,78,82]

The hypotheses discussed above allow us to predict the diagram of spatial-temporal congested patterns at a highway bottleneck (Fig. 4 (a)) [51,78]. Dependent on the flow rate upstream of the bottleneck  $q_{in}$  and on the bottleneck strength  $\Delta q$  different congested patterns can occur at the bottleneck (Fig. 5) [51,67,78].



**Fig. 4.** Diagram of congested patterns at highway bottlenecks on a multi-lane highway in the three-phase traffic theory (a) [51,67,78,82] and possible dependencies of the maximal highway capacity in free flow at the effective location of a bottleneck  $q_{max}^{(free B)}$  on the bottleneck strength  $\Delta q$  (b).



**Fig. 5.** Main types of congested patterns at highway bottlenecks in the three-phase traffic theory [51,67,78,82]: The general pattern (GP) (a), the dissolving general pattern (DGP) (b), the widening synchronized flow pattern (WSP) (c), the localized synchronized flow pattern (LSP) (d), and the moving synchronized flow pattern (MSP) (e). Results of numerical simulations of a microscopic traffic flow model in the frame of the three-phase traffic theory. Taken from [67].

There are two main boundaries in the diagram of congested patterns at highway bottlenecks:  $F_S^{(B)}$  and  $S_J^{(B)}$  [51,67,78]. Below and left of the boundary  $F_S^{(B)}$  free flow is realized (Fig. 4 (a)). Between the boundaries  $F_S^{(B)}$  and  $S_J^{(B)}$  different synchronized flow patterns (SP) occur (Fig. 5 (c-e)) [51,67,78]. SP consists of synchronized flow upstream of the bottleneck where no wide moving jams emerge.

Right of the boundary  $S_J^{(B)}$  wide moving jams spontaneously emerge in synchronized flow upstream of the bottleneck. The spatial-temporal congested pattern upstream of the bottleneck which consists of both traffic phases (the phase

"synchronized flow" and the phase "wide moving jam") is called the general pattern (GP) (Fig. 5 (a)) [51,67,78].

Between the boundaries  $F_S^{(B)}$  and  $S_J^{(B)}$ , the higher  $q_{in}$  is the higher is the probability that the flow rate in synchronized flow in SP is lower than  $q_{in}$  and the length of SP is continuously increasing over time: At higher  $q_{in}$  a widening of SP (WSP) (Fig. 5 (c)) and at lower  $q_{in}$  the localized SP (LSP) (Fig. 5 (d)) occurs (Fig. 4 (a)). The flow rate inside WSP is lower than  $q_{in}$ . Therefore, the upstream WSP front (boundary), which separates free flow upstream and synchronized flow downstream, is continuously widening upstream. The mean flow rate inside LSP is equal to  $q_{in}$ . For this reason, the upstream LSP front is not continuously widening upstream: The width of LSP is spatially limited. However, this LSP width can show oscillations over time [67,70]. The boundary which separates the region of WSP and the region of LSP is marked in the diagram by the letter  $W$  in Fig. 4 (a).

Note that the flow rate in the synchronized flow of any SP can not exceed the characteristic value  $q_{max}^{(syn)}$  (Fig. 2 (a)). Right of the boundary  $F_S^{(B)}$  and left of the line  $M$  (the region marked "MSP" in Fig. 4 (a)) the flow rate  $q_{in}$  in an initial free flow upstream of the on-ramp satisfies the condition  $q_{in} > q_{max}^{(syn)}$  (Fig. 4 (a)). The point where the line  $M$  intersects the curve  $F_S^{(B)}$  is related to the flow rate  $q_{in} = q_{max}^{(syn)}$ . After SP has just appeared at the on-ramp, the flow rate directly upstream of the on-ramp decreases. This flow rate in SP, pinned at the on-ramp can not be higher than  $q_{max}^{(syn)}$ . This is in contrast to the initial condition  $q_{in} > q_{max}^{(syn)}$ . Apparently for this reason, it has been found out [67] that one or a sequence of moving SP (MSP) emerge upstream of the on-ramp (Fig. 5 (e)). After SP has emerged at the on-ramp, SP comes off the on-ramp and transforms into MSP. In MSP both the upstream and the downstream fronts move upstream of the on-ramp, i.e., MSP moves as a whole localized pattern upstream of the on-ramp. In contrast to a wide moving jam, inside MSP both the vehicle speed (40-70 km/h) and the flow rate are high. Besides, the velocity of the downstream front of MSP is *not* a characteristic parameter. This velocity can change in a wide range in the process of the MSP propagation or for different MSP's. In some cases it has been found out that after MSP is far away from the on-ramp, the pinch effect (the self-compression of synchronized flow) occurs inside MSP and a wide moving jam can be formed there [67].

The boundary  $F_S^{(B)}$  (Fig. 4 (a)) is determined by an occurrence of the F→S-transition at the bottleneck. The nature of this boundary is similar to the one for the curve  $F_S$  in Fig. 2 (b) which determines the critical amplitude of local perturbations for the F→S-transition in traffic flow without bottlenecks [49,55,59]. The effective bottleneck acts as a permanent non-homogeneity which causes the related permanent perturbation at the bottleneck [54]. The higher  $\Delta q$ , the higher is the amplitude of this permanent perturbation. Therefore, the higher  $\Delta q$  the lower the flow rate  $q_{in}$  at which the related critical amplitude occurs at the bottleneck. This may explain the decreasing character of the boundary  $F_S^{(B)}$  in the

flow-flow plane in Fig. 4 (a). The limit point of the boundary  $F_S^{(B)}$  at  $\Delta q = 0$  is related to the maximum flow rate in free flow where  $q_{in} = q_{max}^{(free)}$ .

The boundary  $S_J^{(B)}$  is determined by the wide moving jam emergence in synchronized flow (i.e., the S→J-transition) upstream of the on-ramp. The nature of the boundary  $S_J^{(B)}$  is similar to the one for the curve  $S_J$  in Fig. 2 (b). The latter boundary determines the critical amplitude of local perturbation and the related probability for the S→J-transition in synchronized flow of a given vehicle speed [49,53,45,55].

On the one hand, between the boundaries  $F_S^{(B)}$  and  $S_J^{(B)}$  at the same  $q_{in}$  the vehicle speed in SP should decrease when  $\Delta q$  increases. On the other hand, from Fig. 3 (a, b) it can be seen that at the same flow rate  $q$  the lower the vehicle speed in synchronized flow is the lower is the critical amplitude  $\Delta\rho_c$  and therefore the higher is the probability of the S→J-transition (compare  $S_J^{(1)}$  and  $S_J^{(2)}$  for the respective vehicle speeds  $v_{syn}^{(1)}$  and  $v_{syn}^{(2)}$ , where  $v_{syn}^{(1)} > v_{syn}^{(2)}$  in Fig. 3 (a, b)). Thus, in comparison with  $F_S^{(B)}$ , the boundary  $S_J^{(B)}$  should be shifted to the right in the flow-flow plane (Fig. 4 (a)).

The empirical results of the GP formation which have recently been found out in [51,78,82] also allow to suggest that if due to the high value of  $\Delta q$  the strong congestion (for the definition of "strong" and "weak" congestion see in [51]) is achieved in GP, then GP can not exist if  $q_{in} < q_{lim}^{(pinch)}$ , where  $q_{lim}^{(pinch)}$  is the limit flow rate in the pinch region. Therefore, under the strong congestion the boundary  $S_J^{(B)}$  should transform into a horizontal line at  $q_{in} = q_{lim}^{(pinch)}$ .

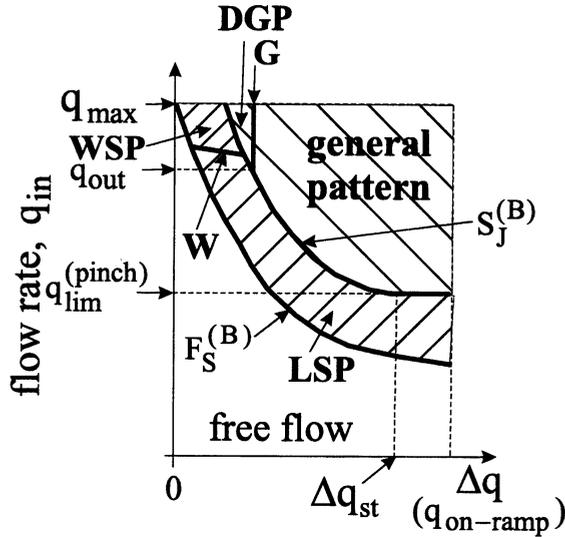
Right of the boundary  $S_J^{(B)}$  and left of the line  $G$  (the region marked "DGP" in Fig. 4 (a)) in an initial (before the congested pattern formation) free flow upstream of the on-ramp the flow rate  $q_{in}$  satisfies the condition  $q_{in} > q_{out}$ . The point where the line  $G$  intersects the curve  $S_J^{(B)}$  is related to the flow rate  $q_{in} = q_{out}$ . Thus, after a wide moving jam in synchronized flow of the congested pattern has been formed, this initial condition  $q_{in} > q_{out}$  can not be satisfied any more because the flow rate in the jam outflow can not be higher than  $q_{out}$ . As a result, the *dissolving general pattern* or DGP for short occurs (Fig. 5 (b)), i.e., the GP which dissolves over time (Fig. 4 (a)). As a result of this GP dissolving process, GP transforms into one of the SP, or free flow occurs at the bottleneck [67,70].

Right of the boundary  $S_J^{(B)}$  and right of the line  $G$ , GP occurs (Fig. 4 (a)). As well as in the empirical results presented in [51,78], GP does *not* transform into another type of pattern if  $\Delta q$  increases. If  $\Delta q$  decreases, GP can transform into one of the SP.

The diagram in Fig. 4 (a) [51,67,78] can also explain the variety of diverse patterns and the transformations between them, which occur in the progress of time under the weak congestion at off-ramps which have recently been found out in [51]. Indeed, the weak congestion in [51] should be related to lower  $\Delta q$ , i.e., to the part of the diagram in Fig. 4 (a) where even relatively small changes in  $q_{in}$  and  $\Delta q$  can cause diverse transitions between MSP, WSP, GP, DGP and free

flow. Note that in the case of the off-ramp the bottleneck strength  $\Delta q$  is related to the percentage of vehicles which want to leave the main road to the off-ramp.

Note that for a one-lane road the diagram of congested patterns should be qualitatively similar to that for a multi-lane highway with one exception: The part of the diagram above the flow rate  $q_{in} = q_{max}^{(syn)}$  in Fig. 4 (a) is not realized in the diagram of congested patterns for the one-lane road (Fig. 6) [51]. This can be linked to the hypotheses of the three-phase traffic theory (Fig. 1 (b)) [49,53,45,55]. Indeed, these hypotheses suggest that the maximal flow rate  $q_{max}$  (Fig. 1 (b)) in hypothetical spatially homogeneous and time-independent (steady) states on the one-lane road is equal to the maximal flow rate in the related states of synchronized flow on the multi-lane highway:  $q_{max} = q_{max}^{(syn)}$  (Fig. 1 (a)). Therefore, corresponding to Fig. 6 MSP should not occur at a bottleneck on the one-lane road.



**Fig. 6.** Diagram of congested patterns at highway bottlenecks on a one-lane road in the three-phase traffic theory [51].

In Fig. 4 (a) and Fig. 6, fluctuations and hysteresis effects have not been taken into account. Fluctuations may lead to the S→F-transition inside WSP, and in particular to an occurrence of MSP also on one-lane roads. This is indeed has recently been found out in the KKW CA models within the three-phase traffic theory [70]. Hysteresis effects may lead to an appearance of regions where dependent on initial conditions several different patterns can occur. In particular, the region of solely MSP may "shrink" so that in a limit case in the whole region marked "MSP" in Fig. 4 (a) WSP can exist on a multi-lane road. The region of

solely DGP may also "shrink" and in a limit case GP can exist at all values  $q_{in}$  right of the boundary  $S_j^{(B)}$ .

### 6.3 Comparison with the Diagram of Congested Patterns in the Fundamental Diagram Approach

The diagram of congested patterns at highway bottlenecks for a wide class of models in the fundamental diagram approach [32,33,36,42] is qualitatively totally different from the diagram in the three-phase traffic theory (Figs. 4 (a) and 6) [51,67,78]:

(i) In the diagram based on the fundamental diagram approach [32,36,42], homogeneous congested patterns (HCT) occur at very *high* flow rates to the on-ramp  $q_{on-ramp}$  (in the general case, the bottleneck strength  $\Delta q$ ) *only*. In contrast, in Figs. 4 (a) and 6 [51,67,78], WSP (Fig. 5 (c)), where as in HCT synchronized flow can be homogeneous can occur in the vicinity of *low*  $q_{on-ramp}$  *only*.

(ii) In [32,36,42], at a given high enough  $q_{in}$  if the flow rate  $q_{on-ramp}$  continuously increases first triggered stop-and-go traffic (TSG) where no synchronized flow can be formed occurs, then oscillating patterns (OCT) where no wide moving jams can be formed appear, and finally HCT occurs. In contrast, in the diagrams in Figs. 4 (a) and 6 [51,67,78] there are neither TSG nor OCT, nor else HCT.

(iii) Near the boundary which separates TSG and OCT a congested pattern which is a "mixture" of TSG, OCT and HCT can occur [42]. This pattern which at first sight looks like GP has been used in [42] for an explanation of the pinch effect and the jam emergence observed in [49]. However, this mixture pattern has no own region in the diagram of states in [32,36,42]. The pattern transforms into TSG if  $q_{on-ramp}$  decreases or into OCT (or else into HCT) if  $q_{on-ramp}$  increases.

In our diagrams (Figs. 4 (a) and 6) [51,67,78] there are no TSG, no OCT, and no HCT. Instead, GP exists in the very large range of the flow rates  $q_{on-ramp}$  and  $q_{in}$ . At a given  $q_{in}$  GP in the three-phase traffic theory does not transform into another congested pattern even if  $q_{on-ramp}$  increases up to the highest possible values. Thus GP in Figs. 4 (a) and 6 [51,67,78] has a qualitatively different nature in comparison with the mixture of TSG, OCT and HCT in [42].

(iv) In [32,36,42] it is claimed that if  $q_{on-ramp}$  decreases, then dependently on  $q_{in}$ , either pinned localized cluster (PLC) or TSG, i.e., either a pinned jam or moving jams of very high density and very low speed occur. This is also in contrast to the diagrams in Figs. 4 (a) and 6 [51,67,78]. If  $q_{on-ramp}$  decreases SP occurs where the density is much lower and the speed is much higher than inside either synchronized flow of the general pattern or inside any jams (for a more detailed comparison see [67]).

The diagram of states in the fundamental diagram approach [32,36,42] predicts the following sequence of the congested pattern transformation if  $q_{on-ramp}$  gradually increases: (i) TSG→OCT→HCT at a high given flow rate  $q_{in}$  and (ii) PLC→OCT→HCT at a lower given  $q_{in}$ . These sequences (or a part of them or

even one of these theoretical congested states) have *not* been observed in the empirical study of congested patterns at isolated bottlenecks (an isolated bottleneck is the effective bottleneck which is far enough from other effective bottlenecks) [51]. In contrast, GP usually spontaneously emerges at the on-ramp. GP does not transform into another congested pattern if  $q_{on-ramp}$  increases [51]. If  $q_{on-ramp}$  decreases, GP transforms into SP where the vehicle speed is considerably higher than in the synchronized flow of GP. This is in accordance with the diagram in Fig. 4 (a) of the three-phase traffic theory [49,55,59,67,78].

## 7 Theory of Highway Capacity

The determination of highway capacity is one of the most important applications of any traffic theory. Empirical observations show that the speed breakdown at a bottleneck (the breakdown phenomenon) is in general accompanied by a drop in highway capacity. If there is free flow rather than congested flow upstream of the bottleneck, highway capacity is usually higher. This phenomenon is called "the capacity drop" (see e.g., [83,84]). Here we give a qualitative theory of highway capacity which follows from the three-phase traffic theory by Kerner. Highway capacity depends on whether a homogeneous road (without bottlenecks) or a highway bottleneck is considered.

### 7.1 Homogeneous Road

On a homogeneous (without bottlenecks) multi-lane road, highway capacity depends on in which traffic phase the traffic is in [49]: (i) The maximal highway capacity in the traffic phase "free flow" is equal to the maximal possible flow rate in free flow,  $q_{max}^{(free)}$ . (ii) The maximal highway capacity in the traffic phase "synchronized flow" is equal to the maximum possible flow rate in synchronized flow,  $q_{max}^{(syn)}$ . (iii) The maximal highway capacity downstream of the traffic phase "wide moving jam" is equal to the flow rate in the wide moving jam outflow,  $q_{out}$ . Because of the first order phase transitions between the traffic phases each of these maximal highway capacities has a probabilistic nature.

In particular, the probabilistic nature of the highway capacity in the traffic phase "free flow" means the following [45]:

(1) The probability of the F→S transition (the spontaneous local phase transition from free flow to synchronized flow) per a chosen time interval  $T_{ob}$  (this probability will be designated  $P_{FS}$ ) reaches one at the flow rate  $q = q_{max}^{(free)}$ :

$$P_{FS} = 1 \Big|_{q=q_{max}^{(free)}} . \quad (1)$$

This means that the flow rate  $q = q_{max}^{(free)}$  is determined as the flow rate in free flow at which the F→S transition occurs during  $T_{ob}$  with the probability equals one. In other words, the maximal capacity  $q = q_{max}^{(free)}$  depends on the time interval  $T_{ob}$  (at least in some range of  $T_{ob}$ ).

(2) There is a threshold flow rate  $q_{th}$  which is lower than  $q = q_{max}^{(free)}$  (Fig. 2 (b)):

$$P_{FS} = 0 \mid_{q=q_{th}} . \quad (2)$$

At  $q < q_{th}$  no spontaneous F→S transition can occur during the time interval  $T_{ob}$ . The threshold flow rate  $q_{th}$  depends on the time interval  $T_{ob}$  (at least in some range of  $T_{ob}$ ).

(3) If the flow rate in free flow  $q$  is within the range  $[q_{th}, q_{max}^{(free)}]$  then the higher  $q$  is the higher is the probability the F→S transition  $P_{FS}$ . Thus, the attribute of this *probabilistic highway capacity* is the probability  $1 - P_{FS}$  that free flow remains on the road during the whole time interval  $T_{ob}$ .

However, if a bottleneck exist on the road, then a much more complicated non-linear phenomena determine highway capacity.

## 7.2 Highway Capacity in Free Flow at Bottleneck

In the three-phase traffic theory, the breakdown phenomenon at a highway bottleneck is explained by the F→S transition at the bottleneck [54] (Sec. 6.1). Due to the bottleneck the road is spatially non-homogeneous, i.e., highway capacity can depend on a highway location. We will consider highway capacity in free flow which is related to the effective location of the bottleneck, i.e., the location where the probability of the F→S transition per a time interval as function of a highway location (the probability of the breakdown phenomenon) has a maximum (see Fig. 1 (b) in [54]). We will designate the flow rate in free flow at this location as  $q_{sum}$ .

The F→S transition occurs at the bottleneck during a given time interval  $T_{ob}$  if the flow rate upstream of the bottleneck on the main road  $q_{in}$  and the bottleneck strength  $\Delta q$  are related to the boundary  $F_S^{(B)}$  in the diagram of congested patterns (Fig. 4 (a)) [51].

Thus, there is *an infinity multitude* of maximal highway capacities of free flow at the bottleneck which are given by the points on the boundary  $F_S^{(B)}$ . These capacities depend on the values  $q_{in}$  and  $\Delta q$  at the boundary  $F_S^{(B)}$  in the diagram of congested patterns at the bottleneck (Fig. 4 (a)). We designate these capacities  $q_{max}^{(free B)}$ :

$$q_{max}^{(free B)} = q_{sum} \mid_{F_S^{(B)}} . \quad (3)$$

These capacities are the maximal capacities related to the time interval  $T_{ob}$ . This means that the probability of the F→S transition at the bottleneck per the time interval  $T_{ob}$  (this probability will be designated  $P_{FS}^{(B)}$ ) reaches one at the flow rate  $q_{sum} = q_{max}^{(free B)}$  (3):

$$P_{FS}^{(B)} = 1 \mid_{q_{sum} = q_{max}^{(free B)}} . \quad (4)$$

This means that the flow rate  $q = q_{max}^{(free B)}$  is determined as the flow rate in free flow at which the F→S transition occurs at the bottleneck during the time

interval  $T_{ob}$  with the probability equals one. In other words, each of the maximal capacities  $q = q_{max}^{(free\ B)}$  depends on the time interval  $T_{ob}$  (at least in some range of  $T_{ob}$ ).

For values  $\Delta q, q_{in}$  which are related to points in some vicinity *left of* the boundary  $F_S^{(B)}$  in the diagram of congested patterns, i.e., in free flow region the F→S transition nevertheless occurs at the bottleneck during the time interval  $T_{ob}$  with a probability  $P_{FS}^{(B)} < 1$ . This probability  $P_{FS}^{(B)}$  the lower is the more is the distance of a point  $(\Delta q, q_{in})$  from the boundary  $F_S^{(B)}$  in the diagram of congested patterns (Fig. 4 (a)).

The region in the diagram of congested patterns where this probabilistic effect occurs is restricted by a threshold boundary which is left of the boundary  $F_S^{(B)}$ : This threshold boundary which depends on the time interval  $T_{ob}$  is related to the *infinity multitude* of threshold flow rates  $q_{th}^{(B)}(\Delta q, q_{in})$  for which

$$P_{FS}^{(B)} = 0 \mid_{q_{sum}=q_{th}^{(B)}} . \quad (5)$$

Usually the higher the bottleneck strength  $\Delta q$  is the lower is the maximal highway capacity in free flow at the bottleneck  $q_{max}^{(free\ B)}$ : The maximal highway capacity  $q_{max}^{(free\ B)}$  is a decreasing function of  $\Delta q$  (curve 1 in Fig. 4 (b)). The highest is the capacity of free flow at  $\Delta q = 0$ . This case is obviously related to the maximal capacity on a homogeneous (without bottlenecks) road (Fig. 2 (a)):

$$q_{max}^{(free\ B)} \mid_{\Delta q=0} = q_{max}^{(free)} . \quad (6)$$

Let us further in Sec. 7 consider the special example of a bottleneck due to an on-ramp. In this case, the bottleneck strength  $\Delta q = q_{on-ramp}$  where  $q_{on-ramp}$  is the flow rate to the on-ramp and  $q_{sum} = q_{on-ramp} + q_{in}$ .

A decrease of the maximal highway capacity  $q_{max}^{(free\ B)}$  at the on-ramp when the flow rate  $q_{on-ramp}$  increases can have a saturation at a high enough flow rate to the on-ramp  $q_{on-ramp}$  (Fig. 4 (b), dotted curve 2) [70]. In this case, the maximal highway capacity  $q_{max}^{(free\ B)}$  does not reduce below some saturation value  $q_{max, sat}^{(free\ B)}$  even at a very high flow rate to the on-ramp. As it follows from (6) the highest is the capacity of free flow at  $q_{on-ramp} = 0$ :

$$q_{max}^{(free\ B)} \mid_{q_{on-ramp}=0} = q_{max}^{(free)} . \quad (7)$$

Corresponding to (4) and (5), if the flow rate in free flow  $q_{sum} = q_{on-ramp} + q_{in}$  at the effective location of the bottleneck is within the range  $[q_{th}^{(B)}, q_{max}^{(free\ B)}]$  then at a given  $q_{on-ramp}$  the higher  $q_{sum}$  is the higher is  $P_{FS}^{(B)}$ . Note that an increasing dependence of the probability  $P_{FS}^{(B)}$  of the breakdown phenomenon at the on-ramp of the flow rate  $q_{sum}$  has indeed been found out in numerical simulations of the KKW CA-model within three-phase traffic theory [70]. This explains empirical results by Persaud *et al.* [85] and empirical features of the breakdown phenomenon at the on-ramp as the first order F→S transition [86,54].

### 7.3 Highway Capacity in Congested Traffic at Bottleneck. Capacity Drop

In order to study the highway capacity downstream of the congested bottleneck one has to consider the outflow from a congested bottleneck  $q_{out}^{(bottle)}$  (the discharge flow rate), which is measured downstream of the bottleneck, where free flow conditions are reached. In the three-phase traffic theory, the discharge flow rate  $q_{out}^{(bottle)}$  is not just a characteristic property of the type of bottleneck under consideration only. It also depends on the type of congested pattern which actually is formed upstream of the bottleneck [54]. Thus, in the three-phase traffic theory, highway capacity in free flow downstream of the congested bottleneck depends on the type of congested pattern upstream of the bottleneck, the pattern characteristics and on parameters of the bottleneck. We call this highway capacity as a *congested pattern capacity*.

In the case of an on-ramp,  $q_{out}^{(bottle)}$  is expected to vary with  $(q_{on-ramp}, q_{in})$ . Obviously,  $q_{out}^{(bottle)}$  only limits the highway capacity, if it is smaller than the traffic demand upstream of the on-ramp,  $q_{sum} = q_{in} + q_{on-ramp}$ , i.e. if the condition

$$q_{out}^{(bottle)}(q_{on-ramp}, q_{in}) < q_{sum}(q_{on-ramp}, q_{in}) \quad (8)$$

is fulfilled. Then the congested pattern upstream from the on-ramp simply expands, while the throughput remains limited by  $q_{out}^{(bottle)}$ . For example, if the general pattern (GP) is formed at the bottleneck, an increase of  $q_{in}$  does not influence the discharge flow rate  $q_{out}^{(bottle)}$ . Instead, the width of the wide moving jam, which is most upstream in the GP, simply grows.

Thus, in the case (8) the congested pattern capacity  $q_{cong}^{(B)}$  is equal to  $q_{out}^{(bottle)}$ :

$$q_{cong}^{(B)} = q_{out}^{(bottle)}. \quad (9)$$

It must be noted that  $q_{out}^{(bottle)}$  can strongly depend on the congested pattern type and congested pattern parameters [54,51,67,70]. The congested pattern type and the pattern parameters depend on initial conditions and the flow rates  $q_{on-ramp}$  and  $q_{in}$  (Fig. 4 (a)). Thus, the congested pattern capacity  $q_{cong}^{(B)}$  implicitly depends on the flow rates  $q_{on-ramp}$  and  $q_{in}$  [54].

The capacity drop is the difference between highway capacity in free flow and in a situation, where there is synchronized flow upstream and free flow downstream of the bottleneck.

Assuming that (8) is fulfilled, the capacity drop is given by

$$\delta q = q_{max}^{(free)} - q_{cong}^{(B)}, \quad (10)$$

where the congested pattern capacity  $q_{cong}^{(B)}$  is given by the fomula (9);  $q_{max}^{(free)}$  is the maximal highway capacity in free flow at  $q_{on-ramp} = 0$ .

The minimum value, which  $q_{out}^{(bottle)}$  can take under the condition (8), if one considers all kinds of congested patterns upstream from a bottleneck, should

be a characteristic quantity for the type of bottleneck under consideration. We denote this quantity by  $q_{min}^{(bottle)}$ . The maximum of  $q_{out}^{(bottle)}$  (denoted by  $q_{max}^{(bottle)}$ ) is predicted to be the maximum flow rate, which can be realized in synchronized flow [54],  $q_{max}^{(bottle)} = q_{max}^{(syn)}$ . To explain the latter condition, recall that the downstream front of a congested pattern at the bottleneck due to the on-ramp separates free flow downstream of the front and synchronized flow upstream of the front. This downstream front is fixed at the bottleneck. Thus, within the front the total flow rate across the road (together with the on-ramp) does not depend on the co-ordinate along the road. Just downstream of the front, i.e., in free flow this flow rate is equal to  $q_{max}^{(bottle)}$ . Just upstream of the front it is suggested that synchronized flow occurs both on the main road and the on-ramp. The maximal possible total flow rate in synchronized flow is equal to  $q_{max}^{(syn)}$ .

Hence, the capacity drop at a bottleneck cannot be smaller than

$$\delta q_{min} = q_{max}^{(free)} - q_{max}^{(syn)}. \quad (11)$$

There are some remarks to this consideration:

(1) There may also be another definition of the capacity drop:

$$\delta q = q_{max}^{(free\ B)} - q_{cong}^{(B)}, \quad (12)$$

where the congested pattern capacity  $q_{cong}^{(B)}$  is given by the formula (9) and  $q_{max}^{(free\ B)}$  is given by (3). However, there could be a difficulty in the application of the definition (12): There is an infinity multitude of different maximal highway capacities in free flow at a bottleneck,  $q_{max}^{(free\ B)}$  (see the formula (3) and its explanation in Sec. 7.2).

(2) There may be one exception of the condition (8): If LSP (the localized synchronized flow pattern) (see Sec. 6.2) occur both on the main road and on the on-ramp upstream of the merge region of the on-ramp then the discharge flow rate is equal to traffic demand:

$$q_{out}^{(bottle)}(q_{on}, q_{in}) = q_{sum}(q_{on}, q_{in}). \quad (13)$$

The congested pattern capacity which is related to this LSP should be determined by the maximal discharge flow rate at which the LSP still exists upstream of the on-ramp.

(3) All flow rates considered above are total flow rates across the whole road (including the flow rates to on- and off-ramps) at the road location where highway capacity is considered.

(4) It has been suggested that an influence of fluctuations on the characteristic flow rates  $q_{th}$ ,  $q_{th}^{(B)}$ ,  $q_{max}^{(free)}$ ,  $q_{max}^{(free\ B)}$  in all formulas of Sec. 7 as well in the diagram of congested patterns (Figs. 4 and 6) is negligible. Otherwise, the characteristic flow rates  $q_{th}$  and  $q_{max}^{(free)}$  for a homogeneous (without bottlenecks) road can be different from the respective values  $q_{th}^{(B)} = q_{th\ lim}^{(B)}$  and  $q_{max}^{(free\ B)} = q_{max\ lim}^{(free\ B)}$  related to the limit case  $\Delta q \rightarrow 0$  ( $q_{on-ramp} \rightarrow 0$ ) for a road

with a bottleneck. Then, in the diagram of congested patterns at a bottleneck (Figs. 4 and 6) and in all corresponding formulas of Secs. 7.2 and 7.3 the condition  $\Delta q = 0$  ( $q_{on-ramp} = 0$ ) should be replaced by  $\Delta q \rightarrow 0$  ( $q_{on-ramp} \rightarrow 0$ ) and the flow rates  $q_{th}, q_{max}^{(free)}$  should be replaced by  $q_{th}^{(B)}, q_{max}^{(free B)}$ , respectively.

This general picture of the congested pattern capacity and the capacity drop proposed by the author for the three-phase traffic theory has been confirmed by empirical investigations [54,51] and in numerical studies of congested patterns at an on-ramp [67,70,87].

## 8 Classification of Congested Patterns at Highway Bottlenecks

An empirical study which has recently been made by the author allows to make the following classification of congested patterns at highway bottlenecks [51]:

- (i) There are two main types of congested patterns at an isolated bottleneck (the effective bottleneck which is far enough from other effective bottlenecks):
  - (1) *The general pattern* or *GP* for short: GP is the congested pattern at the isolated bottleneck where synchronized flow occurs upstream of the bottleneck and wide moving jams spontaneously emerge in that synchronized flow. Thus GP consists of both traffic phases in congested traffic: "synchronized flow" and "wide moving jam". The general pattern (GP) is the most frequent type of congested pattern at isolated bottlenecks.
  - (2) *The synchronized flow pattern* or *SP* for short: SP consists of synchronized flow upstream of the isolated bottleneck *only*, i.e., *no* wide moving jams emerge in that synchronized flow.  
However, dependent on the bottleneck features and on traffic demand, GP and SP show a diverse variety of special cases.
- (ii) In particular, there are three main different types of the synchronized flow patterns (SP) at the isolated bottleneck:
  - (1) The localized SP (LSP).
  - (2) The widening SP (WSP).
  - (3) The moving SP (MSP).
- (iii) In some cases in GP and SP (in LSP and WSP) local regions of free flow which spatially alternate with regions of synchronized flow can occur.
- (iv) If two or more effective bottlenecks are close to one another, the expanded congested pattern (EP) where synchronized flow covers two or more effective bottlenecks can occur (see empirical examples of EP in [51,79,82]).
- (v) For each effective bottleneck or for each set of several effective bottlenecks which are close to one another the spatial-temporal structure of the congested patterns possesses predictable, i.e., characteristic, unique and reproducible features, like the types of patterns which are frequently formed and the mean width of synchronized flow inside GP. These features can be nearly *the same* for different days and years. They can also remain within a large range of flow rates (traffic demand) at which the patterns exist. These results are used for the forecasting of congested patterns at highway bottlenecks [88,89].

## 9 Some Empirical Spatial-Temporal Features of Synchronized Flow

### 9.1 Example of WSP

As it has been mentioned above, there are synchronized flow patterns (SP) where no wide moving jams emerge. One of the SP is the widening SP (WSP). These patterns can occur at a low value of bottleneck strength  $\Delta q$  when the case of the weak congestion is realized [51]. The empirical study made in [51] allows to assume that the latter case often occurs at off-ramps. An importance of the WSP analysis is linked to the possibility to show empirical features of synchronized flow of a relative high vehicle speed. It will be shown, that in this case synchronized flow can exist upstream of the bottleneck on a long stretch of the highway (about 4.5 km) during a long time (more than 60 min) without wide moving jam emergence in that synchronized flow.

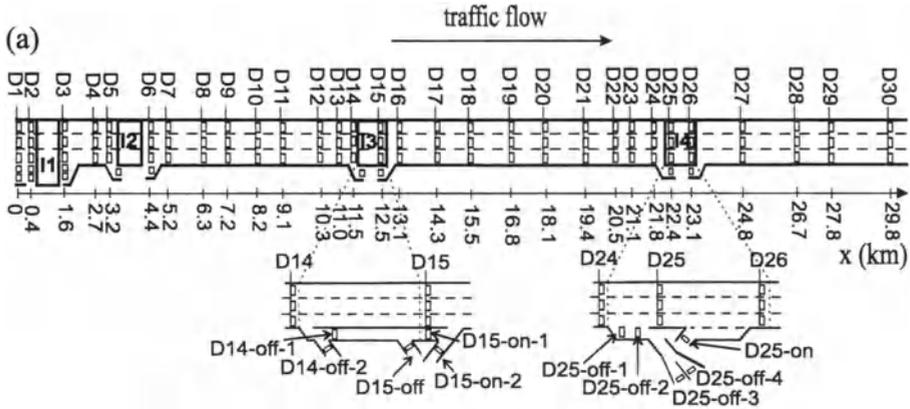
An example of WSP which occurs upstream of the off-ramp D25-off on the section of the highway A5-North is shown in Figs. 7, 8, 9 and 10. The section of the highway A5-North (Fig. 7 (a)) has already been described in [51] (see Fig. 3 (c) in this paper). WSP occurs as a result of the F $\rightarrow$ S transition upstream of the off-ramp (also discussed in [51]). It can be seen in Figs. 7 (b) and 8 that the vehicle speeds slowly decrease within WSP in the upstream direction whereas the flow rate does not change considerably when WSP occurs. This is a peculiarity of synchronized flow.

When the vehicle speed in WSP decreases in the upstream direction, some narrow moving jams emerge in this synchronized flow of low vehicle speed (D17 and D16, Fig. 8). However, D16 is already at the on-ramp (D15-on). For this reason, the synchronized flow propagating upstream, covers this upstream bottleneck at the on-ramp. As a result, an expanded congested pattern (EP) occurs [51] (this is not shown in Fig. 8). Thus, WSP upstream of the off-ramp at D25-off and downstream of the on-ramp at D16 is only a part of this EP. Nevertheless, the consideration of this WSP allows us some important conclusions about synchronized flow features.

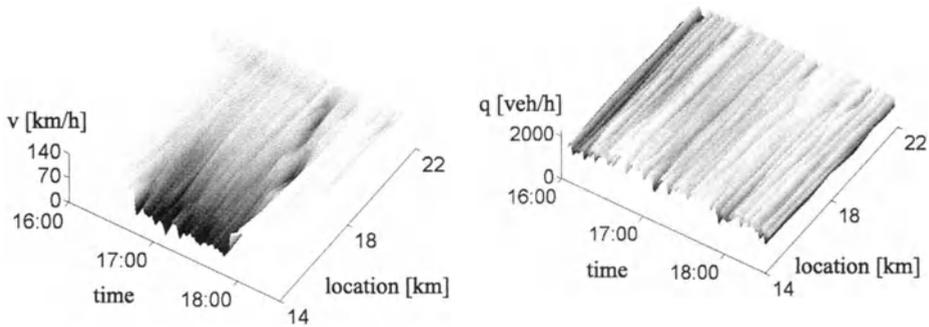
### 9.2 Overlapping of States of Free Flow and Synchronized Flow in Density

In particular, when free flow (black quadrates in Fig. 9) and synchronized flow inside WSP (circles in Fig. 9) are shown in the flow-density-plane it can be seen, that at least at detectors D20-D18 states of synchronized flow partially overlap with free flow in the density.

The same conclusion can be made if the vehicle speed as function of the density is drawn (Fig. 10 (a)). This means that at the same density either a state of synchronized flow or a state of free flow is possible. It is obviously that the mean probability of overtaking is higher in free flow than in synchronized flow. Thus, the empirical fact that states of free flow and synchronized flow overlap in the vehicle density (Figs. 9 and 10 (a)) means that the mean probability of



(b) A5-North, 26.06.1996



**Fig. 7.** The widening synchronized flow pattern (WSP): (a) - Schema of the highway infrastructure and local measurements at the section of the highway A5-North [51]; (b) - the average speed and the flow rate in WSP as functions of time and location [80].

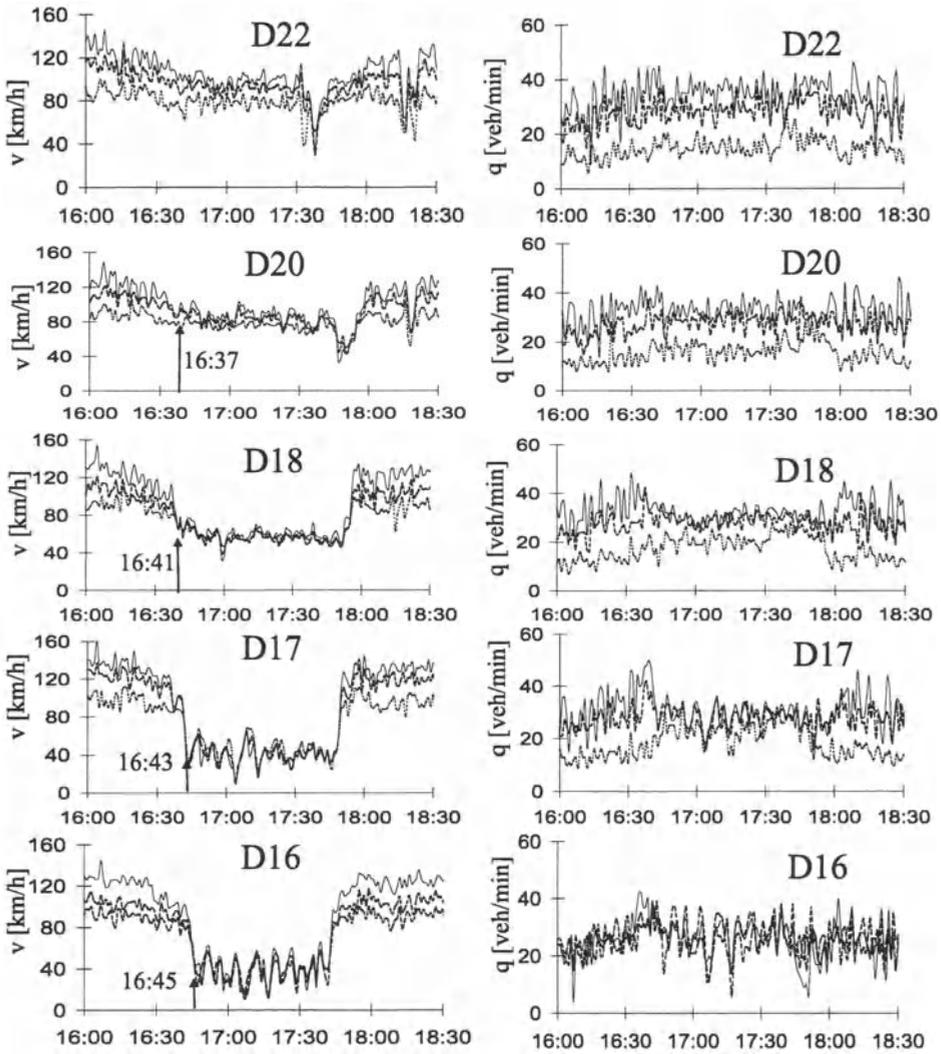
overtaking has a Z-shape. This confirms the hypotheses about the Z-form of the mean probability of overtaking in traffic flow discussed above (Fig. 2 (d)).

If now the average absolute values of the vehicle speed difference between left lane and middle lane  $\Delta v$  for free flow (curve  $F$  in Fig. 10 (b)) and for synchronized flow (curve  $S$  in Fig. 10 (b)) are shown, then this difference as a function of the density has also a Z-shape. This result is in agreement with the above conclusion about the Z-shape of the mean probability for overtaking (Fig. 2 (d)). Indeed, the mean rate of the overtaking the lower is the lower is the mean probability of overtaking  $P$ . The mean rate of the overtaking should decrease when the vehicle speed difference  $\Delta v$  becomes lower.

The physical meaning of the result in Fig. 10 (b) is the following: In free flow the difference in the average vehicle speed on German highways between left (passing) highway lane and middle lane due to the high mean probability of overtaking is considerably higher than that in synchronized flow. However,

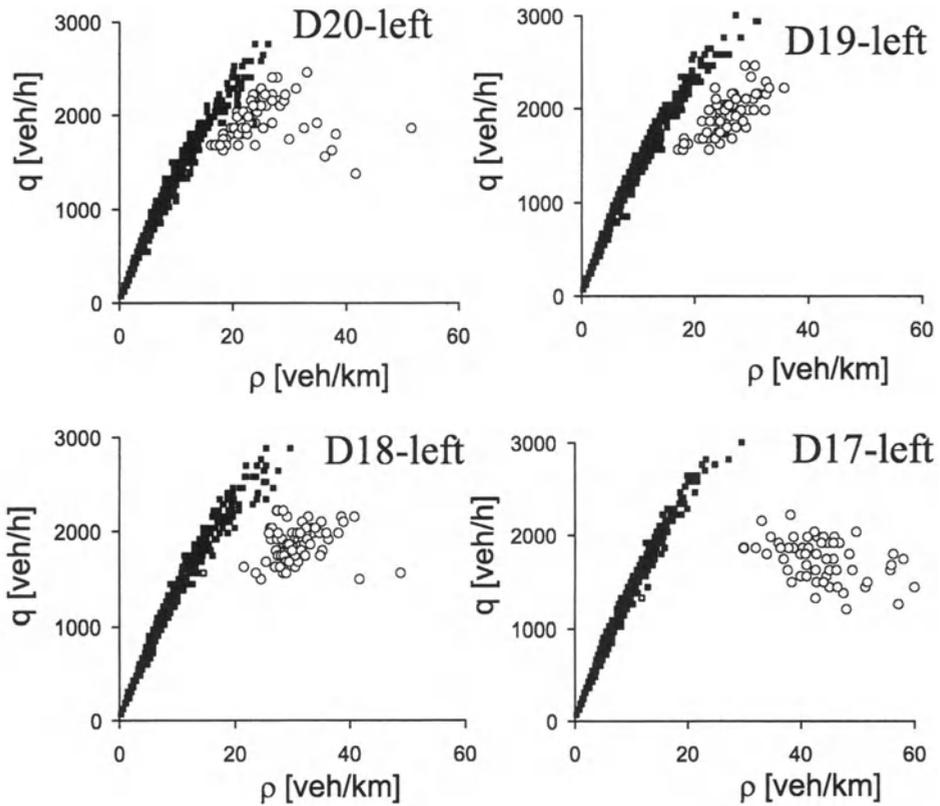
A5-North, 26.06.1996

— left lane    - - middle lane    ··· right lane



**Fig. 8.** The widening synchronized flow pattern (WSP): Time series of the vehicle speed (left) and flow rate (right) for different highway lanes for detectors D22 - D16. The F→S transitions at the related detectors leading to the WSP formation are marked with up arrows.

## A5-North, 26.06.1996



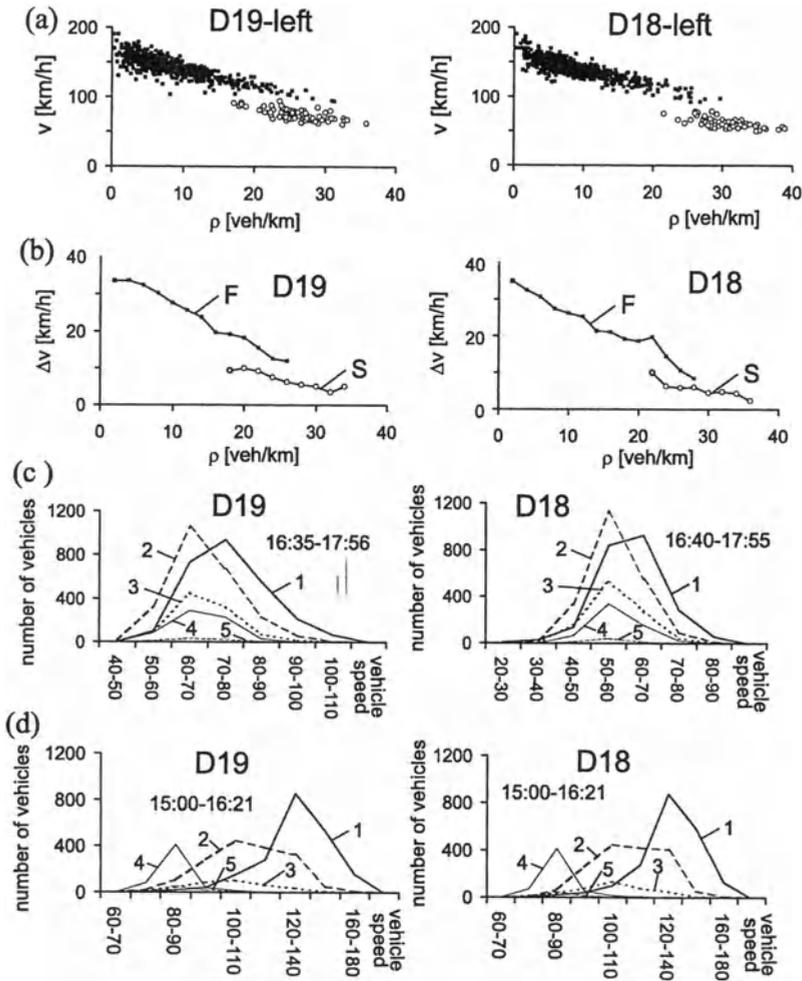
**Fig. 9.** The measurement points in the flow-density plane for WSP shown in Fig. 7 at the detectors D20 - D17 (left line). Free traffic is related to black quadrates, synchronized flow is related to circles. Overlapping of states of free flow and synchronized flow at the detectors D19 is in the density range from about 18 to 36 vehicles/km.

at the same density in a limited range (e.g., at D19 from 18 vehicles/km to 26 vehicles/km) either states of free flow or synchronized flow can exist. This leads to a nearly Z-form of the dependency  $\Delta v$  on the vehicle density. The lower the vehicle speed in synchronized flow is the less is the density range of the overlapping of the curves  $F$  and  $S$  in Fig. 10(b) (D18). This overlapping is fully disappearing if the vehicle speed in synchronized flow further decreases.

### 9.3 Analysis of Individual Vehicle Speeds

To see the difference between free flow and synchronized flow and features of synchronized flow more clearly, distributions of the number of vehicles as a function of the individual vehicle speed for synchronized flow (Fig. 10 (c)) and for free flow (Fig. 10 (d)) are shown. This is possible because the types of vehicles

## A5-North, 26.06.1996



**Fig. 10.** Empirical features of synchronized flow at the detectors D19 (left) and D18 (right) in WSP shown in Figs. 7 (b) and 8: (a) - Measurement points in the speed-density plane (free flow - black quadrates, synchronized flow - circles; left lane), (b) - The average difference in the vehicle speeds between left and middle highway lanes as the function of the density, curve  $F$  for free flow, curve  $S$  for synchronized flow. The speed differences in (b) are averaged for density intervals of 2 vehicles/km; (c, d) - Distribution of the number of vehicles as a function of the different speed classes related to individual single vehicle data for synchronized flow (c) and for free flow (d): The curve 1 is related to vehicles on the left lane, 2 - vehicles on the middle lane, 3 - vehicles on the right lane, 4 - long vehicles on the right lane, 5 - long vehicles on the middle lane (long vehicles may not move on the left (passing) lane of a three-lane (in one direction) highway in Germany). In (c, d), measured single vehicle data are shown whereby the number of vehicles in each of 15 different classes in regard to the vehicle speed is used separately for vehicles and for long vehicles.

and their individual vehicle speed during each of the one minute intervals are also available.

First, it can be seen that in synchronized flow the mean vehicle speed of vehicles and long vehicles are almost the same for different highway lanes whereas for free flow these mean values are strongly shifted to one another.

Second, we see that at the detectors D19 during of 121 min of the observation individual vehicle speeds in synchronized flow were not lower than 40 km/h (Fig. 10 (c), left). At the detectors D18 during 105 min of the observations individual speeds of 6181 vehicles which passed the detectors were measured (Fig. 10 (c), right). Among these 6181 vehicles there were no vehicles which had the speed below 20 km/h, there were only 9 vehicles which had individual speeds between 20 and 30 km/h and 59 vehicles which had individual speeds between 30 and 40 km/h. All other 6113 vehicles had the individual speeds higher than 40 km/h. Thus, there were no narrow moving jams in synchronized flow between D19 and D18. Nevertheless, these states of synchronized flow cover 2D regions in the flow-density plane (Fig. 9, D19-left, D18-left).

## 10 Conclusions

The presented concept of "synchronized flow" in the context of the three-phase traffic theory, both invented by the author himself, describes phase transitions and a diverse variety of spatial-temporal congested patterns both on homogeneous roads and at highway bottlenecks. It seems that the features of these phase transitions and of the spatial-temporal congested patterns at bottlenecks in the three-phase traffic theory [49,51,45,55,59,67,78] are qualitative totally different in comparison with the related results which have been derived in the fundamental diagram approach [41,42]. An exception is only the propagation of wide moving jams whose characteristic parameters and features appear to play an important role (in particular, the flow rate in the outflow from the jam,  $q_{out}$ ) in both the three-phase traffic theory and the traffic theories in the fundamental diagram approach.

Recent empirical results of a study of the congested patterns at on- and off-ramp and of their evolution when the bottleneck strength is gradually changing [51] confirm the discussed results and conclusions of the three-phase-traffic-theory [49,51,45,55,59,67,78] rather than the related results and conclusions of traffic theories in the fundamental diagram approach [41,42,32–36].

The three-phase traffic theory [53,49,55,59,67,78] has also been confirmed by the on-line application in the traffic center of the State Hessen of some recent models "ASDA" (Automatische Staudynamikanalyse: Automatic Tracing of Moving Traffic Jams) [90,91] and "FOTO" (Forecasting of Traffic Objects) [92,93] which are based on this traffic flow theory. These models allow reconstruction, tracing and prediction of spatial-temporal traffic dynamics based on local measurements of traffic. The models ASDA and FOTO perform without validation of model parameters at different traffic conditions (see a recent review about the models ASDA and FOTO in [80]).

A mathematical microscopic traffic flow model in the frame of the three-phase traffic theory which has recently been proposed by Kerner and Klenov [67] shows a considerable potential both for the development of qualitatively new mathematical traffic models and for the traffic flow theory development on the basis of the discussed hypotheses of the three-phase traffic theory. This is also confirmed by results of a numerical study of several new cellular automata traffic flow models which have recently been developed by Kerner, Klenov and Wolf [70].

## References

1. M. J. Lighthill and G. B. Whitham, Proc. R. Soc. A **229**, 317 (1955).
2. D. C. Gazis, R. Herman, and R. W. Rothery, Operations Res. **9**, 545 - 567 (1961).
3. G. F. Newell, Operations Res. **9**, 209 (1961).
4. I. Prigogine, in *Theory of Traffic Flow* (Herman R (ed.)) (Elsevier, Amsterdam, 1961), p. 158.
5. I. Prigogine and R. Herman, *Kinetic Theory of Vehicular Traffic*, (American Elsevier, New York, 1971).
6. D. E. Wolf, M. Schreckenberg, and A. Bachem (editors), *Traffic and Granular Flow*, Proceedings of the International Workshop on Traffic and Granular Flow, October 1995 (World Scientific, Singapore, 1995).
7. M. Schreckenberg and D. E. Wolf (editors), *Traffic and Granular Flow' 97*, Proceedings of the International Workshop on Traffic and Granular Flow, October 1997 (Springer, Singapore, 1998).
8. D. Helbing, H. J. Herrmann, M. Schreckenberg, and D. E. Wolf (editors), *Traffic and Granular Flow' 99*, Proceedings of the International Workshop on Traffic and Granular Flow, October 1999, (Springer, Heidelberg, 2000).
9. K. Nagel and M. Schreckenberg, J Phys. (France) I **2**, 2221 (1992).
10. K. Nagel and M. Paczuski, Phys. Rev. E **51**, 2909 (1995).
11. M. Schreckenberg, A. Schadschneider, K. Nagel, and N. Ito, Phys. Rev. E **51**, 2939 (1995).
12. G.B. Whitham, Proc. R. Soc. London A **428**, 49 (1990).
13. B. S. Kerner and P. Konhäuser, Phys. Rev. E **50**, 54 - 83 (1994).
14. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, Phys. Rev. E **51**, 1035 - 1042 (1995).
15. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, J. Phys. I France **5**, 1389 (1995).
16. Y. Sugiyama and H. Yamada, Phys. Rev. E **55**, 7749 (1997).
17. M. Herrmann and B. S. Kerner, Physica A **255**, 163 - 188 (1998).
18. S. Krauß, P. Wagner, and C. Gawron, Phys. Rev. E **53**, 5597 (1997).
19. R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg, Eur. Phys. J. B. **5**, 793 (1998).
20. T. Nagatani, Phys. Rev. E **58**, 4271 - 4279 (1998); Phys. Rev. E **59**, 4857 - 4864 (1999); Phys. Rev. E **60**, 180 - 187 (1999); Phys. Rev. E **60**, 6395 - 6401 (1999); Phys. Rev. E **61**, 3564 - 3570 (2000); Phys. Rev. E **61**, 3534 - 3540 (2000); Phys. Rev. E **64**, 036115 (2001); Phys. Rev. E **64**, 016106 (2001); Physica A **290**, 501 - 511 (2001); Physica A **280**, 602 - 613 (2000).
21. T. Nagatani, Physica A **258**, 237 (1998).
22. D. Helbing and M. Schreckenberg, Phys. Rev. E **59**, R2505 (1999).

23. M. Treiber, A. Hennecke, and D. Helbing, *Phys. Rev. E* **59**, 239 (1999).
24. R. Mahnke and J. Kaupužs, *Phys. Rev. E* **59**, 117 (1999).
25. G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
26. M. Cremer, *Der Verkehrsfluss auf Schnellstrassen* (Springer, Berlin, 1979).
27. W. Leutzbach, *Introduction to the Theory of Traffic Flow* (Springer, Berlin, 1988).
28. B. S. Kerner, S. L. Klenov, and P. Konhäuser, *Phys. Rev. E* **56**, 4200 - 4216 (1997).
29. B. S. Kerner, P. Konhäuser, and M. Schilke, *Phys. Rev. E* **51**, 6243 - 6246 (1995).
30. H. Y. Lee, H.-W. Lee, and D. Kim, *Phys. Rev. Lett.* **81**, 1130 (1998).
31. D. Helbing and M. Treiber, *Phys. Rev. Lett.* **81**, 3042 - 3045 (1998).
32. D. Helbing, A. Hennecke, and M. Treiber, *Phys. Rev. Lett.* **82**, 4360 (1999).
33. H. Y. Lee, H.-W. Lee, and D. Kim, *Phys. Rev. E* **59**, 5101 (1999).
34. H. Y. Lee, H.-W. Lee, and D. Kim, *Physica A* **281**, 78 (2000).
35. H. Y. Lee, H.-W. Lee, and D. Kim, *Phys. Rev. E* **62**, 4737 (2000).
36. M. Treiber, A. Hennecke, and D. Helbing, *Phys. Rev. E* **62**, 1805 (2000).
37. M. Treiber and D. Helbing *J. Phys. A: Math. Gen.* **32**, L17 (1999).
38. E. Tomer, L. Safonov, and S. Havlin, *Phys. Rev. Lett.* **84**, 382 (2000).
39. W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, *J. Phys. A* **33**, L477 (2000).
40. W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, *Phys. Rev. E* **65**, 015101 (R) (2002).
41. D. Chowdhury, L. Santen, and A. Schadschneider, *Physics Reports* **329**, 199 (2000).
42. D. Helbing, *Rev. Mod. Phys.* **73**, 1067-1141 (2001).
43. M. Koshi, M. Iwasaki, and I. Ohkura, in *Proceedings of 8th International Symposium on Transportation and Traffic Theory*, edited by V. F. Hurdle, et al (University of Toronto Press, Toronto, Ontario, 1983) p. 403.
44. J. Treiterer, Investigation of traffic dynamics by aerial photogrammetry techniques. Ohio State University Technical Report PB 246 094, Columbus, Ohio (1975).
45. B. S. Kerner, in *Transportation and Traffic Theory*, edited by A. Ceder (Elsevier Science, Amsterdam, 1999) p. 147 - 171.
46. C. F. Daganzo, *Fundamentals of Transportation and Traffic Operations* (Elsevier Science Inc., New York, 1997).
47. A. D. May, *Traffic Flow Fundamental* (Prentice Hall, Inc., New Jersey, 1990).
48. *Highway Capacity Manual* (Transportation Research Board, Washington, D.C. 1998).
49. B. S. Kerner, *Phys. Rev. Lett.* **81**, 3797 (1998).
50. B. S. Kerner, *J. of Physics A: Math. Gen.* **33**, L221 (2000).
51. B. S. Kerner, *Phys. Rev. E* **65**, 046138 (2002).
52. B. S. Kerner, and H. Rehborn, *Phys. Rev. E* **53**, R4275 - R4278 (1996).
53. B. S. Kerner, in *Proceedings of the 3<sup>rd</sup> Symposium on Highway Capacity and Level of Service*, edited by R. Rysgaard, Vol. 2 (Road Directorate, Ministry of Transport - Denmark, 1998) p. 621 - 642.
54. B. S. Kerner, *Transportation Research Record* **1710**, 136 - 144 (2000).
55. B. S. Kerner, *Transportation Research Record* **1678**, 160 - 167 (1999).
56. B. S. Kerner, in [8], pp. 253 - 284.
57. B. S. Kerner, *Networks and Spatial Economics*, **1**, 35 (2001).
58. B. S. Kerner, in [7], p. 239 - 267.
59. B. S. Kerner, *Physics World* **12**, (8), 25 - 30 (August 1999).
60. I. Lubashevsky and R. Mahnke, *Phys. Rev. E* **62**, 6082 (2000).
61. I. Lubashevsky, R. Mahnke, P. Wagner, and S. Kalenkov, *Phys. Rev. E* **66**, 016117 (2002).

62. P. Nelson, Phys. Rev. E **61**, R6052 (2000).
63. S. Rosswog and P. Wagner, Phys. Rev. E **65** (2002).
64. D. Helbing, D. Batic, M. Schönhof, and M. Treiber, cond-mat/0108548 (2001).
65. S. Kriso, R. Friedrich, J. Peinke, and P. Wagner, cond-mat/0110084 (2001).
66. E. Tomer, L. Safonov, N. Madar, and S. Havlin, cond-mat/0105493 (2001); Phys. Rev. E **65**, 065101 (R) (2002).
67. B. S. Kerner and S. L. Klenov, J. Phys. A: Math. Gen. **35**, L31 (2002).
68. M. Fukui, K. Nishinari, D. Takahashi, and Y. Ishibashi, Physica A **303**, 226-238 (2002).
69. D. Helbing and M. Treiber, trafficforum/02031301 (2002); Cooper@tive Tr@nsport@tion Dyn@mics **1**, 2.1.-2.24 (2002).
70. B. S. Kerner, S. L. Klenov, and D. E. Wolf, *Cellular automata approach to three-phase traffic theory*, cond-mat/0206370 (2002); J. Phys. A: Math. Gen. **35** (2002).
71. K. Nishinari, and D. Takahashi, J. Phys. A **33**, 7709 (2000).
72. P. G. Gipps, Trans. Res. B. **15**, 105 - 111 (1981).
73. E. Kometani and T. Sasaki, J. Oper. Res. Soc. Jap. **2**, 11 (1958); Oper. Res. **7**, 704 (1959); In: *Theory of Traffic Flow* edited by R. Herman (Elsevier, Amsterdam 1961), pp. 105.
74. M. Takayasu, and H. Takayasu, Fractals **1**, 860 (1993).
75. D. E. Wolf, Physica A **263**, 438 (1999).
76. T. Nagatani and K. Nakanishi, Phys. Rev. E **57**, 6415 (1998).
77. R. Wiedemann, *Simulation des Verkehrsflusses* (University of Karlsruhe, Karlsruhe 1974).
78. B. S. Kerner, in *Preprints of the Transportation Research Board 81 st Annual Meeting*, TRB Paper 02-2918, January 13-17, 2002 (TRB, Washington D.C., 2002); Transportation Research Record (2002) (at press).
79. B. S. Kerner, Mathematical and Computer Modelling **35**, 481-508 (2002).
80. B. S. Kerner, H. Rehborn, M. Aleksic, and A. Haug, in *Human Behaviour and Traffic Networks* edited by M. Schreckenberg and R. Selten (Springer, Berlin, 2002).
81. B. S. Kerner and H. Rehborn, Phys. Rev. E **53**, R1297 - R1300 (1996).
82. B. S. Kerner, in *Transportation and Traffic Theory in the 21th Century* edited by M. A. P. Taylor (Elsevier Science, Amsterdam, 2002) p. 417-439.
83. F. L. Hall, V. F. Hurdle, and J. H. Banks, Transportation Research Record **1365**, 12 - 18 (1992).
84. F. L. Hall and K. Agyemang-Duah, Transportation Research Record **1320**, 91 - 98 (1991).
85. B. Persaud, S. Yagar, and R. Brownlee, Transportation Research Record **1634**, 64 (1998).
86. B. S. Kerner, and H. Rehborn, Phys. Rev. Lett. **79**, 4030 (1997).
87. B. S. Kerner (in preparation).
88. B. S. Kerner, German patent DE 199 44 075 C2.
89. B. S. Kerner, in *2001 IEEE Intelligent Transportation Systems Proceedings* (IEEE, Oakland, USA, 2001), p. 88-93.
90. B. S. Kerner, H. Rehborn, and H. Kirschfink, German patent DE 196 47 127 C2; US-patent US 5861820 (1998).
91. B. S. Kerner, and H. Rehborn, German patent publication DE 198 35 979 A1.
92. B. S. Kerner, M. Aleksic, and U. Denneker, German patent DE 199 44 077 C2.
93. B. S. Kerner, H. Rehborn, M. Aleksic, and A. Haug, Traffic Eng. Control **42**, 282-287 (2001).
94. B. S. Kerner, H. Rehborn, M. Aleksic, A. Haug, and R. Lange, Traffic Eng. Control **42**, 345-350 (2001).

# Spacing-Oriented Analytical Approach to a Middle Traffic Flow CA Model Between FI-Type and NS-Type

B.-H. Wang<sup>1,2</sup>, D. Mao<sup>1,2</sup>, L. Wang<sup>3</sup>, and P.M. Hui<sup>4</sup>

<sup>1</sup> Department of Modern Physics and Nonlinear Science Center,  
University of Science and Technology of China, Hefei, Anhui 230026, P.R. China

<sup>2</sup> CCAST (World Laboratory), PO Box 8730, Beijing, 100080, P.R. China

<sup>3</sup> Physics Department and Center for Nonlinear Studies,  
Hong Kong Baptist University, Kowloon, Hong Kong

<sup>4</sup> Department of Physics, The Chinese University of Hong Kong, New Territories,  
Hong Kong

**Abstract.** We propose and study a one-dimensional traffic flow cellular automaton (CA) model of high speed vehicles with rapid acceleration for all cars as in the Fukui-Ishibashi (FI) model and with stochastic delay applying only to the cars following the trail of the ahead car. The main difference comparing to the Nagel-Schreckenberg (NS) model is that a car with spacing ahead longer than the car velocity limit  $M$  is not delayed in the new model. By using a spacing-oriented mean field theory, we derive a set of equations describing the fundamental diagram, which gives the average speed as a function of the car density, for any high velocity limit  $M$ . Our theoretical results are in excellent agreement with numerical simulation.

## 1 Introduction

Research on the traffic problems began in the beginning of the 20th century. Since the proposal of the first fluid-dynamical model of vehicular traffic [1], many other models have been developed and studied. In the 1970's and 1980's, the application of Cellular Automaton (CA) models [2-7] broadened the study of vehicular traffic. Compared with the fluid-dynamical models, the CA models for numerical simulations can be implemented easily.

The simplest CA model for vehicular traffic is the CA rule 184 in Wolfram's notation [3]. As the extension of this simple model, Nagel and Schreckenberg introduced a one-dimensional traffic flow CA model (NS model) [8,9] in which the possibilities of acceleration and random delay for cars were considered, and the car speed can be greater than 1. Let  $N$  be the total number of cars on a one-dimensional road of length  $L$ . The density of cars is  $\rho = N/L$ . Every car is numbered from left to right,  $n = 1, 2, 3, \dots, N$ . Let  $x_n$  and  $v_n$  denote the position and speed of the  $n$ -th vehicle ( $v_n$  may be  $0, 1, 2, \dots, v_{\max} = M$ ), respectively. The length of the spacing between the  $n$ -th and  $(n + 1)$ -th car is  $C_n = x_{n+1} - x_n - 1$ . In each time step  $t \rightarrow t + 1$ , the arrangement of the  $N$  vehicles is updated in parallel according to the following rules:

- **Step 1: Gradually Acceleration.**  $v_n \rightarrow \min(v_n + 1, v_{\max}, C_n)$ .

- *Step 2: Random Delay for all cars.*  $v_n \rightarrow \max(v_n - 1, 0)$ , with probability  $f$ .

Here, *Step 1* reflects the general tendency of drivers to drive as fast as possible while avoiding collision between vehicles. *Step 2* reflects different behavioral patterns of the individual drivers' reactions to the environment. The stochastic delay is crucially important for the formation of traffic jams by self-organization.

To simplify NS model, Fukui and Ishibashi introduced a new traffic flow CA model (FI model) [10,11] in which the rules of acceleration and randomization are changed as follows:

- *Step 1: Rapid Acceleration.*  $v_n \rightarrow \min(v_{\max}, C_n)$ .
- *Step 2: Random Delay for the cars with maximal velocity.* If  $v_n = v_{\max}$ ,  $v_n \rightarrow v_{\max} - 1$ , with probability  $f$ .

Apparently, when  $v_{\max} = 1$ , the FI model is the same as the NS model. Concerning traffic flow models, we are interested in the fundamental diagram, which gives the vehicle flux as a function of density  $\rho$  [12]. It will be of practical importance if we can find an analytical mean field theory for the fundamental diagram. Thus, attempts have been made to use statistical mechanical approaches to get the fundamental diagram of the NS model [13,14]. However, the problem is difficult since the interaction between cars leads to complex correlation in both space and time [15,16]. On the other hand, for the FI model, we can derive an approximate mean field theory of the fundamental diagram by a car-oriented statistical mechanical approach [17-21] and an exact analytical mean field equation for the fundamental diagram by a spacing-oriented dynamical approach [22,23]. In order to investigate the difficulty in obtaining analytical solution to the NS model, we also consider the gradual acceleration in the FI model and prove that the fundamental diagram is given by that of the FI model [24,25]. Furthermore, we study an intermediate one-dimensional traffic flow CA model situated between FI model and NS model, in which only the car following the trail of the car ahead may be delayed [26]. It includes the factors which are closer to the real vehicular traffic. In this paper, we shall derive the general mean field equations describing the fundamental diagram for any high velocity limit  $M$  using a spacing-oriented mean field theory. Our theoretical results are in excellent agreement with numerical simulation.

The definition of the model and the evolution equation for the inter-car spacings are given in Section 2. In Section 3, some observations are given for the dynamical evolution towards the steady state. In Section 4, we derive a general relation for spacing distribution of asymptotic steady state. In Section 5, we present the analytical solution to the fundamental diagram for the case of  $M = 3$ . In Section 6, we show the excellent agreement between numerical simulation and our spacing-oriented mean field theoretical results.

## 2 Definition of the Model

The acceleration rule of the new model is the same as that of the FI model. i.e., if the inter-car spacing allows, the car speed increases as much as possible up to

a upper bound of velocity  $v_{\max}$ . Random delay is introduced almost in the same way as that of NS model. We define the tracing car as a car whose spacing ahead has the length of no longer than  $M$ , and which will follow the trail of the car ahead if the front car stops. Suppose the  $n$ -th car has the velocity  $v_n(t) = C_n(t)$  at time step  $t$ . If the  $(n + 1)$ -th car stops, the  $n$ -th car will follow the trail of the  $(n + 1)$ -th car in the present model. In this sense, the  $n$ -th car is called a tracing car. Random delay is only applied to the tracing cars. Therefore, the arrangement of cars is updated in parallel according to the following rules:

- **Step 1: Rapid Acceleration.**  $v_n \rightarrow \min(v_{\max}, C_n)$ .
- **Step 2: Random Delay for tracing cars.** If  $0 < C_n \leq v_{\max}$ ,  $v_n \rightarrow v_n - 1$ , with probability  $f$ ; otherwise, it will not be delayed.

### 3 Dynamical Evolution of Inter-Car Spacings

The evolutionary rules of the present model can be expressed in the following form:

$$v_n(t) = F_M(C_n(t)) \quad (1)$$

where

$$F_M(C_n(t)) = \begin{cases} 0 & (C_n(t) = 0) \\ C_n(t) & \text{with probability } (1 - f) \\ C_n(t) - 1 & \text{with probability } f \\ M & (C_n(t) > M) \end{cases} \quad (2)$$

The evolution of the inter-car spacings follows the equation:

$$C_n(t + 1) = C_n(t) + v_{n+1}(t) - v_n(t) \quad (3)$$

From Eqs.(1)-(3), we can make several observations concerning the dynamical evolution of the inter-car spacings.

**Observation I:** If an inter-car spacing is not longer than  $(M + 1)$ , it will not be longer than  $(M + 1)$  as time evolves.

**Proof:** Given  $C_n(t) \leq M + 1$ , it follows that  $C_n(t) - F_M[C_n(t)] \leq 1$ . From  $F_M[C_{n+1}(t)] \leq M$ , we have  $C_n(t + 1) = C_n(t) - F_M[C_n(t)] + F_M[C_{n+1}(t)] \leq M + 1$ .

**Observation II:** Inter-car spacings which are longer than or equal to  $(M + 1)$  will not become longer as time evolves.

**Proof:** Given  $C_n(t) \geq M + 1$ , it follows that  $F_M[C_n(t)] = M$ . From  $F_M[C_{n+1}(t)] \leq M$ , we have  $C_n(t + 1) = C_n(t) - F_M[C_n(t)] + F_M[C_{n+1}(t)] \leq C_n(t)$ .

For convenience, we define the notion of long spacings and short spacings as:

Long-spacing  $C_n(t)$  if  $C_n(t) > (M + 1)$ .

Short-spacing  $C_n(t)$  if  $C_n(t) < (M + 1)$ .

We then define the excessive length of a long-spacing as:

$$L_n(t) = \max(C_n(t) - (M + 1), 0),$$

and the deficient length of a short-spacing as:

$$S_n(t) = \mathbf{max}((M + 1) - C_n(t), 0).$$

The sum of the excessive lengths of all long-spacings is given by

$$L(t) = \sum_n L_n(t),$$

and the sum of the deficient lengths of all short-spacings is given by

$$S(t) = \sum_n S_n(t).$$

From the above definitions, it follows that:

$$\begin{aligned} L(t) - S(t) &= \sum_n [C_n(t) - (M + 1)] \\ &= L - N - N(M + 1) \\ &= L - (M + 2)N \\ &= \text{constant} \end{aligned} \tag{4}$$

Together with **Observations I** and **II**, we have

$$L_n(t + 1) \leq L_n(t) \tag{5}$$

and hence

$$L(t + 1) \leq L(t) \tag{6}$$

From Eqs.(4) and (6), we have

$$S(t + 1) \leq S(t) \tag{7}$$

Therefore  $L$  will never increase. In addition, if a certain  $L_n$  decreases, then  $L$  and  $S$  have to decrease so as to satisfy Eq.(4).

**Observation III:** Long-spacings and short-spacings can not co-exist in the asymptotic steady state.

**Proof:** Suppose in one time step, long and short spacings co-exist and the  $n$ -th car has a long spacing ahead. So  $v_n = v_{\max} = M$ . If  $v_{n+1} = (M - 1)$ ,  $C_n$  will decrease by 1. This happens in only two cases:

**Case 1:** When  $C_{n+1} = M$ , the  $(n + 1)$ -th car moves with the velocity  $v_{n+1} = (M - 1)$  due to delay. This case occurs with the probability  $fP_M$ .

**Case 2:** When  $C_{n+1} = (M - 1)$ , the  $(n + 1)$ -th car moves with the velocity  $M - 1$  without delay. This case occurs with probability  $(1 - f)P_{M-1}$ .

The length of inter-car spacing decreases by 1 with probability  $F_1$  given by

$$F_1 = (1 - f)P_{M-1} + fP_M.$$

Similarly, the probability that the length of the spacing decreases by 2 is

$$F_2 = (1 - f)P_{M-2} + fP_{M-1};$$

and decreases by 3 is

$$F_3 = (1 - f)P_{M-3} + fP_{M-2},$$

and so on. Thus, the probability that the spacing decreases by  $(M - 1)$  is

$$F_{M-1} = (1 - f)P_1 + fP_2;$$

and decreases by  $M$  is

$$F_M = P_0 + fP_1.$$

In one time step, a long spacing will be shortened on the average by:

$$\begin{aligned} \sum_{i=1}^{M-1} iF_i &= \sum_{i=1}^{M-1} i(1 - f)P_{M-i} + \sum_{i=1}^{M-1} ifP_{M-i+1} + MP_0 + MfP_1 \\ &= MP_0 + \sum_{i=1}^{M-1} [i(1 - f)P_{M-i} + (i + 1)fP_{M-i}] + fP_M \\ &= MP_0 + \sum_{i=1}^{M-1} (i + f)P_{M-i} + fP_M \end{aligned} \quad (8)$$

Therefore, if and only if  $P_0 = P_1 = P_2 = \dots = P_{M-1} = P_M = 0$ , i.e.  $S = 0$ , the shortened length will be zero; otherwise, it will be positive.

Hence, the long spacings as well as short spacings, will be shortened until the system reaches its steady state. In the steady state,  $L$  and  $S$  no longer evolve, and either  $L$  or  $S$  vanishes, and Observation III follows.

## 4 Spacing Distribution in the Asymptotic Steady State

### 4.1 Low Density Case

For the low density case ( $\rho < 1/(M + 2)$ ), it is apparent that in the asymptotic steady state, we have  $L > 0$  and  $S = 0$ . Thus,

$$C_n(t) \geq M + 1, (\forall n) \quad (9)$$

Stochastic delays will no longer be effective in this case. All the cars will keep going with the maximum speed  $M$ . So the average car speed is:

$$\langle V(t \rightarrow \infty) \rangle = M \quad (10)$$

### 4.2 High Density Case

For the high density case ( $\rho > 1/(M + 2)$ ), in the asymptotic steady state, we have  $S > 0$  and  $L = 0$ . We can also obtain

$$C_n(t) \leq M + 1, (\forall n) \quad (11)$$

The average car speed can be expressed as:

$$\begin{aligned} \langle V(t \rightarrow \infty) \rangle &= \sum_{i=1}^M P_i [i(1-f) + (i-1)f] + MP_{M+1} \\ &= \sum_{i=1}^M P_i (i-f) + MP_{M+1} \end{aligned} \quad (12)$$

In order to get the average speed of the cars, we need to know the possibility of having different inter-car spacings  $P_i$  ( $i = 1, 2, \dots, M+1$ ).

To obtain  $P_i(t)$ , we introduce the transition possibility  $W_{i \rightarrow j}$  to describe the probability that at one time step the length of an inter-car spacing is changed from  $i$  to  $j$ . The transition probabilities of spacings for the general case  $v_{max} = M$  can be written as follows:

$$\begin{aligned} W_{0 \rightarrow j} &= P_0[(1-f)P_j + fP_{j+1}] (j = 1, \dots, M-1) \\ &\dots\dots \\ W_{0 \rightarrow M} &= P_0[(1-f)P_M + P_{M+1}] \\ W_{0 \rightarrow M+1} &= 0 \\ &\dots\dots \\ W_{i \rightarrow 0} &= P_i[(1-f)P_0 + f(1-f)P_1] \\ W_{i \rightarrow 1} &= P_i[fP_0 + (f^2 + (1-f)^2)P_1 + f(1-f)P_2] (i \neq 1) \\ &\dots\dots \\ W_{i \rightarrow j} &= P_i f(1-f)P_{j-1} + [f^2 + (1-f)^2]P_j + f(1-f)P_{j+1} \\ &(j = 2, \dots, M-1, (i \neq j)) \\ &\dots\dots \\ W_{i \rightarrow M} &= P_i[f(1-f)P_{M-1} + (f^2 + (1-f)^2)P_M + (1-f)P_{M+1}] \\ &\hspace{15em} (i \neq M) \\ W_{i \rightarrow M+1} &= P_i[f(1-f)P_M + fP_{M+1}] \\ &(i = 1, \dots, M) \\ &\dots\dots \end{aligned}$$

$$\begin{aligned} W_{M+1 \rightarrow 0} &= 0 \\ W_{M+1 \rightarrow 1} &= P_{M+1}[P_0 + fP_1] \\ &\dots\dots \\ W_{M+1 \rightarrow j} &= P_{M+1}[(1-f)P_{j-1} + fP_j] (j = 2, \dots, M) \end{aligned} \quad (13)$$

In the asymptotic steady state, all the  $P_i$  will no longer change. The transition probabilities satisfy the following detail equilibrium conditions:

$$\sum_{i \neq m} W_{i \rightarrow m} = \sum_{i \neq m} W_{m \rightarrow i} (\forall m) \quad (14)$$

Substituting the expressions of all  $W_{i \rightarrow j}$  into the above conditions, we get the following set of equations:

$$\begin{aligned} [0] : & P_0[f(P_2 + P_3 + \cdots + P_M) + P_{M+1}] - P_1[f(1-f)(P_1 + P_2 \\ & + P_3 + \cdots + P_M)] = 0 \\ [1] : & P_0[f(P_2 + P_3 + \cdots + P_M) + P_{M+1}] - P_1[2f(1-f)(P_1 + P_2 \\ & + P_3 + \cdots + P_M) - f(1-f)P_2 + (1-f)P_{M+1}] \\ & + P_2[fP_0 + f(1-f)(P_2 + P_3 + \cdots + P_M)] = 0 \\ & \dots\dots \\ [j] : & P_{j-1}[f(1-f)(P_1 + P_2 + P_3 + \cdots + P_M - P_j) + (1-f)P_{M+1}] \\ & - P_j[fP_0 + 2f(1-f)(P_1 + P_2 + P_3 + \cdots + P_M) \\ & + f(1-f)(P_{j-1} + P_{j+1}) + (1-f)P_{M+1}] + P_{j+1}[fP_0 \\ & + f(1-f)(P_1 + P_2 + P_3 + \cdots + P_M - P_j)] = 0 \\ & \dots\dots \\ [M] : & P_{M-1}[f(1-f)(P_1 + P_2 + P_3 + \cdots + P_{M-1}) + (1-f)P_{M+1}] \\ & - P_M[fP_0 + 2f(1-f)(P_1 + P_2 + P_3 + \cdots + P_M) \\ & - f(1-f)P_{M-1}] + P_{M+1}[P_0 + (1-f)(P_1 + P_2 \\ & + P_3 + \cdots + P_{M-1})] = 0 \\ [M+1] : & P_M f(1-f)(P_1 + P_2 + P_3 + \cdots + P_M) - P_{M+1}[P_0 \\ & + (1-f)(P_1 + P_2 + P_3 + \cdots + P_{M-1})] = 0 \end{aligned} \quad (15)$$

These equations are not linearly independent of each other, but satisfy the following conditions, given here symbolically as

$$[0] - [1] - [2] - \cdots - [M] - [M+1] = 0 \quad (16)$$

$$[0] + [2] + 2[3] + 3[4] + \cdots + M[M+1] = 0 \quad (17)$$

We also have the normalization conditions:

$$\sum_{i=0}^{M+1} P_i = 1 \quad (18)$$

$$\sum_{i=1}^{M+1} iP_i = \bar{C} = 1/\rho - 1 \quad (19)$$

Hence we can solve these  $(M+2)$  equations to get all the probabilities  $P_i (i = 0, 1, \dots, M+1)$  for any general case of  $v_{max} = M$  and obtain the average speed analytically.

## 5 The Fundamental Diagram Solution

In Ref.[26], we have derived the asymptotic average speed of traffic flow for  $v_{\max} = M = 1$  as

$$\langle V(t \rightarrow \infty) \rangle = \frac{1}{2} \left( -1 + \frac{1}{\rho} + \frac{-1 + \sqrt{1 + \frac{(2f-1)^2(\rho-1)(3\rho-1)}{\rho^2}}}{2f-1} \right).$$

and for  $v_{\max} = M = 2$  as

$$\langle V(t \rightarrow \infty) \rangle = \sum_{i=1}^2 P_i(i-f) + 2P_3$$

where  $P_1, P_2$  and  $P_3$  can be solved from the following four equations:

$$fP_0P_2 + P_0P_3 - f(1-f)P_1^2 - f(1-f)P_1P_2 = 0$$

$$P_0P_3 - f(1-f)P_1P_2 + (1-f)P_1P_3 - f(1-f)P_2^2 = 0$$

$$\sum_i P_i = P_0 + P_1 + P_2 + P_3 = 1$$

$$\sum_i iP_i = P_1 + 2P_2 + 3P_3 = \bar{C} = 1/\rho - 1,$$

To illustrate the method established in Section 4, we consider the special case of  $M = 3$ . For  $M = 3$ , the low density regime corresponds to  $\rho \leq 1/5$ . The average speed is

$$\langle V(t \rightarrow \infty) \rangle = M = 3.$$

In the high density regime ( $\rho > 1/5$ ), since  $P_i = 0$  ( $\forall i > 4$ ), we only need to consider  $P_0, P_1, P_2, P_3$  and  $P_4$ . All the non-zero transition probabilities for

different inter-car spacings  $W_{i \rightarrow j}$  are given by:

$$W_{0 \rightarrow 1} = P_0[(1-f)P_1 + fP_2]$$

$$W_{0 \rightarrow 2} = P_0[(1-f)P_2 + fP_3]$$

$$W_{0 \rightarrow 3} = P_0[(1-f)P_3 + P_4]$$

$$W_{0 \rightarrow 4} = 0$$

$$W_{1 \rightarrow 0} = P_1[(1-f)P_0 + f(1-f)P_1]$$

$$W_{1 \rightarrow 2} = P_1[f(1-f)P_1 + (f^2 + (1-f)^2)P_2 + f(1-f)P_3]$$

$$W_{1 \rightarrow 3} = P_1[f(1-f)P_2 + (f^2 + (1-f)^2)P_3 + (1-f)P_4]$$

$$W_{1 \rightarrow 4} = P_1[f(1-f)P_3 + fP_4]$$

$$W_{2 \rightarrow 0} = P_2[(1-f)P_0 + f(1-f)P_1]$$

$$W_{2 \rightarrow 1} = P_2[fP_0 + (f^2 + (1-f)^2)P_1 + f(1-f)P_2]$$

$$W_{2 \rightarrow 3} = P_2[f(1-f)P_2 + (f^2 + (1-f)^2)P_3 + (1-f)P_4]$$

$$W_{2 \rightarrow 4} = P_2[f(1-f)P_3 + fP_4]$$

$$W_{3 \rightarrow 0} = P_3[(1-f)P_0 + f(1-f)P_1]$$

$$W_{3 \rightarrow 1} = P_3[fP_0 + (f^2 + (1-f)^2)P_1 + f(1-f)P_2]$$

$$W_{3 \rightarrow 2} = P_3[f(1-f)P_1 + (f^2 + (1-f)^2)P_2 + f(1-f)P_3]$$

$$W_{3 \rightarrow 4} = P_3[f(1-f)P_3 + fP_4]$$

$$W_{4 \rightarrow 0} = 0$$

$$W_{4 \rightarrow 1} = P_4[P_0 + fP_1]$$

$$W_{4 \rightarrow 2} = P_4[(1-f)P_1 + fP_2]$$

$$W_{4 \rightarrow 3} = P_4[(1-f)P_2 + fP_3]$$

Substituting the expressions of  $W_{i \rightarrow j}$  into the equilibrium conditions (14), we get a set of five equations, in which only three of them are linearly independent of each other. Together with the normalization conditions (18)(19), we have:

$$\begin{aligned} fP_0(P_2 + P_3) + P_0P_4 - f(1-f)P_1(P_1 + P_2 + P_3) &= 0, \\ 2fP_0P_2 + fP_0P_3 + P_0P_4 - f(1-f)(2P_1 - P_2)(P_1 + P_2 + P_3) \\ - (1-f)P_1P_4 &= 0, \\ P_0P_4 - f(1-f)P_3(P_1 + P_2 + P_3) + (1-f)(P_1 + P_2)P_4 &= 0, \\ P_0 + P_1 + P_2 + P_3 + P_4 &= 1, \\ P_1 + 2P_2 + 3P_3 + 4P_4 &= 1/\rho - 1. \end{aligned} \quad (20)$$

Numerically solving these equations, we can obtain  $P_0, P_1, P_2, P_3, P_4$  as the functions of the vehicle density  $\rho$ . Therefore, we have

$$\langle V(t \rightarrow \infty) \rangle = P_1 + 2P_2 + 3(P_3 + P_4) - f(P_1 + P_2 + P_3), \quad (21)$$

which gives the fundamental diagram curve of  $M = 3$  model.

## 6 Comparison Between Analytical Results and Numerical Simulation

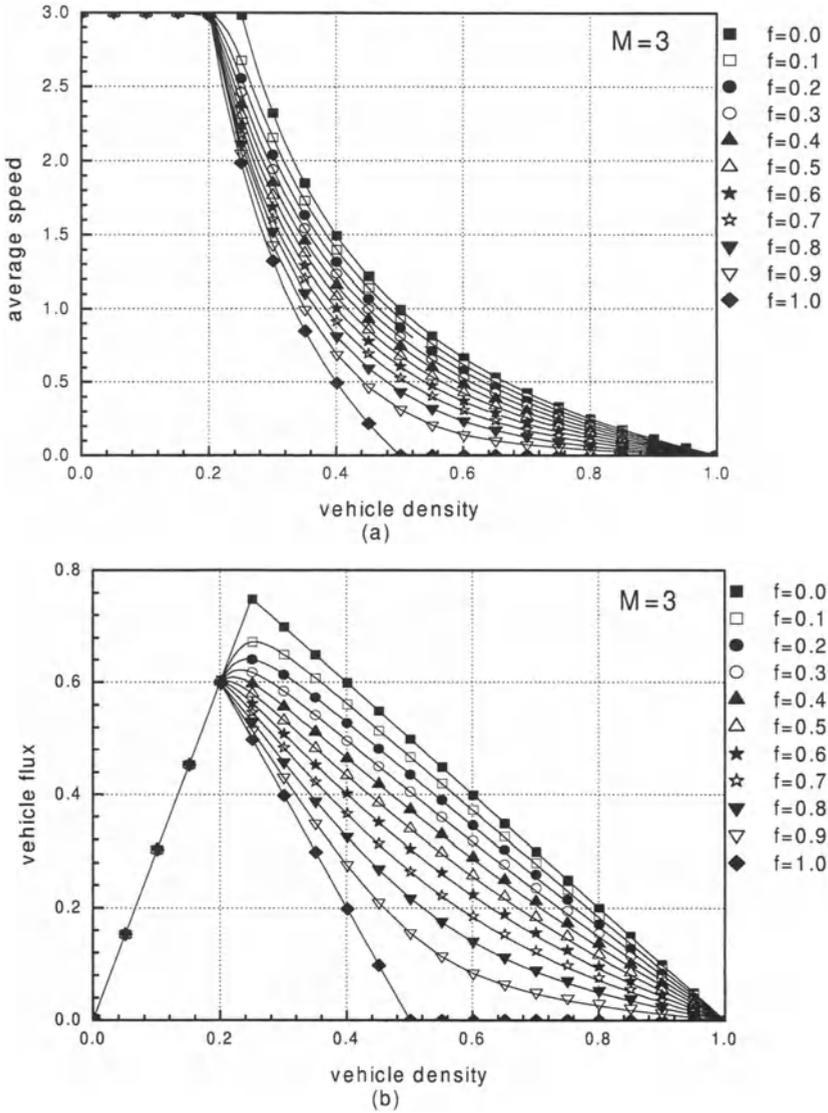
We carried out numerical simulations on a one-dimensional lattice with 1000 cells and periodic boundary conditions. The first 1000 time steps were excluded from the averaging procedure in order to get to the asymptotic steady state. Averages of the vehicle speed and flux were taken over the next 5000 time steps. Figure 1 shows the comparison between numerical simulation results and our analytical results for the average speed  $\langle V \rangle$  (a) and flux  $\rho \langle V \rangle$  (b) as the function of the density  $\rho$  for  $M = 3$  and for different delay probability  $f$ . Excellent agreements between simulation and our theory are found.

The fundamental diagram shows that when the vehicle density is low ( $\rho \leq 1/(M + 2)$ ), all the inter-car spacings will not be shorter than  $(M + 1)$  and the vehicles will not be delayed; the average speed is thus equal to the maximum velocity  $M$ . When the vehicle density is high ( $\rho \geq 1/(M + 2)$ ), most of the vehicles drive with a lower velocity due to traffic jam. In Figure 2, we present a space-time evolution picture of traffic flow for  $M = 3$ . It shows a transition in the present model from free moving phase in low density to a jamming phase in high density as a result of the self organization of the vehicles.

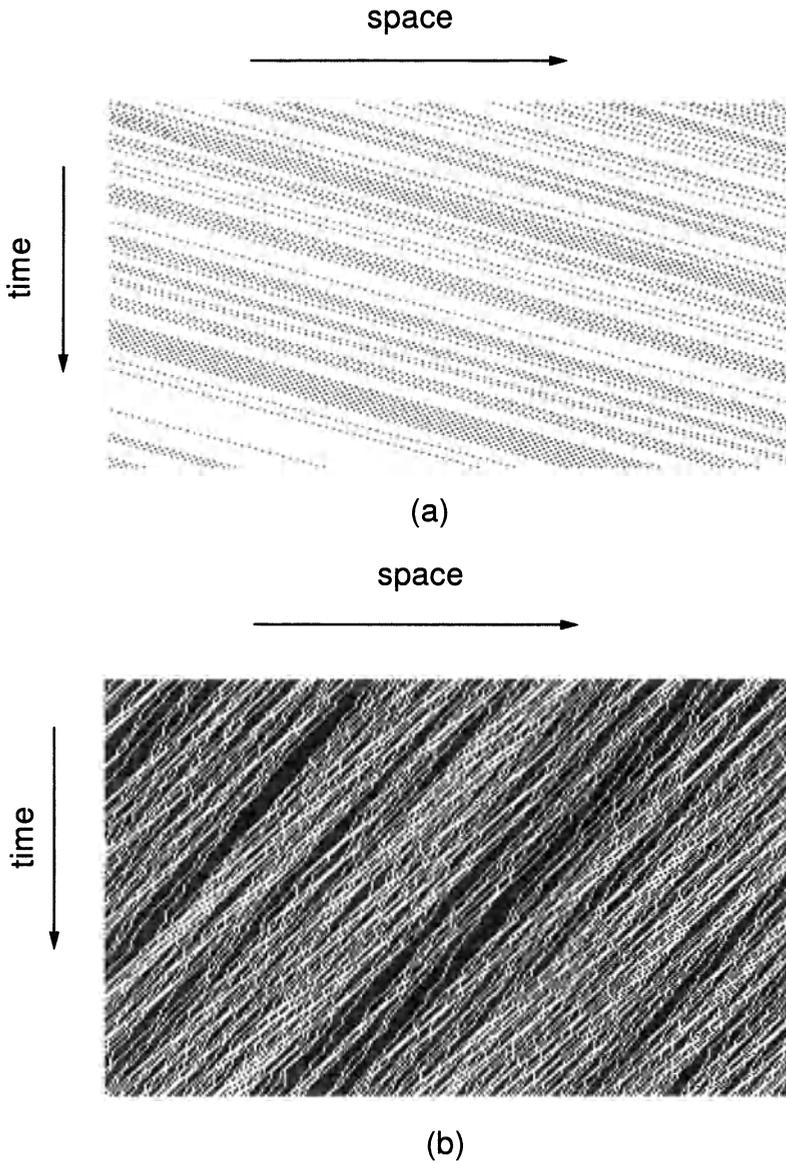
In summary, we introduced a new model with stochastic delays applied only to vehicles which are following the trail of the vehicle ahead. We studied the time evolution of the inter-car spacings and obtained the fundamental diagram analytically by a spacing-oriented mean field theory. The results show an exact agreement with numerical simulations.

## Acknowledgements

We acknowledge the support from a Special Fund for National Basic Research Major Project in China (973 Project), the National Basic Research Climbing-Up Project "Nonlinear Science" and the Natural Science Foundation of China under Key Project Grant No.19932020 and General Project Grant No.19974039 and No.59876039). PMH acknowledges the support from the research Grants Council of the Hong Kong SAR Government through the grant CUHK 4191/97P and the grant CUHK 4241/01P.



**Fig. 1.** The fundamental diagram of the average speed (a) and the average flux (b) as the function of the car density for  $M = 3$ . The continuous curves represent the results of the mean field theory. The points with different symbols represent the results of numerical simulations. The curves from the top down correspond to different  $f$  ranging from  $f = 0$  to  $f = 1$  in steps of 0.1.



**Fig. 2.** Space-time evolution of traffic flow for the present model with  $M = 3$  and  $f = 0.3$ . The evolution of the first 200 time steps in 500 cells of a total of  $L = 5000$  sites are shown. A black point represents a car. (a) Low density regime:  $N = 500$ ,  $\rho = N/L = 0.1 < \rho_c = 1/5$ ; (b) High density regime:  $N = 3000$ ,  $\rho = N/L = 0.6 > \rho_c = 1/5$ .

## References

1. M.J. Lighthill and G.B. Whitham, Proc. Roy. Soc. of London A **229**, 317 (1955).
2. E.F. Codd, *Cellular Automata*, Academic Press, New York, 1968.
3. S. Wolfram, *Theory and Application of Cellular Automata*, World Scientific, Singapore, 1986.
4. D. Stauffer, J.Phys. A **24**, 909 (1991).
5. K. Preston and M. Duff, *Modern Cellular Automata: Theory and Applications*, Plenum, New York, 1984.
6. B. Chopard and M. Droz, *Cellular Automata Modelling of Physical Systems*, Cambridge University Press, Cambridge, 1998.
7. J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*, Cambridge University Press, Cambridge, 1999.
8. K. Nagel and M. Schreckenberg, J. Phys. I (France) **2**, 2221 (1992).
9. M. Schreckenberg, A. Schadschneider, K. Nagel, and N. Ito, Phys. Rev. E **51**, 2939 (1995).
10. Y. Ishibashi and M. Fukui, J.Phys. Soc. Japan **63**, 2882 (1994) 2882.
11. M. Fukui and Y. Ishibashi, J. Phys. Soc. Japan **65**, 1868 (1996).
12. N.H. Gartner and N.H.M. Wilson, (ed.) *Transportation and Traffic Theory*, Elsevier, New York, 1987.
13. B.H. Wang, P.M. Hui, and G.Q. Gu, Chin. Phys. Lett. **14**, 202 (1997).
14. B.H. Wang and P.M. Hui, J. Phys. Soc. Japan **66**, 1238 (1997).
15. A. Schadschneider and M. Schreckenberg, J. Phys. A: Math. Gen. **30**,L69 (1997).
16. D.E. Wolf, M. Schreckenberg, and A. Bachem, (eds.) *Traffic and Granular Flow*, World Scientific, Singapore, 1996.
17. B.H. Wang, P.M. Hui, and G.Q. Gu, Chin.Phys.Lett. **14**, 3, 202 (1997).
18. B.H. Wang and P.M. Hui, J.Phys.Soc. Japan **66**, 5, 1238 (1997).
19. B.H. Wang, Y.R. Kwong, and P.M. Hui, Physica A **254**, 1-2, 122 (1998).
20. B.H. Wang, Y.R. Kwong, and P.M. Hui, Phys. Rev. E **57**, 3, 2568 (1998).
21. B.H. Wang, Y.R. Kwong, P.M. Hui, and B.Hu, Physical Review E **60**, 1, 149 (1999).
22. B.H. Wang, L. Wang, and P.M. Hui, Journal of the Physical Society of Japan **66**, 3683 (1997).
23. B.H. Wang, L. Wang, P.M. Hui, and B. Hu, Physical Review E **58**, 2876 (1998).
24. B.H. Wang, L. Wang, P.M. Hui, and B. Hu, Physica B **279**, 1-3, 237 (2000).
25. B.H. Wang, L. Wang, P.M. Hui, and B. Hu, International Journal of Numerical Simulation and Nonlinear Science, **1**, 4, 255 (2000).
26. L. Wang, B.H. Wang, and B. Hu, Physical Review E **63**, 5, No.056117 (2001).

# Continuum Traffic Equations from Microscopic Car-Following Models

H.-W. Lee<sup>1</sup>, H.K. Lee<sup>2</sup>, H.Y. Lee<sup>3</sup>, and D. Kim<sup>2</sup>

<sup>1</sup> School of Physics, Korea Institute for Advanced Study,  
207-43 Cheongryangri-dong, Dongdaemun-gu, Seoul 130-012, Korea

<sup>2</sup> School of Physics, Seoul National University, Seoul 151-742, Korea

<sup>3</sup> Chemical Engineering Department, University of California,  
Los Angeles, CA 90095-1592, U.S.A.

**Abstract.** We present a method to derive macroscopic traffic models through a direct coarse graining procedure from microscopic car-following models. The method is similar to that developed by Helbing and coworkers to derive the gas-kinetic theory (GKT), but our resulting macroscopic equations are local and contain no empirical parameters other than those inherited from the microscopic models. Fluctuation effects are not included. When we work in the so-called linear regime, the derived momentum equation for the optimal velocity model consists of the relaxation term, the anticipation term, and the diffusion term, as in the previous local continuum model of Kerner and Konhäuser (KK). However, the origin of the last two terms are different and their density dependent coefficients are fixed from the microscopic behavior of individual vehicles. Properties of the resulting macroscopic models are compared with those of the optimal velocity model through numerical simulations, and reasonable agreement is found although there are deviations in the quantitative level. The phase diagram in the presence of an on-ramp also shows qualitatively similar features as in the case of the GKT and KK models.

## 1 Introduction

Numerous traffic models have been investigated [1–3] and considerable progress has been achieved [4–9] toward an understanding of various traffic phenomena observed empirically [10–14]. Depending on the mathematical formulation, traffic models may be categorized into one of the following types: car-following models, particle-hopping models, coupled-map lattice models, gas-kinetic models, and fluid-dynamic models. The first three types use a microscopic approach while the last type uses a macroscopic one. The approach used in the gas-kinetic models is intermediate and may be called mesoscopic.

Recently it was suggested [15,16] that different types of traffic models may belong to the same “universality” class in the sense that they share qualitatively similar properties. More recently, a nonlocal fluid-dynamic model was derived from a gas-kinetic model [17]. These reports motivate further studies on mutual relationship between different types of traffic models.

In this paper, we address the relationship between microscopic car-following models and macroscopic fluid-dynamic models. Specifically we use a coarse-graining procedure (Sect. 2) to derive a macroscopic model (Sect. 3) from the

microscopic optimal velocity model, a particular case of the car-following-type model [18]. In Sect. 4, the derivation is extended to general car-following-type models. In Sect. 5, the macroscopic model derived from the microscopic optimal velocity model is examined numerically in comparison with the optimal velocity model. Section 6 concludes the paper. Technical details can be found in [19].

## 2 General Formulation

We first introduce microscopic density field  $\hat{\rho}(x, t)$  and flux field  $\hat{q}(x, t)$ ,

$$\hat{\rho}(x, t) \equiv \sum_n \delta(y_n(t) - x) , \quad \hat{q}(x, t) \equiv \sum_n \dot{y}_n(t) \delta(y_n(t) - x) , \quad (1)$$

where  $y_n(t)$  is the coordinate of the  $n$ th vehicle at time  $t$  with  $y_1 < y_2 < \dots < y_{n-1} < y_n < y_{n+1} < \dots$ . A natural way to obtain macroscopic description is to coarse-grain these fields. We introduce a coarse graining envelope function  $\phi(x, t)$  which is non-negative valued, peaked at  $(x, t) = (0, 0)$ , and normalized as  $\int dx dt \phi(x, t) = 1$ . The coarse grained density  $\rho(x, t)$  and flux  $q(x, t)$  can be defined as

$$\begin{aligned} \rho(x, t) &\equiv \int dx' dt' \phi(x - x', t - t') \hat{\rho}(x', t') , \\ q(x, t) &\equiv \int dx' dt' \phi(x - x', t - t') \hat{q}(x', t') . \end{aligned} \quad (2)$$

Next we derive equations that govern the time evolution of  $\rho(x, t)$  and  $q(x, t)$ . For the evolution of  $\rho(x, t)$ , one finds

$$\frac{\partial}{\partial t} \rho(x, t) + \frac{\partial}{\partial x} q(x, t) = 0 , \quad (3)$$

which describes the local conservation of vehicles. One also obtains the dynamic equation for  $q(x, t)$  after some algebra,

$$\frac{\partial}{\partial t} q(x, t) = \rho(x, t) \langle \ddot{y}_n(t') \rangle_{(x,t)} - \frac{\partial}{\partial x} \left[ \rho(x, t) \langle \dot{y}_n^2(t') \rangle_{(x,t)} \right] , \quad (4)$$

where the bracketed average of a quantity  $O_n(x', t')$  is defined as follows;

$$\langle O_n(x', t') \rangle_{(x,t)} \equiv \frac{1}{\rho(x, t)} \int dx' dt' \phi(x - x', t - t') \sum_n O_n(x', t') \delta(y_n(t') - x') . \quad (5)$$

Note that  $x'$ ,  $t'$ , and  $n$  inside the brackets are dummy variables, while the label  $(x, t)$  in the subscript of the bracket notation represents a spatiotemporal position where the average is evaluated. This label will be omitted in the rest of the paper when its omission does not cause confusion.

Here it is useful to introduce another macroscopic field  $v(x, t)$ ,

$$v(x, t) \equiv \langle \dot{y}_n(x', t') \rangle = q(x, t) / \rho(x, t) , \quad (6)$$

which represents some kind of macroscopic velocity, whose precise meaning depends on  $\phi(x, t)$ . Two particular coarse graining schemes are good for illustration: spatial coarse graining  $\phi(x, t) = \delta(t)\Theta(X/2 - |x|)/X$  and temporal coarse graining  $\phi(x, t) = \delta(x)\Theta(T/2 - |t|)/T$ , where  $\Theta(x)$  is the step function which is one for  $x > 0$  and zero for  $x < 0$ . For the spatial coarse graining,  $v(x, t)$  becomes  $v(x, t) = \sum'_n \dot{y}_n(t) / \sum'_n 1$ , where the primed summation runs over the vehicles in the range  $(x - X/2, x + X/2)$  at time  $t$ . The denominator is equal to the total number of vehicles within the range and thus  $v(x, t)$  represents the *arithmetic mean velocity*. For the temporal coarse graining, on the other hand, it can be verified that  $v(x, t) = \sum'_n 1 / \sum'_n [\dot{y}_n(t_n(x))]^{-1}$ , where the primed summation now runs over the vehicles that reach the point  $x$  within the time interval  $(t - T/2, t + T/2)$ , and  $t_n(x)$  represents the time at which the  $n$ th vehicle reaches the position  $x$ . Here  $\dot{y}_n(t) \geq 0$  is assumed. Thus  $v(x, t)$  represents the *harmonic mean velocity* measured at local detectors.

It is straightforward to rewrite (3) in terms of  $\rho$  and  $v$  instead of  $\rho$  and  $q$ . Also expressing (4) in terms of  $\rho$  and  $v$ , one obtains

$$\rho \left( \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \right) = \rho \langle \ddot{y}_n(t') \rangle - \frac{\partial}{\partial x} (\rho \theta), \quad (7)$$

where  $\theta(x, t) \equiv \langle \dot{y}_n^2(t') \rangle - v^2(x, t)$  measures the degree of microscopic velocity variation. The remaining job is to express the right-hand-side of (7) in terms of  $\rho$  and  $v$ . This procedure is illustrated for the optimal velocity model in Sect. 3 and for general car-following models in Sect. 4. In both sections, traffic states are assumed to be almost homogeneous, so that products of differentiated quantities such as  $\prod_{m=1}^M (\partial^{l_m} O_m / \partial x^{l_m})$  become progressively smaller as  $M$  increases, where  $l_m$  are integers and  $O_m$  are arbitrary functions of  $\rho$  and  $v$ . Then it is sufficient to retain terms with  $M = 0$  or 1 only, which simplifies the construction of a macroscopic description considerably. In this sense, terms with  $M = 0$  or 1 can be called *linearly relevant* terms, and terms with  $M \geq 2$  *linearly irrelevant* terms. Effects of the linearly irrelevant terms with  $M = 2$  are briefly discussed in [19].

### 3 Optimal Velocity Model

We first study the optimal velocity model [20]

$$\ddot{y}_n(t) = \lambda [V_{op}(\Delta y_n(t)) - \dot{y}_n(t)], \quad (8)$$

where the constant  $\lambda$  represents a driver's sensitivity and  $\Delta y_n \equiv y_{n+1} - y_n$  is the distance to the preceding vehicle  $n + 1$ .  $V_{op}(\Delta y)$  is the optimal velocity to which drivers want to adjust their speed.

The coarse graining of (8) leads to

$$\langle \ddot{y}_n \rangle = \lambda [\langle V_{op}(\Delta y_n) \rangle - v]. \quad (9)$$

The expansion of  $\langle V_{op}(\Delta y_n) \rangle$  with respect to  $\langle \Delta y_n \rangle$  gives

$$\langle V_{op}(\Delta y_n) \rangle = \sum_{m=0}^{\infty} \frac{V_{op}^{(m)}(\langle \Delta y_n \rangle)}{m!} \langle (\Delta y_n - \langle \Delta y_n \rangle)^m \rangle \equiv V_{op}(\langle \Delta y_n \rangle) + \sum_{m=2}^{\infty} I_m, \quad (10)$$

where  $I_m$  is the term that is proportional to  $\langle(\Delta y_n - \langle\Delta y_n\rangle)^m\rangle$ . Here  $I_1$  is absent since  $\langle(\Delta y_n - \langle\Delta y_n\rangle)\rangle = 0$ . The leading correction  $I_2$  compensates for the difference  $\langle V_{op}(\Delta y_n) \rangle - V_{op}(\langle\Delta y_n\rangle)$ , which is positive (negative) when  $V_{op}$  is a convex (concave) function. In [19], it is shown that all corrections  $I_m$  ( $m \geq 2$ ) and the second term on the right-hand side of (7) can be ignored in the linear regime. Therefore, the derivation of a macroscopic description in the linear regime is reduced to developing a proper approximation of  $\langle\Delta y_n\rangle$ .

### 3.1 Directed Influence

In a homogeneous state,  $\langle\Delta y_n\rangle_{(x,t)}$  equals  $\rho^{-1}(x, t)$ . However when the system is not in a homogeneous state, a controlled approximation in the linear regime [19] shows corrections,

$$\langle\Delta y_n\rangle = \rho^{-1} + \frac{1}{2\rho} \frac{\partial\rho^{-1}}{\partial x} + \Sigma, \quad (11)$$

where  $\Sigma$  represents the sum of all terms with second or higher order derivatives. The leading contribution to  $\Sigma$  is given by [21,19]

$$\Sigma = \frac{1}{6\rho^2} \frac{\partial^2\rho^{-1}}{\partial x^2} + O\left(\frac{\partial^3\rho^{-1}}{\partial x^3}\right). \quad (12)$$

For more general optimal velocity models, it can be shown that the sign of the first correction in (11) varies depending on whether each vehicle responds to the vehicle ahead of or behind it. Thus the first correction term represents the directionality of vehicular interaction, or directed influence.

By using (11), the first term in (10) can be expanded as

$$V_{op}(\langle\Delta y_n\rangle) = V_{op}(\rho^{-1}) + V'_{op}(\rho^{-1}) \left[ \frac{1}{2\rho} \frac{\partial\rho^{-1}}{\partial x} + \Sigma \right], \quad (13)$$

where linearly irrelevant terms are ignored. From (7), (9), (10), and (13), one obtains

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \lambda [V_{op}(\rho^{-1}) - v] + \frac{\lambda}{2\rho} V'_{op}(\rho^{-1}) \frac{\partial\rho^{-1}}{\partial x} + \lambda V'_{op}(\rho^{-1}) \Sigma. \quad (14)$$

Note that the second term proportional to the density gradient arises from the directed influence, while conventional derivations of fluid-dynamic models [4] attribute the density gradient term to the velocity variance term in (7). We will call the second term the anticipation term. The first term is often called the relaxation term.

It is interesting to compare the dispersion relations of microscopic and macroscopic models. In a microscopic description, small perturbations with respect to the homogeneous state can be written as

$$y_n(t) = v_h t + \rho_h^{-1} n + \delta y \exp(i\kappa n + \gamma t), \quad (15)$$

where  $v_h = V_{op}(\rho_h^{-1})$ . By linearizing (8), one obtains the dispersion relation

$$\gamma_{\pm} = \frac{\lambda}{2} \left[ -1 \pm \sqrt{1 + \frac{4V'_{op}}{\lambda} (e^{i\kappa} - 1)} \right]. \quad (16)$$

On the other hand, small perturbations in the macroscopic description can be written as

$$\rho(x, t) = \rho_h + \delta\rho \exp(ikx + \omega t), \quad v(x, t) = v_h + \delta v \exp(ikx + \omega t), \quad (17)$$

where  $k\rho_h^{-1}$  is the macroscopic counterpart of  $\kappa$  since both represent the phase difference between two successive vehicles, and  $\omega + ikv_h$  is the macroscopic counterpart of  $\gamma$ . To see the origin of the additional term  $ikv_h$ , note that  $\gamma$  is the frequency measured in the *moving* reference frame with the velocity  $v_h$ , while  $\omega$  is the frequency measured in the stationary frame. By linearizing (3) and (14), one finds

$$\omega_{\pm} + ikv_h = \frac{\lambda}{2} \left[ -1 \pm \sqrt{1 + \frac{4V'_{op}}{\lambda} A(k\rho_h^{-1})} \right], \quad (18)$$

where  $A(x) = ix + (ix)^2/2$ , when the last term in (14) proportional to  $\Sigma$  is ignored and  $A(x) = ix + (ix)^2/2 + (ix)^3/6$ , when the leading contribution to  $\Sigma$  in (12) is included. The macroscopic model derived in [21] amounts to the latter approximation of  $\Sigma$ . Note that  $A(x)$  agrees with the Taylor expansion of the factor  $(e^{i\kappa} - 1)$  in (16). Thus the macroscopic momentum equation (14), combined with the continuity equation (3), describes the long wavelength behavior of the microscopic model (8) properly. Recalling that the spirit of macroscopic models is to describe long wavelength properties of corresponding microscopic models, this feature is encouraging. However it turns out that this feature alone is not sufficient since this agreement in the long wavelength components can be completely destroyed by the dynamics of short wavelength components, as demonstrated below.

### 3.2 Artificial Instabilities

We compare the linear stability of the microscopic and macroscopic models. From (16), one finds that small fluctuations of the mode  $\kappa$  in the microscopic model become linearly unstable when

$$V'_{op}(\rho_h^{-1}) > \frac{\lambda}{1 + \cos \kappa}. \quad (19)$$

Note that the  $\kappa = 0$  mode shows the strongest instability and at the critical density where the instability first sets in, only the  $\kappa = 0$  mode becomes unstable.

A proper macroscopic model is expected to share the same stability/instability with respect to infinitesimal fluctuations. From (18), it is found that the mode  $k$  in the macroscopic model becomes unstable when

$$V'_{op}(\rho_h^{-1}) > \lambda \frac{-\text{Re}A(k\rho_h^{-1})}{\text{Im}^2 A(k\rho_h^{-1})}. \quad (20)$$

When  $\Sigma$  in (14) is ignored completely, (20) becomes  $V'_{op}(\rho_h^{-1}) > \lambda/2$ . Note that this inequality does not contain  $k$ . Thus as soon as  $\rho_h$  satisfies this inequality, fluctuations of *all* wavelengths become unstable simultaneously. This is qualitatively different from the behavior in the microscopic model, where only the long wavelength components become unstable. When the leading contribution to  $\Sigma$  in (12) is retained, on the other hand, (20) becomes  $V'_{op}(\rho_h^{-1}) > (\lambda/2)[1 - (k\rho_h^{-1})^2/6]^{-2}$ . Note that the right-hand side vanishes as  $k\rho_h^{-1} \rightarrow \infty$  and thus the homogeneous state is *always* unstable with respect to small wavelength fluctuations [21,19]. This should be contrasted with the existence of linearly stable density ranges in the microscopic model. This *artificial* instability in the macroscopic model cannot be cured by merely using higher order approximations of  $\Sigma$ . For example, if we assume that the next order contribution to  $\Sigma$  is  $(1/4!\rho^3)(\partial^3\rho^{-1}/\partial x^3)$ , which generates the correct next order in  $A(x)$ , (20) becomes  $V'_{op}(\rho_h^{-1}) > (\lambda/2)[1 - (k\rho_h^{-1})^2/12]/[1 - (k\rho_h^{-1})^2/6]^2$ , which again shows an artificial instability for the short wavelength components.

Therefore the macroscopic models defined by the mentioned approximations of  $\Sigma$  *cannot* be a proper macroscopic description of the microscopic model. An insight can be gained from the microscopic dispersion relation (16) by recalling that the mentioned approximations of  $\Sigma$  are equivalent to the truncation of the series  $e^{i\kappa} - 1 = i\kappa + (i\kappa)^2/2 + (i\kappa)^3/3! + (i\kappa)^4/4! + \dots$  at a certain order. It can be verified that when the series is truncated at a *finite* order, highest order terms dominate the physics for large  $\kappa \gg 1$  and generate the artificial instability for short wavelength components.

What would happen if the series is summed up to infinite order within a macroscopic description? The infinite order summation can be implemented for example by introducing a headway field  $b(x, t)$ , which is related to  $\rho(x, t)$  via the *nonlocal* relation [21],

$$\int_x^{x+b(x,t)} dx' \rho(x', t) = 1. \quad (21)$$

With this additional field, the momentum equation becomes

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \lambda [V_{op}(b) - v], \quad (22)$$

and the dispersion relation of this macroscopic model in terms of *three* macroscopic fields  $\rho$ ,  $v$ , and  $b$  is given by (18) with

$$A(k\rho_h^{-1}) = \exp(ik\rho_h^{-1}) - 1. \quad (23)$$

Thus the macroscopic dispersion relation is identical to the microscopic dispersion (16) when  $k\rho_h^{-1}$  is identified with  $\kappa$ . This confirms the infinite order summation in this model. However from the periodicity of  $A(k\rho_h^{-1})$ , a different problem arises:  $\text{Re}(\omega_+)$  vanishes for all modes  $k = 2n\pi\rho_h$ , where  $n$  is an arbitrary integer. Thus the homogeneous state in the macroscopic model has an infinite number of marginal modes at all densities, and when  $V'_{op}(\rho_h^{-1}) > \lambda/2$ , instabilities develop near *all* these marginal modes. In the microscopic description, on the other hand, this problem does not arise since only  $\kappa$  modulo  $2\pi$  is

meaningful, and thus  $\kappa$  and  $\kappa + 2\pi$  modes are not physically distinct modes. Hence the microscopic model is free from such marginal modes and it becomes evident that this infinite-order macroscopic model is not acceptable.

### 3.3 Effective Diffusion

Here we develop a new approximation of  $\Sigma$ , which generates neither artificial instabilities nor an infinite number of artificial marginal modes. A key observation is that modes with  $k\rho_h^{-1} \gg 1$  are *unphysical* since fluctuations on length scales shorter than the vehicle spacing are not defined in the original microscopic model. Motivated by this observation, we transform the leading order term of  $\Sigma$  in (12) in such a way that it preserves the long wavelength behavior but suppresses fluctuations in short wavelength components,  $k \gg \rho_h$ . To implement this idea, one first notes that (3) relates small fluctuations of  $\rho$  and  $v$  as follows:

$$\delta\rho = -\frac{ik\rho_h}{\omega + ikv_h} \delta v. \quad (24)$$

One then exploits the correspondence between  $\omega + ikv_h$  and  $\gamma$ , and between  $k\rho_h^{-1}$  and  $\kappa$ . From the result  $\gamma_+ \approx V'_{op}(\rho_h^{-1}) i\kappa$  for small  $\kappa$ , one obtains

$$\delta\rho^{-1} \approx \frac{1}{V'_{op}(\rho_h^{-1})} \delta v.$$

In this derivation, the  $\gamma_-$  mode is ignored since it always decays with time. Note that the resulting relation amounts to a variational form of  $v = V_{op}(\rho^{-1})$ . Corrections to the relation will be ignored since they introduce third or higher order derivatives to the new approximation of  $\Sigma$  (25). This way, we construct an approximation

$$V'_{op}(\rho^{-1}) \Sigma \approx V'_{op}(\rho^{-1}) \frac{1}{6\rho^2} \frac{\partial^2 \rho^{-1}}{\partial x^2} \approx \frac{1}{6\rho^2} \frac{\partial^2 v}{\partial x^2}. \quad (25)$$

The momentum equation becomes

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \lambda [V_{op}(\rho^{-1}) - v] - \frac{\lambda V'_{op}}{2\rho^3} \frac{\partial \rho}{\partial x} + \frac{\lambda}{6\rho^2} \frac{\partial^2 v}{\partial x^2}. \quad (26)$$

Note that our approximation of  $\Sigma$  results in a *diffusion* term, which tends to suppress short wavelength fluctuations. Indeed, the linear instability criterion from (3) and (26) becomes  $V'_{op}(\rho_h^{-1}) > \lambda(1 + k^2/6\rho_h^2)^2/2$ , which confirms the suppression of modes with  $k \gg \rho_h$ . In addition, it can be verified that the macroscopic and microscopic dispersion relations agree up to order  $k^3$ . Thus we conclude that (26) is a satisfactory macroscopic momentum equation in the linear regime.

## 4 General Car-Following Models

In this section, we consider a general car-following equation,

$$\ddot{y}_n = A_{op}(\Delta y_n, \Delta \dot{y}_n, \dot{y}_n). \quad (27)$$

Coarse graining leads to

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \approx A_{op}(\langle \Delta y_n \rangle, \langle \Delta \dot{y}_n \rangle, v), \quad (28)$$

where  $\langle \Delta y_n \rangle$  can be approximated by (11) and (12), and

$$\langle \Delta \dot{y}_n \rangle_{(x,t)} \approx \frac{1}{\rho} \frac{\partial v}{\partial x} + \frac{1}{2\rho^2} \frac{\partial^2 v}{\partial x^2}. \quad (29)$$

See [19] for a derivation of (29). We further expand  $A_{op}(\dots)$  as

$$\begin{aligned} A_{op}(\dots) &\approx A_{op}(\rho^{-1}, 0, v) + A_{op,1} \left( \frac{1}{2\rho} \frac{\partial \rho^{-1}}{\partial x} + \frac{1}{6\rho^2} \frac{\partial^2 \rho^{-1}}{\partial x^2} \right) \\ &\quad + A_{op,2} \left( \frac{1}{\rho} \frac{\partial v}{\partial x} + \frac{1}{2\rho^2} \frac{\partial^2 v}{\partial x^2} \right), \end{aligned} \quad (30)$$

where  $A_{op,i} \equiv \partial_{z_i} A_{op}(z_1, z_2, z_3)|_{(z_1, z_2, z_3)=(\rho^{-1}, 0, v)}$ . In real traffic systems,  $A_{op,1}$  and  $A_{op,2}$  are expected to be positive while  $A_{op,3}$  is expected to be negative. Linearly irrelevant terms are ignored in (30). The macroscopic momentum equation then becomes

$$\begin{aligned} \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} &= A_{op}(\rho^{-1}, 0, v) + \frac{A_{op,1}}{2\rho} \frac{\partial \rho^{-1}}{\partial x} + \frac{A_{op,1}}{6\rho^2} \frac{\partial^2 \rho^{-1}}{\partial x^2} \\ &\quad + \frac{A_{op,2}}{\rho} \frac{\partial v}{\partial x} + \frac{A_{op,2}}{2\rho^2} \frac{\partial^2 v}{\partial x^2}. \end{aligned} \quad (31)$$

Note that the dependence of  $A_{op}$  on  $\Delta \dot{y}_n$  gives rise to an explicit diffusion term.

Despite the explicit diffusion term, the artificial instability at short wavelength components may still arise when  $A_{op,1}$  is sufficiently large since the term proportional to  $\partial^2 \rho^{-1} / \partial x^2$  tends to generate the artificial instability, as demonstrated in Sect. 3. Thus we follow the procedure in Sect. 3.3 to obtain

$$\frac{\partial^2 \rho^{-1}}{\partial x^2} \approx -\frac{A_{op,3}}{A_{op,1}} \frac{\partial^2 v}{\partial x^2}, \quad (32)$$

which is a generalization of (25). The resulting momentum equation is

$$\begin{aligned} \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} &= A_{op}(\rho^{-1}, 0, v) + \frac{A_{op,1}}{2\rho} \frac{\partial \rho^{-1}}{\partial x} \\ &\quad + \frac{A_{op,2}}{\rho} \frac{\partial v}{\partial x} + \frac{3A_{op,2} - A_{op,3}}{6\rho^2} \frac{\partial^2 v}{\partial x^2}. \end{aligned} \quad (33)$$

Note that the factor  $3A_{op,2} - A_{op,3}$  in front of the diffusion term is manifestly positive. This equation is free from the artificial instability.

To elucidate the relation with (26), it is useful to define an effective optimal velocity  $V_{op,eff}(\rho^{-1})$  in an implicit way as a solution of

$$A_{op}(\rho^{-1}, 0, V_{op,eff}) = 0. \quad (34)$$

When  $A_{op,3} < 0$  for all  $v$ , the solution is unique and there is no ambiguity in  $V_{op,eff}(\rho^{-1})$ . One also defines

$$\lambda_{eff}(\rho^{-1}, v) \equiv \frac{A_{op}(\rho^{-1}, 0, v)}{V_{op,eff}(\rho^{-1}) - v}, \quad (35)$$

which is positive for all  $\rho$  and  $v$  if  $A_{op,3} < 0$  always. Thus the first term in (33) can be interpreted as a generalized relaxation term:

$$A_{op}(\rho^{-1}, 0, v) = \lambda_{eff}(\rho^{-1}, v)[V_{op,eff}(\rho^{-1}) - v]. \quad (36)$$

## 5 Micro vs. Macro

We numerically compare properties of the microscopic optimal velocity model (8) and the macroscopic model [(3) and (26)]. For definiteness, we use

$$V_{op}(\Delta y) = \frac{v_{max}}{2} \left[ \tanh \left( 2 \frac{\Delta y - x_{neutral}}{x_{width}} \right) + c_{bias} \right],$$

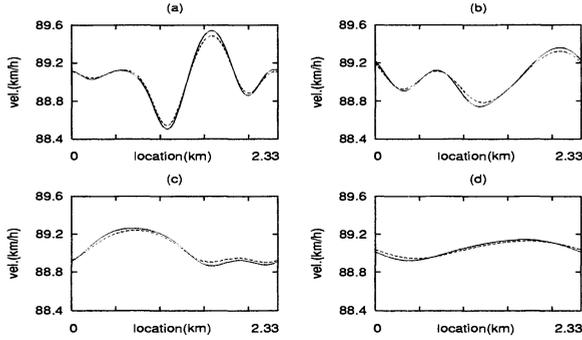
with  $v_{max} = 33.6$  m/s,  $x_{neutral} = 25.0$  m,  $x_{width} = 23.3$  m,  $c_{bias} = 0.913$ , and  $\lambda = 2 \text{ sec}^{-1}$  as in [22]. A system size  $L = 2.33$  km is simulated with  $N$  vehicles ( $\rho_h \equiv N/L$ ), and the following microscopic initial conditions are used:

$$\begin{aligned} y_n(0) &= n\rho_h^{-1} + A \sin(6\pi n\rho_h^{-1}/L), & 1 \leq n < N/3, \\ y_n(0) &= n\rho_h^{-1}, & N/3 \leq n \leq 2N/3, \\ \dot{y}_n(0) &= V_{op}(\Delta y_n(0)), & \text{for all } n. \end{aligned} \quad (37)$$

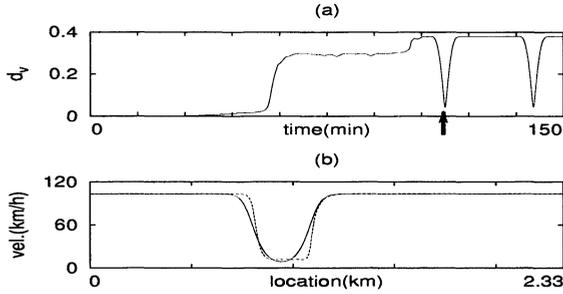
The corresponding macroscopic initial condition is prepared by coarse-graining the microscopic initial condition with the coarse graining function  $\phi(x, t) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)\delta(t)$ , where we choose  $\sigma = 46.4$  m. The periodic boundary condition is imposed for both the microscopic and macroscopic systems.

We first verify that the density range  $\rho_{c1} < \rho < \rho_{c2}$ , in which the homogeneous traffic state becomes unstable with respect to infinitesimal perturbations, is essentially identical for the microscopic and macroscopic models. This implies that, in the linear regime, the macroscopic model describes the long wavelength behavior of the microscopic model very accurately. To quantify the accuracy of the macroscopic model, we introduce the space-averaged relative deviation  $d_v(t)$ ,

$$d_v(t) \equiv \frac{\sqrt{\langle [v_{macro}(x, t) - v_{micro}(x, t)]^2 \rangle_{space}}}{\langle v_{micro}(x, t) \rangle_{space}},$$



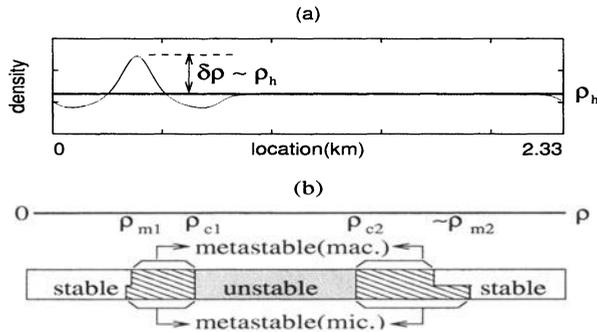
**Fig. 1.** The velocity profiles for  $N = 72$ . The initial condition in (37) is used with  $A = 1.165$  m. (a)  $t \approx 10$  min, (b)  $t \approx 30$  min, (c)  $t \approx 1$  h, and (d)  $t \approx 4$  h. The *solid* (*dashed*) line shows the microscopic (macroscopic) velocity profile in each plot. The vertical scale is magnified for clarity. After [19]



**Fig. 2.** (a) The time evolution of the space-averaged relative deviation of velocity for  $A = 1.165$  m and  $N = 73$ . (b)  $v_{micro}$  (*solid line*) vs  $v_{macro}$  (*dashed line*) near 115 min [marked by the arrow in (a)]. After [19]

where  $\langle \dots \rangle_{space}$  represents the spatial average. Here  $v_{macro}(x, t)$  is calculated from the macroscopic model, while  $v_{micro}(x, t)$  is obtained by coarse-graining the microscopic configuration at the time  $t$ . When the initial perturbation from homogeneous flow is small, say  $A = 1.165$  m, we find that  $d_v(t)$  is negligible for all density outside the linearly unstable density range. A typical velocity profile is shown in Fig. 1. Note that the macroscopic profiles are almost indistinguishable from the microscopic ones. Even when  $N = 72$  (131), which corresponds to a density slightly below (above) the lower (upper) critical density  $\rho_{c1(c2)} \approx 73$  (130)/2.33 km,  $d_v(t)$  remains  $\sim 2 \times 10^{-4}$  during several hours of simulation time.

The accuracy in the linearly unstable density range is also examined for  $A = 1.165$  m and  $N = 73$ , which is the smallest  $N$  that demonstrates the linear instability. The microscopic simulation shows that the initially smooth profile becomes “rough” as short wavelength fluctuations develop. An almost

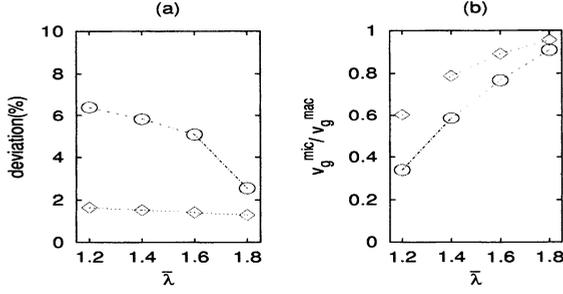


**Fig. 3.** (a) The density profile for the initial condition in (37).  $\delta\rho$  depends on  $A$  and  $\rho_h$ . For  $A = 74.56$  m and  $N = 100$ ,  $\delta\rho \simeq 1.5\rho_h$ . (b) Schematic phase diagrams for the microscopic and macroscopic models. After [19]

identical roughening is found in the macroscopic simulation, and  $d_v(t)$  is almost negligible initially [Fig. 2(a)]. However, the growth rate of the short wavelength fluctuations is faster in the microscopic simulation compared to the macroscopic simulation. This difference is responsible for the rapid growth of  $d_v(t)$  near  $t \approx 55$  min. The growth of  $d_v(t)$  occurs at an earlier time for the density with stronger linear instability. Both in microscopic and macroscopic simulations, after a sufficient time interval ( $\sim 120$  min) all short wavelength fluctuations merge into a single large traffic jam, which moves backward at a constant speed without further evolution in its shape. Thus this jam corresponds to the final steady state. Figure 2(b) compares the velocity profiles of the jams from the microscopic and macroscopic simulations. The jam propagation speeds in the microscopic and macroscopic simulations are different and in a periodic system, the locations of the jams coincide periodically in time, resulting in the periodic dips in Fig. 2(a).

Next we choose  $A = 74.56$  m in (37), and examine the performance of the macroscopic model for large perturbations. Figure 3(a) shows the initial density profile. After a sufficiently long time, the initial condition may evolve to a homogeneous state or to a congested state. The evolution to a congested state is realized for  $65 \leq N \leq 156$  when the microscopic model is used and for  $66 \leq N \leq 147$  when the macroscopic model is used. Thus the lower critical density is in good agreement while the upper critical density shows about 6% deviation. The comparison with the linear critical densities shows that the metastable density ranges exist for both microscopic and macroscopic models. The phase diagram in Fig. 3(b) summarizes the result.

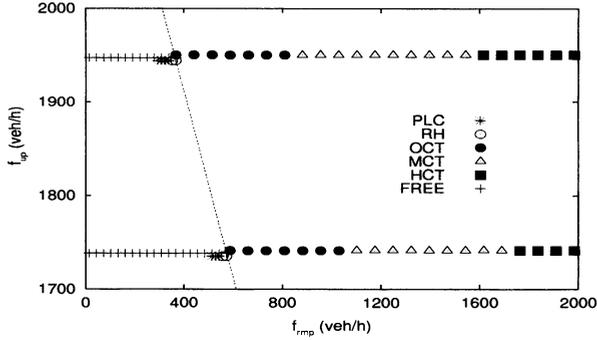
We also investigate the dependence of the critical density on  $\lambda$  for fixed  $A = 74.56$  m. It is convenient to introduce a dimensionless parameter  $\bar{\lambda} \equiv (x_{width}/v_{max})\lambda$ , which is about 1.387 for  $\lambda = 2 \text{ sec}^{-1}$ . Figure 4(a) shows the relative deviations of the macroscopic critical densities with respect to the microscopic ones. For the lower critical density, the macroscopic result is in good



**Fig. 4.** (a) The relative deviations of the macroscopic lower (*diamonds*) and upper (*circles*) critical densities with respect to the microscopic counterparts for the initial condition (37) with  $A = 74.56$  m. Note that the relative deviations shrink as  $\bar{\lambda}$  increases. (b) The ratio  $v_g^{\text{mic}}/v_g^{\text{mac}}$  as a function of  $\bar{\lambda}$  for the macroscopic model [(3) and (26)] (*diamonds*) and for the modified macroscopic model [19] that takes into account the effects of some linearly irrelevant terms (*circles*). After [19]

agreement with the microscopic one for general  $\bar{\lambda}$ . For the upper critical density, on the other hand, the deviation of about 6% at  $\bar{\lambda} \approx 1.387$  shrinks with the increase of  $\bar{\lambda}$  and good agreement is achieved near  $\bar{\lambda} = 2$ . Thus the difference between the microscopic and macroscopic metastable regions in Fig. 3(b) shrinks as  $\bar{\lambda} \rightarrow 2$ . The velocity  $-v_g$  of a backward propagating traffic jam cluster ( $v_g > 0$ ) is also investigated. Since  $v_g$  is almost independent of  $N$ , we fix  $N = 100$  ( $\rho_h \approx 42.9 \text{ km}^{-1}$ ) for simplicity, and examine  $v_g$  as a function of  $\bar{\lambda}$ . Figure 4(b) (*diamonds*) shows the ratio between the microscopic value  $v_g^{\text{mic}}$  and the macroscopic value  $v_g^{\text{mac}}$ . Note that  $v_g^{\text{mic}}/v_g^{\text{mac}} \approx 1$  when  $\bar{\lambda}$  is close to 2. This agreement is notable considering that the macroscopic model does not have any free parameter. The agreement, however, becomes less satisfactory as  $\bar{\lambda}$  becomes smaller. A crude understanding for the good agreement near  $\bar{\lambda} = 2$  can be achieved via the linear analysis, although the given initial condition is not in the linear regime. For the general optimal velocity model, the linear instability develops when  $\bar{V}'_{op} > \bar{\lambda}/(1 + \cos \kappa)$ ; here we introduce  $\bar{V}'_{op} \equiv (x_{width}/v_{max})V'_{op}$ . This inequality sets an upper limit  $\kappa_c$ , above which the instability does not appear. Note that  $\kappa_c$  shrinks to zero as  $\bar{\lambda}/2$  approaches  $max(\bar{V}'_{op})$ , which is 1. Thus the characteristic length scale of the instability becomes longer as  $\bar{\lambda} \rightarrow 2$ . This may explain the excellent agreement near  $\bar{\lambda} = 2$ , since the macroscopic model becomes more precise as the characteristic length scale grows.

From these comparisons, we conclude that the macroscopic model [(3) and (26)] is quite accurate in the linear regime, and provides a reasonable description of fully developed jam clusters in the nonlinear regime, although there are deviations in the quantitative level. To construct more accurate macroscopic models, one needs to take into account effects of various terms ignored in the macroscopic momentum equation derivation. As a first trial, [19] extends the derivation to the nonlinear regime by including effects of all terms proportional to  $(\partial v/\partial x)^2$ ,



**Fig. 5.** Phase diagram in the presence of an on-ramp, obtained from the macroscopic model derived from the optimal velocity model. The dotted line represents  $f_{up} + f_{rm} = \rho_{c1} V_{op}(\rho_{c1}^{-1})$ , where  $\rho_{c1}$  is the lower critical density for the linear instability

$(\partial\rho^{-1}/\partial x)^2$ , and  $(\partial\rho^{-1}/\partial x)(\partial v/\partial x)$ . The resulting equation for the same optimal velocity model is examined and as expected, the linearly unstable density region is identical to that by (26). However, the ratio  $v_g^{mic}/v_g^{mac}$  deviates further from one [circles in Fig. 4(b)]. Thus it appears that naive inclusion of linearly irrelevant terms does not improve the accuracy.

Lastly we examine effects of an on-ramp. The continuity equation becomes

$$\frac{\partial}{\partial t}\rho + \frac{\partial}{\partial x}(\rho v) = f_{rm}\varphi(x), \quad (38)$$

where  $f_{rm}$  represents the total external flux through an on-ramp and the normalized spatial distribution function  $\varphi(x)$  is localized near  $x = 0$ . For (26), we choose  $\lambda = 1 \text{ sec}^{-1}$  and use  $V(\rho)$  in [6] for  $V_{op}(\rho^{-1})$ . It is found that various traffic phases reported in [6], such as pinned localized cluster (PLC) state, recurring hump (RH) states, oscillating congested traffic (OCT) states, mixed congested traffic (MCT) states, and homogeneous congested traffic (HCT) states, are all reproduced by the macroscopic model. Moreover the phase diagram (Fig. 5) of traffic phases in the plane of the upstream flux  $f_{up}$  and  $f_{rm}$  has a qualitatively similar structure as the corresponding diagram in [6].

## 6 Summary

A local macroscopic fluid-dynamic model is derived from a microscopic car-following model, which establishes the link between the two types of traffic models. For the optimal velocity model, the corresponding macroscopic momentum equation consists of a relaxation term, an anticipation term (proportional to the density gradient), and a diffusion term. Thus it has a structure similar to the fluid-dynamic model in [4]. However, the density gradient term is found to arise from the directed influence rather than the velocity variance. It is demonstrated that the diffusion term also arises from the directed influence. The derivation

provides an unambiguous way to determine the density-dependent coefficients of the anticipation term and the diffusion term. The macroscopic model derived from the optimal velocity model is examined numerically, and its properties are found to be in reasonable agreement with those of the microscopic model.

## Acknowledgments

H. K. L. thanks G. S. Jeon for discussions. This work was supported by the Brain Korea 21 Project in 2001.

## References

1. K. Nagel, J. Esser, M. Rickert: In *Annu. Rev. Comp. Phys.*, ed. by D. Stauffer (World Scientific, 1999).
2. D. Chowdhury, L. Santen, A. Schadschneider: *Phys. Rep.* **329**, 199 (2000).
3. D. Helbing: To appear in *Rev. Mod. Phys.* (cond-mat/0012229).
4. B.S. Kerner, P. Konhäuser: *Phys. Rev. E* **48**, R2335 (1993).
5. T. Nagatani: *J. Phys. Soc. Jpn.* **66**, 1928 (1997).
6. H.Y. Lee, H.-W. Lee, D. Kim: *Phys. Rev. Lett.* **81**, 1130 (1998); *Phys. Rev. E* **59**, 5101 (1999).
7. D. Helbing, M. Treiber: *Phys. Rev. Lett.* **81**, 3042 (1998); D. Helbing, A. Hennecke, M. Treiber: *ibid.* **82**, 4360 (1999).
8. N. Mitarai, H. Nakanishi: *J. Phys. Soc. Jpn.* **68**, 2475 (1999); *Phys. Rev. Lett.* **85**, 1766 (2000).
9. E. Tomer, L. Safonov, S. Havlin: *Phys. Rev. Lett.* **84**, 382 (2000).
10. J. Treiterer, J.A. Myers: In *Proc. 6th Int. Symp. on Transportation and Traffic Theory*, ed. by D. J. Buckley (Elsevier, New York 1974); M. Koshi, M. Iwasaki, I. Ohkura: In *Proceedings of the Eighth International Symposium on Transportation and Traffic Flow*, ed. by V. F. Hurdle, E. Hauer, G. N. Stewart (University of Toronto Press, Toronto 1983).
11. B.S. Kerner, H. Rehborn: *Phys. Rev. E* **53**, R4275 (1996); *Phys. Rev. Lett.* **79**, 4030 (1997); B.S. Kerner: *ibid.* **81**, 1130 (1998); *J. Phys. A* **33**, L221 (2000).
12. L. Neubert et al.: *Phys. Rev. E* **60**, 6480 (1999).
13. M. Treiber, A. Hennecke, D. Helbing: *Phys. Rev. E* **62**, 1805 (2000).
14. H.Y. Lee, H.-W. Lee, D. Kim: *Phys. Rev. E* **62**, 4737 (2000).
15. M. Hermann, B.S. Kerner: *Physica A* **255**, 163 (1998).
16. H. Hayakawa, K. Nakanishi: *Prog. Theor. Phys. Suppl.* no. 130, 57 (1998).
17. M. Treiber, A. Hennecke, D. Helbing: *Phys. Rev. E* **59**, 239 (1999).
18. We remark that a different scheme to construct macroscopic models from microscopic car-following models was proposed recently [D. Helbing, A. Hennecke, V. Shvetsov, M. Treiber: To appear in *Mathematical and Computer Modelling* (cond-mat/0003269)]. The macroscopic fields  $\rho$  and  $v$  are defined via an interpolation procedure instead of a coarse-graining procedure. The resulting momentum equation is *nonlocal*, while our momentum equation is local. Also  $\rho$  and  $v$  defined in such a way do not strictly satisfy the continuity equation (3), while the continuity equation is an exact identity in the coarse-graining-based scheme.
19. H.K. Lee, H.-W. Lee, D. Kim: *Phys. Rev. E* **64**, 056126 (2001).
20. M. Bando et al.: *Phys. Rev. E* **51**, 1035 (1995).
21. P. Berg, A. Mason, A. Woods: *Phys. Rev. E* **61**, 1056 (2000).
22. S.-I. Tadaki et al.: *J. Phys. Soc. Jpn.* **67**, 2270 (1998).

# Metastable Flows in an Extended Burgers Cellular Automaton Traffic Model

M. Fukui<sup>1</sup>, K. Nishinari<sup>2</sup>, D. Takahashi<sup>3</sup>, and Y. Ishibashi<sup>4</sup>

<sup>1</sup>Nakanihon Automotive College, Sakahogi-cho, Gifu-ken 505-0077, Japan

<sup>2</sup>Department of Applied Mathematics and Informatics, Ryukoku University, Seta, Ohtsu 520-2194, Japan

<sup>3</sup>Department of Mathematical Sciences, Waseda University, Ohkubo 3-4-1, Shinjuku-ku, Tokyo 169-8555, Japan

<sup>4</sup>School of Communications, Aichi Shukutoku University, Nagakute-cho, Aichi-ken 480-1197, Japan

**Abstract.** Traffic flow in the extended Burgers cellular automaton (EBCA1) traffic model is studied. The flow and configuration of vehicles on two lanes in the model are simulated and three kinds of metastable local congested states are found in a two-dimensional region on the density-flow diagram. The first is that cars advance by stop and go-flow on their own lanes without lane-change and values of the flow are stable in time. The second is that cars change the lane periodically with several time-steps. The third is that they advance changing the lane that induces fluctuation of the flow with extremely long period. This fluctuation flow exists in a wide range between car densities  $5/12$  and  $3/4$ . The metastable flow is discussed in connection with the synchronized states observed in the traffic flow on an expressway.

## 1 Introduction

Recently, a new CA traffic model [1] has been proposed as a transformation of the Burgers equation by the ultradiscrete method, which is called Burgers cellular automaton (BCA). The multiple states in the BCA can be interpreted to represent a multi-lane expressway in a traffic model. The BCA is further extended to several CA models [2]. In these extended BCA (EBCA) models, EBCA1 is a multi-value (multi-lane) and multi-neighbor (high-speed) extension model of the rule-184. The most notable character of EBCA1 is an appearance of new branches of metastable states in the density-flow diagram.

In the recent years a lot of traffic data of expressways are accumulated and characteristic features of traffic flow are analyzed by Kerner and Rehborn [3-6]. They have focused on metastable synchronized flow and proposed that there are three phases in traffic flow: free flow, synchronized flow and wide moving jam [3]. The jams usually emerge in free flow through a sequence of two phase transitions: “free flow  $\rightarrow$  synchronized flow  $\rightarrow$  jams”. Thus the synchronized phase becomes an essential key word in the study of traffic flow.

In the present paper, a two-lane traffic model equivalent to the EBCA1 is proposed and many metastable branches found in fundamental diagram are analyzed. Furthermore lane-changing phenomena and configurations of vehicles on each lane in the congested states are investigated numerically.

## 2 Two-Lane Traffic Model Equivalent to EBCA1 Model

In EBCA1, vehicles advance by following two successive procedures.

- a) Vehicles move to their next site if the site is not fully occupied.
- b) Only vehicles moved in the procedure a) can move one more site if their next site is not fully occupied after the procedure a).

The procedure b) expresses the slow start rule where vehicles that could not move in the procedure a) stay at that site even if their next site is not fully occupied. These procedures have been expressed by following an evolution equation [2]:

$$U_j^{t+1} = U_j^t + b_{j-1}^t - b_j^t + \min(b_{j-2}^t, L - U_j^t - b_{j-1}^t + b_j^t) - \min(b_{j-1}^t, L - U_{j+1}^t - b_j^t + b_{j+1}^t), \quad (1)$$

where  $U_j^t (\in \{Z \mid 0 \leq U_j^t \leq L\})$  is the number of states at the site  $j$  and time  $t$ .  $L$  is assumed as a capacity in each site. The multiple state expresses that each site can hold  $L$  vehicles at maximum. Now the multiple states are interpreted as a multiple-lane model. Let us consider the case  $L=2$  in this paper and name each lane A-lane or B-lane.  $A_j^t$  and  $B_j^t$  ( $U_j^t = A_j^t + B_j^t$ ) denote the numbers of vehicles in the A-lane and B-lane at site  $j$  and time  $t$ , respectively.

We can rewrite (1) separately to two equations for vehicles on A and B-lanes as follows,

$$A_j^{t+1} = A_j^t + \min(A_{j-1}^{*t}, 1 - A_j^t) - \min(A_j^{*t}, 1 - A_{j+1}^t) + \min(1 - A_{j-1}^{*t}, 1 - A_j^t, B_{j-1}^{*t}, B_j^t) - \min(A_j^{*t}, A_{j+1}^t, 1 - B_j^{*t}, 1 - B_{j+1}^t), \quad (2)$$

$$B_j^{t+1} = B_j^t + \min(B_{j-1}^{*t}, 1 - B_j^t) - \min(B_j^{*t}, 1 - B_{j+1}^t) + \min(1 - B_{j-1}^{*t}, 1 - B_j^t, A_{j-1}^{*t}, A_j^t) - \min(B_j^{*t}, B_{j+1}^t, 1 - A_j^{*t}, 1 - A_{j+1}^t), \quad (3)$$

where

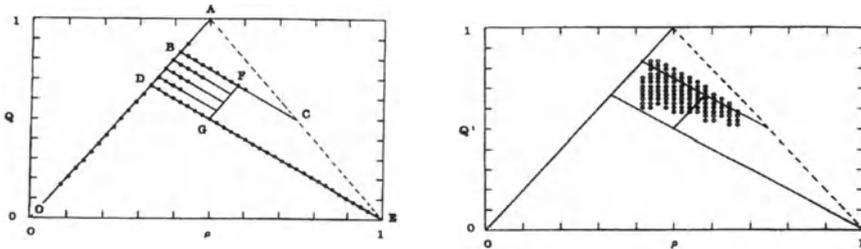
$$A_j^{*t} = A_j^t - [A_j^{t-1} - \min(A_j^{t-1}, 1 - A_{j+1}^{t-1}, \max(1 - A_j^t, 1 - A_{j+1}^t))] - \min(A_j^{t-1}, A_{j+1}^{t-1}, 1 - B_{j+1}^{t-1}, \max(1 - B_j^{t-1}, 1 - A_j^t)), \quad (4)$$

$$B_j^{*t} = B_j^t - [B_j^{t-1} - \min(B_j^{t-1}, 1 - B_{j+1}^{t-1}, \max(1 - B_j^t, 1 - B_{j+1}^t))] - \min(B_j^{t-1}, B_{j+1}^{t-1}, 1 - A_{j+1}^{t-1}, \max(1 - A_j^{t-1}, 1 - B_j^t)), \quad (5)$$

The last two terms in the first line in (2) and (3) express the number of vehicles going straight on their own lane, and the last two lines show that of vehicles changing lane. The asterisked term  $A^*$  or  $B^*$  includes the slow start effect.

### 3 Metastable Branches in Fundamental Diagram

Two-lane traffic flows ( $Q^h$ ,  $Q_A^t$  (A-lane) and  $Q_B^t$  (B-lane)) on a periodic road are simulated and the fundamental diagram of the EBCA1 is discussed. We got a density-flow diagram of the total average flow  $Q^t$  of the EBCA1 with  $L=2$  and  $K=24$  ( $K$  is road length) at various vehicle densities  $\rho$  ( $=N/KL$ :  $N$  is the number of vehicles). Above the critical density  $\rho_c$  ( $=1/3$ ), many different values of  $Q^t$  are obtained for different initial distributions for a given density. It is found that these multi-value states can be classified into three types. The first one is an asymptotically steady state whose value converges to a constant quickly in time. The cars advance by stop-and-go-flow on their own lanes without lane-change. The second is that cars change the lane periodically after several time-steps. The third is that they advance changing the lane that induces fluctuation of the flow with extremely long period. This fluctuating flow exists in a wide range between car densities  $\rho = 5/12$  and  $\rho = 3/4$ . According to this classification, the first and third type flows are separately shown in Figs. 1(a) and 1(b).

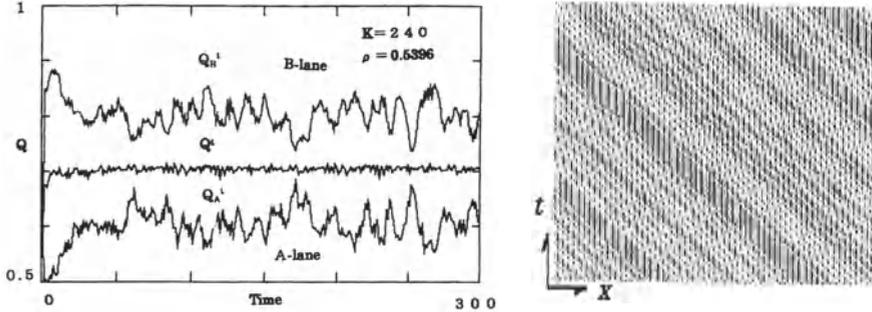


**Fig. 1.** (a) Density-flow diagram of the steady flow. (b) Density-flow diagram of the fluctuating flow.

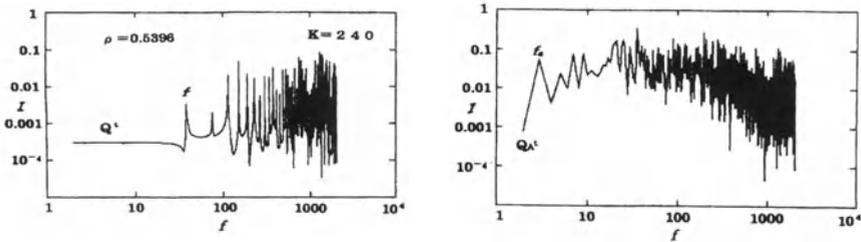
All points in Fig. 1(a) are on straight lines and the flow quickly relaxes to a steady state in each case. A part of the AO line, namely AD, and the branch BC are metastable [2]. Apart from this main frame, there are many metastable branches in a parallelogram region BDGF. In the region, there are  $K/6-1$  straight branches parallel to BC with an equal interval. The points are sited on these straight branches and make a two-dimensional lattice. As  $K$  increases, the interval between the points decreases and the parallelogram BDGF is filled with the points quasi-continuously. All these points are also metastable.

Fig. 1(b) shows the third case of the multi-value states. The values fluctuate in time and distribute in the parallelogram region and around the branch BC. Fig. 2 shows variation in time of the flows  $Q^h$ ,  $Q_A^t$  and  $Q_B^t$  at  $K=240$  and  $\rho=0.5396$ , whose values fluctuate with long period, extremely long in  $Q_A^t$  and  $Q_B^t$ . The fluc-

tuating state exists between two critical densities,  $\rho=5/12$  and  $\rho=3/4$ . In the region between  $\rho=1/3$  and  $\rho=5/12$ , the state relaxes to the free state through “metastable→ free state” transition or to jam through “metastable→ jam state” transition. Figs. 3(a) and 3(b) show power spectra  $I$  versus frequency  $f$  defined by  $I=|\sum_{t=1}^T Q^t \exp(2\pi i f t/T)|/T$  for  $Q^t$  and  $Q_A^t$ , respectively. The corresponding periods are marked by  $f(=1/T)$  or  $f_a(=1/nT)$  in each figure. The low frequency terms in  $Q_A^t$  and  $Q_B^t$  cancel out each other.  $T$  becomes long to almost infinite as  $K$  increases. This suggests that the flow in this fluctuating state is chaotic.



**Fig. 2.** (a) Fluctuation of the flows  $Q_A^t$ ,  $Q_B^t$  in the A- and B-lanes and the total flow  $Q^t$ . (b)Traces of cars on both lanes in space-time at  $\rho=0.5396$ . A large dot corresponds to two cars in a site and a small dot to one car.



**Fig. 3.** Power spectrum of (a):  $Q^t$  and (b)  $Q_A^t$ .

#### 4 Flow and Configuration of Vehicles on the Two-Lane Road

The total flow on linear branches in the density-flow diagram is stable in time and all vehicles advance regularly on their own lane without lane-change. The configuration of the point  $D_{nm}$  of  $m$ -th point toward B and  $n$ -th toward E from D point sited in BDGF can be expressed by exchanging  $m$  blocks of “100100” for “101010” on the A-lane and  $n$  blocks of “100” for “101” on the B-lane from the basic configuration  $D_{00}$ , which is...100100100... on A-lane, ...100100100... on

B-lane, and then  $A+B=\dots 200200200\dots$  on the total lane. Vehicles in this state advance on their own lane with stop-and-go flow.

There are many points fluctuating around the two-dimensional lattice or the straight lines. This fluctuation is induced by the lane-change process. We introduce a variable  $C$  as the lane-change rate of a vehicle as follows,

$$C = 1/TN \sum_{t=1}^T \sum_{j=1}^K \{ \min(1-A_{j-1}^{*t}, 1-A_j^t, B_{j-1}^{*t}, B_j^t) + \min(A_j^{*t}, A_{j+1}^t, 1-B_j^{*t}, 1-B_{j+1}^t) \}, \quad (6)$$

where the first and second terms in the braces indicate numbers of vehicles changing from B to A lane and from A to B lane, respectively. Fig. 4 shows the variation of  $C$  for various vehicle densities at  $K=240$ . It has the critical density ( $\rho=5/12$  and  $\rho=3/4$ ).

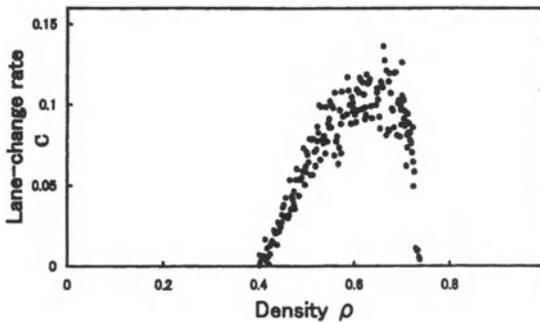


Fig. 4. Lane-change rate of cars in the fluctuating flow.

## 5 Discussions

In this paper, a two-lane traffic model equivalent to the EBCA1 model is studied. We have obtained the evolution equations for each lane and found many metastable sub-branches besides the main branches in the flow-density diagram. Many two-lane traffic models including more realistic lane-change models and considering asymmetric roles of two lanes have been studied [7-10]. Compared with them, the EBCA1 will be a too simple model. Nevertheless, it can provide many kinds of metastable states. The fact may suggest that *slow-start*, *high-velocity*, and *multiple-lane* models are essential processes to bring metastable states into existence in a traffic model.

Recent experimental investigations by Kerner and Rehborn [3-6] show that in synchronized traffic flow cars could move with nearly the same velocities in different lanes near on-ramps. They say that there are three phases: free flow, synchronized flow and wide moving jam. The jams emerge in free flow through a se-

quence of two phase transitions: “free flow→ synchronized flow→ jams”. They moreover say that the synchronized flow can further be characterized into three different kinds of phases [4]: 1) both the velocity and the flow are stationary, 2) the velocity is only stationary and the flow is not stationary and 3) both the velocity and the flow abruptly change. Thus the synchronized phase seems to be a key word to clarify the traffic flow, although it has not been fully established. In many metastable states found in the EBCA1 model, the stable flow in linear branches with constant velocity may correspond to the above type 1) synchronized flow. The type 2) and 3) synchronized flows may be found in the present fluctuating flows. Especially, we think that type 3) has deep relation with our fluctuating flow. In further studies, the local and dynamical configuration of cars in those flows should be studied and behaviors of individual car should be clarified in connection with the synchronized states.

## References

1. K. Nishinari and D. Takahashi, *J. Phys. A: Math. Gen.* **32**, 93 (1998).
2. K. Nishinari and D. Takahashi, *J. Phys. A: Math. Gen.* **33**, 7709 (2000).
3. B.S. Kerner and H. Rehborn, *Phys. Rev.* **E53**, R4275 (1996).
4. B.S. Kerner and H. Rehborn, *Phys. Rev. Lett.* **79**, 4030 (1997).
5. B.S. Kerner, *Phys. Rev. Lett.* **81**, 3797 (1998).
6. B.S. Kerner, *J. Phys. A: Math. Gen.* **33**, L221 (2000).
7. T. Nagatani, *J. Phys. Soc. Jpn.* **63**, 52 (1994).
8. P. Wanger, K. Nagel, and D.E. Wolf, *Physica* **A234**, 687 (1997).
9. D. Chowdhury, D.E. Wolf, and M. Schreckenberg, *Physica* **A235**, 417 (1997).
10. W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, *Physica* **A265**, 614 (1999).

# Existence and Classification of Travelling Wave Solutions to Second Order Highway Traffic Models

R.E. Wilson<sup>\*1</sup> and P. Berg<sup>2</sup>

<sup>1</sup> University of Bristol, Department of Engineering Mathematics, Bristol BS8 1TR, United Kingdom

<sup>2</sup> Simon Fraser University, Pacific Institute for the Mathematical Sciences, Vancouver, Canada

**Abstract.** The Bando optimal velocity model has a rich structure of wave solutions of travelling, composite and expansion types, even in the stable regime. We analyse these waves via a continuum limit which resembles the Kerner-Konhäuser PDE model, to which our analysis also applies. Geometrical properties of the fundamental diagram, in combination with global bifurcation theory of the travelling wave phase plane, lead to a complete classification of wave types which applies to several second order highway traffic models.

## 1 Introduction and Scope

The aim of this paper is to start the development of a universal mathematical theory which explains the qualitative types of wave solution exhibited by optimal velocity highway traffic models.

We are concerned with *second order* models, where the density  $\rho(x, t)$  and velocity  $v(x, t)$  solve a partial differential equation system of the form

$$\rho_t + (\rho v)_x = 0, \quad (1a)$$

$$v_t + vv_x = \alpha \{V(\rho) - v\} + N(\rho, \rho_x, \rho_{xx}, v, v_x, v_{xx}). \quad (1b)$$

Here (1a) is known as the *continuity equation* and expresses the conservation of vehicles; the *momentum equation* (1b) describes acceleration. Further, the *optimal velocity function* is given by  $V$ , and  $\alpha > 0$  is the *sensitivity* parameter. We suppose that  $N$  models anticipation and dispersive effects, and includes nontrivial dependence on at least one of the second derivatives  $\rho_{xx}$  or  $v_{xx}$ . A minimum constraint is that  $N(\rho, 0, 0, v, 0, 0) = 0$  for any  $\rho, v$ , so that uniform flow  $v = V(\rho)$ ,  $\rho$  independent of  $x, t$ , is a solution of the model.

In this paper, we consider only models and parameter choices for which all uniform flows are linearly stable. However, despite this simplifying assumption, our goal of a universal theory is still far from achieved, and an open question

---

\* R. Eddie Wilson wishes to thank the Royal Society and the Royal Academy of Engineering for their generous travel grant support.

concerns the minimum conditions to be imposed on  $N$  for qualitatively equivalent phase diagrams. For this reason, we focus here on the two models given by

$$N = -\beta \frac{\rho_x}{\rho} + \mu \frac{v_{xx}}{\rho}, \quad \text{and} \quad (2)$$

$$N = +\alpha V'(\rho) \left[ \frac{\rho_x}{2\rho} - \frac{\rho_x^2}{2\rho^3} + \frac{\rho_{xx}}{6\rho^2} \right]. \quad (3)$$

Formula (2) was introduced by Kerner and Konhäuser [4] and (3) was derived as a continuum limit of the discrete Bando model [1], by Berg, Mason and Woods [2]. Thus to prove equivalence of the dynamics of highway traffic models, one should convert all models to continuum variables, which are analytically more tractable.

Our approach now is to seek travelling wave solutions of speed  $c$ ,

$$v = W(z), \quad \rho = X(z), \quad z = x - ct, \quad (4)$$

to (1a,1b) on an infinite road, with uniform flow conditions

$$X(\pm\infty) = \rho_{\pm}, \quad W(\pm\infty) = V(\rho_{\pm}), \quad (5)$$

at its ends.

The continuity equation (1a) thus becomes  $-cX' + (WX)' = 0$ , where  $'$  denotes differentiation with respect to the travelling wave variable  $z$ , and we integrate directly to obtain  $X(W - c) = q$ , where  $q$  is a constant of integration with units flux. Thus we find

$$W = \frac{q}{X} + c, \quad (6)$$

and henceforth eliminate the velocity variable  $W$  from our calculations.

When we apply the boundary conditions at  $z = \pm\infty$ , we obtain  $V(\rho_{\pm}) = q/\rho_{\pm} + c$ , and eliminations yield

$$c = \frac{\rho_+ V(\rho_+) - \rho_- V(\rho_-)}{\rho_+ - \rho_-}, \quad (7)$$

and noting

$$q = \rho_- V(\rho_-) - c\rho_-, \quad (8)$$

we obtain the (well-known) result that if a wave exists, then its speed  $c$  is the gradient of the chord cutting the *fundamental diagram*  $Q(\rho) := \rho V(\rho)$  at  $(\rho_-, Q(\rho_-))$  and  $(\rho_+, Q(\rho_+))$ . Further,  $q$  is clearly the height of the chord's intercept with the  $\rho = 0$  axis.

It is thus useful to study the geometry of the fundamental diagram (details in Section 2), which describes the flow rate of equilibrium solutions. Later we outline technical assumptions concerning  $Q$  which are employed to ensure that chords with negative gradient intersect the graph of  $Q$  in three places.

The next point (and the main trick of this paper) is to note that if the travelling wave ansatz (4) is substituted in the momentum equation (1b), and

(6) used to eliminate the velocity variable, a second order ODE for the density variable  $X$  is found. This may be re-cast as a pair of first order ODEs: i.e., we consider the *travelling wave phase plane* defined by

$$X' = Y, \quad (9a)$$

$$Y' = h_1(X) \left[ \frac{q^2 Y}{X^3} + \frac{\alpha}{X} \{Q(X) - (q + cX)\} \right] + Y h_2(X, Y). \quad (9b)$$

Here the functions  $h_1(X)$ ,  $h_2(X, Y)$  contain all the details of the dynamics relating to the pressure and diffusion terms in the momentum equation. We find that

$$h_1(X) = \begin{cases} \frac{X^3}{\mu q} \\ -\frac{6X^2}{\alpha V'(X)} \end{cases} \quad \text{and} \quad h_2(X, Y) = \begin{cases} -\frac{\beta X^2}{\mu q} + \frac{2Y}{X} \\ -3X + \frac{3Y}{X} \end{cases} \quad (10)$$

for (2) and (3) respectively: similar expressions may be derived for other second order traffic models.

Our plan is now as follows. In Section 2, we analyse the geometrical structure of the fundamental diagram and its chords. In Section 3, we consider the fixed points and topology of the phase plane of (9a,9b). This assists us in building a phase diagram which describes wave types.

## 2 Geometry of the Fundamental Diagram

We now analyse the geometry of the *fundamental diagram*  $Q(\rho) := \rho V(\rho)$  which describes uniform flow solutions. We suppose that

1.  $Q$  is twice continuously differentiable with  $Q(0) = 0$ ,  $Q(\rho) > 0$  for  $\rho > 0$ .
2.  $\lim_{\rho \rightarrow \infty} Q(\rho) = Q_\infty \geq 0$  and  $\lim_{\rho \rightarrow \infty} \rho Q'(\rho) = 0$ .
3. There exists a unique turning point at  $\rho = \phi > 0$  such that  $Q'(\rho) > 0$  for  $\rho < \phi$  and  $Q'(\rho) < 0$  for  $\rho > \phi$ .
4. There exists a unique point of inflexion at  $\rho = \psi > \phi$ , such that  $Q''(\rho) < 0$  for  $\rho < \psi$  and  $Q''(\rho) > 0$  for  $\rho > \psi$ .

In particular, these assumptions ensure that chords with negative gradient intersect the graph of  $Q$  in three places, provided tangential intersections are counted as to their degree; we refer to the three intersection abscissae as  $L < C < R$ . A key question is as follows: given arbitrary  $\rho_-, \rho_+ > 0$ , is the gradient of the chord through  $(\rho_\pm, Q(\rho_\pm))$  positive or negative, and if negative, which of  $\rho_\pm$  correspond to which of  $L, C$  and  $R$ ? To assist our answer, we define the *level set* and *tangent chord* functions as follows.

1. The *level set* function  $F$  yields  $\sigma = F(\rho) \neq \rho$  such that  $Q(\sigma) = Q(\rho)$ ; that is,  $\sigma \neq \rho$  is the zero of the auxiliary function

$$f(\sigma) := Q(\sigma) - Q(\rho). \quad (11)$$

By applying the intermediate and mean value theorems to  $f$ , it may be shown that (i) for  $\rho > \rho_*$ ,  $F$  is well and uniquely defined, provided we assign  $F(\phi) = \phi$ . Here  $0 \leq \rho_* < \phi$ , and  $Q(\rho_*) = Q_\infty$ . For  $\rho < \phi$ ,  $F(\rho) > \phi$ , and vice-versa. (ii)  $F$  is self-inverse. (iii)  $F$  is strictly decreasing and in fact differentiable. (Thus from (ii),  $F'(\phi) = -1$ .) (iv)  $\lim_{\rho \downarrow \rho_*} F(\rho) = +\infty$ , and  $\lim_{\rho \uparrow +\infty} F(\rho) = \rho_*$ .

2. The *tangent chord* function  $G$  yields  $\sigma = G(\rho) \neq \rho$  such that the tangent to  $Q$  at  $(\sigma, Q(\sigma))$  intersects the fundamental diagram again at  $(\rho, Q(\rho))$ ; that is,  $\sigma \neq \rho$  is the zero of the auxiliary function

$$g(\sigma) := Q(\sigma) - Q(\rho) - (\sigma - \rho)Q'(\sigma). \tag{12}$$

Note that  $Q'(G(\rho))$  has a natural interpretation as the speed of the fastest possible wave that the system might connect to the uniform flow  $(\rho, V(\rho))$ . Again, the details of this definition may be checked with the intermediate and mean value theorems. We find that (i) for  $\rho > \rho_*$ ,  $G$  is well and uniquely defined, provided we assign  $G(\psi) = \psi$ . For  $\rho < \psi$ ,  $G(\rho) > \psi$ , and vice-versa. (ii) In contrast to  $F$ ,  $G$  is *not* self-inverse. (iii)  $G$  is strictly decreasing and in fact differentiable. (iv)  $\lim_{\rho \downarrow \rho_*} G(\rho) = +\infty$ , and  $\lim_{\rho \uparrow +\infty} G(\rho) = \phi$ .

Given up- and down- stream densities  $\rho_-, \rho_+$ , the *level set* function may be used as follows. If  $\rho_+ < F(\rho_-)$  (equivalent to  $\rho_- < F(\rho_+)$ ), then the chord through  $(\rho_\pm, Q(\rho_\pm))$  has positive a gradient (i.e., putative waves propagate downstream).

However, more interesting is  $\rho_+ > F(\rho_-)$  (equivalent to  $\rho_- > F(\rho_+)$ ) where waves propagate upstream and the chord through  $(\rho_\pm, Q(\rho_\pm))$  has a third intersection. Firstly, if  $\rho_\pm < \psi$ , then  $\min(\rho_\pm)$  corresponds to L and  $\max(\rho_\pm)$  corresponds to C. Similarly, if  $\rho_\pm > \psi$ , then  $\min(\rho_\pm)$  corresponds to C and  $\max(\rho_\pm)$  corresponds to R.

The interesting case is thus when  $\rho_\pm$  straddle  $\psi$ . If  $\min(\rho_\pm) < G(\max(\rho_\pm))$ , then  $\min(\rho_\pm)$  corresponds to L, else it corresponds to C; similarly, if  $\max(\rho_\pm) > G(\min(\rho_\pm))$ , then  $\max(\rho_\pm)$  corresponds to R, else it corresponds to C.

Thus given up- and down- stream densities  $\rho_-$  and  $\rho_+$  (no order assumed), we may use the graphs of  $F$ ,  $G$  and  $G^{-1}$  to classify the chord gradient and intersection types. In particular, the region of  $(\rho_-, \rho_+)$  space corresponding to chord gradient  $c < 0$  is split into six regions which we label LC, CL, LR, RL, CR, RC, depending on how  $\rho_-$  and  $\rho_+$  correspond to the three chord intersection abscissae  $L < C < R$ . These regions are depicted in Fig. 2.

### 3 Phase Plane Analysis

Our attention now turns to dynamics and the phase plane defined by (9a,9b). Firstly, the fixed points are clearly given by

$$Y = 0 \quad \text{and} \quad Q(X) = q + cX; \tag{13}$$

i.e., the positions correspond to chord intersection abscissae, and two of these  $\rho_-$  and  $\rho_+$  have been prescribed (and in fact define  $q$  and  $c$ ). However, as explained

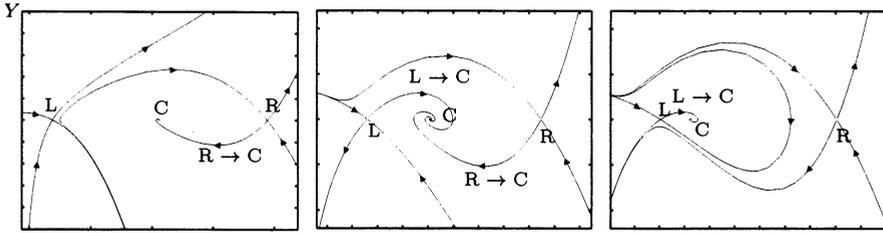


Fig. 1. Possible connection topologies for phase plane (9a,9b).

in Section 2, there are three intersections  $L < C < R$  when  $c < 0$ . We therefore classify the fixed points  $(L, 0)$ ,  $(C, 0)$ ,  $(R, 0)$  via the Jacobian of (9a,9b).

The general form of the Jacobian is quite complicated, but at fixed points  $(X, 0)$  where (13) hold, expressions simplify, and we find

$$\frac{\partial Y'}{\partial X} = h_1(X) \frac{\alpha}{X} \{Q'(X) - c\}. \tag{14}$$

The sign of  $\partial Y'/\partial X$  at fixed points is thus governed by the direction in which the chord cuts the fundamental diagram. Since  $Q'(L) > c$ ,  $Q'(C) < c$  and  $Q'(R) > c$ , and  $h_1(X) > 0$  for  $X = L, C, R$ , we find that  $(L, 0)$ ,  $(R, 0)$  are saddle points, whereas  $(C, 0)$  is either a spiral or node, whose stability is governed by the sign of  $\partial Y'/\partial Y$ .

The full expression for  $\partial Y'/\partial Y$  is also quite complicated, but at fixed points where (13) holds it simplifies to

$$\frac{\partial Y'}{\partial Y} = h_1(X) \frac{q^2}{X^3} + h_2(X, 0). \tag{15}$$

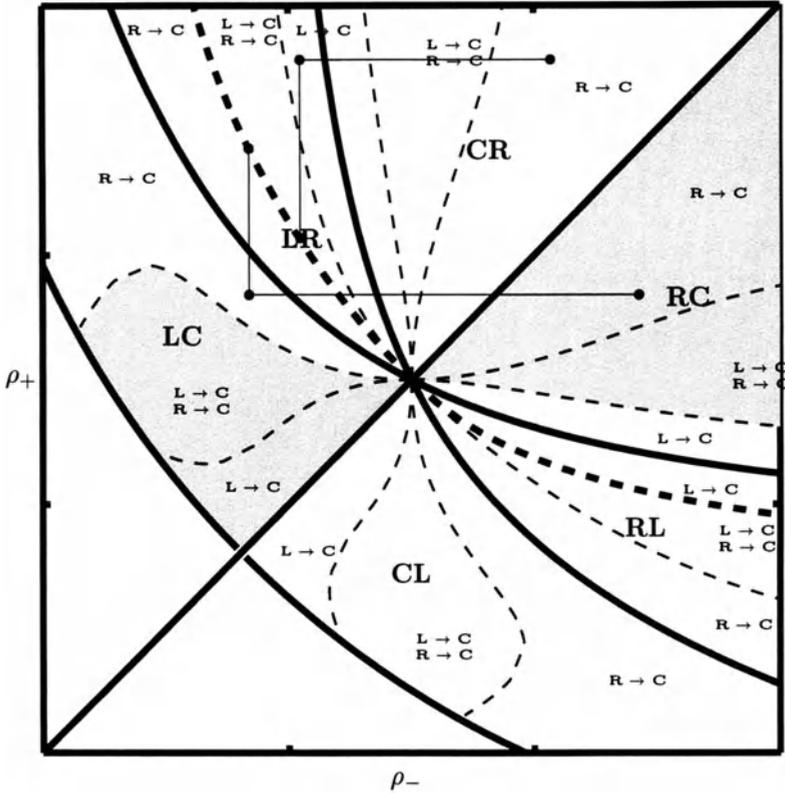
For models (2) and (3) this yields

$$\frac{\partial Y'}{\partial Y} = \frac{q^2 - \beta X^2}{\mu q}, \quad \text{and} \quad -\frac{6q^2}{\alpha X V'(X)} - 3X \tag{16}$$

respectively. In each case, it may be shown that  $\partial Y'/\partial Y < 0$  when uniform flows are stable in the underlying model, and thus that C is a stable node/spiral.

We now turn to the global connection topology of the  $(X, Y)$  phase plane. There are three generic possibilities which are illustrated in Fig. 1. These cases are separated in parameter space by the co-dimension one *gluing bifurcations*, where either the left hand saddle connects to the right (called  $L \rightarrow R$ ) or vice versa (called  $R \rightarrow L$ ).

It is possible to trace out the gluing bifurcations using numerical continuation software, add these curves to Fig. 2 (as we have done), and hence draw conclusions about when travelling waves are (im)possible. Further, when travelling waves of permanent form do not occur, the diagram can explain the compound waves found by Berg and Woods [3].



**Fig. 2.** Phase diagram  $(\rho_-, \rho_+)$  of up- and down- stream densities. Heavy solid curves relate to the geometry of the fundamental diagram, and dashed curves describe gluing bifurcations. The different connection behaviour has also been catalogued. On the heavy dashed curve  $/$  in the shaded areas, the connections are of the correct type and travelling waves of permanent form are possible. Outside of these regions, it is possible to make a connection via a pair of waves  $(\rho_-, \rho_+)$  and  $(\rho_+, \rho_+)$ , and this construction is depicted via the right-angled sets of lines. Note a separate asymptotic argument establishes expansion waves in the CR region.

Although the continuation is a numerical procedure which must be repeated for any change in the model, we have strong evidence that the results are qualitatively the same for both (2) and (3), and we have partial proofs of this claim. However, this stage of our unification theory is far from complete.

### References

1. M. Bando, K. Hasebe et al.: Phys. Rev. E **51**, 1035–1042 (1995).
2. P. Berg, A. Mason, A.W. Woods: Phys. Rev. E **61**, 1056–1066 (2000).
3. P. Berg, A.W. Woods: Phys. Rev. E **63**, art. no. 036107 (2001).
4. B.S. Kerner, P. Konhäuser: Phys. Rev. E **48** 2335–2338 (1993).

# Improved Optimal Velocity Model for Traffic

A. Shibata

Computing Research Center, High Energy Accelerator Research Organization (KEK)  
Oho 1-1, Tsukuba, Ibaraki 305-0801, Japan

**Abstract.** The OV (optimal velocity) model for traffic succeeds in dynamical generating a traffic jam and in reproducing real data qualitatively. However, it has a problem on realistic driving simulation such as over shooting. We study the improvements of the OV model so that improvements keep the profile of the model as much as possible. We also apply the model to the problem of bottleneck traffic flow such as a tunnel.

## 1 Introduction

Recently, lots of studies on traffic flow have been developed in a view point of physics such as the cellular automaton, fouled dynamics, granular flow in a pipe and so on. The car following models have a long history from 1960's and the optimal velocity (OV) model took the first success in the dynamical and spontaneous generation of a traffic jam[1]. The OV model is simple but realistic enough to reproduce a traffic jam with a few model parameters. However, the OV model has a problem for the realistic simulation of driving since it sometimes takes an unacceptable huge acceleration. There are several studies [2,6] of the extensions and improvements. Most of the models have no explicit mechanism for limiting the acceleration within the realistic region, however, the acceleration are controlled by tuning of the sensitivity parameter in simulations which also determines critical density for the congestion. We study the improvements of the OV model, so that the profile of the original OV model are kept as much as possible. We introduce the term of limiting acceleration explicitly in the model and the term of controlling by the relative velocity. We compare the profile of the improved model with the original one and also apply the model to the simulation of a bottleneck traffic problem.

This paper is organized as follows; in the next section we propose the improved OV model. In section 3, the stability of the homogeneous flow is studied. In section 4, we have numerical simulations, and in section 5 we apply the model to the bottleneck problem of tunnel. The final section gives the summary and discussion.

## 2 Improved Model

We consider the one-lane traffic model in a high-way. We introduce an improvement so that profiles of the OV model are inherited. Thus the improvements should include the parameter region which produces the original model. Let us consider the situation that the vehicle  $i$  with position  $x_i$  and velocity  $v_i$  follows the vehicle

$i + 1$ , and each vehicle obeys the following motion of the equation;

$$\frac{d}{dt}x_i = v_i, \quad (1)$$

$$\frac{d}{dt}v_i = \alpha_1 \tanh\left(\frac{\beta}{\alpha_1}\{OV(\Delta x_i) - v_i\}\right) + \alpha_2 \Delta v_i S(\Delta x_i) \quad (2)$$

$$OV(y) = \frac{V_{\max}}{1 + c_0} \left\{ \tanh\left(2\frac{y - d}{w_1}\right) + c_0 \right\}, \quad c_0 = -\tanh\left(2\frac{l_0 - d}{w_1}\right), \quad (3)$$

$$S(y) = \left[ V_{\max} \left( 1 + \exp\left(\frac{y - d_2}{w_2}\right) \right) \right]^{-1}, \quad (4)$$

where  $\Delta x_i := x_{i+1} - x_i$  and  $\Delta v_i := v_{i+1} - v_i$  are headway and relative velocity between the forwarding vehicle, respectively. We chose the parameterization of the OV function Eq.(3) according to [5];  $V_{\max}$  is the maximum velocity of the vehicle,  $d_1$  and  $w_1$  determine the shape of OV curve and  $l_0$  the minimum headway or vehicle length, respectively. We improve the model by focusing on the acceleration term in Eq.(2). We introduce the explicit limitation of the acceleration  $\alpha_1$  for the car control obeying OV function and the control term by a relative velocity. The control by a relative velocity is screened by a kind of step function  $S(\Delta x)$ , which is flat within headway  $d_2 + w_2$  and quickly dumped outside it. So the maximal acceleration is limited by  $\alpha_1 + \alpha_2$ . The parameter  $\beta$  is a sensitivity for driving, which will determine critical density to generate the congested flow from homogeneous flow. It should be noticed, that the improved model includes the original OV model itself, that is, in the limit of  $\alpha_2 \rightarrow 0$  and  $\alpha_1 \gg 1$  the original model is obtained.

### 3 Stability of Homogenous Flows

We first study the stability of the homogeneous flow. Let us consider the circuit with length  $L$  for the convenient of the analysis. We obtain the homogeneous flow solution  $\Delta x_i = L/N$  ( $= 1/\rho$ ) and  $v_i = V(L/N)$  for Eqs.(1)–(4). When we rewrite the variable  $x_i(t)$  using a small difference from homogeneous flow  $y_i(t)$ ;

$$y_i(t) = x_i(t) - \{i(L/N) + v_i t + x_0(t = 0)\},$$

we can reduce the equations of motion to that for  $y_i(t)$ ,

$$\frac{d^2}{dx^2}y_i = \beta \left( f(\rho)(y_{i+1} - y_i) - \frac{d}{dt}y_i \right) + \gamma(\rho) \left( \frac{d}{dt}y_{i+1} - \frac{d}{dt}y_i \right), \quad (5)$$

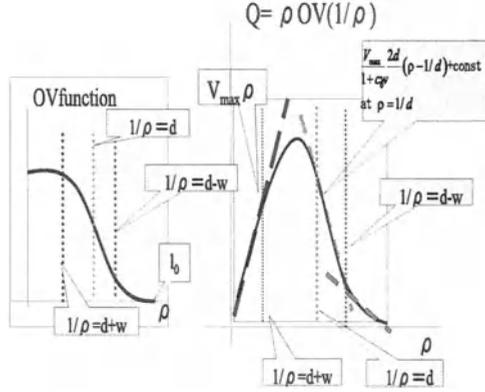
where the parameters  $f(\rho)$  and  $\gamma(\rho)$  are defined as follows;

$$f(\rho) = \frac{d}{dy}V_{opt}(1/\rho), \quad (6)$$

$$\gamma(\rho) = \frac{\alpha_2}{V_{\max}} \frac{1}{1 + \exp((1/\rho - d_2)/w_2)}. \quad (7)$$

When we put  $y_i(t) = \exp(zt + \sqrt{-1}\alpha(k)i)$  as a special solution, the equation of the motion is converted to the algebraic equation for the complex variable  $z$ :

$$z^2 + (\beta + \gamma(\rho)(1 - e^{\sqrt{-1}\alpha(k)}))z + f(\rho)(1 - e^{\sqrt{-1}\alpha(k)}) = 0, \quad (8)$$



**Fig. 1.** Relation between OV function and fundamental graph.

where the phase  $\alpha(k) = \frac{2\pi}{N}k$ , ( $k \in \mathbb{N}$ ) is determined from the boundary condition of the circuit, ( $x_N \equiv x_0$ ). The real part of  $z$  determines the stability of the homogeneous flow, i.e., if the mode is greater than zero  $\text{Re}(z) > 0$  for each  $\alpha$  the deviation from homogeneous flow diverges and if the mode is smaller than zero  $\text{Re}(z) < 0$  the deviation converges. The  $\text{Re}(z) = 0$  gives the critical curve,

$$f(\rho) = \frac{\beta + 2\gamma(\rho)}{1 + \cos \alpha} \left( 1 + \frac{\gamma(\rho)}{\beta} (1 - \cos \alpha) \right) \quad (\beta f \neq 0). \quad (9)$$

For a possible phase  $\alpha$  the density of  $\text{Re}(z) < 0$  produces the steady state flow, and that of  $\text{Re}(z) > 0$  produces an unstable state flow. Then we finally obtain the stability condition as follows:

$$f(\rho) - \gamma(\rho) \begin{cases} > \beta/2 \text{ unstable state flow} \\ = \beta/2 \text{ marginal} \\ < \beta/2 \text{ steady state flow} \end{cases} \quad (10)$$

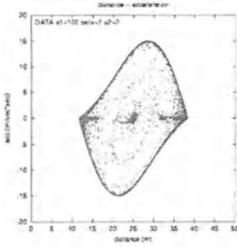
This shows that the relative velocity term relaxes the unstable state flow region of the OV model.

In the case of  $\beta f = 0$ , the Eq.(8) becomes  $z(z + \gamma(1 - e^{\sqrt{-1}\alpha})) = 0$ , and the condition  $\text{Re}(z) \leq 0$  is always satisfied. That means the car control only by a relative velocity brings steady state flow and no traffic jam grows.<sup>1</sup>

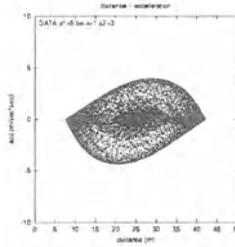
### 4 Numerical Simulations

We then move to the numerical simulation of the model. First, we discuss the relation between the fundamental graph and the OV function to determine the model parameters. In the steady state flow region, the control term by relative velocity does not contribute much, and only from the OV function the relation

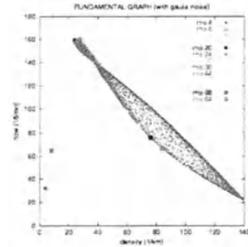
<sup>1</sup> The most of the car following models in 1960's need the time delay term since the car controls were based on the relative velocity.



**Fig. 2.** Acceleration from the OV model.



**Fig. 3.** Acceleration from the improved OV model.



**Fig. 4.** Fundamental graph from the OV model.

between distance and the velocity are determined. So the flow is obtained by  $Q = \rho V_{opt}(1/\rho)$  for the low and the high density region. In Fig.1, we show the OV function and the fundamental graph as a function of the density  $\rho$ . In the low density region vehicles run with maximum velocities and we have the flow relation  $Q = \rho V_{max}$ . For the high density region the flow goes to zero according to the OV function. In the middle density region, where the flow density reaches its maximum an unstable state flow region also appears. The point of inflection in the OV function has the maximal gradient and in the  $Q - k$  curve the gradient becomes  $-\frac{V_{max}}{1+c_0} \frac{2d}{w}$ . We use these relations for the parameter fitting.

We next discuss the constraints for the parameters in the OV function. In [8], the observed data of high-way gives the relation between the minimal distance  $s_m$  and the velocity  $v$ ;

$$s_m = 5.7 + 0.14v + 0.0022v^2. \tag{11}$$

The OV function should satisfy this constraint for the high speed region. Using the 170.6kp data of the Tomei Express Way at Aug. 1998 [10], we obtain the following parameters;  $d = 25m$ ,  $w = 40m$ ,  $V_{max} = 27m/sec^2$ ,  $\beta = 1Hz$ ,  $l_0 = 5.7m$ . For the acceleration parameter there is not enough information in the fundamental graph. But the maximum acceleration must be smaller than the acceleration caused by the friction between load and tires. We can determine a maximal one from the data of the running distance after the emergency breaking [7]. It shows that maximal acceleration when breaking is  $5m/sec^2$ . Thus we here chose  $\alpha_1 = 5m/sec^2$  and  $\alpha_2 = 3m/sec^2$ .

We now show the result of the simulations. Figure 2 shows the distance acceleration graph obtained by the original OV model ( $\alpha_1 = 100$ ,  $\alpha_2 = 0$ ), where we use the fitted parameter by the Chuou-Motorway [1,5]  $d = 25m$ ,  $w = 23.3m$ ,  $V_{max} = 33.6m/sec^2$ ,  $\beta = 2Hz$ ,  $l_0 = 5.7m$ ). This gives a huge acceleration of more than  $15m/sec^2$ . Usually, we need to chose bigger sensitivity for the simulation. On the other hand, Fig.3 shows the simulations of the improved OV model with the above parameters. This shows that the driving with small acceleration is achieved for a small sensitivity with the help of the relative velocity controlling.

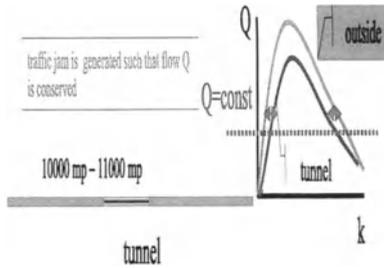


Fig. 5. Simulation of the bottleneck.

## 5 Application to Bottleneck Traffic

We now apply the improved OV model to the bottleneck traffic problem. In the simulation we introduce noise effects to the acceleration in Eq.(2) which plays the role of a trigger to generate or to ease the congested flow;

$$\frac{d}{dt}v_i = \alpha_1 \tanh \left( \frac{\beta}{\alpha_1} \{OV(\Delta x_i) - v_i\} \right) + \alpha_2 \Delta v_i S(\Delta x_i) + \eta(t), \quad (12)$$

where  $\eta(t)$  is Gaussian noise of zero average  $\langle \eta(t) \rangle = 0$  and variance  $\langle \eta(t)\eta(t') \rangle = \sigma^2 \delta(t - t')$ . In the following simulation we chose  $\sigma = 0.2 (m/s^2)$ . Fig.4 shows the result of the simulation with a random noise Eq.(12) in the circuit. The noise does not change the profile of the fundamental graphs but lead to wider distributions according to  $\sigma$ .

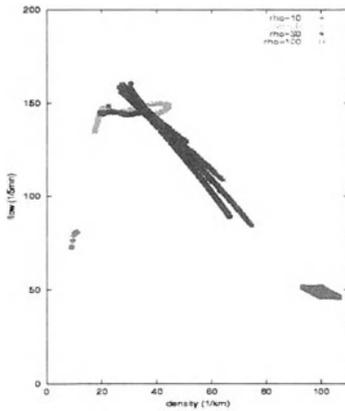
Next we pick up the bottleneck by a "tunnel". We consider the following situation. Setting up a load as an open boundary and the tunnel is at  $10000 - 11000mp$  (see Fig. 5). Though there will be several effects from the tunnel. We will only introduce the maximum velocity limitation as a bottleneck effect,

$$V_{\max}^{(\text{tunnel})} = (1 - c)V_{\max} .$$

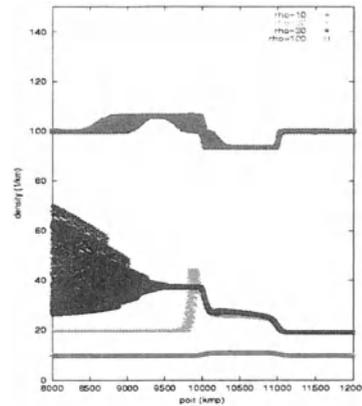
In the following simulation we set  $c = 0.2$ . The vehicles are injected at  $0mp$  with several densities and velocities which satisfy the OV function. As in the initial state the vehicles are set, with the same density for all regions, with their optimal velocity. We observe the velocity and density (headway) within  $8000 - 12000mp$ . Fig.6 shows the fundamental graph of the simulation for several injections with the density  $\rho(1/km)$  given by 10, 20, 30, 100. The fundamental graph shows that the flow is conserved between inside and outside of the tunnel for the steady flow region. We summarize that the traffic jam in Fig.7 (density–position graph) shows four types of the traffic flow according to the injection density; the free flow, the finite length of the traffic jam, the back propagating congested flow and the heavy congested homogeneous flow.

## 6 Conclusion and Discussion

We proposed an improved OV model for traffic. The OV model was successfully improved concerning huge acceleration without changing the fundamental profile.



**Fig. 6.** Fundamental graph.



**Fig. 7.** Position–density relation.

Through the analysis of homogeneous flow, the car control only with the relative velocity never produce congestion. The OV model is one of the minimal set to produce the dynamical generation of congestions. The relative velocity term relaxes the critical points of generation of congestion and brings soft acceleration and breaking. In the simulation of the bottleneck the traffic jams around the tunnel are reproduced, however, we need more detailed data of acceleration and back propagation of jams for further studies.

## References

1. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev.* **E51**, 1035 (1995); M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Jpn. J. Ind. Appl. Math.* **11**, 203 (1994); M. Bando, K. Hasebe, K. Nakanishi, A. Nakayama, A. Shibata, and Y. Sugiyama, *J. Physics I (France)* **5**, 1389 (1995).
2. Y. Igarashi, K. Itoh, and K. Nakanishi, pat-sol/9810007.
3. M. Kikuchi, Y. Sugiyama, S. Tadaki, and S. Yukawa "Asymmetric optimal velocity model for traffic flow", talk at the 5th International Conference on Computing Physics (ICCP5) at Kanazawa (1999); talk at 6th symposium on simulation of traffic flow.
4. D. Helbing and B. Tilch, *Phys. Rev.* **E58**, 133 (1998).
5. S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Jpn.* **67** (1998), no.7; S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Jpn.* **68** (1999), no.9.
6. S. Sawada, "Generalized Optimal velocity Model for traffic Flow", e-print nlin.PS/0105054.
7. Text book of drivers, (in Japanese).
8. K. Tanaka, K. Watanabe, and Sigemitsu, "On observation of distance of automobile traffic", Technical data of transportation, general 9 (1963), (in Japanese).
9. text book on "Transportation technology" by Shogo Kawakami and Hiroshi Matsui, Morikami publish, ISBN4-627-48350-3.
10. K. Nishinari and N. Hayashi, "Traffic Statistics in Tomei Express Way", Mathematical Society of Traffic Flow (Printed in Japanese) (1999).

# Microscopic Modeling of Synchronized Traffic

W. Knospe<sup>1</sup>, L. Santen<sup>2</sup>, A. Schadschneider<sup>3</sup>, and M. Schreckenberg<sup>1</sup>

<sup>1</sup> Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg,  
Lotharstr. 1, 47048 Duisburg, Germany

<sup>2</sup> Theoretical Physics, Saarland University, PO Box 151150, 66041 Saarbrücken,  
Germany

<sup>3</sup> Institute for Theoretical Physics, University of Cologne, Zùlpicher Str. 77, 50937  
Cologne, Germany

**Abstract.** Based on a recently proposed cellular automaton model for traffic flow essential properties of synchronized traffic like the synchronization of the lanes, the pinning of a synchronized region at an onramp and the propagation of a wide jam through synchronized traffic are analyzed. The used model is able to reproduce the three phases (free flow, synchronized, and stop-and-go) observed in real traffic even on a single-lane road and shows a good agreement with detailed empirical single-vehicle data in all phases.

## 1 Introduction

Empirical observations of highway traffic [1] have revealed the existence of three traffic phases, namely (i) free flow, (ii) synchronized traffic and (iii) wide moving jams<sup>1</sup>. The characteristics of free flow and wide moving jams are intuitively clear. While free flow traffic is characterized by a high average velocity of the vehicles, e.g., close to the eventually applied speed limit, wide moving traffic jams are upstream moving structures consisting of two fronts separated by a region of negligible velocity and flow.

Our understanding of synchronized traffic is far behind that of free flow and wide jams. In synchronized traffic, which is mainly observed at on- and offramps, the velocity is considerably smaller than in free flow but the flow can still have large values. Moreover, a large variance of density and flow measurements leads to an irregular pattern in the fundamental diagram which can be characterized by vanishing cross-correlations [2]. If synchronized traffic is generated at a bottleneck its downstream front is pinned there. In contrast, the downstream front of a wide moving jam can propagate undisturbed in upstream direction with a velocity of about 15 km/h through either free flow and synchronized traffic without disturbing these states. The propagation velocity is only determined by the density inside the jam and the outflow from the jam [3]. The name giving effect of synchronized traffic is that measurements of the velocity of neighboring lanes show strong correlations. However, empirical observations reveal that this is the case even for wide moving jams [4,5].

---

<sup>1</sup> Synchronized traffic and wide moving jams are summarized as congested traffic.

By means of single-vehicle data [2,6] it is possible to analyze the microscopic driving behavior of the vehicles. In free flow, platoons of vehicles driving bumper-to-bumper with a large velocity and a time-headway comparable to the reaction time can be found. Within these platoons the vehicles are driving with nearly the same velocity leading to a large stability of these clusters. In synchronized traffic these clusters are dissolved due to the increased traffic density. Nevertheless, large correlations of the velocity of consecutive vehicles can be found while the distance-headways are not correlated – the vehicles are moving in a synchronous manner. As a result, the velocity-headway curve saturates for even small headways at a small asymptotic velocity and depends not only on the density but also on the traffic state. In a wide jam, the retarded acceleration behavior of the vehicles at rest (a vehicle needs about 2 s to accelerate) results in a considerably reduced outflow from a jam compared with the maximum possible flow.

Here, we present a recently introduced generalization [7,8] of the Nagel and Schreckenberg cellular automaton model for traffic flow [9] that is able to reproduce the three empirically observed traffic states even on a microscopic level. Our previous analyses [7,8] were only performed on an one-lane road with periodic boundary conditions. A comparison with empirical data, however, often requires a more realistic simulation setup. By introducing disorder in a two-lane system, it is shown that the lanes synchronize and a wide jam propagates through synchronized traffic according to the empirics. Thus, the three traffic states can clearly be identified and are recovered even in more complex simulation topologies.

## 2 Modeling Approach

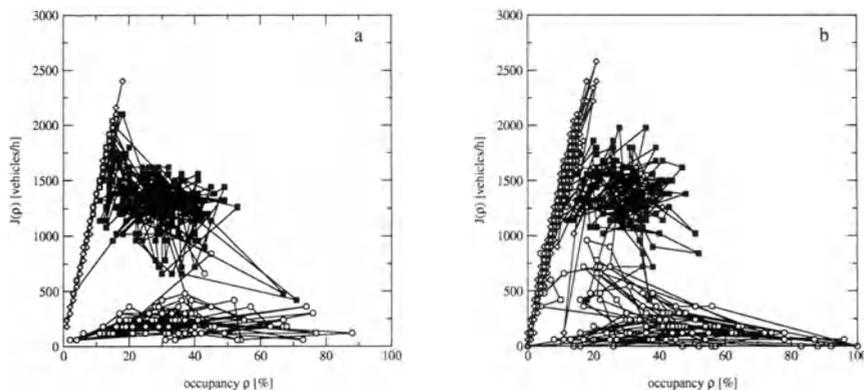
The empirical observations of single-vehicle data can be summarized in the following driving strategy: In free flow the anticipation of the predecessors movement reduces velocity fluctuations and allows driving bumper-to-bumper. In synchronized traffic, large headways are a result of the retarded acceleration of the vehicles and allow a timely adjustment of the vehicle's velocity. Furthermore, vehicles at rest need about 2 s to escape from a jam leading to a reduced outflow from a jam.

This desire for smooth and comfortable driving, e.g., the desire to avoid strong accelerations and abrupt brakings, is implemented in a cellular automaton model based on the model of Nagel and Schreckenberg [9] by considering anticipation effects. Anticipation of the predecessors movement allows the reduction of the headway at a given velocity while brake lights visualize acceleration changes. This event-driven anticipation allows a timely adjustment to the velocity of downstream vehicles and increases the distance between the vehicles (see [8] for a detailed description of the model).

The simulations were performed on a two-lane highway with open boundary conditions. Vehicles were inserted at the left boundary. For the sake of simplicity we used only one type of vehicles which leads to a smaller variance of the data points in the free flow regime compared to empirical data. Analogously to the empirical setup the simulation data are evaluated by a virtual induction loop,

i.e., we counted the number and measured the speed of vehicles passing a given link of the lattice. This allows to calculate minute averages of the velocity, the flow and the density which is given by the occupancy of the detector.

The simulations show that we can recover the fundamental diagram found in empirical observations (Fig. 1). The flow and the density measurements of the synchronized regime cover a two-dimensional region in the fundamental diagram. Moreover, wide jams can be identified by a triangular shape which is a consequence of the event-driven measurement process of the inductive loops. On one hand, small gaps inside a wide jam lead to small values of the density and the flow. On the other hand, standing vehicles can block the detector for a certain time which results in large densities but small flows.



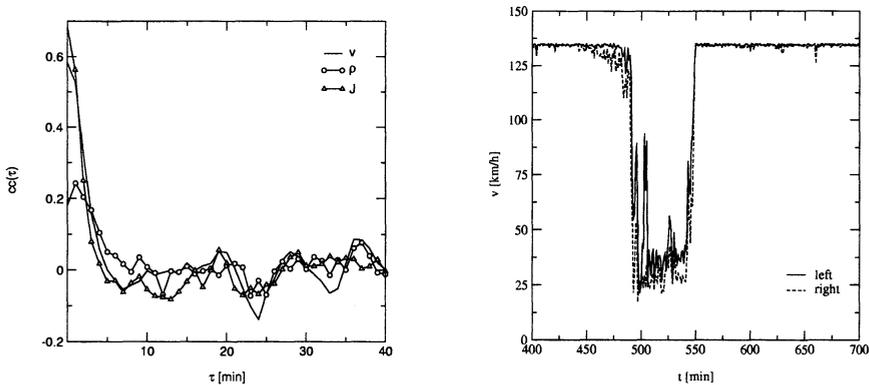
**Fig. 1.** Comparison between empirical results (b) for the fundamental diagram, and simulation results (a). Each data point corresponds to an average over an one-minute interval. Consecutive measurements are connected by lines. Part (a) shows that we recover the three empirically observed phases of highway traffic: free flow (diamonds), synchronized traffic (squares) and wide jams (circles). The empirical data are one-minute averages of detector data from the German freeway A40 near Moers junction at 2000-12-12 (synchronized state) and near Bochum-Werne junction at 2001-02-14 (wide moving jam of about 2 hours duration). The density is given by the occupancy which gives the percentage of the measurement time a detector is covered by cars.

Empirical observations suggest a strong coupling of the lanes in the synchronized state [2]. In order to quantify the interaction of the lanes, we calculated the cross-correlation  $cc(\tau)$  between the time-series of measurements of the density, the flow and the velocity on different lanes:

$$cc(x_i, x_j) = \frac{\langle x_i(t)x_j(t+\tau) \rangle - \langle x_i(t) \rangle \langle x_j(t+\tau) \rangle}{\sqrt{\langle x_i^2(t) \rangle - \langle x_i(t) \rangle^2} \sqrt{\langle x_j^2(t+\tau) \rangle - \langle x_j(t+\tau) \rangle^2}}. \quad (1)$$

Here  $x_i$  denotes either the flow, the density or the velocity in lane  $i$ .

In the free flow state the weak interaction of the lanes results in a small correlation of all quantities. Since the flow is mainly controlled by density fluctuations the density and the flow measurements show the same correlations. In contrast, at larger densities the velocity and the flow on both lanes is strongly correlated, i.e., indicating that the system is in the synchronized regime. However, due to the large variance of the density measurements in the synchronized regime the density on both lanes is not correlated (Fig. 2).



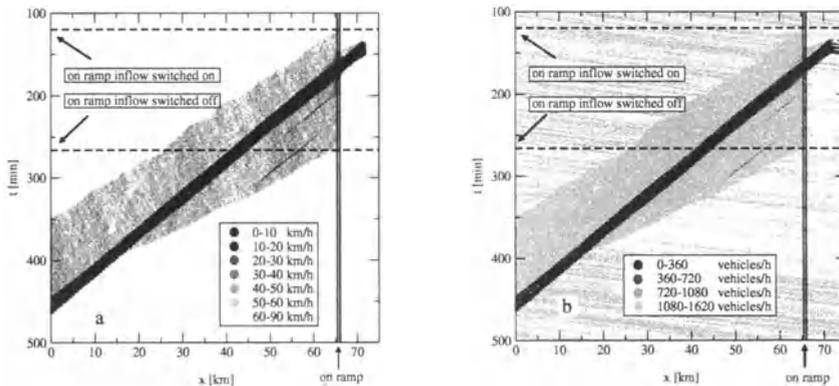
**Fig. 2.** Left: Cross-correlation of the local measured flow, density and velocity of neighboring lanes in the synchronized state ( $\rho = 50 \text{ veh/km}$ ). Right: Time-series of the velocity. A large flow from the onramp induces a jam on both lanes.

Lane changes are responsible for the synchronization of the lanes velocity [1,10]. In case of a velocity gradient between the lanes, vehicles are changing to the slower lanes thus decreasing the velocity difference. This leads to a strong coupling of the lanes in the congested regime. In particular, in the vicinity of on- and offramps cars enter or leave only the right lane, but both lanes are disturbed at large in- or outflows.

In order to verify this, we generated a jam on the right lane by a large input of an onramp. As a result, the jam spreads on the left lane, so that the time-series of the velocity measurements of both, the left and the right lane are highly correlated. Moreover, the traffic breakdown occurs at the same time (Fig. 2) leading to a synchronization of the lanes.

A large input rate of the onramp in combination with a large flow on the highway segment generates a dense traffic region whose downstream front is pinned at the onramp. The region shows a flow comparable to free flow but a velocity considerable below the free flow velocity and can be identified as synchronized traffic.

The stability of the synchronized region allows the propagation of a wide jam. Therefore, a jam was generated by an obstacle at the downstream end of the highway section. Figure 3 demonstrates the coexistence of the traffic states and illustrates how the jam wave travels through the free flow region with constant velocity of about 13.5 km/h and also passes the section where the synchronized traffic is localized. Thus, the model is able to reproduce the three traffic phases since the downstream front of synchronized traffic is fixed at the onramp and the jam propagates with a constant velocity upstreams and passes the free flow, onramp and synchronized regions without being disturbed [1].



**Fig. 3.** Coexistence of wide moving jams and synchronized states. Space/time evolution of the velocity (a) and of the flow (b). Downstream the onramp a jam has been generated which moves in upstream direction and passes the segment with free flow and synchronized states. One clearly observes that the synchronized state is recovered directly after the jam has passed the onramp.

### 3 Summary and Discussion

The model proposed in [8] is able to reproduce the three empirically observed traffic states not only in an artificial simulation setup but also in the context of more realistic and complex highway topologies. The undisturbed propagation of a wide jam through synchronized traffic and free flow demonstrates the empirically observed coexistence of the traffic states and underscores the stability of the synchronized state. In [11] this stability is traced back to the human desire of smooth and comfortable driving which reduces the impact of velocity fluctuations.

Moreover, it is shown that the synchronization of the velocity on different lanes is a consequence of the lane changes [1,10]. Not only the negative velocity

gradient between neighboring lanes is reduced, but also information on the traffic state can spread to all lanes. As a result, a jam generated on only one lane exists on both lanes.

From a theoretical point of view, our results shed some light on the mechanisms that are responsible for the formation of synchronized traffic. Moreover, it turns out, that the role of the boundaries is restricted to a selection rather than a to generation of the various traffic states, which are equally well observed in periodic systems. From a practical point of view, the degree of realism of the simulation allows a more detailed description of highway traffic and opens the door for a forecast of highway traffic [12].

**Acknowledgments.** The authors are grateful to the Landesbetrieb Straßenbau NRW for the data support and to the Ministry of Economics and Mid-size Businesses, Technology and Transport for the financial support. L. S. acknowledges support from the Deutsche Forschungsgemeinschaft under Grant No. SA864/1-2.

## References

1. B.S. Kerner, *Networks and Spatial Economics* **1**, 35 (2001).
2. L. Neubert, L. Santen, A. Schadschneider, and M. Schreckenberg, *Phys. Rev. E* **60**, 6480 (1999).
3. B.S. Kerner, *Phys. Rev. Lett.* **81**, 3797 (1998).
4. B.S. Kerner, *Phys. Rev. E* **53**, R4275 (1996).
5. M. Koshi, M. Iwasaki, and I. Ohkura, in: *Proceedings of the 8th International Symposium on Transportation and Traffic Theory*, V.F. Hurdle, E. Hauer, and G.N. Stewart, (Eds.), p. 403 (University of Toronto Press, Toronto, Ontario, 1983).
6. B. Tilch and D. Helbing, in: *Traffic and Granular Flow '99*, D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf, (Eds.), p. 333 (Springer, Heidelberg, 2000).
7. W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, in: *Traffic and Granular Flow '99*, D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf, (Eds.), p. 431 (Springer, Heidelberg, 2000).
8. W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, *J. Phys. A* **33**, L477 (2000).
9. K. Nagel and M. Schreckenberg, *J. Phys. I (France)* **2**, 2221 (1992).
10. B.S. Kerner, *Physics World*, **8**, 25 (1999).
11. W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, *Phys. Rev. E*, (2001), in press.
12. O. Kaumann, K. Froese, R. Chrobok, J. Wahle, L. Neubert, and M. Schreckenberg, in: *Traffic and Granular Flow '99*, D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf, (Eds.), pp. 351-356 (Springer, 2000).

# Limit Sets and the Rate of Convergence for One-Dimensional Cellular Automata Traffic Models

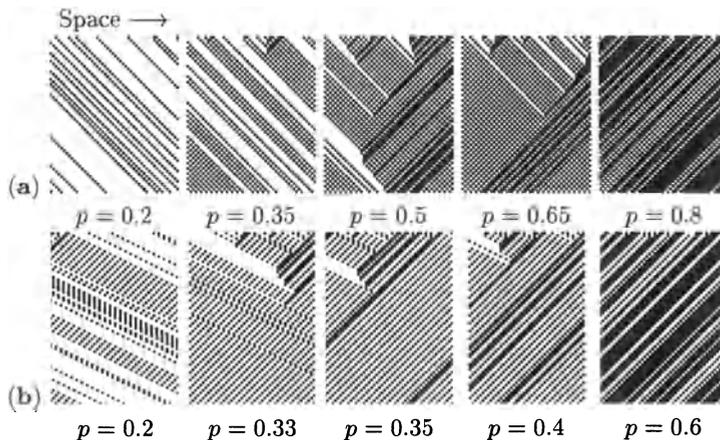
T. Namiki

Division of Mathematics, Hokkaido University, Sapporo 060-0810, Japan

**Abstract.** A critical exponent of phase transition for one-dimensional cellular automata with conservative law is proposed and it is shown that the exponent represents the asymptotic behavior of the dynamics.

## 1 Introduction

For one-dimensional cellular automata with conservative law, ECA184 [9] and so on, there exists many rules exhibit phase transition in its dynamics according to the initial distribution shown in Fig. 1.



**Fig. 1.** Typical orbits of traffic models. Initial distribution of “1” is denoted by  $p$ . (a) ECA184. (b) EBCA1.

The case of  $p = 1/2$ , the “speed” of convergence to limit set becomes significantly slow comparing to the case  $p \neq 1/2$ . Numerical studies for Fukui-Ishibashi model are in [3] and they observe that for  $p = p_c$  the relaxation time diverges. We estimate such phenomena, that is the asymptotic behavior of one-dimensional cellular automata traffic model from initial distribution to the limit distribution with respect to Gibbs measure [1] on the limit set.

In [4] the equivalent condition that cellular automata have additive conserved quantity. Many cases of traffic models, shown in [2,7], the cellular automata satisfy its condition. With the condition we show the existence and the vanishing case of exponential rate in asymptotics.

## 2 Preliminary

**Definition 1.** Let  $S$  be a finite set and  $\Lambda$  be a finite subset of  $\mathbb{Z}$ . For a given map  $f : S^\Lambda \rightarrow S$ , we can define a map  $\tau : S^\mathbb{Z} \rightarrow S^\mathbb{Z}$  for all  $x = (x_i)_{i \in \mathbb{Z}} \in S^\mathbb{Z}$  by

$$(\tau x)_i = f(x_{i+\lambda}; \lambda \in \Lambda).$$

We call the dynamical system  $(S^\mathbb{Z}, \tau)$  one-dimensional cellular automata with rule  $f$ .

It is clear that  $\tau$  is shift commute continuous map with respect to the product topology. After here we denote the shift map by  $\sigma$ .

For one-dimensional cellular automata  $\tau : S^\mathbb{Z} \mapsto S^\mathbb{Z}$ , the limit set  $X_\infty = \bigcap_{n \geq 0} \tau^n X_0$ ,  $X_0 = S^\mathbb{Z}$  is always well defined. According to the convention in symbolic dynamics, we call  $w \in S^* := \bigcup_{n \in \mathbb{N}} S^n$  a word and  $|w|$  means the length of word  $w$ . For a countable word set  $W$ ,  $\langle W \rangle$  means the shift invariant closed set the words  $w \in W$  generate by their concatenations. The shift map is denoted by  $\sigma$ .

Now we define the additive conservative quantity for cellular automata. Take a real-valued function  $U$  on  $S^\mathbb{Z}$  depends only on finite number of indices, i.e.  $U : S^m \rightarrow \mathbb{R}$ ,  $U(x) = U(x_i, \dots, x_{i+m-1})$ .

**Definition 2.** We call  $U$  be additive conserved quantity for  $\tau$  if  $\rho_n(x) = \frac{1}{n} S_n U(x) := \frac{1}{n} \sum_{i=0}^{n-1} U(\sigma^i x)$  is invariant under  $\tau$  for all  $x \in \text{Fix}(S^\mathbb{Z}, \sigma^n)$ , where  $\text{Fix}(S^\mathbb{Z}, \sigma^n)$  is a set of all  $n$ -periodic configurations that is fixed under  $n$ -th iteration of  $\sigma$ . If  $\tau$  has such  $U$ , we call  $\tau$  is conservative.

*Remark 1.* For an additive conserved  $U$ , there exists finite range function  $q$  and a constant  $C$  such that

$$U(\tau x) = U(x) + q(x) - q(\sigma x) + C$$

for all  $x \in S^\mathbb{Z}$ . This fact is in [1,4]. From physical viewpoint as in [4,7]  $q$  corresponds to the flow.

We denote the average of flow as  $Q_n(x) := \frac{1}{n} S_n q(x) = \frac{1}{n} \sum_{i=0}^{n-1} q(\sigma^i x)$  for all  $x \in \text{Fix}(S^\mathbb{Z}, \sigma^n)$ . From ergodic theorems, there exist the limit  $\rho := \lim_{n \rightarrow \infty} \rho_n(x) = \int_{S^\mathbb{Z}} U d\mu$  and  $Q := \lim_{n \rightarrow \infty} Q_n(x) = \int_{S^\mathbb{Z}} q d\mu$  for almost every  $x$  with respect to any shift invariant probability ergodic measure  $\mu$  on  $S^\mathbb{Z}$ .

The relation with  $\rho$  and  $Q$  on the limit set is called fundamental diagram like Fig. 2. From numerical experiments the critical density  $p_c$  corresponds

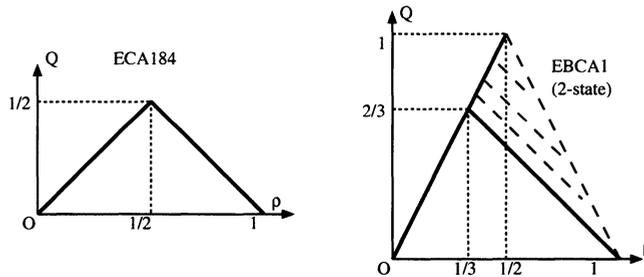


Fig. 2. Fundamental diagram for ECA184 and EBCA1 [6,7].

to the critical points of the Fig., i.e.  $\rho = 1/2$  in ECA184 and  $\rho = 1/3$  in EBCA1.

Though it is well known that the conservative case contains rich and various phenomena, we concentrate the study to the phenomena of “phase transition” related to initial distribution described in Section 1.

### 3 Main Result

After here we assume  $S = \{1, 0\}$ . At first we show the main result on the case of ECA184. We can see the essential situation here.

#### 3.1 ECA184

The rule  $f$  of ECA184 is the following:

$$A = \{-1, 0, 1\} \text{ and } f(x_{-1}x_0x_1) = \begin{cases} 1 & \text{if } x_{-1}x_0x_1 = 111, 101, 100, 011 \\ 0 & \text{elsewhere.} \end{cases}$$

In this case,  $X_0 = \{0, 1\}^{\mathbb{Z}}$ ,  $X_{\infty} = \langle 0, 10 \rangle \cup \langle 1, 10 \rangle = X_- \cup X_+$ . Set  $x_c = (x_c)_i \in \mathbb{Z}$ , if  $i$  is even  $(x_c)_i = 1$  and otherwise  $(x_c)_i = 0$ . Note that  $X_- \cap X_+ = \{x_c, \sigma x_c\}$ .

This rule can be rewritten using boolean algebra as follows:

$$f(x_{-1}x_0x_1) = x_0 + (x_{-1}(x_1 + 1)) - (x_0(x_1 + 1)).$$

So the function  $U(x) = x_0$  is a additive conservative quantity and the flow is  $q(x_{-1}x_0) = x_{-1}(x_1 + 1)$ .

Let  $\mu^p$  be a Bernoulli measure on  $(X_0, \sigma)$  with respect to a probability vector  $p = (1 - p, p)$ ,  $\nu$  be a simple Markov measure on  $(X_-, \sigma)$  or  $(X_+, \sigma)$ , and  $a = (1 - a, a)$  be a left invariant probability vector for transition matrix of  $\nu$ . Note that  $\nu$  is a special case of Gibbs measure.

For  $x \in X_{\infty}$  and set  $S_n(x) = x_{-n} + \dots + x_{n-1}$ , we get the following result about the rate of convergence from  $X_0$  to  $X_{\infty}$ :

**Proposition 1.**

$$R(a, p, \nu) := \lim_{n \rightarrow \infty} -\frac{1}{n} \log \sum_{\frac{1}{2^n} S_n(x) = a} \mu^p(\tau^{-n}[x_{-n} \cdots x_{n-1}]) \nu[x_{-n} \cdots x_{n-1}]$$

$$= H(a|p) + h_\nu(X_\infty, \sigma),$$

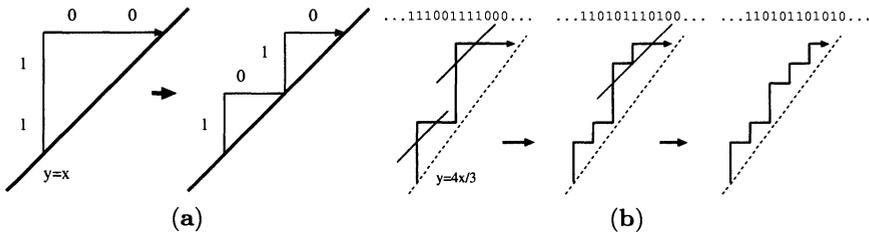
where  $[x_{-n} \cdots x_{n-1}] = \{y \in X_\infty; y_i = x_i - n \leq i \leq n-1\}$  is called cylinder set,  $H(a|p)$  is a relative entropy of two probability vectors and  $h_\nu(X_\infty, \sigma)$  is the metrical entropy of  $(X_\infty, \sigma, \nu)$ . For  $\nu$  is supported on  $X_+$  or  $X_-$ ,  $h_\nu(X_\infty, \sigma)$  is well defined.

*Remark 2.* The result shows that the rate is vanished if and only if  $p = a = (1/2, 1/2)$ ,  $\nu = \frac{1}{2}(\delta_{x_c} + \delta_{\sigma x_c})$  and it means that there are two cases,  $R = 0$  and  $R > 0$ .

*Remark 3.* In [3] of Fig. 4 they study the mean convergence time from random initial configurations to stable states. From numerical experiment if  $p \neq p_c$  the value remains finite and if  $p = p_c$  it shows the divergence phenomena. Proposition 1 explains these numerical result.

If  $R > 0$  we have exponential order of convergence, as appear in the fast dynamics of orbits. On the contrary, the case of  $R = 0$  we have sub-exponential order, really polynomial that is appear in slow dynamics of orbits. This shows the existence of phase transition on  $p = p_c$ .

*Proof (sketch).* The essential idea of the proof is random walk representation of configurations.



**Fig. 3.** ECA184: (a) The essential part of the rule in path space. (b) Its typical dynamics.

Let  $W_a = \{w \in S^{2n}; \frac{1}{n} S_n(w^\infty) = a\}$ . We count the initial configuration of length  $4n$  which converges to  $W_a$  within  $n$ -steps. It is easy to see that the set of path of simple random walk never touch the line  $y = ax$  of length  $4n$  and return to the line at the last step is essential to the term.

For  $S = \{0, 1\}$  and rational  $a$ , it is clear with certain amount of efforts to count up the number as in [5]

$$C_n \frac{2n!}{(2an)!(2(1-a)n)!} p^{2an} (1-p)^{2(1-a)n},$$

where  $\lim_{n \rightarrow \infty} \log C_n/n = 0$ . Using Stirling's formula, we have

$$e^{2an \log a + (2(1-a)n) \log(1-a)} e^{-2an \log p - 2(1-a)n \log(1-p)}$$

From this we get the relative entropy term. The term  $h_\nu$  comes from the uniformness of  $\nu$ .

We complete for rational  $a$ . For irrational  $a$ , take the sequence  $\{a_i\}$  which converges to  $a$ . By the continuity of  $H(a|p)$  with respect to  $a$ ,  $H(a_i|p)$  also converges to  $H(a|p)$  and  $\nu$  is well defined whether  $a$  is rational or not. So we have the proposition.

*Remark 4.* More precisely we use harmonic analysis — generating function method. We get  $e^{-H(a|p)}$  as the convergence radius of generating function for first return time to the given line.

### 3.2 EBCA1

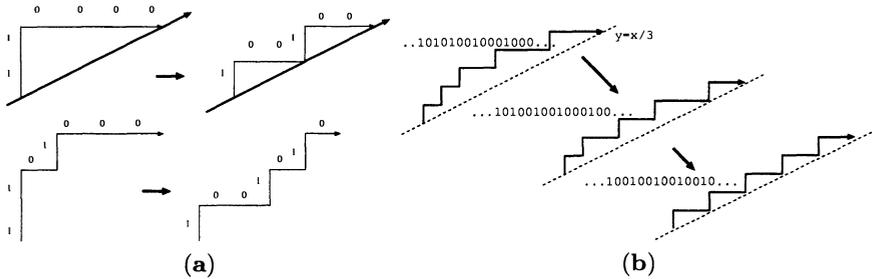
EBCA1 is a traffic model studied in [6,7]. The rule is described as following:

$$x_i^{t+1} = x_i^t + \min(b_{i-1}^t + b_{i-2}^t, 1 - x_i^t + b_i^t) - \min(b_i^t + b_{i-1}^t, 1 - x_{i+1}^t + b_{i+1}^t)$$

where  $b_i = \min(x_i^t, 1 - x_{i+1}^t)$ .

As ECA184, the limit set of EBCA1 has decomposition  $X = X_- \cup X_+$  and essentially  $X = X_- \cap X_+ = \{(100)^\infty\}$  where  $X_- = \langle 0, 10 \rangle$  and  $X_+ = \langle 1, 10, 100 \rangle$ .

Its rule in path space and the dynamics is shown in Fig. 4.



**Fig. 4.** EBCA1: (a) The essential part of the rule in path space. That is more complicated with ECA184. (b) Its typical dynamics.

The problem for EBCA1 is existence of the metastable state as [6,7]. If the preimages of metastable state in  $X_0$  has the positive measure with respect to  $\mu^p$ , the assumption of our method is broken. Fortunately, we can show that the measure of metastable state is 0, so our method is still valid.

### 3.3 Generalization

We can generalize this more typical traffic models on  $S = \{0, 1\}$  because our proof does not depend on the detail of its rule.

Our proof depends on the fact that the case of  $p = p_c$  the limit set consists of finite number of configurations. That is the case that the intersection  $X_+ \cap X_-$  has only finite number of points.

## 4 Discussion

The exponent  $R$  discussed the paper seems to describe well the phase transition of dynamics in one-dimensional cellular automata traffic models. This is the result not only for the “stable state” of the dynamics but for the “asymptoticity” of the dynamics. We think these point of view is important from the application to the real traffic or flow.

For our results, the essential part for a given rule is to determine the structure of attractor or stable states  $X_\infty$ . Though it is difficult to set the general algorithms, there exists a class having “Lyapunov function  $Q(x)$ ”.

In [7] they proved that  $Q(x)$  increases monotonously. The behavior corresponds to the “Lyapunov function” for the theory of ordinary differentiable equation. If we get the fact, the set  $\{w ; q(w) = 1\}$  is a essential word set for attractor in  $S = \{0, 1\}$ . Though this point is discussed in [10] for steady state, the case of congestive state is remained yet.

## References

1. R. Bowen, *Equilibrium states and the ergodic theory of Anosov Diffeomorphisms*, (Springer, New York 1974).
2. H. Fuk s, Phys. Rev. E **60**, 197–202 (1999).
3. M. Fukui and Y. Ishibashi, J. Phys. Soc. Japan **65**, 6 June, 1868–1870 (1996).
4. T. Hattori and S. Takesue, Physica D **49**, 295–322 (1991).
5. SRI G. Mohanty, *Lattice Path Counting and Applications*, (Academic Press, New York 1979).
6. K. Nishinari and D. Takahashi, J. Phys. A: Math. Gen. **32**, 93–104 (1999).
7. K. Nishinari and D. Takahashi, J. Phys. A: Math. Gen. **33**, 7709–7720 (2000).
8. D.A.Lind and B.Marcus, *An introduction to Symbolic Dynamics and Coding*, (Cambridge University Press 1995).
9. S. Wolfram, *Theory and Applications of Cellular Automata*, (World Scientific, Singapore 1986).
10. M. Blank, Markov Processes and Related Fields, 6:3, 287–304 (2000).

# Localized Defects in a Cellular Automaton Model for Traffic Flow with Phase Separation

A. Pottmeier<sup>1</sup>, R. Barlovic<sup>1</sup>, W. Knosp<sup>1</sup>, A. Schadschneider<sup>2</sup>, and M. Schreckenberg<sup>1</sup>

<sup>1</sup> Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany

<sup>2</sup> Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany

**Abstract.** The impact of a localized defect in a cellular automaton model for traffic flow which exhibits metastable states and phase separation is studied. The defect is implemented by locally increasing the deceleration probability. Depending on the magnitude three phases can be identified in the system. One of them shows the characteristics of stop-and-go traffic which can not be found in the model without lattice defect. From a physical point of view the model describes the competition between two mechanisms of phase separation.

## 1 Introduction

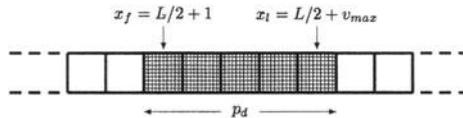
Nowadays mobility, mostly realized by means of vehicular traffic, is an integral part of any modern society. Unfortunately, the capacity of the existing road networks is often exceeded in densely populated areas. In many cases it seems as if the occurrence of traffic states with high density like jammed or synchronized traffic can be linked to external influences, e.g., on- and off-ramps, bottlenecks, lane reductions or road works (see [1–6]). Such local perturbations can act as a seed for the formation of jammed traffic states.

In recent years various theoretical approaches for the description of traffic flow have been suggested (for reviews, see e.g. [7,8]). Here we focus our investigations on the Nagel-Schreckenberg (NaSch) model [9], a microscopic cellular automata (CA) approach that is capable to perform real-time simulations of large networks with access to individual vehicles. As an enhancement we take velocity-dependent randomization into account (VDR model) [10] leading to metastable states and phase separation into free flow and wide jams. Due to the fact that only wide jams appear in this model, any additional high density patterns induced by external forces (defects) can be identified easily. The impact of defects in the NaSch model (and related models, e.g., the asymmetric simple exclusion process) is by now well understood. Basically two types of defects can be distinguished which can be characterized as particlewise and sitewise disorder. In both cases a parameter regime exists where the global behavior of the system is controlled by the defect which acts as a bottleneck. Generically it induces phase separation into a high and a low density region separated by a sharp discontinuity (“shock”). This behavior has been found in a variety of models and for different defect realizations. However, none of the models investigated so far

exhibits phase separation through the existence of high-flow metastable states which is an important ingredient for any realistic traffic model.

## 2 Definition of the Model

In the spirit of modeling complex phenomena in statistical physics one is interested in keeping the model as simple as possible. Obviously, the modeling of traffic flow in the language of cellular automata (CA) always implies an extreme simplification of real world conditions. Hence space, speed, acceleration, and even time are treated as discrete variables and the motion is realized in the ideal case by a minimal set of local rules. In this manner we tried to find a straightforward representation for local defects on a one-lane street in the VDR model [11]. Since the model contains a stochastic parameter which is needed to implement various phenomena found in real traffic, e.g., spontaneous jam formation and reduced outflow from a jam, it seems obvious to implement a local defect by increasing this parameter in a limited area. The length of the defect is chosen to  $L_d$  cells and the stochastic noise  $p(v)$ , which is part of the update rules of the VDR model, is replaced by a defect noise  $p_d$ . The defect length  $L_d$  itself is set to  $L_d = v_{max}$  to ensure that each car will participate at least once in an update with the enhanced breaking probability  $p_d$ , i.e. in the defect area the randomization is given by  $\max(p(v), p_d)$ . Note, that this also implies that slow vehicles undergo a stronger influence. They need more than one timestep to cross the defect region and thus the defect deceleration rule has to be applied more often than for fast cars which can cross the defect in one timestep. Given that the stochastic noise in the VDR model depends on the velocity of vehicles, the choice of the stochastic parameter inside the defect must be seen with respect to this. Our strategy is to choose the stochastic noise in a way that it is maximal. In Fig. 1 a schematic representation of the local defect with all its parameters is depicted. Obviously, the defect width as well as  $p_d$  control the strength of the defect. For the sake of simplicity, we keep the width of the defect fixed and choose  $p_d$  as control parameter.



**Fig. 1.** Schematic representation of the local defect. The defect itself is placed at a fixed position on a periodic one lane street. Its width is chosen to  $v_{max}$  cells.

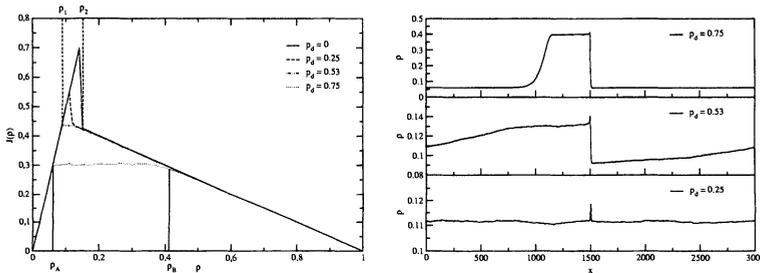
## 3 Numerical Results

In this section we discuss the impact of the local defect scenario introduced in the previous section on the basis of numerical results [11]. We use the stochastic noise

$p_d$  inside the local area belonging to the defect as control parameter. The other parameters of the model are kept fixed ( $L = 3000, v_{max} = 5, p_0 = 0.5, p = 0.01$ ). Starting from low  $p_d$  three different phases can be distinguished in the system.

### 3.1 Small Defect Noise: $p_d \ll p$

For a low stochastic noise inside the defect  $p_d \ll 1$ , the influence on the overall dynamics of the system is negligible. The only exception is the lifetime of the metastable states occurring in the VDR model which is decreasing strongly. The reason is that the metastable states of the VDR model are very sensitive to perturbations [12]. However, the term “small noise” should not be related to the lifetimes of the metastable states but rather to the impact on the systems dynamics after the transition from a metastable high-flow state to a jammed state. Thus a defect noise is “small” if the jammed state of the system is nearly unaffected by it. This phase will be denoted as the *VDR phase* since it matches with the jammed state of the VDR model. It is characterized by a wide jam, which moves backwards and is able to pass the defect area uninfluenced, and free flowing vehicles. In Fig. 2 (right) the density profile of the considered system is shown for various  $p_d$ . The VDR phase shows a constant density profile with a small peak in the area of the defect. This peak is created by the additional velocity fluctuations which lead to increased travel times. Besides this peak there is no markable difference to the jammed state of the VDR model.

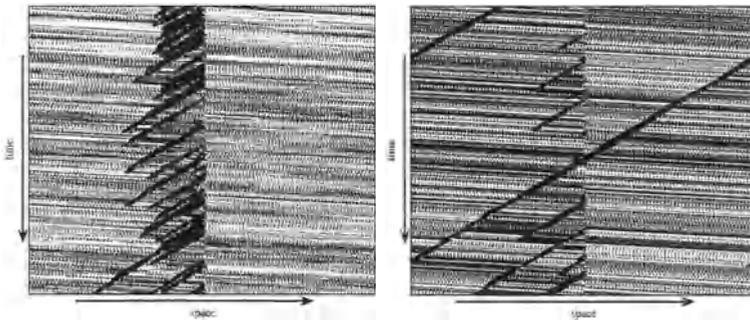


**Fig. 2.** **Left:** Typical fundamental diagram of the VDR model with a lattice defect. The different defect noise parameters  $p_d$  cover the occurring phases. **Right:** Density profiles of the analyzed system for a density of  $\rho = \frac{1}{9}$  in the different phases.

### 3.2 Large Defect Noise: $p_d \gg p$

As expected, large defect noises  $p_d$  have a significant influence on the flow of the system. Three different density regimes can be distinguished corresponding to the ones found in the NaSch model under the same circumstances. At low densities the average distance between the vehicles is large enough to compensate velocity fluctuations induced by the defect. Similarly for high densities the system is

dominated by jams whose movement is nearly unhindered. Thus the fundamental diagram in Fig. 2 coincides for these two density regimes with the one of the undisturbed model. However, the most interesting density regime is situated in the middle of the fundamental diagram and can be identified by a plateau. This plateau is formed since the capacity of the defect limits the global flow in the system. It can not exceed the maximal flow  $J_{\text{def}}$  through the defect which therefore cuts off the fundamental diagram at  $J_{\text{def}}$  and leads to the formation of the plateau. As one can see in the corresponding space-time plot Fig. 3 (left), a considerable amount of vehicles is gathered at the defect forming a high density region. Fig. 2 (right) shows the density profile for a density within the plateau region. The system self-organizes into a macroscopic high density region pinned at the defect and a low density region determined by the capacity of the defect. So far the macroscopic properties are comparable to results obtained by the NaSch model with a local defect. However, a look at the microscopic structure of the high density region reveals some interesting differences. In contrast to the NaSch model, where the high density region at the defect consists of a compact congested region, the high density region in the VDR model is characterized by small compact jams which are separated by free flow regimes (see Fig. 3 (left)). The term “small compact jams” means that the jams at the defect are significantly smaller than the width of the high density region. This specific jam pattern shows some similarities with stop-and-go traffic and will therefore be called *stop-and-go* phase. The *stop-and-go* phase is characterized by a relatively large jammed region (high density) consisting of jams alternating with free flow sections. It must be stressed that the local defect is an essential ingredient for the occurrence of stop-and-go traffic in the VDR model (an undisturbed system shows only free flow or one single wide jam) while in the NaSch model jams of various sizes can occur even in an undisturbed system [8].



**Fig. 3.** Space-time plot of the analyzed system for a density of  $\rho = \frac{1}{8}$ . **Left:** The high defect noise ( $p_d = 0.75$ ) state is characterized by a high density region consisting of small compact jams which are separated by small free flow regions. **Right:** For an intermediate defect noise  $p_d = 0.53$  a mixture of a wide jam moving nearly undisturbed through the system and small jams which are formed at the defect is observable.

### 3.3 Transition Regime: $p_d \approx p$

In the following we focus our investigations on intermediate values of the defect noise parameter  $p_d$ . The space-time plot in Fig. 3 (right) exhibits the microscopic structure of the stationary state for this parameter region. In particular a wide jam moves backwards through the system and additionally some small jams with a limited lifetime are formed at the defect. Note, that as in the case of large and small  $p_d$  the defect has almost no influence on the dynamics of the system for low and high densities. Therefore we will focus on intermediate densities to demonstrate effects caused by the additional localized noise parameter  $p_d$ . As one can see in Fig. 2 the described mixture of a wide jam and some small ones cannot be identified easily in the fundamental diagram. The flow is almost identical to that in the model without defect except for the missing metastable branch. This suggests that for an intermediate  $p_d$  the capacity of the defect is close to the maximum possible flow in the system. Furthermore, taking a look at the density profile for a corresponding intermediate defect noise (respectively  $p_d = 0.53$ ) another difference to the large  $p_d$  case is observable. For a large  $p_d$  the system self-organizes into a macroscopic high density region pinned at the defect and a low density region. In contrast, for intermediate values of  $p_d$  the density profile decreases approximately linear in upstream direction at the defect. This behavior can be traced back to the different lifetimes of the small compact jams. The dynamics of such a small jam can be described analytically by random walk arguments as shown in [12].

This marks an important difference to the behavior of the NaSch model with a defect where a similar state does not exist. This finding is also interesting for the interpretation of empirical results. Traffic states consisting of wide jams passing a localized region with a flow comparable to free flow and small mean velocity (often denoted as synchronized traffic) are observable in real traffic. However, the system state for intermediate  $p_d$  has to be interpreted as a crossover phase. For small  $p_d$  it was shown that only one single wide jam moves undisturbed through the system while for a high  $p_d$  no single wide jam can exist. In contrast a region with many small jams is formed at the defect (stop-and-go traffic). Starting from small  $p_d$  without any small jams in the system one can observe the occurrence of small jams at the defect with an increasing frequency if the defect noise  $p_d$  is increased. Further increasing the defect noise finally leads to the complete dissolving of the large jam. Now the system shows only stop-and-go traffic in the vicinity of the defect. The transition between the crossover phase containing one large and various small jams and the stop-and-go phase (large  $p_d$ ) is determined using an autocorrelation function to  $p_d = 0.57$  [11].

## 4 Summary and Discussion

We have analyzed the impact of a localized lattice defect in the VDR model. An important aspect is the competition between two mechanisms of phase separation. The dynamics of the VDR model leads to a phase separation at high

densities into a large moving jam and a free flow region. In contrast, a localized defect triggers the formation of a high density region pinned at the defect. From a practical point of view our aim was to obtain a deeper insight into the formation of jam patterns due to topological peculiarities. The local defect itself was implemented by increasing the stochastic noise of vehicles within a certain area. Three different system states (phases) can be observed as the defect noise  $p_d$  is varied. Small defect noises  $p_d$  reduce the lifetimes of the metastable states in the VDR model which show a strong sensitivity to disturbances. The vehicles in the jammed state of the system, consisting of a single wide jam and free flow, can pass nearly undisturbed through the defect. We denoted this phase as *VDR phase* since there is almost no difference to the jammed state of the VDR model without defect. In contrast to the low  $p_d$  case, for large  $p_d$  a pinned high density region is formed at the defect limiting the overall system flow. The microscopic structure of this high density region reveals the occurrence of small compact jams which are separated by small free flow regions. This phase is called *stop-and-go phase* since the jam pattern shows strong similarities to stop-and-go traffic. An important point is that stop-and-go traffic cannot be found in the VDR model without a lattice defect. Furthermore, we found crossover behavior for an intermediate defect noise  $p_d$ . Here a wide jam moves backwards through the system. Additionally small jams are formed at the defect which have a limited lifetime.

## References

1. C.F. Daganzo, M.J. Cassidy, and R.L. Bertini, *Transp. Res. A* **33**, 365 (1999).
2. B.S. Kerner, *Phys. Rev. Lett.* **81**, 3797 (1998).
3. D. Helbing, A. Hennecke, and M. Treiber, *Phys. Rev. Lett.* **82**, 4360 (1999).
4. L. Neubert, L. Santen, A. Schadschneider, and M. Schreckenberg, *Phys. Rev. E* **60**, 6480 (1999).
5. H.Y. Lee, H.-W. Lee, and D. Kim, *Phys. Rev. Lett.* **81**, 1130 (1998).
6. V. Popkov, L. Santen, A. Schadschneider, and G. Schütz, *J. Phys. A* **34**, L45 (2001).
7. D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf (eds.), *Traffic and Granular Flow '99* (Springer, 2000).
8. D. Chowdhury, L. Santen, and A. Schadschneider, *Phys. Rep.* **329**, 199 (2000).
9. K. Nagel and M. Schreckenberg, *J. Physique I* **2**, 2221 (1992).
10. R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg, *Eur. Phys. J. B* **5**, 793 (1998).
11. A. Pottmeier, R. Barlovic, W. Knospe, A. Schadschneider, and M. Schreckenberg, *Physica A* (in press).
12. R. Barlovic, A. Schadschneider, and M. Schreckenberg, *Physica A* **294**, 525 (2001).

# Asymmetric Optimal Velocity Model

A. Okumura<sup>1</sup> and S. Tadaki<sup>2</sup>

<sup>1</sup> Department of Information Science, Saga University, Saga 840-8502, Japan

<sup>2</sup> Computer and Network Center, Saga University, Saga 840-8502, Japan

**Abstract.** The optimal velocity (OV) model is one of the car-following traffic flow models, which can reproduce some observed features. The acceleration and braking abilities are symmetrically introduced into the OV model, besides the asymmetry in real vehicles. We introduce the asymmetry between acceleration and braking through the susceptibility in the OV model. As a result, the density contrast becomes large between inside and outside of jam clusters. Jam clusters propagate slower than in the symmetric case.

## 1 Introduction

The optimal velocity (OV) model is one of the traffic flow models, which succeeds to describe some realistic features [1–3]. It is a type of car-following models, in which the motion of cars are controlled according to stimuli from the preceding car. In the OV model, each car controls the acceleration to fit its velocity to the optimal (safety) velocity  $V_{\text{optimal}}$  which depends on the headway distance  $\Delta x$  to the preceding car. The position  $x$  of a car obeys the equation:

$$\frac{d^2x}{dt^2} = \alpha \left[ V_{\text{optimal}}(\Delta x) - \frac{dx}{dt} \right], \quad (1)$$

where the constant  $\alpha$  is a susceptibility.

In the original optimal velocity model, the acceleration and deceleration processes are treated symmetrically. A real vehicle, however, has a stronger deceleration ability than an acceleration. We will discuss the effect of the asymmetry between the acceleration and deceleration processes.

## 2 Model and Analytical Results

There are some methods to take the asymmetry into account. We here introduce the asymmetry into the susceptibility  $\alpha$ . We switch the value  $\alpha$  to  $\alpha_+$  or  $\alpha_-$  depending on the signature of  $V_{\text{optimal}}(\Delta x) - dx/dt$ . We study the case that  $\alpha_-$  takes the same value as in the symmetric case and  $\alpha_+ < \alpha_-$ . We observe the properties of congested states by analytical and numerical methods.

We can obtain an exact solution of (1) for the step optimal velocity function [4,5] even for asymmetric cases.

$$V_{\text{optimal}}(\Delta x) = v_{\text{max}} \theta(\Delta x - d). \quad (2)$$

Consider the  $n$ -th car escaping from a jam cluster. By solving (1) with (2), the headway  $\Delta x_n(t) = x_{n-1}(t) - x_n(t)$  to the preceding car is described as

$$\Delta x_n(t) = \Delta x_J + v_{\max} \left[ \tau + \alpha_+^{-1} \left( e^{-\alpha_+(t-t_{n-1})} - e^{-\alpha_+(t-t_n)} \right) \right], \quad (3)$$

where  $\Delta x_J$  is a headway in the jam cluster and  $t_j$  is the time at which the  $j$ -th car starts to escape from the jam. It is stressed that each car follows the behavior of its preceding car with the delay

$$\tau = t_n - t_{n-1} = t_{n-1} - t_{n-2} = \dots \quad (4)$$

For the  $n$ -th car catching up a jam cluster, the headway obeys the equation:

$$\Delta x_n(t) = \Delta x_F + v_{\max} \left[ -\tau' - \alpha_-^{-1} \left( e^{-\alpha_-(t-t_{n-1})} - e^{-\alpha_-(t-t_n)} \right) \right], \quad (5)$$

where  $\Delta x_F$  is a headway in the freely-moving region upstream of the jam,  $t_j$  is a time at which the  $j$ -th car starts to decelerate and  $\tau' = t_{n-1} - t_n$ .

The headway converges to an asymptotic value exponentially. So we consider the case of an infinitely long system with periodic boundaries. This means the equality  $\tau = \tau'$ . So the following relations are obtained:

$$\Delta x_J = d - \frac{v_{\max}}{2} T(\alpha_-; \tau), \quad (6)$$

$$\Delta x_F = d + \frac{v_{\max}}{2} T(\alpha_+; \tau), \quad (7)$$

$$\tau = \frac{1}{2} [T(\alpha_+; \tau) + T(\alpha_-; \tau)], \quad (8)$$

where the function  $T$  is defined as

$$T(\alpha; \tau) = \frac{2}{\alpha} (1 - e^{-\alpha\tau}). \quad (9)$$

Figure 1 shows the difference  $\Delta x_F - \Delta x_J$  for  $(\alpha_+, \alpha_-)$ . By introducing the slow acceleration ( $\alpha_+ \leq \alpha_-$ ), the density in a freely-moving region becomes lower than that in the symmetric case. As a result, the density in a jam cluster becomes larger. The velocity of jam clusters, which is given by  $-\Delta x_J/\tau$ , also becomes smaller.

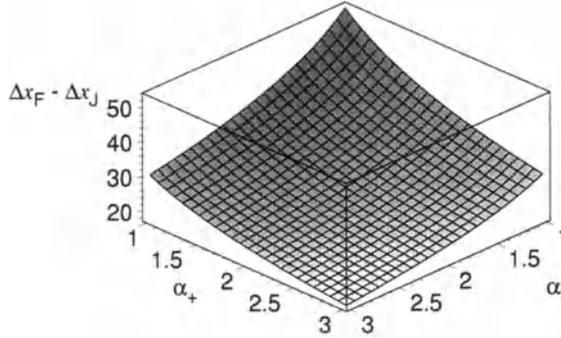
### 3 Simulation

A realistic optimal velocity function has been introduced as the following form:

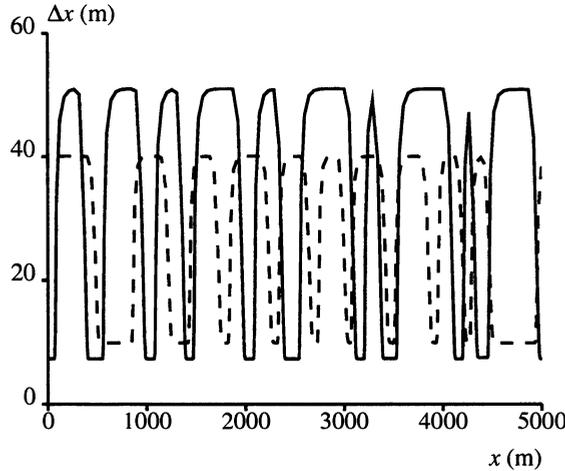
$$V_{\text{optimal}}(\Delta x) = \frac{v_{\max}}{2} \left[ \tanh \left( \frac{\Delta x - d}{w} \right) + c \right], \quad (10)$$

where parameters can be chosen to fit the observed data [6]. We observe numerical simulation results for a periodic system with the Coupled Map Optimal Velocity Model [6,7], which is a discretized OV model.

Figure 2 shows positions and headway for all cars. As discussed in the previous section, the headway clusters becomes shorter inside jam and longer outside jam than the symmetric case.



**Fig. 1.** The difference  $\Delta x_F - \Delta x_J$  is shown for  $(\alpha_+, \alpha_-)$ ,  $v_{\max} = 33.6$  m/sec and  $d = 25$  m with the step OV function.



**Fig. 2.** Positions and headway for all cars are plotted for the realistic OV function. The solid line is for the asymmetric case  $\alpha_+ = 1 \text{ sec}^{-2}$  and  $\alpha_- = 2 \text{ sec}^{-2}$  and the broken line for the symmetric case  $\alpha_{\pm} = 2 \text{ sec}^{-2}$ .

## 4 Summary

We investigate the effects of the asymmetry between acceleration and deceleration. The asymmetry is introduced by switching the susceptibility depending on the difference between the optimal velocity and the current velocity.

By analyzing the exactly solvable case with the step OV function, we found that the headway is shorter inside jam and longer outside jam for the asymmetric case than for the symmetric case. This can be understood by the following simple reasoning. For cars escaping from a jam, the headway becomes larger because of slow acceleration. Next consider a car catching up a jam cluster. If the headway outside jam becomes longer, the relative speed becomes larger at the car starts

to decelerate ( $\Delta x = d$  for the case with a step OV function). So the headway becomes shorter in the jam.

The range of the density becomes wider for the asymmetric case than for the symmetric one, because of the change of the headway distance. The phase transition from the freely-moving phase to the jam phase occurs at the lower density for the asymmetric case than for the symmetric case, because the headway outside jam becomes larger. The effects of the asymmetry upon the transition between jam and freely-moving states should be discussed.

By the asymmetry between acceleration and deceleration, the behavior of catching up a jam and of escaping from a jam will be different. The periodic boundary condition, however, may introduce the balance between them. The behavior in open boundary systems will be interesting.

## Acknowledgment

A part of this work is financially supported by Grant-in-aid No. 12680446 from Ministry of Education, Science, Sports and Culture, Japan, CIRSE in Nagoya University, and by Namura Shipbuilding, Co. Ltd.

## References

1. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Jap. J. Ind. Appl. Math.* **11**, 203 (1994).
2. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995).
3. M. Bando, K. Hasebe, K. Nakanishi, A. Nakayama, A. Shibata, and Y. Sugiyama, *J. Phys. I France* **5**, 1389 (1995).
4. Y. Sugiyama and H. Yamada, *Phys. Rev. E* **55**, 7749 (1997).
5. M. Kikuchi, Y. Sugiyama, S. Tadaki, and S. Yukawa, *Asymmetric Optimal Velocity Model for Traffic Flow*, In *Proceedings of The 5th International Conference on Computational Physics* (Progress of Theoretical Physics Supplement, No. 138), ed. by Y. Hiwatari, Y. Oyanagi, Y. Okabe, and H. Takayama (2000).
6. S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Japan* **67**, 2270 (1998).
7. S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Japan* **68**, 3110 (1999).

# Chaos and Multifractality in a Time-Delay Car-Following Traffic Model

L.A. Safonov<sup>1,2</sup>, E. Tomer<sup>1</sup>, V.V. Strygin<sup>2</sup>, Y. Ashkenazy<sup>3</sup>, and S. Havlin<sup>1</sup>

<sup>1</sup> Minerva Center and Department of Physics, Bar-Ilan University,  
52900 Ramat-Gan, Israel

<sup>2</sup> Department of Applied Mathematics and Mechanics, Voronezh State University,  
394693 Voronezh, Russia

<sup>3</sup> Center for Global Change Science, Massachusetts Institute of Technology,  
Cambridge, Massachusetts 02139, USA

**Abstract.** The presence of chaos in traffic is studied using a car-following model based on a system of delay-differential equations. We find that above a certain time delay and for intermediate density values the system passes to chaos following the Ruelle-Takens-Newhouse scenario (fixed point – limit cycles – two-tori – three-tori – chaos). Exponential decay of the power spectrum and positive Lyapunov exponents support the existence of chaos. We find that the chaotic attractors are multifractal.

## 1 Introduction

Traffic flow often exhibits irregular and complex behavior. It was observed experimentally (e.g. [1]), that although for low and high cars density the motion is relatively simple, for intermediate density values (in the so called “synchronized flow phase” [1]) the motion is characterized by abrupt changes in cars velocities and flow flux. Here we study a model based on a system of delay-differential equations, which for sufficiently large delay and intermediate density values demonstrates complex behavior, attributed to the presence of chaos.

The presence of chaotic phenomena in traffic models has been reported in recent studies. Addison and Low [2] observed chaos in a single-lane car-following model in which a leading car has oscillating velocity. Nagatani [3] reported the presence of a chaotic jam phase in a lattice hydrodynamic model derived from the optimal velocity model [4].

Unlike the above studies, our model is based on a system of autonomous delay-differential equations, and the transition to chaos is possible only in the presence of delay. We show that the system can pass to chaos via many similar routes, and that many different non-chaotic and chaotic attractors may coexist for the same parameter values. We also observe multifractality of the chaotic attractor, which is novel in traffic studies.

We generalize the model proposed and studied in [5–7] by introducing time delay  $\tau$  in the driver’s reaction. The preliminary results of the present study can be found in [8].

The model is based on the assumption that  $N$  cars move in a single lane and the  $n$ th car motion is described by the delay-differential equation

$$\frac{d^2 x_n(t)}{dt^2} = A \left( 1 - \frac{\Delta x_n^0(t - \tau)}{\Delta x_n(t - \tau)} \right) - \frac{Z^2(-\Delta v_n(t - \tau))}{2(\Delta x_n(t - \tau) - D)} - kZ(v_n(t - \tau) - v_{per}), \quad (1)$$

where  $n = 1, \dots, N$ ,  $x_n$  is the car's coordinate,  $v_n$  – its velocity,  $A$  and  $k$  – sensitivity parameters,  $D$  – minimal distance between consecutive cars,  $v_{per}$  – permitted velocity,  $T$  – safety time gap,  $\Delta x_n^0 = v_n T + D$  the safety distance,  $\Delta x_n = x_{n+1} - x_n$  and  $\Delta v_n = v_{n+1} - v_n$ . The function  $Z$  is defined as  $Z(x) = x$  for  $x > 0$  and  $Z(x) = 0$  for  $x \leq 0$ . In our computations we use the parameters values  $v_{per} = 25(m/s)$ ,  $T = 2(s)$ ,  $D = 5(m)$ ,  $A = 3(m/s^2)$ ,  $k = 2(s^{-1})$  and  $N = 100$ . The boundary conditions are periodic, i. e.  $x_{N+1} = x_1 + L$ ,  $v_{N+1} = v_1$ , where  $L$  is the road length.

In this study we find that for sufficiently large delay the system behaves in a complex manner. We show numerically, that the above mentioned limit cycles may bifurcate into two-dimensional tori. With further change of density the tori bifurcate into three-dimensional tori, which are subsequently destroyed forming chaotic attractors. This scenario is known as Ruelle-Takens-Newhouse route to chaos [9].

## 2 Transition to Chaos

To study the transition to chaos we first consider the following solution of system (1)

$$v_n^0 = v^0 = \begin{cases} \frac{A(1-D\rho) + kv_{per}}{A\rho T + k}, & \rho \leq \frac{1}{D+Tv_{per}}, \\ \frac{1-D\rho}{\rho T}, & \rho > \frac{1}{D+Tv_{per}}, \end{cases} \quad x_n^0 = \frac{n-1}{\rho} + v^0 t, \quad (2)$$

where  $\rho = N/L$  is the cars density. This solution corresponds to the homogeneous flow, in which all cars have the same velocity and headways are equal. We introduce a new variable  $\xi_n = \Delta x_n - 1/\rho$  in (1). By this change of variables the homogeneous flow solution (2) is mapped to zero. Its stability can be analyzed using the linearization of (1)

$$\ddot{\xi}_n^0(t) = -p\dot{\xi}_n^0(t - \tau) + q(\xi_{n+1}^0(t - \tau) - \xi_n^0(t - \tau)), \quad (3)$$

where  $p = AT\rho + k$ ,  $q = \frac{AT+kTv_{per}+kD}{AT\rho+k} \cdot A\rho^2$  if  $\rho \leq \frac{1}{D+Tv_{per}}$  and  $p = AT\rho$ ,  $q = A\rho$  otherwise.

Following [4,5], we look for a solution of (3) in the form

$$\xi_n^0 = \exp(i\alpha_\kappa n + \lambda t), \quad (4)$$

where  $\alpha_\kappa = \frac{2\pi}{N}\kappa$  ( $\kappa = 0, \dots, N - 1$ ) and  $\lambda$  is a complex number. Substituting (4) into (3) we obtain a set of algebraic equations for  $\lambda$

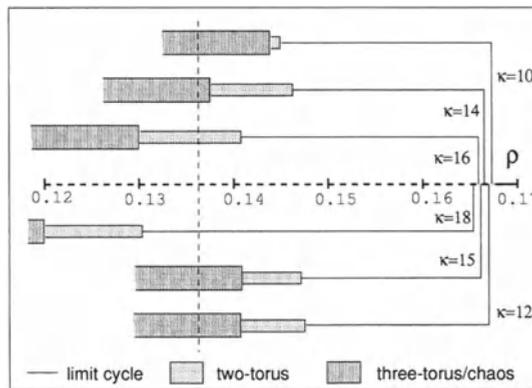
$$\lambda^2 + [p\lambda - q(e^{i\alpha_\kappa} - 1)]e^{-\lambda\tau} = 0. \tag{5}$$

The solutions of (5) are the eigenvalues of system (3). One of these solutions (for  $\kappa = 0$ ) is zero. The others have negative real part for sufficiently high and sufficiently low values of  $\rho$ , which indicates the stability of the homogeneous flow solution. As  $\rho$  decreases (increases), pairs of complex eigenvalues may cross the imaginary axis, causing the formation of small limit cycles (Hopf bifurcations).

We study the formation of limit cycles with the density decreasing from high to intermediate values. Let for some  $\rho = \rho_0$  (5) with some  $\kappa$  have a pair of purely imaginary solutions (a Hopf bifurcation point). For  $\rho = \rho_0 - \varepsilon$  we find the newly-born limit cycle in the form  $\xi_n(t) \sim \varepsilon \xi_n^0(t)$ , using an approximate technique similar to that used in [7] for the non-delay case. Obviously, the flow state corresponding to this limit cycle is a wave with the wavelength equal to  $L/\kappa$  (in length units) or  $N/\kappa$  (in number of cars).

After the small limit cycle for density close to the Hopf bifurcation value is found analytically, its global continuation is performed numerically in the following manner. For  $\rho \approx \rho_0 - \varepsilon$  we take the analytically found approximate periodic solution as an initial condition and solve (1) numerically. After the solution has reached an attracting set, we decrease  $\rho$  with a small step and solve the equations numerically again, taking the results from the previous step as initial conditions. This procedure is iterated further. In this way we keep the track of the particular limit cycle.

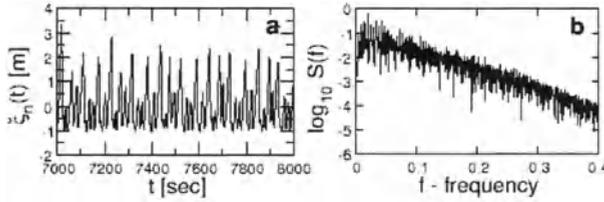
For the non-delay case [7] we have not found any other attracting sets than limit cycles and fixed points. With a small delay the system's behavior



**Fig. 1.** A schematic bifurcation diagram, showing transition to chaos from six different limit cycles. The figure shows that limit cycles, two-tori and chaotic attractors can coexist for the same parameter values (see the vertical dashed line).

does not change qualitatively. For higher values of  $\tau$  (above approximately  $0.5(s)$ ) the cycles may undergo bifurcations leading to the transition to chaos.

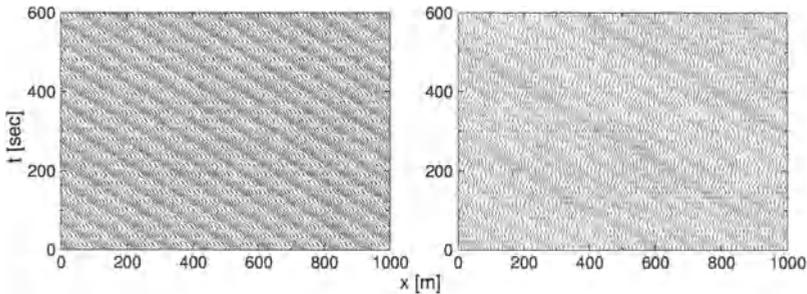
A bifurcation diagram obtained from global continuation of six different limit cycles for  $\tau = 0.59(s)$  is sketched in Fig. 1. The figure shows the transition from each of these cycles to chaos via a two-torus phase. It is important to note that chaotic and non-chaotic attractors coexist for the same parameters values.



**Fig. 2.** **a** -  $\xi_n(t)$  for  $n = 10$ ,  $\kappa = 15$ ,  $\tau = 0.59(s)$  and  $\rho = 0.1387(m^{-1})$ . **b** - the corresponding power spectrum.

Fig. 2a shows the  $\xi_n(t)$  time series for  $\kappa = 15$ ,  $\rho = 0.1387(m^{-1})$  and an arbitrarily chosen  $n$ . Fig. 2b presents the corresponding power spectrum. Its exponential decay is a sign of chaotic behavior of our system [10].

Another indication of chaos is the existence of positive Lyapunov exponents. A direct calculation of the largest Lyapunov exponent [11] and the Lyapunov exponents spectrum [12] yields three positive Lyapunov exponents of order  $10^{-4}$ .



**Fig. 3.** Space-time diagrams of the traffic flow for  $\kappa = 10$  and  $\rho = 0.1(m^{-1})$ . Left -  $\tau = 0.4(s)$  (a limit cycle); right -  $\tau = 0.59(s)$  (chaos). Each dot corresponds to a car.

As observed in real traffic, jams usually move in upstream direction. Our model reproduces this phenomenon. Fig. 3 presents space-time diagrams for the traffic flow for the cases of the system is on a limit cycle and on a chaotic attractor. One can clearly observe jams moving upstream, but in the chaotic

case they are not regular and perhaps look more similar to what can be observed on the road.

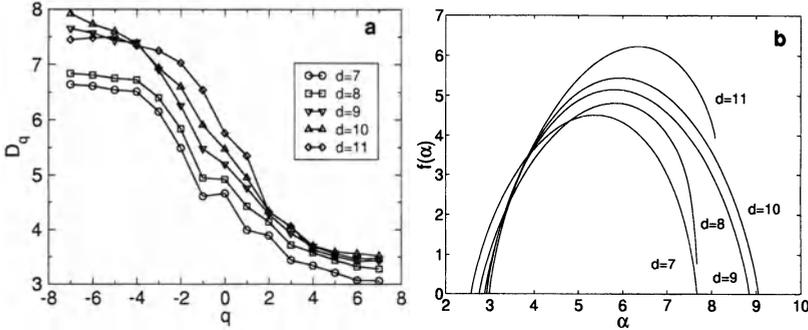
### 3 Multifractality Analysis

The chaoticity of an attractor can also be characterized by its fractal dimensions. We consider the correlation function of the moment  $q$

$$C_q(r) = \left[ \frac{1}{M} \sum_{i=1}^M \left[ \frac{1}{M} \sum_{j=1}^M \Theta(r - |\mathbf{x}_i - \mathbf{x}_j|) \right]^{q-1} \right]^{\frac{1}{q-1}}, \tag{6}$$

where  $\mathbf{x}_i = (\xi_n(t_i), \xi_n(t_i + \Delta t), \dots, \xi_n(t_i + (d-1)\Delta t))$  are  $d$ -dimensional vectors ( $d$  is called embedding dimension),  $\Delta t$  is the first zero of the time series autocorrelation function,  $\Theta$  is the Heaviside step function and  $M$  is the time series length, which should be sufficiently large. The correlation dimension  $D_q$  is defined by the relation  $C_q(r) \sim r^{D_q}$  (see [13,14] and references therein for more details).

To find the correlation dimensions numerically we use the algorithm proposed in [14]. Using these dimensions one can obtain the  $f(\alpha)$  function [15], which represents the spectrum of fractal dimensions of the attractor.



**Fig. 4.** a. Results of measurement of the correlation dimension  $D_q$  for  $q = -7..7$  and  $d = 7..11$  (bottom to top). b. Approximate  $f(\alpha)$  fits for these data.

Fig. 4a shows the results of measurements of  $D_q$  for  $q = -7..7$  and  $d = 7..11$ . Fitting these data with a continuous function  $D(q)$  for each  $d$ , we find the  $f(\alpha)$  function according to the formula

$$f(\alpha(q)) = q\alpha(q) - \tau(q),$$

where  $\tau(q) = (q - 1)D(q)$  and  $\alpha(q) = d\tau/dq$  (see e.g. [15]).

As can be observed from Fig. 4a, the values of  $D_q$  show weak convergence with growing  $q$ , especially for  $d$  close to 0. Therefore, presented values of dimensions may be underestimated. Nevertheless, shapes of  $f(\alpha)$  for each  $d$  enable us to claim the multifractality of the considered attractor.

## References

1. B.S. Kerner, H. Rehborn: *Phys. Rev. E* **53**, R4275 (1996); B.S. Kerner, H. Rehborn: *Phys. Rev. Lett.* **79**, 4030 (1997); B.S. Kerner: *Physics World* 12 (August), 25 (1999).
2. P.S. Addison, D.J. Low: *Chaos* **8**, 791 (1998); D.J. Low, P.S. Addison: *Nonlinear Dynamics* **16**, 127 (1998).
3. T. Nagatani: *Phys. Rev. E* **60**, 1535 (1999).
4. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995); Y. Sugiyama. In: *Proceedings of Workshop on Traffic and Granular Flow (1995, Julich)*, ed. by D.E. Wolf, M. Schreckenberg, A. Bachem (World Scientific, Singapore 1996).
5. E. Tomer, L. Safonov, S.Havlin: *Phys. Rev. Lett.* **84**, 382 (2000).
6. E. Tomer, L. Safonov, and S. Havlin. In: *Traffic and Granular Flow '99: Social, Traffic and Granular Dynamics*, ed. by D. Helbing, H.J. Herrmann, M. Schreckenberg, D.E. Wolf, pp. 419-424 (Springer, Heidelberg/Berlin 2000).
7. L.A. Safonov, E. Tomer, V.V. Strygin, S. Havlin: *Physica A* **285**, 147 (2000).
8. L.A. Safonov, E. Tomer, V.V. Strygin, Y. Ashkenazy, S. Havlin: *Europhys. Lett.*, in press (2001).
9. S. Newhouse, D. Ruelle, F. Takens: *Commun. Math. Phys.* **64**, 35 (1978); D. Ruelle, F. Takens, *Commun. Math. Phys.* **20**, 167 (1971).
10. H.G. Schuster, *Deterministic Chaos: An Introduction* (VCH, Weinheim 1995).
11. M.T. Rosenstein, J.J. Collins, C.J. De Luca: *Physica D* **65**, 117 (1993).
12. A. Wolf, J.B. Swift, H.L. Swinney, A. Vastano: *Physica D* **16**, 285 (1985).
13. P. Grassberger, I. Procaccia: *Physica D* **9**, 189 (1983); K. Pawelzik, H.G. Schuster: *Phys. Rev. A* **35**, 481 (1987).
14. Y. Ashkenazy: *Physica A* **271**, 427 (1999).
15. C.J.G. Evertsz, B.B. Mandelbrot, 'Multifractal measures.' In: H.-O. Peitgen, H. Jürgens, D. Saupe *Chaos and Fractals. New Frontiers of Science*, pp. 921-944 (Springer-Verlag, New York 1992).

# **Traffic Models: Application**

# Optimal Velocity Model and its Applications

A. Nakayama<sup>1</sup>, K. Hasebe<sup>2</sup>, and Y. Sugiyama<sup>3</sup>

<sup>1</sup> Gifu Keizai University, Ohgaki, Gifu 503-8550, Japan

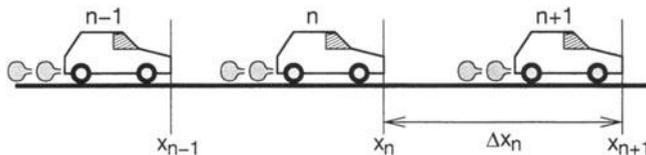
<sup>2</sup> Faculty of Business Administration, Aichi University, Miyoshi, Aichi 470-0296, Japan

<sup>3</sup> Division of Mathematical Science, City College of Mie, Tsu, Mie 514-0112, Japan

**Abstract.** We review some aspects of the optimal velocity (OV) model and compare these aspects with those of other car-following models. We also show two applications of the OV model. One is an application to the intelligent transport system, especially how to suppress the emergence of traffic congestion. The other is an application to pedestrian flow. For these purposes, we propose some extensions of the OV model.

## 1 Introduction

The motivation to study traffic flow is to understand the property of traffic flow, mainly why traffic congestion occurs. The study of traffic flow has a long history and many models have been proposed for engineering and physical study [1,2]. Physical models of traffic flow are roughly divided into two groups. One is called the macroscopic model, and the other is called the microscopic model. Macroscopic models are based on the analogy between traffic flow and fluid or gas [3,4]. In these models, traffic congestion is expressed as high density or high pressure regions. On the other hand, microscopic models describe the motion of each vehicle and include several types of models. Car-following models, which have been studied for a long time, use the continuous variables, time, position and velocity. The optimal velocity (OV) model is one of the car-following models [5,6]. There are models which use discrete variables, for example, cellular automata models [7] and coupled map lattice models [8,9].



**Fig. 1.** Definitions of car number and headway.

Generally the microscopic model with continuous variables is expressed by the equation of motion

$$\ddot{x}_n(t) = aF( \begin{array}{c} \cdots, \Delta x_{n+1}, \Delta x_n, \Delta x_{n-1}, \cdots, \\ \cdots, \dot{x}_{n+1}, \dot{x}_n, \dot{x}_{n-1}, \cdots, \\ \cdots, x_{n+1}, x_n, x_{n-1}, \cdots \end{array} ), \quad (1)$$

where  $x_n$ ,  $\dot{x}_n$  and  $\Delta x_n = x_{n+1} - x_n$  are position, velocity, and headway of the  $n$ th car respectively. The coefficient  $a$  is called ‘‘sensitivity’’ and represents the magnitude of the driver’s response to a stimulus, for example, the change of the headway. Equation (1) simply states that the acceleration of the  $n$ th car is decided by the motion of the  $n$ th car and the surrounding traffic, i.e. the headway, velocity, and position of the preceding car and those of the car that follows and so on.

The car-following models are based on the idea that the dominant part of stimulus comes from the preceding car. Along this idea, the equation of motion (1) can be simplified. The following three models are examples, that are constructed in three different viewpoints with respect to the behavior of a driver: How a driver controls his car.

**Example 1 : traditional car-following model (CFM)**

This model is based on the viewpoint that a driver is going to adjust his velocity to that of the preceding car. The acceleration is proportional to the difference between them, and the equation of motion has the form

$$\ddot{x}_n(t + \tau) \propto \dot{x}_{n+1}(t) - \dot{x}_n(t) , \quad (2)$$

where  $\tau$  is the delay of response. If  $\tau$  is zero, (2) is almost trivial. Therefore the delay plays an essential role in this model to describe the behavior of traffic flow.

**Example 2 : optimal velocity model (OVM)**

In this model, a driver is supposed to maintain an optimal velocity which depends on his headway. The equation of motion is

$$\ddot{x}_n \propto V(\Delta x_n) - \dot{x}_n , \quad (3)$$

where  $V$  is called the optimal velocity (OV) function, which represents the relationship between headway and velocity of the driver. The delay  $\tau$  is not necessary. Moreover the delay has only a small effect even if it is introduced [6].

**Example 3 : optimal headway model (OHM)**

In this model, a driver is supposed to maintain an optimal headway which depends on his velocity. The equation of motion is

$$\ddot{x}_n \propto H(\dot{x}_n) - \Delta x_n , \quad (4)$$

where  $H$  is the optimal headway function, which is the inverse function of  $V$ .

Proportional factors of these equations and the OV function are determined as follows. In OVM and OHM, the OV function  $V (= H^{-1})$  is a essential information to construct the models. In principle,  $V$  should be determined by a experiment, which measures headway and velocity of a car simultaneously, and proportional factors, which are determined by measuring the magnitude of acceleration of the car. For simplicity, we suppose that a certain function expresses the relation between velocity and headway and the proportional factors is a constant. Then we can fix the equation of motion for OVM and OHM.

CFM reduces to a too simple model, if we choose a constant proportional factor. Then the proportional factor has been thought to be a function of headway

and velocity. Here we suppose that the OV function is given by some observations.

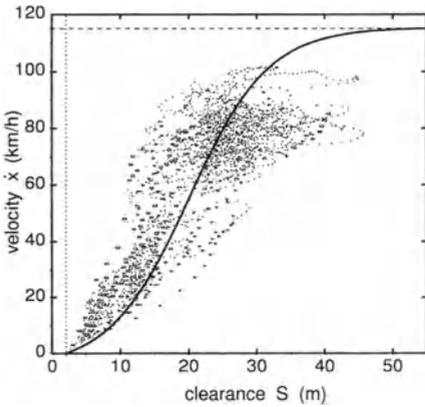
$$\dot{x}_n(t) = V(x_{n+1}(t) - x_n(t)) . \tag{5}$$

By differentiate (5) with respect to  $t$ , we obtain

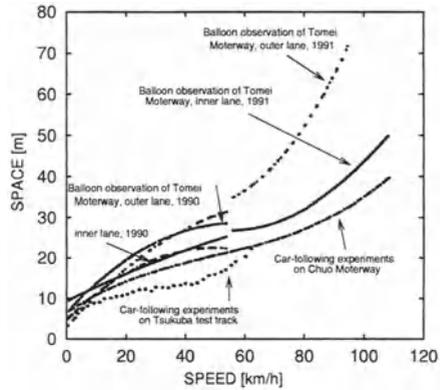
$$\ddot{x}_n(t) = V'(\Delta x_n)\{\dot{x}_{n+1}(t) - \dot{x}_n(t)\} . \tag{6}$$

From Eq.(6), we can induce a class of car-following models by introducing the delay  $\tau$ . Though the proportional factor often have a dependence on the velocity in traditional car-following models, here we concentrate on the model, in which the factor is a function of the headway only. In these constructions of car-following models and other dynamical models, the relation between headway and velocity is a fundamental information and plays an essential role.

The OV function can be fixed by experiments or observations. Figure 2 shows the result of an experiment, where headway and velocity of a single car moving on a highway are measured simultaneously [10]. The data are randomly scattered due to road and traffic conditions. We simply can choose a hyperbolic-tangent function and can fit the parameters to the data. Figure 3 shows observational



**Fig. 2.** Dots represent measured data. The solid line is a fitted OV function.



**Fig. 3.** Each line represents observational or experimental data.

data on Japanese highways and experimental data on a test circuit [11]. From these data, we can conclude that the similar behavior is observed universally and the hyperbolic-tangent function is a best choice of a simple OV function.

Then we obtain the concrete forms of (2), (3) and (4).

$$\ddot{x}_n(t + \tau) = V'(\Delta x_n)\{\dot{x}_{n+1}(t) - \dot{x}_n(t)\} , \tag{7}$$

$$\ddot{x}_n = a\{V(\Delta x_n) - \dot{x}_n\} , \tag{8}$$

$$\ddot{x}_n = a\{H(\dot{x}_n) - \Delta x_n\} , \tag{9}$$

where

$$V(\Delta x_n) = v_0 \{ \tanh \beta(\Delta x_n - b) + c \}, \quad H(\dot{x}_n) = V^{-1}(\dot{x}_n). \quad (10)$$

In (8) and (9), the proportional factors are chosen as a constant parameter, that is, sensitivity  $a$  for simplicity. From the data of Fig. 2, we can fix the OV function as

$$V(\Delta x) = 16.8 \{ \tanh 0.086(\Delta x - 25) + 0.913 \} \quad (\text{m/s}), \quad (11)$$

and we use this in the simulations. Obviously the values of the parameters must be changed in order to fit  $V(\Delta x)$  to each data. However, these values are not important from the physical viewpoint. Finally we note that each model includes only one free parameter  $\tau$  or  $a$ . Sensitivity  $a$  in the models (8) and (9) plays a similar role to  $1/\tau$  in the model (7) [6].

## 2 Review of OV Model

In this section, the properties of the OV model (8) with (10) are reviewed and the comparison between three models is shown. The OV model can explain the mechanism of the emergence of traffic congestion and can reproduce the behavior of real traffic flow. In the framework of the OV model, the change from free flow to congested flow can be understood as a kind of phase transition.

Equation (8) has a trivial, homogeneous flow solution, where all cars move with a constant headway and a constant velocity.

$$\Delta x_n = h, \quad \dot{x}_n = V(h) \quad \text{for all } n. \quad (12)$$

We identify this solution with free flow in real traffic. By a linear analysis, we can investigate the stability of the solution (12). The stability condition is simply given by the sensitivity and the derivative of the OV function as

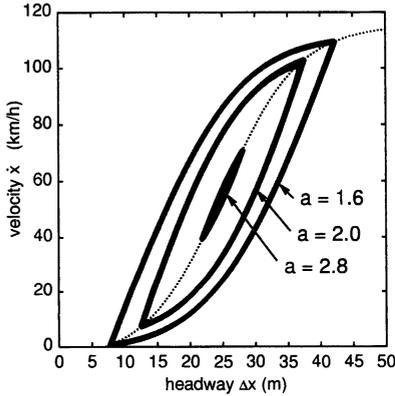
$$a > 2V'(h) \quad : \text{ stable}, \quad (13)$$

$$a < 2V'(h) \quad : \text{ unstable}. \quad (14)$$

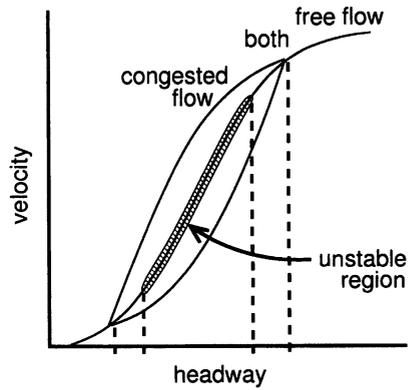
For the case  $a > 2V'(b)$ , the homogeneous flow is stable for any headway, because the derivative of the OV function (10) takes the maximum value at  $\Delta x = b$ . In other cases, the homogeneous flow is stable for a large headway (and sufficiently small headway). If the headway is smaller than a certain value, that is, the density of cars exceeds a certain critical value, the homogeneous flow becomes unstable and transits to congested flow spontaneously. This phenomenon can be understood as a bifurcation or a phase transition.

The behavior of traffic flow in the unstable case is investigated by numerical simulations. Global features of the formation of traffic congestion can be seen by a spacetime diagram. A typical example is shown in Fig. 8. The behavior of each car is clearly shown in the headway-velocity phase space, which is the same as in Fig. 2. The homogeneous flow state is represented by a single point on the OV function, because all cars have an identical headway and velocity. In unstable

case, only the congested flow is stable and any initial state develops to congested flow states [5]. Final congested flow states are represented by a limit cycle or a hysteresis loop in the phase space (see Fig.4). The size of the loop is dynamically



**Fig. 4.** Each curve represent a limit cycle or a hysteresis loop. Typical values  $a = 1.6, 2.0, 2.8$  are chosen as the sensitivity. The dashed line represents the OV function.



**Fig. 5.** Typical relation among the hysteresis loop, the unstable region, and the flow state which is realized finally.

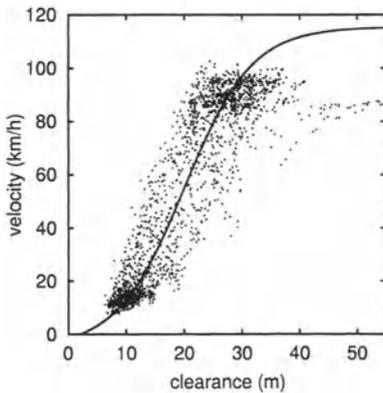
decided by the sensitivity. If the sensitivity is large, the size of the loop is small. It corresponds to the fact that if sensitivity is large, the unstable region is small. As mentioned above, for sufficiently large sensitivity, the homogeneous flow is completely stable. In the case of Fig. 4 where we use the OV function (11), the critical sensitivity  $2V'(b)$  is 2.88.

Two endpoints of the hysteresis loop always exist outside the unstable region. Two regions, one of which is the width of the loop and the other is the unstable region, do not coincide generally (see Fig. 5). Therefore, there are three regions. For a large headway, only the homogeneous flow (the free flow) is stable. Congestion clusters will disappear even if they initially exist. For a small headway, only the congested flow is stable and any flow transits to congested flow. Between them, there exists a region where both flows are stable. It depends on the initial condition which flow is realized.

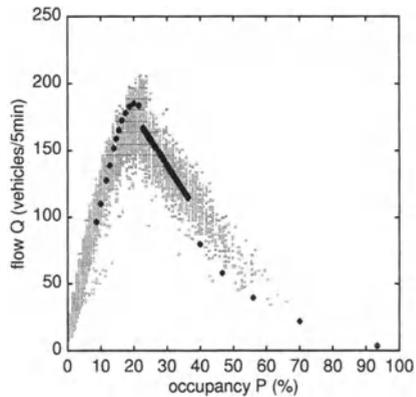
The properties of congested flow are also investigated by simulations and summarized as follows. Small congestion clusters disappear and closely separated two clusters merge. Only large clusters separated by large distances can survive for a long time (see Fig. 8). The bifurcation of a cluster is never observed in any simulation. The two endpoints of the hysteresis loop in Fig. 4 represent the states of cars inside congestion clusters and of cars in the region between clusters. Two curves connecting the endpoints represent the motion of the cars going out from the clusters and coming into the clusters. In the simulations of Fig. 4,

several clusters with different length are usually created. The uniqueness of the hysteresis loop shows that there exists only one species of congestion clusters and that they are distinguished by their length.

Next, we show whether the OV model can reproduce the behavior of real traffic or not. As a first example, we investigate the motion of cars shown in Fig.2. Here, we introduce the randomness in the personalities of drivers instead of the randomness of the road or the traffic conditions. Each driver has a OV function, which is scaled  $\pm 20\%$  randomly. Then, the maximum velocity is  $90 \sim 140\text{km/h}$  ( $115\text{km/h} \pm 20\%$ ). This is similar for the headway. The result of the simulation shown in Fig.6 is in good agreement with the observed data in Fig.2. The OV model can reproduce the observed motion of cars by use of an adequate OV function.



**Fig. 6.** The motion of cars is reproduced by a simulation.

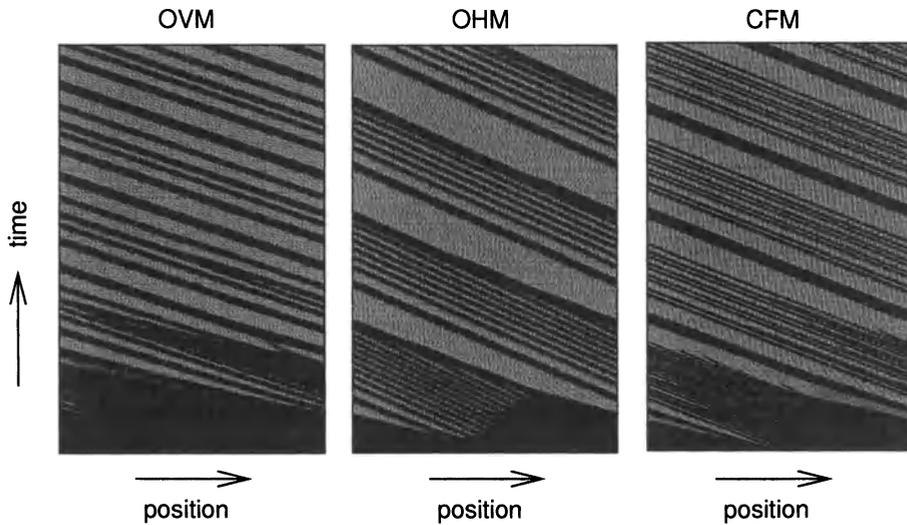


**Fig. 7.** Small dots represent observational data on a Japanese highway and diamond marks represent results of simulations.

Figure 7 shows the fundamental diagram, which expresses the relation between car density and the flow [12]. Simulations are carried out without introducing the randomness in order to clarify the behavior and the result is shown by diamond marks in Fig. 7. The global shape of the diagram obtained by simulations is quite similar to the observed one. We can find a possible jump of the flow at the occupancy  $P \simeq 25\%$ . In the framework of the OV model, this point is identified as a critical density, which is the transition point from homogeneous flow to congested flow. It can be easily understood that the result of simulations becomes much more similar to the observations, if the randomness of traffic condition is introduced.

Finally, we clarify the difference among the OV model (OVM), the car-following model (CFM), and the optimal headway model (OHM) of section 1. For this purpose, we carry out the similar investigation to OVM. The main subject is whether only OVM can explain the emergence of traffic congestion or not.

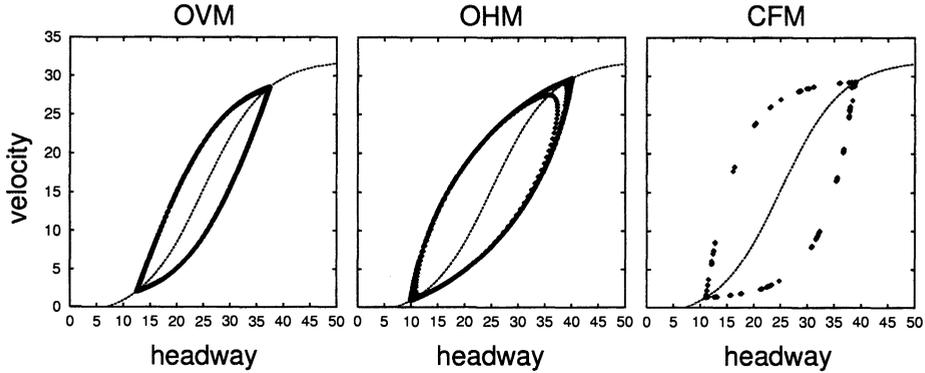
Figures 8 and 9 are the results of simulations for OVM, OHM and CFM. Figure 8 shows the process of congestion formation. At first sight, three models seem to produce almost the same result. However, the details of the dynamics are different. In OVM, small clusters combine or disappear and only large clusters survive. The dynamics of other models are much more simple. In OHM, small clusters survive until the combination with a large cluster. In CFM, very small clusters are created and roughly a half of them disappear. The other half of the clusters can exist stable but they consist of only four or five cars. These behaviors seem to be unnatural. We conclude that OVM should be thought as a basic model to reproduce the behavior of real traffic.



**Fig. 8.** Positions of all cars in a circuit are plotted. All cars move from the left to the right. The initial state is a homogeneous flow state and a small disturbance is added at the center. The headway is chosen so that the homogeneous flow is unstable. Dark bands express congestion clusters.

Figure 9 shows the motion of cars in phase space. Hysteresis loops for three models take roughly the same shape. In CFM, the sparseness of data points is an artificial effect caused by the delay time and the time step of the simulation and has no physical meaning.

As a result, the transition itself from free flow to congested flow can be explained by almost any model. We think that the essence of traffic congestion exists in the OV function as follows. When the OV function is steep, a driver must react sensitively to a deviation of the headway, that is, a driver must change his velocity rapidly. However, the small sensitivity indicates small acceleration and therefore such a reaction cannot be realized. In such case, the deviation is amplified in the chain of reactions and the flow becomes unstable. Therefore, the



**Fig. 9.** The motion of cars at the final stage of simulation are shown. Dashed lines represent the OV function which is common to all models.

congestion phenomenon can be thought as a common and unavoidable property for similar systems. In fact, traffic congestion in CFM can be understood as a kink-soliton and in a special case the system is related to a network of electric circuit [13].

Here we show another example. This is a story of zero fighter pilots in World War II. When they start from a base, they intend to maintain a short distance between the airplanes. But during the flight, their distances continue to change, larger and smaller. Finally, the airplanes spread out when they arrive at the destination. This phenomenon can be understood in the framework of the 1-dimensional OV model in spite of the 3-dimensional motion of airplanes [14].

### 3 Application to ITS

In this section, we present an application to the realistic problem, that is, how to control cars in order to avoid the formation of congestion. As stated in the previous section, traffic congestion is unavoidable in natural cases. However, by use of the driving control system which is a part of ITS, we can suppress the emergence of congestion. We will search a candidate of a basic model for such a control system.

From the physical viewpoint, a simple model or a model with few free parameters is thought to be a good model. We can understand the mechanism of a phenomenon and can predict the results of experiments by such a model. The OV model has a simple form with only one parameter (sensitivity  $a$ ). This is an advantage for the physical study, but is a disadvantage as a model for the control system of cars, because there is a little allowance to tune up the model. The model for the control system does not have any restriction and we investigate some extended models for this purpose.

Here we discuss the case where each car is controlled by a certain system independently. In order to extend the model, we also suppose that the control

system can obtain more information in addition to that obtained by itself. For example, ITS has the ability to communicate with other cars and also a certain control center. When such a system is realized, each control system can use the information of not only its state but also the states of other cars. The extension of the OV model is carried out under these suppositions.

We concentrate on the simplest extension, where the equation is the same as the OV model but the OV function depends on the headway of other cars. We investigate two examples. The effects of these extensions are clarified and we show a suggestion how to incorporate the information of other cars.

One example is a model where the OV function depends also on the headway of the preceding car:  $V(\Delta x_{n+1}, \Delta x_n)$  [15], and the other is a model where the OV function depends also on the headway of the car that follows:  $V(\Delta x_n, \Delta x_{n-1})$  [16]. Hereafter, we refer the former by the forward looking OV (FL-OV) model and the latter by the backward looking OV (BL-OV) model. We can search a better model in a wide class of the OV function  $V(\dots, \Delta x_{n+2}, \Delta x_{n+1}, \Delta x_n, \Delta x_{n-1}, \Delta x_{n-2}, \dots)$ . The comprehensive survey of this widely extended OV model is shown in Ref. [17].

To compare the two models, we assume the condition that they give the same homogeneous flow state as the OV model in any headway. This condition is simply realized by the following OV functions. For the FL-OV model

$$V(\Delta x_{n+1}, \Delta x_n) = \alpha V(\Delta x_{n+1}) + (1 - \alpha)V(\Delta x_n) , \quad (15)$$

and for the BL-OV model

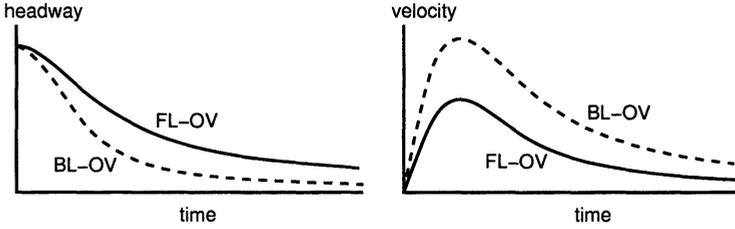
$$V(\Delta x_n, \Delta x_{n-1}) = (1 + \alpha)V(\Delta x_n) - \alpha V(\Delta x_{n-1}) , \quad (16)$$

where  $\alpha > 0$  and  $V(\Delta x)$  is the same function as (10).

By the linear analysis, we can investigate the stability of the homogeneous flow solutions of FL-OV and BL-OV models [15,16]. The stability in both models is improved compared to that in the OV model. Obviously, this is not so surprising result, because these models include the OV model as a special case.

The two models stabilize the homogeneous flow in different ways compared to the OV model. We can clearly see the difference between two models in the damping behavior of a disturbance (see Fig.10). The FL-OV model is insensitive to the deviation of headway. As a result of this, the change of velocity is small. If we control cars based on this model, the deviation is slowly reduced. This control method may be comfortable for a user of the car, but is a little dangerous because the deviation remains for a long time. In the BL-OV model, the deviation of headway is quickly reduced and the change of velocity is large. This control method may be safe but is uncomfortable for the user.

The difference can be seen also in the process of dissolution of traffic congestion, for example, in simulations starting from a congested flow state. We can observe the shrinking behavior of the hysteresis loop by choosing parameters to make the homogeneous flow stable. In the FL-OV model, the hysteresis loop shrinks in the direction of velocity first, that is, the velocity deviation reduces



**Fig. 10.** Damping behaviors of the headway and velocity are shown. The disturbance is added by changing the headway of a car. Each line shows the sum of deviations of all cars.

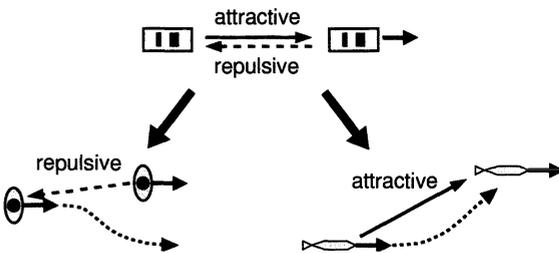
first and then its whole size gradually reduces. In the BL-OV model, the hysteresis loop shrinks in the direction of the headway first, that is, the headway deviation reduces first.

In the OV model, a large value of sensitivity is the only way to stabilize the homogeneous flow. The extended models can stabilize the flow even for a small sensitivity by taking into account the information of the preceding car or the car that follows. The effect of each extension is different from each other. The best way may be a hybrid model which incorporates both extensions [17].

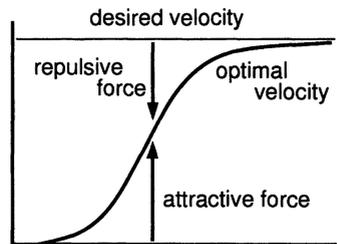
### 4 Extension to 2-Dimensional Model

In this section, we propose an extension of the OV model to a 2-dimensional model, which can be applied to multilane traffic or the motion of pedestrians. The extension is realized by replacing all quantities in (8) by 2-dimensional vectors. However, there are two ways to extend the OV function depending on our viewpoints for the 1-dimensional OV model.

From one viewpoint, the follower in the upper part of Fig. 11 wants to follow the leading car, that is, the follower is attracted by the leading car. If we extend this situation into a 2-dimensional one straightforwardly, the follower behaves like that in the right part of Fig. 11. This model is suitable for the motion of fish, for example, clustering.



**Fig. 11.** Differences between two viewpoints.



**Fig. 12.** Rough relation.

From the other viewpoint, the follower wants to move at his desired velocity. But he is forced to reduce the velocity because of the existence of the preceding car, that is, the follower receives a repulsive force from the preceding car. If we extend the situation into a 2-dimensional one, the follower behaves like that in the left part of Fig.11. This model is suitable for multilane traffic or the motion of pedestrians or granular flow. Figure 12 shows the rough relation among the attractive force, the repulsive force, the desired velocity, and the optimal velocity. Here the desired velocity is defined as the velocity that he moves when there are no restrictions and no other people. We investigate the property of this model and compare it with the result in the 1-dimensional case. Specific phenomena in two dimension like a lane formation are shown in Ref. [18].

The equation of motion in the 2-dimensional OV model is written as<sup>1</sup>

$$\ddot{\vec{x}}_n = a \{ \vec{V}(\dots) - \dot{\vec{x}}_n \}, \tag{17}$$

$$\vec{V}(\dots) = \vec{V}_0 - \sum_k \frac{\overrightarrow{\Delta x}_{kn}}{\Delta x_{kn}} f(\theta) \{ c - \tanh \beta(\Delta x_{kn} - b) \}, \tag{18}$$

where  $\overrightarrow{\Delta x}_{kn} = \vec{x}_k - \vec{x}_n$  and  $\Delta x_{kn} = |\overrightarrow{\Delta x}_{kn}|$ . The summation is taken for cars (or pedestrians or something else) in the range where the distance  $\Delta x_{kn}$  is smaller than a certain value.  $\vec{V}_0$  is the desired velocity and the second term in the right-hand side of (18) expresses the repulsive force. Figure 13 shows the relation among vectors  $\vec{V}_0$ ,  $\overrightarrow{\Delta x}_{kn}$ ,  $\vec{V}$  and the repulsive force.  $f(\theta)$  is an arbitrary function

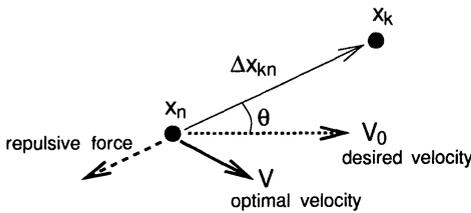


Fig. 13. Relation among vectors in (18).

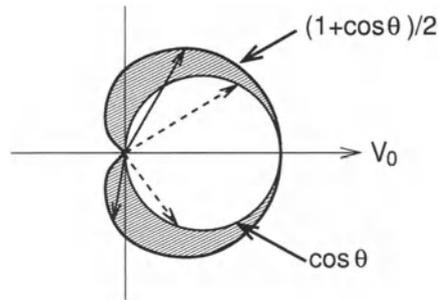


Fig. 14. Angular dependence of the 2-dimensional OV function.

which expresses the angular dependence of the 2-dimensional OV function. At present, there is no principle to decide  $f(\theta)$  and we show two examples:

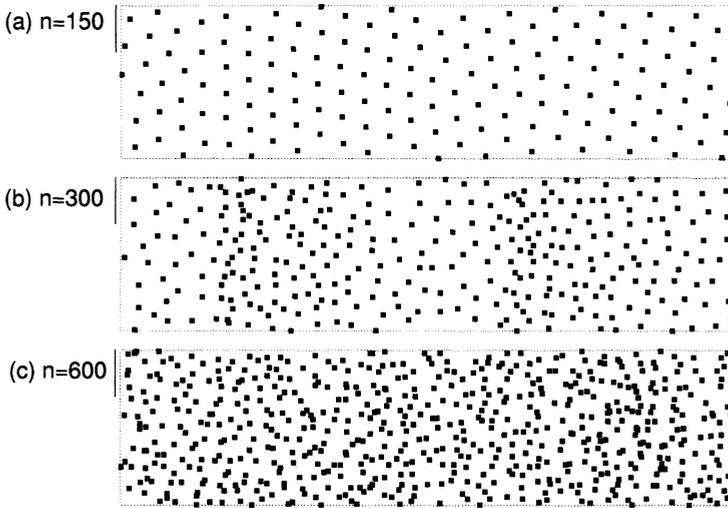
$$f(\theta) \propto \frac{1}{2}(1 + \cos \theta) \quad (-\pi < \theta < \pi), \tag{19}$$

$$f(\theta) \propto \cos \theta \quad \left(-\frac{\pi}{2} < \theta < \frac{\pi}{2}\right). \tag{20}$$

<sup>1</sup> Details of the definitions of the factors, parameters, and the summation are shown in Ref. [18].

Equation (20) may be a candidate for the model of pedestrian flow (or multilane traffic), where a pedestrian looks only at people in front. Equation (19) is a candidate for a model of granular flow, where a particle receives forces from all particles around it. It may be possible that there exists the lateral force between pedestrians. In such case we may use (19) instead of (20).

We investigate the behavior of flow by simulations in the case of  $f(\theta) \sim (1 + \cos \theta)$ . Figure 15 shows typical snapshots of flow states. Each particle (or pedestrian) is supposed to be a point particle for simplicity. At low density (Fig.15a), the state has a hexagonal structure. We identify this state as the homogeneous flow. At middle density (Fig. 15b), the flow changes to congested



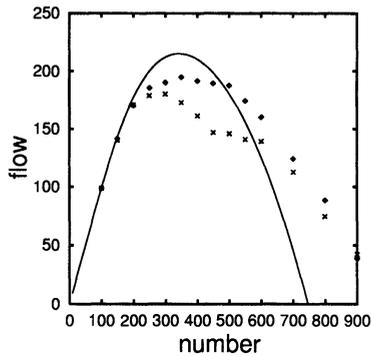
**Fig. 15.** Snapshots of flow states are shown at (a) low density, (b) middle density and (c) high density. Each small square represents a position of one particle (or pedestrian).

flow. High density regions move forward or backward case by case. The behavior is quite similar to that of granular flow in media. We have no data about whether a density wave exists in pedestrian flow or not. At high density (Fig. 15c), each particle seems to move randomly and no structure exists.

We can easily find the correspondence between the first two states and those in 1-dimensional traffic flow. By use of simulations, we can confirm that the stability of homogeneous flow is controlled by the sensitivity. The system is reduced to 1-dimensional one effectively. Here we note that the density wave is not observed clearly in the case of  $f(\theta) \sim \cos \theta$ . In the reduction of the system, the existence of the lateral force, which homogenizes density transversally, seems to play an important role.

As a result, the properties of the flow in the 2-dimensional OV model are quite similar to those of 1-dimensional flow. This result is also confirmed by

the fundamental diagram. Figure 16 shows the results of simulations and the



**Fig. 16.** Solid line represents the flow of the homogeneous flow state. Diamond and cross marks represent the results of simulations in the cases  $f(\theta) \sim \cos \theta$  and  $f(\theta) \sim (1 + \cos \theta)$  respectively.

analytical result for the flow of the homogeneous flow state. We have taken into account up to next-nearest interaction. We cannot find any qualitative difference between the case of  $f(\theta) \sim \cos \theta$  and the case of the homogeneous flow. However, in the case of  $f(\theta) \sim (1 + \cos \theta)$ , the behavior of the flow is different from others. This may indicate the existence of a transition to congested flow.

The 2-dimensional OV model can be easily applied to mixed flow, where fast moving and slowly moving pedestrians exist, and counter flow. For example, at low density, we can observe the lane formation in both cases of mixed flow and counter flow by the simulations. At high density, we observe a congested flow similar to Fig. 15 in the case of mixed flow, and a blocking state in the case of counter flow. A detailed analysis of this system is shown in Ref. [18].

## Acknowledgements

This work is partly supported by a Grant-in-Aid for Scientific Research (C) (No.12650065) and (C) (No.13640409) of the Japanese Ministry of Education, Science, Sports and Culture, and by the Center for Integrated Research in Science and Engineering in Nagoya University.

## References

1. B.D. Greenshields, H.R.B. Proceedings **14**, 468 (1935); L.A. Pipes, J. Appl. Phys. **24**, 274 (1953); M.J. Lighthill and G.B. Whitham, Proc. R. Soc. Lond. **A229**, 281 (1955); ib. 317 (1955); E.S. Olcott, J. Opns. Res. Soc. Am. **3**, 147 (1955); E. Kometani and T. Sasaki, J. Opns. Res. Japan **2**, 11 (1958); R. Herman,

- E.W. Montroll, R.B. Potts, and R.W. Rothery, *Opns. Res.* **7**, 86 (1959). H. Greenberg, *J. Opns. Res. Soc. Am.* **7**, 79 (1959); G.F. Newell, *Opns. Res.* **9**, 209 (1961); D.C. Gazis, R. Herman, and R.W. Rothery, *Opns. Res.* **9**, 545 (1961); L.C. Edie, *J. Opns. Res. Soc. Am.* **9**, 66 (1961); J. Drake, A.D. May, and J.L. Shofer, *Highway Research Record* **154**, 53 (1967).
2. D.E. Wolf, M. Schreckenberg and A. Bachem (Eds.), *Traffic and Granular Flow* (World Scientific, Singapore, 1996); M. Schreckenberg and D.E. Wolf (Eds.), *Traffic and Granular Flow '97* (Springer-Verlag, Singapore, 1998); D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf (Eds.), *Traffic and Granular Flow '99* (Springer-Verlag, Berlin Heidelberg, 2000); D. Chowdhury, L. Santen, and A. Schadschneider, *Phys. Rep.* **329**, 199 (2000).
  3. I. Prigogine and F.C. Andrews, *Opns. Res.* **8**, 789 (1960); I. Prigogine and R. Herman, *Kinetic theory of vehicular traffic* (Elsevier, New York, 1971).
  4. B.S. Kerner and P. Konhäuser, *Phys. Rev. E* **48**, 2335 (1993).
  5. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995); *Japan J. Indust. Appl. Math.* **11**, 203 (1994); M. Bando, K. Hasebe, K. Nakanishi, A. Nakayama, A. Shibata, and Y. Sugiyama, *J. Phys. I France* **5**, 1389 (1995); T.S. Komatsu and S.-i. Sasa, *Phys. Rev. E* **52**, 5574 (1995). Y. Sugiyama and H. Yamada, *Phys. Rev. E* **55**, 7749 (1997).
  6. M. Bando, K. Hasebe, K. Nakanishi, and A. Nakayama, *Phys. Rev. E* **58**, 5429 (1998); *Japan J. Indust. Appl. Math.* **17**, 275 (2000).
  7. K. Nagel and M. Schreckenberg, *J. Phys. I France* **2**, 2221 (1992); T. Nagatani, *Phys. Rev. E* **48**, 3290 (1993); M. Fukui and Y. Ishibashi, *J. Phys. Soc. Jpn.* **62**, 3841 (1993).
  8. S. Yukawa, M. Kikuchi, and S. Tadaki, *J. Phys. Soc. Jpn.* **63**, 3609 (1994); S. Yukawa and M. Kikuchi, *J. Phys. Soc. Jpn.* **64**, 35 (1995).
  9. S.-i. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Jpn.* **67**, 2270 (1998); Y. Sugiyama, M. Kikuchi S.-i. Tadaki, and S. Yukawa, *Prog. Thor. Phys. Supplement No.138*, 549 (2000); *ib.* 594 (2000).
  10. T. Oba, *An Experimental Study on Car-following Behavior*, Thesis of Master of Engineering, Univ. of Tokyo, (1988).
  11. J. Xing, *A Study on the Bottleneck Phenomenon and Car-Following Behavior on Motorways*, Thesis of Doctor of Engineering, Univ. of Tokyo, (1992); M. Koshi, M. Iwasaki, and I. Ohkura, *Proc. 8th Intl. Symp. on Transp. Traffic Theory* (edited by V. F. Hurdle etc.), 403 (1983).
  12. T. Sasaki and Y. Iida, *Traffic Engineering* (in Japanese), Kokumin Kagakusha, Tokyo, 1992.
  13. G. B. Whitham, *Proc. R. Soc. Lond.* **A 428**, 49 (1990). Y. Igarashi, K. Itoh, and K. Nakanishi, *J. Phys. Soc. Jpn.* **68**, 791 (1999). K. Hasebe, A. Nakayama, and Y. Sugiyama, *Phys. Lett.* **A259**, 135 (1999); *Prog. Theor. Phys. Supplement No.138*, 602 (2000).
  14. S. Sakai, *Zero-sen no Shinjitsu* (in Japanese), Kodansha, Tokyo, 1992.
  15. T. Nagatani, *Phys. Rev.* **E60**, 6395 (1999); S. Sawada, e-print nlin.PS/0105054.
  16. K. Hasebe, in: *Proceedings of the 3rd Symposium on Simulation of Traffic Flow*, M. Fukui and Y. Ishibashi (Eds.) pp.6 (The mathematical society of traffic flow, Japan, 1997)(in Japanese); K. Hasebe, A. Nakayama, and Y. Sugiyama, to appear in *Phys. Rev. E*.
  17. K. Hasebe, A. Nakayama, and Y. Sugiyama, *Widely extended optimal velocity model of traffic flow and their linear stability*, in this volume.
  18. K. Hasebe, A. Nakayama, and Y. Sugiyama, *Modeling pedestrians and granular flow in 2-dimensional optimal velocity models*, in this volume.

# Breakdown and Recovery in Traffic Flow Models

K. Nagel<sup>1</sup>, C. Kayatz<sup>1</sup>, and P. Wagner<sup>2</sup>

<sup>1</sup> Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland

<sup>2</sup> Institute for Transportation Research, German Aerospace Centre, 12489 Berlin, Germany

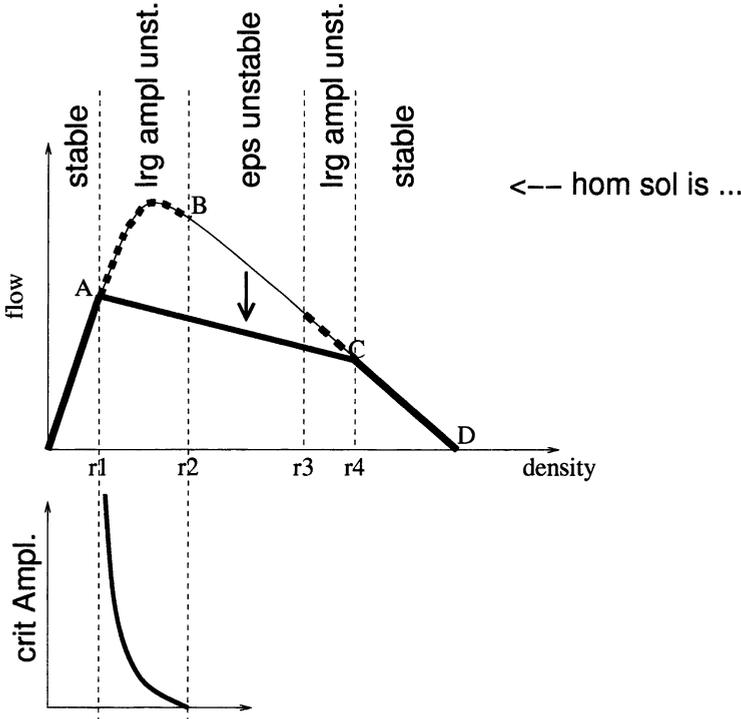
**Abstract.** Most car-following models show a transition from laminar to “congested” flow and vice versa. Deterministic models often have a density range where a disturbance needs a sufficiently large critical amplitude to move the flow from the laminar into the congested phase. In stochastic models, it may be assumed that the size of this amplitude gets translated into a waiting time, i.e. until fluctuations sufficiently add up to trigger the transition. A recently introduced model of traffic flow however does not show this behavior: in the density regime where the jam solution co-exists with the high-flow state, the intrinsic stochasticity of the model is not sufficient to cause a transition into the jammed regime, at least not within relevant time scales. In addition, models can be differentiated by the stability of the outflow interface. We demonstrate that this additional criterion is not related to the stability of the flow. The combination of these criteria makes it possible to characterize similarities and differences between many existing models for traffic in a new way.

## 1 Introduction

Car traffic is not always homogeneous. For example, stop-and-go waves are a frequently observed phenomenon. Correspondingly, most traffic models show a transition from laminar to “congested” flow and vice versa. For many deterministic models, this mechanism is well understood (e.g. [1,2], see Fig. 1): For certain densities, the homogeneous solution is linearly unstable, meaning that any tiny disturbance will destroy the homogeneity and lead to another state, typically to one or more waves. For other densities, the homogeneous state may be linearly stable, but unstable against large amplitude disturbances.

In stochastic models, one would intuitively assume (see Fig. 2) that linear instability gets translated into plain instability – meaning that, for the corresponding densities, the homogeneous state breaks down immediately – and that for large amplitude instability the large amplitude instability gets translated into meta-stability – meaning that, for the corresponding densities, one has to wait some time until the noise conspires in a way that a critical disturbance is generated and the instability is triggered. This is exactly the topic for this paper, where we will demonstrate that this speculation is correct in some cases but not in others.

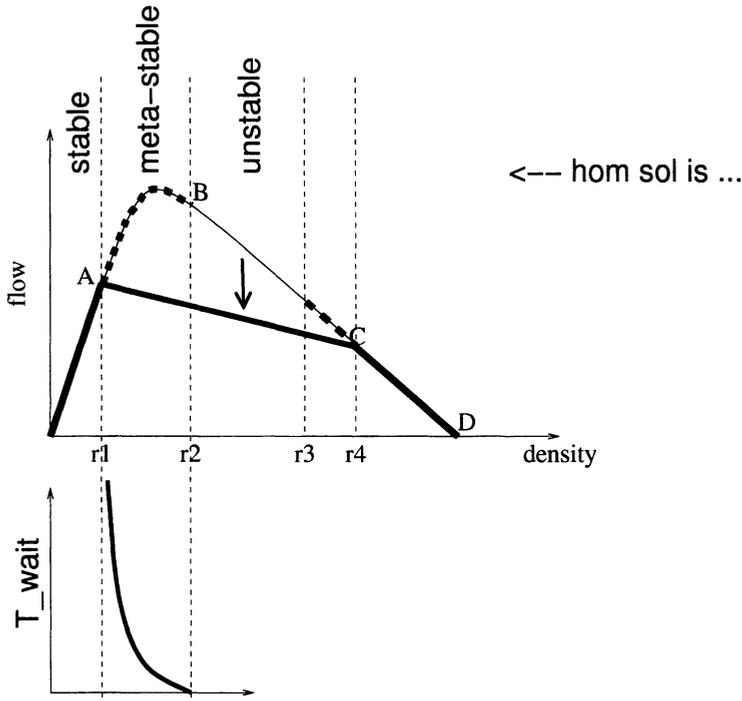
Recent field measurements identify additional dynamic phenomena, such as oscillations and so-called synchronized traffic [3,4]. It is under discussion in how far these additional phenomena can be explained by the above model instabilities in conjunction with geometrical constraints (such as bottlenecks) [5,6], or if additional features are necessary in the models [7]. Given this, it seems desirable to understand as much as we can about existing models.



**Fig. 1.** Schematic fundamental diagram for deterministic models. “lrg ampl unst.” means that the homogeneous solution can be kicked into an inhomogeneous state by a large enough amplitude; the bottom plot gives a schematic graph for the necessary size of that critical amplitude. “eps unstable” means that the homogeneous solution is linearly unstable.

Indeed, Krauss [8] introduces model classes which he names type I, type II, and type III. Type III refers to a viscous, syrup-like behavior without breakdown, and is not of relevance here. Type II displays jam formation, but jams have a typical size, meaning that the system is macroscopically homogeneous, and that there is no true phase transition. Type I displays true, macroscopic structure formation and therefore a first order phase transition. In this paper, we will argue that the Krauss characterization is incomplete. We will demonstrate that models can be stable or unstable at maximum flow, and that the jams can have a stable or unstable interface. The difference to Krauss in Ref. [8] is that he implicitly assumes that stable maximum flow goes together with a stable interface, and that unstable maximum flow goes together with an unstable interface. Introducing our additional characterization means that we have  $2 \times 2 = 4$  different classes, instead of just I and II.

In order to demonstrate this, we will first review what expectation one has for traffic flow breakdown in analogy to a gas-liquid transition (Sec. 2). We then, in Sec. 3, describe the traffic model that we use. The central Sections 4 and 5 describe results with respect to transition times, and with respect to interface dynamics. Sec. 6 is a longer dis-



**Fig. 2.** Schematic fundamental diagram for stochastic models. “meta-stable” means that the homogeneous solution will be kicked into an inhomogeneous state after a certain waiting time; the bottom plot gives a schematic graph for that waiting time.

cussion of our results, including speculations, conjectures, and some simulation results for other models. The paper is concluded by a summary.

## 2 Traffic Breakdown and the Gas-Liquid Transition

The breakdown of laminar traffic, i.e. the transition from homogeneous traffic to stop-and-go waves, can be compared to a gas-liquid transition, i.e. the transition from the homogeneous gas state to the inhomogeneous gas/liquid coexistence state (e.g. [9,8,10]). As is well known, if one compresses a gas beyond a certain critical density, then it becomes super-critical, and small fluctuations will lead to droplet formation and thus into the coexistence state [11]. Similarly, we would expect for homogeneous traffic that, once compressed beyond a certain critical density, small fluctuations will lead to jam formation and thus into the coexistence state.

And conversely, one knows that all droplets vanish once the mixture is expanded beyond the critical density. Similarly, one would expect that all traffic jams vanish once the system is expanded beyond a critical density.

This leads to predictions about the statistics of jam formation and jam dissolution. Krauss [8] gives the following for the probability in a given time step that a jam starts

somewhere in the system:

$$p_{\downarrow} \sim L \exp\left(\frac{\hat{\alpha}}{\langle g \rangle - g_0}\right),$$

where  $\langle g \rangle = 1/\rho - \ell = L/N - \ell$  is the average gap,  $g_0$  and  $\hat{\alpha}$  are free parameters, and  $L$  is the system size.  $N$  is the number of cars in the system;  $\ell$  is the length that a vehicle occupies in a dense jam.

The above is a bulk effect; jam formation can happen anywhere in the system. Jam dissolution, in contrast, is an interface effect: A jam with  $N$  vehicles dissolves if the random numbers come out the correct way to let all  $N_{jam}$  vehicles make the ‘‘correct’’ type of movement. This leads to a recovery probability of  $p_{\uparrow} \sim \exp(-N_{jam})$ . Since  $N_{jam} \sim L(\rho - \rho_c)$ , we obtain

$$p_{\uparrow} \sim e^{-L(\rho - \rho_c)}.$$

Note that  $p_{\uparrow}$  is larger than zero also for  $\rho > \rho_c$ , that is, spontaneous recovery should be possible in super-critical systems although it becomes exponentially improbable with increasing system size.

In both cases, the time to recovery would be the inverse of the above probabilities. When setting the two equations equal, one obtains the condition for a system to fluctuate back and forth between the homogeneous and the coexistence state. This would occur above  $\rho_c$ ; however, for any given  $\rho > \rho_c$ , in the case of  $L \rightarrow \infty$ ,  $p_{\downarrow}$  would go to infinity while  $p_{\uparrow}$  would go to zero, meaning that above  $\rho_c$  only the coexistence state is stable in the limit of  $L \rightarrow \infty$ .

### 3 The Model

The model to be used in the following was introduced by Krauss [8]. The basic idea is that cars drive as fast as possible, but avoid crashes. Therefore, they have to choose their velocity  $v \leq v_{safe}$  which takes into account the braking distance  $d(v)$  of the following and the braking distance  $d(\tilde{v})$  of the preceding car. That means that the velocity has to fulfill  $d(v) + v\tau \leq d(\tilde{v}) + g$ . Here,  $g$  is the space headway between the cars given by  $g = \tilde{x} - x - \ell$ . The braking capabilities of the cars are the same for all cars and are parametrized by the maximum deceleration  $b$ .  $\tau$  is uniformly set to one throughout this paper. This safety condition can be transformed into a set of update rules as follows:

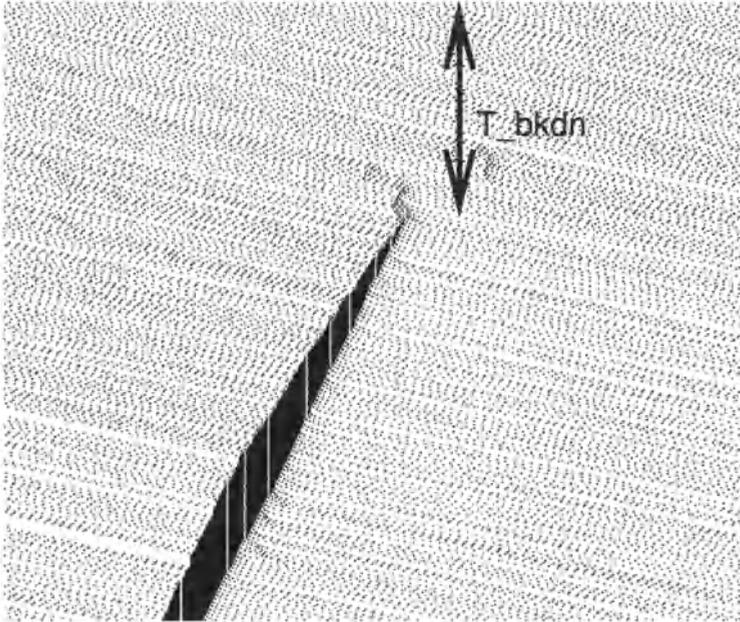
$$v_{safe} = \tilde{v}_t + 2b \frac{g_t - \tilde{v}_t}{2b + v_t + \tilde{v}_t}, \quad (1)$$

$$v_{des} = \min\{v_t + a, v_{safe}, v_{max}\}, \quad (2)$$

$$v_{t+1} = \max\{v_{des} - a\epsilon\xi, 0\}, \quad (3)$$

$$x_{t+1} = x_t + v_{t+1}. \quad (4)$$

with index  $t$  counting integer time. The parameter  $a$  is the maximum acceleration, the parameter  $\epsilon$  measures the degree of randomness,  $\xi$  is a random number,  $\xi \in [0, 1]$ , while  $v_{max}$  is the maximum velocity. We will use  $v_{max} = 3$  throughout this paper. Ref. [8] discusses what our selection of parameters means in terms of real world units; let us state that our specific values have a reasonably close relation to the real world.



**Fig. 3.** Space-time plot for the breakdown time measurement. Space is horizontal; time increases downward; each line is a snapshot; vehicles move from left to right. Initially, all vehicles are lined up equidistant with the specified density. Time is measured until one vehicle in the simulation comes to a complete stop ( $v_i = 0$ ). Once a jam is started, it typically keeps growing until inflow is reduced, either by another jam upstream, or by the effect of periodic boundary conditions.

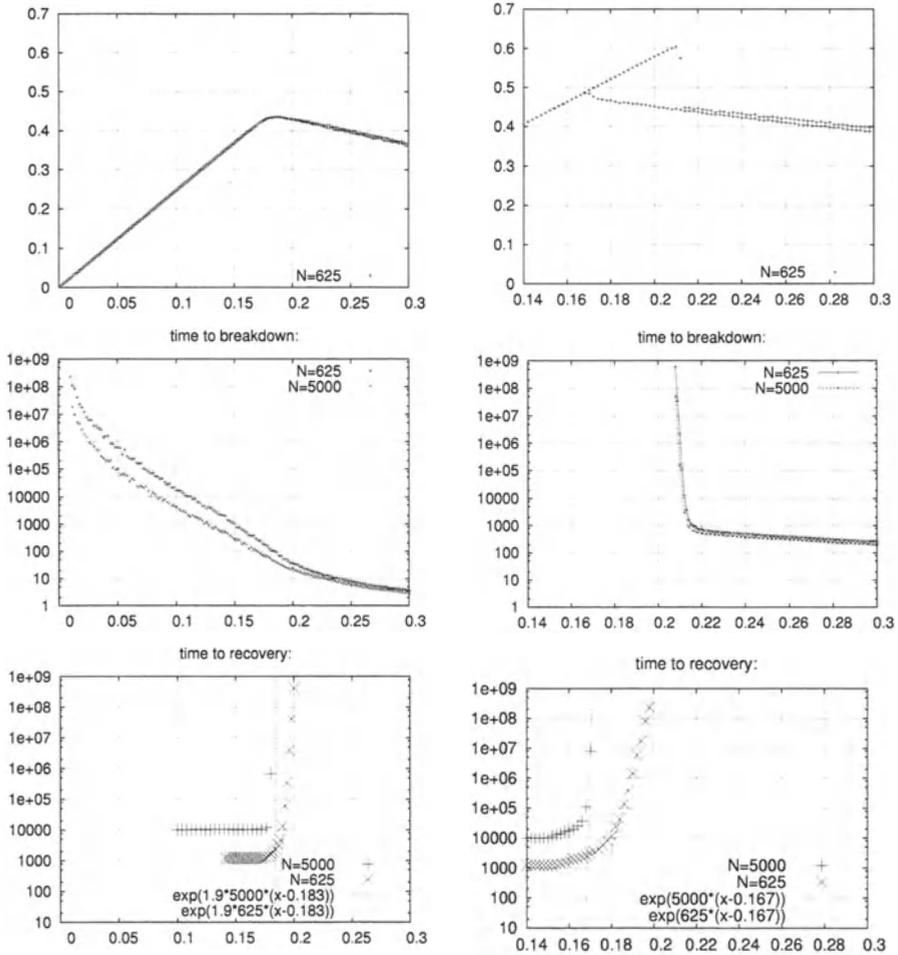
## 4 Transition Times

Fig. 4 shows the breakdown and the recovery times for two different sets of parameters:  $(a, b) = (1, \infty)$  and  $(a, b) = (0.2, 0.6)$ . Recall that  $a$  and  $b$  are the acceleration and braking capabilities, respectively. The simulations are run with a fixed number  $N$  of vehicles; different densities are obtained by adapting the system size  $L$  via  $L = N/\rho_L$ . The times are obtained as follows:

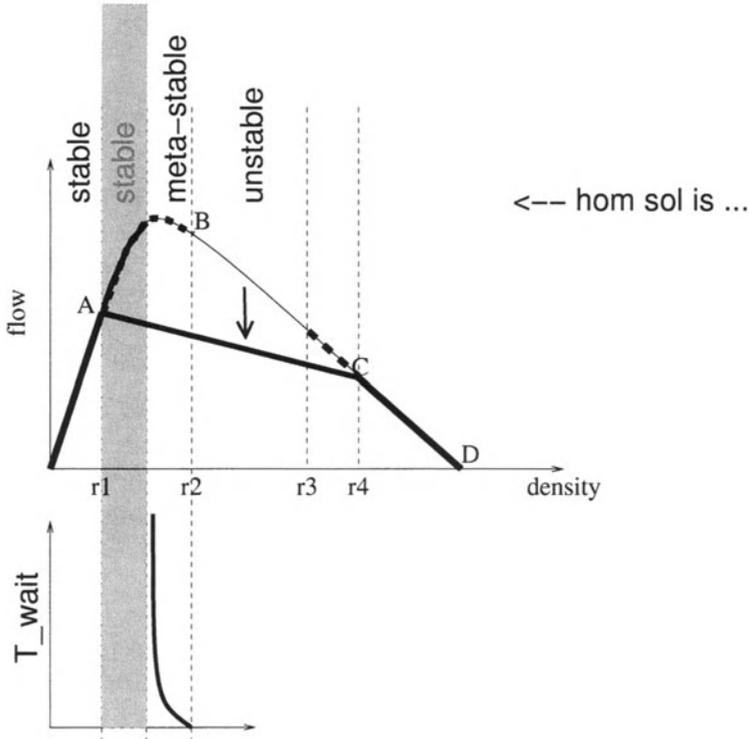
- **Breakdown times:** The system is started with all vehicles at equal distance  $g = 1/\rho_L - 1$  and with the initial velocity taken from the laminar branch of the fundamental diagram.<sup>1</sup> The time is measured until the first vehicle in the system shows  $v = 0$  (see Fig. 3).
- **Recovery times:** The system is started with all vehicles except the leading one at distance one, i.e. gap equal to zero, and velocity zero. The time is measured until no vehicle with velocity zero is left in the system.

Each data point is an average of at least 50 runs.

<sup>1</sup> Annoyingly, in the transition regime, for some parameters of  $a$  and  $b$  different initial conditions lead to significantly different breakdown times. Outside the transition regime, the results are robust.



**Fig. 4.** Breakdown (middle) and recovery (bottom) times as function of density. Left column:  $(a, b) = (1, \infty)$ . Right column:  $(a, b) = (0.2, 0.6)$ . The straight lines in the bottom plots are proportional to  $\exp(AN\rho)$ , where  $N$  is the number of cars in the system, and  $A$  is a free parameter. – In the right column, we see that for  $N = 5000$  there is a gap from  $\rho \approx 0.17$  to  $\rho \approx 0.205$ , where the system is, up to  $10^9$  time steps, stable both against breakdown and against recovery. – The corresponding sections of the fundamental diagram (throughput as function of density; top) for  $N = 625$  are given for orientation. Each value of the fundamental diagram is obtained at 5000 time steps; this is done once for homogeneous and once for jammed initial condition, resulting in two branches for bi-stable models.



**Fig. 5.** Schematic fundamental diagram for stochastic models including the new regime. Compare to Fig. 2.

One observes from Fig. 4 that, for both cases, the recovery behavior is qualitatively consistent with the gas-liquid transition picture: Above a certain  $\rho_c$ , the waiting time until recovery (i.e. until a system with jams transitions to a system without jams) shows exponential growth, which increases with system size.

Similarly, for  $(a, b) = (1, \infty)$  (Fig. 4 left), the breakdown results are qualitatively consistent with the gas-liquid transition picture: The time to breakdown *decreases* both with increasing system size and with increasing density. Putting breakdown and recovery together, one obtains that for  $L \rightarrow \infty$  and in equilibrium, a system with  $\rho > \rho_c$  should always be in the coexistence state. For smaller  $L$ , the system can jump back and forth between coexistence and the homogeneous state.

For parameters  $(a, b) = (0.2, 0.6)$ , a possibly different picture emerges. Here, the breakdown times seem to diverge at  $\rho^* \approx 0.2$ , meaning that, for large  $L$  and possibly for  $L \rightarrow \infty$ , we have a density range where besides the transition from coexistence to homogeneous also the inverse transition from homogeneous to coexistence is extremely improbable. *That is, we may have a stable supercritical homogeneous phase under an update rule that includes noise.* Fig. 5 shows a schematic fundamental diagram with the new region. – It is however very difficult to get good numerical results for such fast growth as we find here: from  $\rho = 0.22$  to  $\rho = 0.21$ , the breakdown times grow between 2 and 5 orders of magnitude, depending on the system size. Several function fits were

tried out, without convincing success; for example  $T \sim \exp((\rho - \rho_*)^{-\alpha})$  (implying divergence) or  $T \sim \exp(-\gamma(\rho - \rho_*))$  (implying no divergence). We cannot rule out the second functional form; however, note that from Fig. 4 we have, at  $N = 5000$ , a gap from  $\rho \approx 0.17$  up to  $\rho \approx 0.21$  where both branches are stable within times of  $10^9$ .

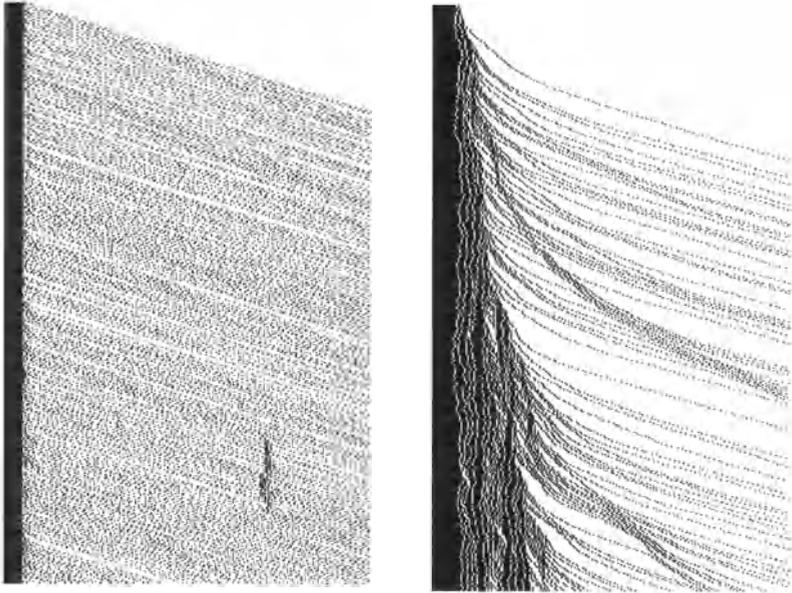
Also, we note that breakdown times in general do not follow well the analytical expectations. Although qualitatively the dependency is as expected (breakdown time decreasing with increasing system size), the quantitative behavior is different. The breakdown times for a given density but different system sizes do not lie on a straight line on a log-log plot (not shown), meaning that neither  $T_{bkdn} \sim 1/L$  nor any other algebraic form is applicable. In contrast, for recovery times, the simulation results are at least not inconsistent with  $T_{rcov} \sim \exp(-L(\rho - \rho_*))$ . One should in general note that very few system sizes were simulated; in particular, the simulations do not extend to – computationally difficult – very large systems sizes where different behavior might be found.

Why is the Krauss model with  $(a, b, \epsilon) = (0.2, 0.6, 1)$  so much different from the “standard” phase transition picture, where noise will relatively quickly add up in a way that a super-critical homogeneous state will break down? We suspect that in many traffic models, because of the parallel update, noise is introduced in a special way. In particular, the amount of noise per spatial and temporal unit is bounded. In conjunction with a dynamics which dissipates noise fast enough, it make sense to obtain states which are absolutely stable under this kind of noise. It is unclear to us if the continuous variables used in the model here are a necessary ingredient or not; preliminary simulation results indicate that the same type of behavior can be obtained by a discrete model, but see Ref. [12] for a similar model where the continuousness of the variables seems to play a crucial role.

## 5 Interface Dynamics

The nature of the transition (e.g. crossover vs. true phase transition) is however not given by the time it takes until the first fluctuation happens, but by how this fluctuation develops further, in particular, if it spreads into the rest of the system or not. In order to further understand the nature of the transition, we will now look at the dynamics of the interface between jam and outflow. That is, we start with an infinitely large mega-jam with  $g = 0$  and thus  $\rho = 1$  in the half space from  $x = -\infty$  to zero. We collect data for the development of the density profile as a function of time. While doing that, we translate the zero of the coordinate system always to the leftmost moving car, i.e. to the rightmost car in the mega-jam which has not moved so far. To the left from this point, density is always one; in consequence, we look at the question if the interface to the right will grow in time or if it will develop a characteristic, time-independent profile.

Fig. 6 contains space-time plots of this interface for two different systems. In the left plot, the interface is stable, whereas in the right plot, it keeps growing throughout the plot. The left plot is obtained with  $(a, b, \epsilon) = (1, \infty, 1)$ , which is one of the two models for which we have investigated the transition times in more detail above. A plot for  $(a, b, \epsilon) = (0.2, 0.6, 1)$  looks similar (not shown). In contrast, the plot on the right



**Fig. 6.** Space-time plot of interface dynamics. As before, space is pointing go the right and time is pointing down. Space coordinates are translated such that the leftmost edge of the moving traffic is always at the same position. LEFT: Krauss with  $(a, b, \epsilon) = (1, \infty, 1)$ . Example for stable interface. RIGHT: Krauss with  $(a, b, \epsilon) = (0.2, 0.6, 1.5)$ . Example for unstable interface. – Note that for the left example,  $a$  and  $b$  are selected in the range typically considered unstable, while for the right example,  $a$  and  $b$  are selected in the range typically considered stable.

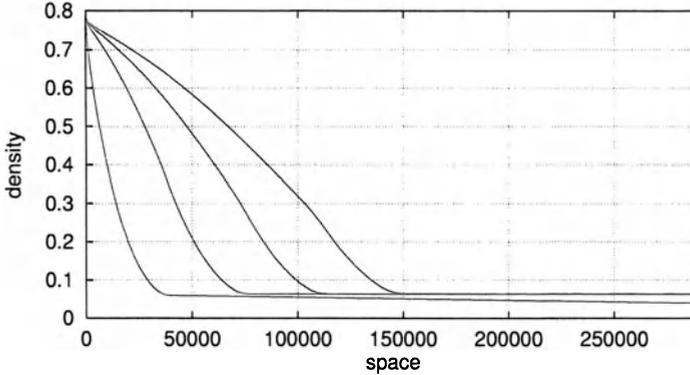
with the growing interface is obtained with a larger noise amplitude, i.e.  $(a, b, \epsilon) = (0.2, 0.6, 1.5)$ .

In order to investigate the long-term behavior, we also plotted density profiles at different times. A stable interface is characterized by a density profile which eventually becomes stationary; an unstable interface keeps growing. Fig. 7 contains a result for the model of Fig. 6 right, i.e. with  $(a, b, \epsilon) = (0.2, 0.6, 1.5)$ . The plot contains density profiles at times 250 000, 500 000, 750 000, and 1 000 000. Each curve is the average of 60 runs. Clearly, the plot shows that the interface grows with time. In fact, when looking at a fixed density value, say  $\rho = 0.2$ , it seems that the interface width is growing linearly in time.

What this means is that *the stability of the interface is a property which is separate from the stability of the flow*. The following table lays out the resulting four cases:

	stable outflow	unstable outflow
stable i-face	E.g. $(a, b, \epsilon) = (0.2, 0.6, 1)$ “Krauss type I”	E.g. $(a, b, \epsilon) = (1, \infty, 1)$
unstable i-face	E.g. $(a, b, \epsilon) = (0.2, 0.6, 1.5)$	E.g. $(a, b, \epsilon) = (1, \infty, 1.5)$

The finding of  $2 \times 2$  criteria goes beyond the findings of Krauss [8], who only differentiates between stable (“Type I”) and unstable (“Type II”) maximum flow. Krauss



**Fig. 7.** Density profiles for  $(a, b, \epsilon) = (0.2, 0.6, 1.5)$  at times 250 000, 500 000, 750 000, and 1 000 000. Clearly, the interface keeps growing with time. For the system on the left in Fig. 6, we obtain a completely stationary interface profile, which also does not extend far into the system (not shown).

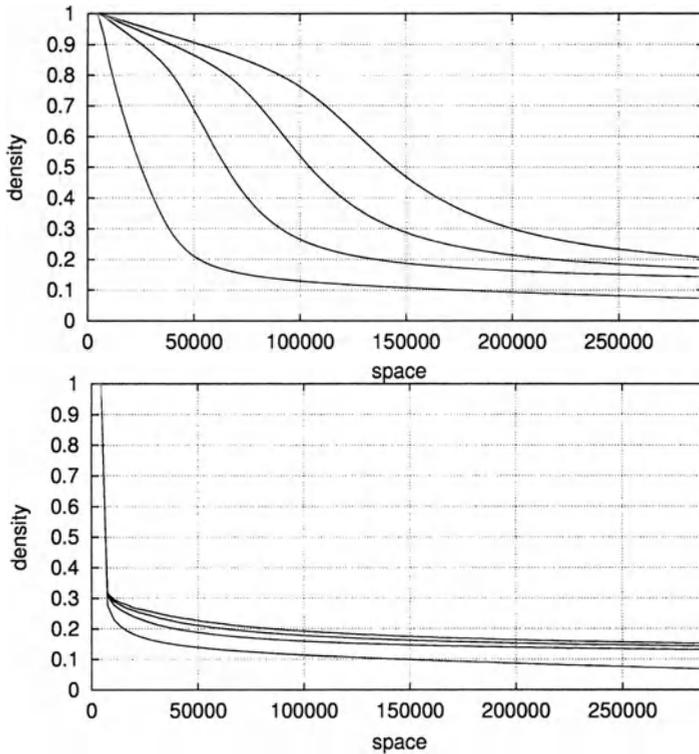
mentions “branching”, but more or less explicitly assumes that branching goes along with unstable maximum flow. In addition, the example of Ref. [8] for “branching”  $((a, b, \epsilon) = (1, \infty, 1))$  in fact has a stable interface as demonstrated in Fig. 6.

## 6 Discussion and Open Questions

There is controversy if cellular automata (CA) models for traffic show a first order phase transition,<sup>2</sup> a true critical phase transition [15], or none at all [16,17]. The discussion was seriously hampered by the fact that no parameter was known to change the possibly critical behavior of the system. Our findings, with a different type of model, shed new light on this discussion. It is plausible to assume that models with an unstable outflow interface display a crossover behavior, because any phase separation in the initial conditions will spread through the system – in a finite system, there would eventually be a macroscopically homogeneous state although there would be structure on the microscopic scale. Conversely, models with a stable outflow interface will display true macroscopic phases. Since the different phases are obtained by variations of continuous parameters, it should be possible (albeit computationally expensive) to find the line in phase space which separates the two regimes.

Additional simulations show that the stochastic traffic cellular automaton (STCA) of Ref. [13] has indeed both an unstable outflow and an unstable interface (Fig. 8 top). The so-called cruise control limit of this model [13] also has an unstable interface but a stable outflow, although marginally so. We also tested the so-called slow-to-start model [14] and found that it has, for the parameters that we tested, an unstable interface (Fig. 8 bottom). This puts it in a class separate from Krauss type I, in contrast to the original motivation that it would display the same kind of “meta”-stability as a Krauss type I model.

<sup>2</sup> In Ref. [10], Wolf reviews evidence for and against true phase coexistence, without making a judgment.



**Fig. 8.** Density profiles at times 250 000, 500 000, 750 000, and 1 000 000 for the stochastic traffic cellular automaton (STCA) of Ref. [13] (top) and for the slow-to-start model of Ref. [14] (bottom). Clearly, in both cases the interface is non-stationary.

Wolf, in Ref. [10], describes a so-called Galileo-invariant CA traffic model, where he observes a different type of meta-stability than the slow-to-start models. It is open into which of our four classes that model belongs.

In summary, it seems that our findings are finally the starting point of a more complete classification of the different models for traffic. Also from an engineering/applications perspective, it is necessary to solve these questions because of their consequences for real world applications. For example, the existence of stable high-flow states under noise would mean that it should be possible to stabilize these states in the real world. And an unstable outflow interface would imply different interpretations of real world data, which are typically averages over 1 minute or longer.

An implication of our findings is that, in contrast to earlier claims, outflow is not constant in STCA-type models: It is constant only outside the boundary region, which however grows to infinity. On the other hand, for both Krauss models of this paper, with  $(a, b) = (1, \infty)$  and  $(a, b) = (0.2, 0.6)$ , and noise  $\epsilon$  small enough, outflow is indeed a constant.

This implies that our theory about breakdown behavior in microscopic models needs to be revised. That theory was that there is a characteristic jam outflow, and any homogeneous solution with higher densities would be unstable against large amplitude

disturbances, such as stopping a vehicle and releasing it later. For models where the outflow is not well defined this is obviously too simplistic.

In addition, it seems that also for models with stable interfaces the situation can be more complicated. Our own simulations show that, essentially, the density between jams can be “compressed” in models with continuous variables. This is not discussed further here.

## 7 Summary

We have demonstrated that the breakdown of the homogeneous state in stochastic traffic models is characterized by two properties: (i) stability or not of the high flow states; (ii) stability or not of the outflow interface of jams. This is different from earlier findings, where it was assumed that the two go together. This is important, since it will allow to characterize the different existing traffic models according to these properties. It should also allow to eventually settle the controversy over the nature of the transition from homogeneous to congested flow. Engineering applications should benefit from these findings by being able to pick the model type which closest reflects reality. And finally, it is an interesting physical question since we are looking at simple one-dimensional driven systems which display interesting dynamics and which can be analyzed using the methods of statistical physics.

## 8 Acknowledgments

We had the suspicion that the Krauss model behaved differently from the slow-to-start rules of CA models for some time. This suspicion was reinforced by discussions with D. Helbing and with D. Wolf, who reported similar observations from some of their models. Additionally, simulations performed by N. Eissfeldt helped to clarify some of the questions related to this different behavior. Computing time of about 100 000 CPU-hours on the clusters Asgard and Xibalba, both at ETHZ, is acknowledged.

## References

1. B.S. Kerner and P. Konhäuser. *Structure and parameters of clusters in traffic flow*. Physical Review E **50**(1):54 (1994).
2. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama. *Dynamical model of traffic congestion and numerical simulation*. Physical Review E **51**(2):1035 (1995).
3. B.S. Kerner and H. Rehborn. *Experimental properties of complexity in traffic flow*. Physical Review E **53**(5):R4275–R4278 (1996).
4. D. Helbing. *Traffic and related self-driven many-particle systems*. Reviews of Modern Physics **73**:1067 – 1141 (2001).
5. H.Y. Lee, H.-W. Lee, and D. Kim. *Origin of synchronized traffic flow on highways and its dynamic phase transitions*. Physical Review Letters **81**(5):1130–1133 (1998).

6. D. Helbing and M. Treiber. *Gas-kinetic-based traffic model explaining observed hysteretic phase transition*. Physical Review Letters **81**(14):3042–3045 (1998).
7. B.S. Kerner. Presented at Traffic and Granular Flow '01 in Nagoya, Japan.
8. S. Krauß. *Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics*. PhD thesis, University of Cologne, Germany, 1997. See [www.zpr.uni-koeln.de](http://www.zpr.uni-koeln.de).
9. H. Reiss, A.D. Hammerich, and E.W. Montroll. *Thermodynamic treatment of nonphysical systems: Formalism and an example (single-lane traffic)*. Journal of Statistical Physics **42**(3/4):647–687 (1986).
10. D.E. Wolf. *Cellular automata for traffic simulations*. Physica A **263**:438–451 (1999).
11. E.M. Lifschitz and L.P. Pitajewski. Statistische Physik, Teil 1. Lehrbuch der Theoretischen Physik. Akademie-Verlag (1987).
12. F. Zielen and A. Schadschneider. *Nonsymmetric ergodicity breaking in a stochastic model on continuous phase space*. preprint (2001).
13. K. Nagel and M. Schreckenberg. *A cellular automaton model for freeway traffic*. Journal de Physique I France **2**:2221 (1992).
14. R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg. *Metastable states in cellular automata*. European Physical Journal B **5**(3):793–800, 10 (1998).
15. L. Roters, S. Lübeck, and K.D. Usadel. *Critical behavior of a traffic flow model*. Physical Review E **59**:2672 (1999).
16. M. Sasvari and J. Kertesz. *Cellular automata models of single lane traffic*. Physical Review E **56**(4):4104–4110 (1997).
17. D. Chowdhury et al. *Comment on: "Critical behavior of a traffic flow model"*. Physical Review E **61**(3):3270–3271 (2000).
18. K. Nagel and M. Paczuski. *Emergent traffic jams*. Physical Review E **51**:2909 (1995).

# Kinetic Theory of Traffic Flows

E. Ben-Naim<sup>1</sup> and P.L. Krapivsky<sup>2</sup>

<sup>1</sup> Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545

<sup>2</sup> Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215

**Abstract.** We describe traffic flows in one lane roadways using a kinetic theory, with special emphasis on the role of quenched randomness in the velocity distributions. When passing is forbidden, growing clusters are formed behind slow cars and the cluster velocity distribution is governed by an exact Boltzmann equation which is linear and has an infinite memory. The distributions of the cluster size and the cluster velocity exhibit scaling behaviors, with exponents dominated solely by extremal characteristics of the intrinsic velocity distribution. When passing is allowed, the system approaches a steady state, whose nature is determined by a single dimensionless number, the ratio of the passing time to the collision time, the two time scales in the problem. The flow exhibits two regimes, a laminar flow regime, and a congested regime where large slow clusters dominate the flow. A phase transition separates these two regimes when only the next-to-leading car can pass.

## 1 Introduction

Traffic flows are strongly interacting many-body systems. Therefore, theoretical techniques such as kinetic theory and hydrodynamics are useful in describing the rich phenomenology of traffic flows which includes shock waves, phase transitions, clustering, metastability, hysteresis, etc. Traffic is typically modeled within macroscopic descriptions such as hydrodynamics and kinetic theory [1–5] or microscopic approaches, e.g., cellular automata [6–12] and car-following models [13–17]. The large body of recent work on the physics of traffic flows is surveyed in Refs. [18–20].

In this review, we describe how quenched randomness in the car velocities leads to formation of clusters in one lane roadways. We assume ballistic motion with infinite memory, namely that each car has a preferred “intrinsic” velocity by which it drives in the absence of other cars [21–26]. While the emerging behavior is quite similar to that found in stochastic particle hopping processes with quenched disorder [27–33], these ballistic motion models often admit deeper analytical treatment. Our starting point is an idealized no passing flow, where an exact analytical solution is possible, and an exact kinetic theory can be constructed [22]. We then treat more realistic generalizations where passing is allowed using approximate kinetic theories [23–26]. Our goal is to provide a concise summary where key features are emphasized and outstanding open issues are highlighted.

## 2 No Passing Zones

Our basic traffic model mimics cluster formation (often also called platoon formation) in one-lane roadways where passing is forbidden [21,22]. In this model, each car moves ballistically at its initial velocity until it overtakes the preceding car or cluster. After this encounter, the incident car assumes the velocity of the cluster which it has just joined. Cars are taken to be size-less, and collisions to be instantaneous. We primarily consider spatially homogeneous situations where the positions and the velocities of the cars are initially uncorrelated. Specifically, cars are distributed randomly in space with a concentration  $c_0$ , and their velocities are drawn from the initial velocity distribution  $P_0(v)$ . Remarkably, analytic expressions can be obtained for the velocity distribution and the joint size-velocity distribution of clusters. Furthermore, it is also possible to describe analytically spatial inhomogeneities and even input of cars into the system.

### 2.1 The Velocity Distribution

We first consider the cluster velocity distribution. In this description, only the lead car is relevant and trailing cars in a cluster can be ignored. Let  $P(v, t)$  be the distribution of clusters with velocity  $v$  at time  $t$ . Initially, all cars are lead cars and the cluster (or lead-car) velocity distribution equals  $P(v, t) = P_0(v)S(v, t)$ , a product of the initial velocity distribution and the survival probability  $S(v, t)$ . The survival probability is the probability that a car with initial velocity  $v$  avoids “collisions” with slower cars up to time  $t$ , and hence, is still moving at the same velocity. Consider a car of initial velocity  $v$ . To ensure that it would not overtake slower cars of velocity  $v' < v$ , an interval of size  $(v - v')t$  ahead of it must not contain  $v'$ -cars initially. For the velocity distribution,  $P_0(v)$ , and a Poissonian initial spatial distribution, the probability for such an event is  $\exp[-t(v - v')P_0(v')]$ . For a car to survive to time  $t$ , this exclusion probability should be taken into account for every  $v' < v$ , and taking the product over all possible slower cars yields the survival probability,  $S(v, t) = \exp[-t \int_0^v dv' (v - v')P_0(v')]$ . Consequently, the cluster velocity distribution is found for arbitrary initial conditions

$$P(v, t) = P_0(v) \exp \left[ -t \int_0^v dv' (v - v')P_0(v') \right]. \quad (1)$$

The process is deterministic (the initial condition is the only source of randomness) and given the initial positions and velocities of the cars, the state of the system at any later time follows. This is reflected nicely in (1).

The exact solution (1) satisfies the following Boltzmann equation,

$$\frac{\partial P(v, t)}{\partial t} = -P(v, t) \int_0^v dv' (v - v')P(v', 0). \quad (2)$$

Interestingly, this rate equation is *linear* in the velocity distribution  $P(v, t)$ . The collision rate is proportional to the relative velocity,  $v - v'$ , and the *initial* velocity distribution of the slower cars, a signature of the infinite memory in the system. This is a unique case where the hierarchy of evolution equations corresponding to the velocity distributions terminates at the first order.

Whereas the steady state of one-lane traffic with no passing is trivial, viz. all cars will eventually join a cluster led by the slowest car in the system, the time dependent behavior is interesting. We concentrate on the long time behavior, which is largely independent of the initial distribution of fast cars, as follows from (1). For discrete velocity distributions, the time dependence is exponential, and we focus on continuous distributions. In this case, we find directly from (1) that both the cluster concentration,  $c(t) = \int dv P(v, t)$ , and the average velocity,  $\langle v(t) \rangle = c^{-1} \int dv v P(v, t)$ , decay algebraically with time

$$\begin{aligned} c(t) &\sim t^{-\alpha} & \alpha &= \frac{\mu + 1}{\mu + 2}, \\ \langle v(t) \rangle &\sim t^{-\beta} & \beta &= \frac{1}{\mu + 2}. \end{aligned} \quad (3)$$

The scaling exponents  $\alpha$  and  $\beta$  depend only on the small- $v$  extremal statistics [34] of the initial velocity distribution,

$$P_0(v) \simeq av^\mu \quad v \rightarrow 0, \quad (4)$$

via the cutoff exponent  $\mu > -1$ . The two scaling exponents are related by  $\alpha + \beta = 1$  as dictated by an elementary mean free path argument:  $cv t \sim 1$ .

Since the number of cars is conserved, the average cluster size is inversely proportional to the concentration,  $\langle m \rangle \sim 1/c$ , and the size growth law is  $\langle m \rangle \sim t^\alpha$ . In the limit  $\mu \rightarrow \infty$ , the size grows linearly with time. In contrast, when  $\mu \rightarrow -1$ , the size remains roughly constant, since the velocity distribution becomes effectively unimodal and collisions become exceedingly rare. This qualitative dependence on the form of the initial velocity distribution is reminiscent of ballistic annihilation processes, where ballistically moving particles annihilate upon collision [35–38]. The above clustering process can be viewed as a ballistic aggregation process that possesses a single mass conservation law. The sensitive dependence on the initial conditions is in contrast with momentum conserving ballistic agglomeration processes (that mimics large scale formation of matter in the universe) where a universal scaling asymptotic behavior emerges [39–41].

The average velocity is the only relevant velocity scale in the problem and asymptotically the velocity distribution follows the scaling form

$$P(v, t) \simeq t^{\beta-\alpha} \Phi(vt^{-\beta}). \quad (5)$$

From (1) and (4), the scaled distribution is  $\Phi(z) = az^\mu \exp(-bz^{\mu+2})$ , with  $b = a/[(\mu + 1)(\mu + 2)]$ . Therefore, the small- $v$  asymptotics of the initial velocity distributions governs not only the scaling exponents but also the entire

shape (including the large velocity tail) of the scaling function  $\bar{\Phi}(z)$ . This scaling behavior indicates that at time  $t$ , most cars moving initially with velocities larger than the typical velocity scale,  $\langle v(t) \rangle \sim t^{-\beta}$ , have already joined clusters led by slower cars, while cars slower than this velocity scale are still driving with their initial velocity.

## 2.2 The Size-Velocity Distribution

Given the nature of the model, a car is only affected by the initial configuration of cars *ahead* of it. This key feature enables solution of the cluster velocity distribution and it allows treatment of a more detailed quantity, the joint size-velocity distribution. To obtain  $P_m(v, t)$ , the density of clusters of size  $m$  and velocity  $v$ , it is useful to introduce the cumulative distribution,  $Q_m(v, t)$ , the distribution of clusters of velocity  $v$  containing at least  $m$  cars. Knowledge of this cumulative distribution yields the joint size-velocity distribution via differencing,  $P_m(v, t) = Q_m(v, t) - Q_{m+1}(v, t)$ .

Consider the first nontrivial quantity,  $Q_2(v, t)$ , the probability distribution of clusters of velocity  $v$  with at least two cars at time  $t$ . This quantity is equal to the product of the probability that lead car has survived up to time  $t$ ,  $P(v, t)$ , and the probability that the car trailing it actually experiences a collision prior to time  $t$ . Let  $x_1$  and  $v_1$  be the initial position and the initial velocity of the trailing car, respectively. For such a collision to occur, the trailing car must be faster than the lead car,  $v_1 > v$ , and the interval separating the two cars must be initially free of other cars. The probability for this composite event is the product of the probabilities of each individual event. Given a random (Poisson) spatial distribution, the probability an interval is empty is exponential in its length, and the collision probability is

$$Q_2(v, t) = P(v, t) \int_v^\infty dv_1 P_0(v_1) \int_{x_1 < (v_1 - v)t} dx_1 \exp(-x_1). \quad (6)$$

The fact that the trailing car cannot be slowed down by any other car before colliding with the lead car is crucial in obtaining this solution. This solution can be generalized to arbitrary cluster sizes. Following the two-car cluster case, one simply integrates over all the initial positions and velocities of the consecutive cars to eventually collide with the lead car in the cluster to give

$$Q_m(v, t) = P(v, t) \prod_{i=1}^{m-1} \int_v^\infty dv_i P_0(v_i) \int_{x_1 + \dots + x_i < (v_i - v)t} dx_i \exp(-x_i). \quad (7)$$

The integration limits reflect the fact that all the colliding cars must be faster than the lead car, and the restriction on the integration limits ensures that cars are sufficiently close to the lead car so that collisions indeed occur.

Given the cumulative car distribution, the joint size-velocity distribution can be formally obtained. Since  $\langle m \rangle \sim t^\alpha$  and  $\langle v \rangle \sim t^{-\beta}$ , we anticipate the

following scaling behavior:  $P_m(v, t) \simeq t^{\beta-2\alpha} \Psi(mt^{-\alpha}, vt^\beta)$ . This indeed holds and from (7) one obtains the scaled joint distribution

$$\Psi(x, z) = cz^\mu(x+z)^{\mu+1} \exp[-b(x+z)^{\mu+2}], \quad (8)$$

with  $c = a^2/(\mu+1)$ . Again only two parameters,  $a$  and  $\mu$  characterizing the small velocity characteristics of the initial conditions, are needed to fully describe the asymptotic state of the system. The joint distribution (8) provides a comprehensive description of the traffic clustering process. It may be considered as the counterpart of the well-known result for diffusion-controlled aggregation in one-dimension [42].

Integration of the scaling function with respect to  $x$  reproduces the scaled velocity distribution  $\Phi(z)$ . The complementary scaled size distribution cannot be found in a closed elementary form, except for the special case of asymptotically flat distributions ( $\mu = 0$ ) where both of the single variable scaling functions are purely Gaussian.

### 2.3 Generalizations

A natural generalization is to spatially heterogeneous initial velocity distributions,  $P_0(x, v)$ . The time and space dependent cluster velocity distribution,  $P(x, v, t)$ , follows from a straightforward generalization of the basic derivation in the homogeneous case

$$P(x, v, t) = P_0(x - vt, v) \exp \left[ - \int_0^v dv' \int_{x-vt}^{x-v't} dx' P_0(x', v') \right]. \quad (9)$$

For instance, consider the special case where cars are uniformly distributed in the region  $x \leq 0$  while the region ahead is empty. Here, one finds a governing length scale  $x \sim vt \sim t^\alpha$  with the same exponent  $\alpha$  as in (3). This length scale characterizes a propagating front of clusters, and the space dependent concentration  $c(x, t)$  becomes a function of the scaling variable  $X = xt^{-\alpha}$ , namely  $c(x, t) = t^{-\alpha} \mathcal{C}(X)$ . Far from the origin, the scaled density decays as  $\mathcal{C}(X) \sim X^{-1}$  implying  $c(x, t) \sim x^{-1}$  for  $x \gg t^\alpha$ . Consequently, the total number of clusters in the originally empty region,  $N(t) = \int_0^\infty dx c(x, t)$ , grows logarithmically slow with time

$$N(t) \sim \ln t. \quad (10)$$

This growth law is universal as the dependence on the details of the initial velocity distribution is secondary, entering only via the prefactor.

In summary, a scaling asymptotic behavior characterizes the kinetics of clustering in no-passing zones of one lane roadways. The corresponding scaling exponents and scaling functions are characterized by the small-velocity statistics of the initial velocity distributions. Remarkably, it is possible to derive the exact Boltzmann equation in this case.

### 3 Passing Zones

We now describe the complementary case of passing zones where fast cars can pass slow cars. The model we consider is a straightforward generalization of the no-passing case. The initial conditions are identical: cars are distributed randomly in space with concentration  $c_0$  and their velocity is drawn from the intrinsic velocity distribution  $P_0(v)$ . The characteristic velocity scale is taken to be  $v_0$ . In the absence of other cars, cars drive ballistically with their intrinsic velocity. In the presence of other cars, two competing mechanisms may cause a change in the car velocity. Collisions lead to slowing down: when a cluster overtakes a slower cluster, a larger cluster moving with the smaller of the two velocities forms. Passing leads to a velocity increase: every car inside a cluster may spontaneously pass the lead car and resume driving with its intrinsic velocity. The corresponding passing rate equals a constant,  $t_0^{-1}$ . This is a significant simplification; in realistic situations only the first few trailing cars may be able to pass.

It proves convenient to introduce dimensionless velocity, space, and time variables:  $v/v_0 \rightarrow v$ ,  $xc_0 \rightarrow x$ ,  $c_0v_0t \rightarrow t$ . This rescales the passing rate,  $t_0^{-1} \rightarrow R^{-1}$ , where

$$R = \frac{t_{\text{pas}}}{t_{\text{col}}} = c_0v_0t_0 \quad (11)$$

is the ratio of the passing time  $t_{\text{pas}} = t_0$  to the collision time  $t_{\text{col}} = (c_0v_0)^{-1}$ . We term this fundamental dimensionless quantity the ‘‘collision number’’ and denote it  $R$  as it is reminiscent of the Reynolds number – the small  $R$  limit is straightforward and the large  $R$  limit is characterized by boundary layers.

The starting point for kinetic theory is again the cluster velocity distribution  $P(v, t)$ . The approximate Boltzmann equation reads

$$\frac{\partial P(v, t)}{\partial t} = R^{-1} [P_0(v) - P(v, t)] - P(v, t) \int_0^v dv' (v - v') P(v', t). \quad (12)$$

This evolution equation assumes molecular chaos, namely that the stochastic passing events effectively mix the velocities, and therefore, spatial correlations can be neglected. This is clearly an approximation, as the collision integral does not coincide with the exact collision integral (2) derived in the no-passing case (the  $R \rightarrow \infty$  limit). Still, this term reflects the fact that collisions occur only with slower clusters and that the collision rate is proportional to the velocity difference. The passing term is exact since the concentration of slowed down cars with intrinsic velocity  $v$  equals  $P_0(v) - P(v, t)$ .

In contrast with the no-passing case, the process is now stochastic in nature and the system approaches a nontrivial steady state. Setting the time derivative in (12) to zero we see that the steady state cluster velocity distribution  $P(v) \equiv P(v, t = \infty)$  satisfies the integral equation

$$P(v) \left[ 1 + R \int_0^v dv' (v - v') P(v') \right] = P_0(v). \quad (13)$$

Given the intrinsic velocity distribution this relation gives the final cluster velocity distribution only implicitly. In contrast, the inverse problem is simpler as knowledge of the final distribution, the observed quantity in real traffic flows, gives explicitly the intrinsic distribution. We confirm that in the limit  $R \rightarrow \infty$ , all clusters move with the minimal velocity, while in the limit  $R \rightarrow 0$ , all cars move with their intrinsic velocity  $P(v) \rightarrow P_0(v)$ .

The integral equation (13) can be transformed into a differential one using the auxiliary function  $Q(v) = R^{-1} + \int_0^v dv'(v - v')P(v')$ , from which  $P(v) = Q''(v)$ . Thus (13) becomes

$$Q(v)Q''(v) = R^{-1}P_0(v). \quad (14)$$

The boundary conditions are  $Q(0) = R^{-1}$  and  $Q'(0) = 0$ . The auxiliary function  $Q(v)$  gives a comprehensive description of the steady state. Calculation of important quantities such as the flux  $J$  requires knowledge of  $G(v)$ , the car velocity distribution. This quantity satisfies  $1 = \int dv G(v)$  and  $J = \int dv v G(v)$ . Following an involved calculation that requires solution of a higher order velocity distribution [23], the car velocity distribution is derived explicitly in terms of  $Q(v)$

$$G(v) = P(v) \left[ 1 + R \int_v^\infty dw P_0(w) \int_v^w \frac{du}{[RQ(u)]^2} \right]. \quad (15)$$

Hence, for arbitrary intrinsic velocity distributions, the entire steady state problem is reduced to the nonlinear second order differential equation (14). Given  $Q(v)$ , steady state distributions such as  $P(v)$  and  $G(v)$  can be calculated using the explicit formulas above.

Except for a few special cases, one can not solve the differential equation (14) analytically. Nevertheless, the formal solution above can be used to evaluate generic features of the flow. The dimensionless collision number  $R$  is extremely useful. For low collision numbers, a perturbation solution in powers of  $R$  can be constructed, as the steady state differs weakly from the initial state. For high collision numbers, a boundary layer analysis is possible as sufficiently small velocities are not affected by collisions. These two limits are quantitatively analyzed as follows.

### 3.1 Low Collision Numbers

The flow characteristics in the collision-controlled regime,  $R \ll 1$ , can be analyzed systematically as a perturbation series in  $R$ . For example, the cluster velocity distribution and the car velocity distribution read

$$\begin{aligned} P(v) &\cong P_0(v) \left[ 1 - R \int_0^v dv'(v - v')P_0(v') \right], \\ G(v) &\cong P_0(v) \left[ 1 + R \int_0^\infty dv'(v' - v)P_0(v') \right]. \end{aligned} \quad (16)$$

Consequently, average quantities such as the flux and the average cluster size vary linearly in  $R$  in this free flow regime,  $J = J_0 - \text{const} \times R$ , and  $\langle m \rangle = 1 + \text{const} \times R$ . The proportionality constant in the case of the flux equals the variance in the initial velocity distribution, indicating that the larger the initial velocity fluctuations, the larger the reduction in the flux.

Therefore, weakly interacting “laminar” flows arise in the  $R \rightarrow 0$  limit. Technically, the steady state remains close to the initial state and a perturbation series in the collision number is possible. Here, the assumptions made in our theory are justified, as the cluster sizes are small, and at the leading order, a simplified model where all cars in the cluster can pass coincides with a more realistic model where only the first few cars may pass. In fact, a basic prediction of the model, namely linear growth of the average cluster size with the flux is consistent with empirical data, obtained from observations of traffic flows in a secondary rural road in Los Alamos, New Mexico [23].

### 3.2 High Collision Numbers

The limit of high collision numbers corresponds to dense, congested flows where large clusters form. The analysis in this passing-controlled regime is more subtle since the condition  $R \int_0^v dv' (v - v') P_0(v') \ll 1$  is satisfied only for small velocities. No matter how large  $R$  is, sufficiently slow cars are not affected by collisions, and  $P(v)$  is still given by (16) when  $v \ll v^*$ . The threshold velocity  $v^* \equiv v^*(R)$  is estimated from  $R \int_0^{v^*} dv (v^* - v) P_0(v) \sim 1$ .

In the limit  $R \rightarrow \infty$  limit, statistics of the slowest cars dominate the flow. Again, it is useful to consider intrinsic distributions with an algebraic small velocity form (4). For such distributions, the threshold velocity decreases with growing  $R$  according to

$$v^* \sim R^{-\frac{1}{\mu+2}}. \quad (17)$$

One can show that the flux is proportional to this velocity  $J \sim v^*$ . For  $v \gg v^*$ , the collision integral in (13) dominates over the constant factor and  $RP(v) \int_0^v dv' (v - v') P(v') \sim v^\mu$ . Anticipating an algebraic behavior for the cluster velocity distribution,  $P(v) \sim R^\sigma v^\delta$  when  $v \gg v^*$ , gives different answers dependent on whether the cutoff exponent  $\mu$  is positive or negative. The leading behavior for  $v \gg v^*$  can be summarized as follows

$$P(v) \sim \begin{cases} (v^*)^\mu (v/v^*)^{\mu-1} & \mu < 0; \\ (v/v^*)^{-1} [\ln(v/v^*)]^{-\frac{1}{2}} & \mu = 0; \\ (v^*)^\mu (v/v^*)^{\frac{\mu}{2}-1} & \mu > 0. \end{cases} \quad (18)$$

On the other hand,  $P(v) \cong P_0(v)$  for  $v \ll v^*$ . This shows that the velocity distribution develops a boundary layer structure, the size of which vanishes in the infinite collision number limit. Inside the boundary layer, the velocity distribution is only marginally lower than its initial values, while the bulk of the velocities are strongly suppressed. Similar to the threshold velocity

$v^*$ , macroscopic characteristics of the flow depend algebraically on  $R$ . For example, the average cluster size is

$$\langle m \rangle \sim \begin{cases} R^{(\mu+1)/(\mu+2)} & \mu < 0; \\ (R/\ln R)^{1/2} & \mu = 0; \\ R^{1/2} & \mu > 0. \end{cases} \quad (19)$$

Two distinct regimes of behavior emerge. For  $\mu > 0$ , car-cluster collisions dominate while for  $\mu < 0$  cluster-cluster collisions dominate. Interestingly, in the cluster-cluster dominated regime,  $\langle m \rangle \sim R^\alpha$  with the scaling exponent  $\alpha = (\mu + 1)/(\mu + 2)$  as in the no-passing case (3). Thus in the passing case the cutoff exponent  $\mu$  also plays an important role in characterizing the behavior. Moreover, the steady state behavior is much richer than that found for the clustering kinetics.

Despite the simplifying assumptions, the model results in realistic behavior. The overall picture is both familiar and intuitive: due to the presence of slower cars, clusters form and the overall flux is reduced. For heavy traffic, the characteristics of the flow are solely determined by the distribution of slow cars. A single dimensionless parameter, the collision number  $R$ , ultimately determines the nature of the steady state.

## 4 The Maxwell Model

While a comprehensive analysis of the steady state velocity distributions is possible using the approximate kinetic theory (12), other important questions such as the relaxation toward the steady state and the nature of the cluster size distribution [43] remain unanswered. To address these issues we consider a further approximation where the collision rate is taken to be uniform [25,26]. This approximation, known as the Maxwell model, is very useful in kinetic theory [44] and it has been recently applied to granular gases as well [45–47]. In our case, it allows for a complete exact solution of the time dependent behavior, and additionally, it leads to closed evolution equations for the cluster-size distribution.

### 4.1 Relaxation

In the Maxwell approximation, the collision rate  $v - v'$  in the Boltzmann equation (12) is replaced by a constant factor which we set equal to unity. The corresponding rate equation for the cluster velocity distribution reads

$$\frac{\partial P(v, t)}{\partial t} = R^{-1} [P_0(v) - P(v, t)] - P(v, t) \int_0^v dv' P(v', t). \quad (20)$$

Again, the analysis is performed via a properly defined auxiliary function,  $Q(v, t) = \int_0^v dv' P(v', t)$ . The constant collision rate results in simpler differential equations, that are only first order in the velocity. The analog of (14)

is the integrable steady state equation  $Q(v)Q'(v) = R^{-1}P_0(v)$ . The resulting steady state properties are governed by  $R$ , with a boundary layer structure in the large  $R$  regime. The quantitative characteristics are somewhat different and for example the threshold velocity decays with  $R$  according to  $v^* \sim R^{-1/(\mu+1)}$  rather than (17). If, however, the collision rate is properly chosen, namely set equal to  $\langle v \rangle$  rather than unity, we recover (17).

Furthermore, the complete time dependence can be obtained analytically by integrating the partial differential equation  $Q_t = R^{-1}Q_v - QQ_v$ . In general, the relaxation is exponential  $P(v, t) - P(v, t = \infty) \sim f(v)e^{-t/\tau(v)}$ , with  $\tau(v) = R[1 + 2RI_0(v)]^{-1/2}$  where  $I_0(v) = \int_0^v dv' P_0(v')$ . The relaxation time depends on the velocity and the collision number according to

$$\tau(v) \sim \begin{cases} R & v \ll v^*; \\ [R/I_0(v)]^{1/2} & v \gg v^*. \end{cases} \quad (21)$$

While small velocities are governed by practically fixed relaxation times, large velocities are characterized by velocity dependent decay rates. Furthermore, a large range of relaxation scales exists,  $R^{1/2} < \tau < R$ , with larger scales corresponding to smaller velocities. Further analysis shows that the same relaxation times underlie the car velocity distribution. We expect that while the predictions of the Maxwell model are only approximate, it correctly predicts the existence of a spectrum of relaxation time scales, and that the qualitative nature of the time dependent behavior generally holds.

## 4.2 The Size Distribution

The size distribution obeys closed evolution equations in the Maxwell model and can be solved exactly [25,26]. It can also be used to address the nature of the passing mechanism. To demonstrate this we consider the model where only the next-to-leading car in the cluster may pass and resume driving with its intrinsic velocity. From numerical simulations of this model, we find two distinct phases. In the laminar regime, clusters are generally small, specifically the cluster size distribution is exponentially suppressed for sufficiently large sizes. When the collision number exceeds a certain threshold, an infinite cluster is formed, i.e., a finite fraction of the cars in the system are in the cluster behind the slowest car. Furthermore, in this jammed phase the size distribution of finite clusters has a fat tail close to a power-law,  $P_m \sim m^{-\tau}$ , with  $\tau \approx 2$ .

In the Maxwell model framework, the cluster size distribution  $P_m(t)$  obeys a closed system of rate equations

$$\frac{dP_m}{dt} = R^{-1}[P_{m+1} - P_m] - cP_m + \frac{1}{2} \sum_{i+j=m} P_i P_j, \quad m \geq 2, \quad (22)$$

$$\frac{dP_1}{dt} = R^{-1}[P_2 - P_1 + c] - cP_1. \quad (23)$$

These equations were derived by enumerating all possible ways in which clusters evolve. For instance, consider (23). Collisions reduce the density of single cars, and the collision rate is clearly equal to  $c$ , as it is velocity-independent. The escape term in (23) is understood by observing that the rate of return of single cars into the system is  $2P_2 + \sum_{j \geq 3} P_j = P_2 - P_1 + c$ . Here  $P_2$  is singled out since passing transforms it into two single cars while an escape from larger clusters produces only one freely moving car.

Similar equations were previously studied in the context of aggregation-fragmentation processes [48–50]. Utilizing the approach of Ref. [48] we find that a phase transition occurs at  $R_c = 1$  [26]. For large  $m$ , the steady state size distribution is

$$P_m \sim \begin{cases} m^{-3/2} [1 - (1 - R)^2]^m & R < 1, \\ m^{-5/2} & R \geq 1. \end{cases} \quad (24)$$

Hence in the laminar regime, the size distribution decays exponentially in the large size limit. In the congested phase, the size distribution has a power law tail, and in addition there is an infinite cluster that contains the following finite fraction of cars in the system:

$$I = \begin{cases} 0, & R < 1; \\ 1 - R^{-1}, & R > 1. \end{cases} \quad (25)$$

Interestingly, this phase transition is similar to phase transitions in driven diffusive systems *without* passing [28–32]. Furthermore, the formation of an infinite cluster is reminiscent of Bose-Einstein condensation [30,49].

## 5 Discussion

The most important question raised by the above results concerns the validity of the “mean-field” Boltzmann equation (12). Although passing is a stochastic mixing mechanism that diminishes correlations between the velocities and the positions of the cars, such correlations do exist, and it will be interesting to determine whether quantitative predictions such as the scaling behaviors (17) and (19) are altered by spatial correlations. Similarly, the collision term in (22) is written in a mean-field spirit and that may be the reason for the discrepancy between the theoretical prediction  $\tau = 5/2$  and the numerically observed value  $\tau \approx 2$  of the decay exponent  $P_m \sim m^{-\tau}$ .

The primary feature of our model is *quenched disorder*, which manifests itself in the random assignment of intrinsic velocities. Road conditions (construction zones, intermittent passing zones, turns, hills, etc.) present another source of quenched randomness in actual roads [51], which is ignored in our model. Quenched disorder significantly affects characteristics of many-particle systems, especially in low spatial dimensions [52]. We have seen that this general conclusion clearly applies to our one-dimensional traffic model. Little is known analytically on the influence of the spatial disorder.

Finally, one may modify the passing rule so that when a car overtakes a slow car, it acquires a new velocity drawn from the distribution  $P_0(v)$  rather than a pre-assigned velocity [53]. This elementary zero-memory model remains highly non-trivial even in the collision-controlled limit  $R \rightarrow 0$  where clustering can be disregarded. The fate of the system is again determined by the behavior of the intrinsic velocity distribution near its lower cutoff. If  $P_0(v)$  vanishes in this limit, the system reaches a steady state, otherwise, the system evolves indefinitely. Specifically, for intrinsic distributions with an algebraic small velocity tail (4) the long-time asymptotics of the average velocity reads

$$\langle v(t) \rangle \sim \begin{cases} \text{const} & \mu > 0; \\ (\ln t)^{-1} & \mu = 0; \\ t^\mu & -1 < \mu < 0. \end{cases} \quad (26)$$

These results were derived in a simplified Boltzmann framework. In particular, the most interesting behavior in the evolving regime was obtained by assuming that as  $t \rightarrow \infty$ , cars can be divided into two groups, the small group of “active” cars which move with velocities  $v \sim 1$  and the vast majority of “creeping” cars that hardly move at all. We then ignored collisions between creeping cars (since their relative velocity is very small) and collisions between active cars (since their density is small). Thence, the velocity distribution of active cars obeys a linear Boltzmann-Lorentz equation which was solved to give (26). Comparison with results of molecular dynamics simulations suggests that the mean-field theory description is asymptotically exact. It will be interesting to confirm this result rigorously.

We are thankful to our collaborators Slava Ispolatov and Sid Redner. This research was supported by DOE (W-7405-ENG-36) and NSF (DMR9978902).

## References

1. M. J. Lighthill and G. B. Whitham, Proc. Roy. Soc. London Ser. A **229**, 281 (1955); *ibid* **229**, 317 (1955).
2. R. Haberman, *Mathematical Models in Mechanical Vibrations, Population Dynamics, and Traffic Flow* (Prentice-Hall, Englewood Cliffs, NJ, 1977).
3. I. Prigogine and R. Herman, *Kinetic Theory of Vehicular Traffic* (Elsevier, New York, 1971).
4. S. L. Paveri-Fontana, Transp. Res. **9**, 225 (1975).
5. B. S. Kerner and P. Konh user, Phys. Rev. E **48**, 2335 (1993); Phys. Rev. E **50**, 54 (1994).
6. K. Nagel and M. Schreckenberg, J. de Physique I **2**, 2221 (1992).
7. O. Biham, A. Middleton, and D. Levine, Phys. Rev. A **46**, 6124 (1992).
8. T. Nagatani, Phys. Rev. E **48**, 3290 (1993).
9. A. Schadschneider and M. Schreckenberg, J. Phys. A **26**, L679 (1993).
10. K. Nagel and M. Paczuski, Phys. Rev. E **51**, 2909 (1995).
11. K. Nagel, Phys. Rev. E **53**, 4655 (1996).

12. J. G. Brankov, V. B. Priezhev, A. Schadschneider, and M. Schreckenberg, *J. Phys. A* **29**, L229 (1996).
13. L. A. Pipes, *J. Appl. Phys.* **24**, 274 (1953).
14. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995).
15. T. S. Komatsu and S. Sasa, *Phys. Rev. E* **52**, 5574 (1995).
16. Y. Sugiyama and H. Yamada, *Phys. Rev. E* **55**, 7749 (1997).
17. T. Nagatani, *J. Phys. Soc. Jpn.* **66**, 1219 (1997).
18. *Traffic and Granular Flow II*, edited by D. E. Wolf and M. Schreckenberg (World Scientific, Singapore, 1998).
19. D. Chowdhury, L. Santen, and A. Schadschneider, *Phys. Rep.* **329**, 199 (2000).
20. D. Helbing, *Rev. Mod. Phys.* **73**, 1067 (2001).
21. G. F. Newell, *Oper. Res.* **7**, 589 (1959).
22. E. Ben-Naim, P. L. Krapivsky, and S. Redner, *Phys. Rev. E* **50**, 822 (1994).
23. E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **56**, 6680 (1997).
24. E. Ben-Naim and P. L. Krapivsky, *J. Phys. A* **31**, 8073 (1998).
25. E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **59**, 88 (1999).
26. I. Ispolatov and P. L. Krapivsky, *Phys. Rev. E* **62**, 5935 (2000).
27. T. Nagatani, *Phys. Rev. E* **51**, 922 (1995).
28. I. Benjamini, P. A. Ferrari, and C. Landim, *Stoch. Proc. Appl.* **61**, 181 (1996).
29. J. Krug and P. A. Ferrari, *J. Phys. A* **29**, L465 (1996).
30. M. R. Evans, *Europhys. Lett.* **36**, 13 (1996); *J. Phys. A* **30**, 5669 (1997).
31. D. V. Kvitarev, D. Chowdhury, and D. E. Wolf, *J. Phys. A* **30**, L221 (1997).
32. J. Krug, in Ref. [18], p. 285.
33. T. Seppäläinen and J. Krug, *J. Stat. Phys.* **95**, 525 (1999).
34. The importance of the extreme statistics on cluster formation in tunnels was first noticed by G. F. Newell in Ref. [21].
35. Y. Elskens and D. L. Frisch, *Phys. Rev. A* **31**, 3812 (1985).
36. E. Ben-Naim, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **70**, 1890 (1993).
37. M. Droz, P. Rey, L. Frachebourg, and J. Piasecki, *Phys. Rev. E* **51**, 5541 (1995).
38. P. L. Krapivsky and C. Sire, *Phys. Rev. Lett.* **86**, 2494 (2001).
39. S. F. Shandarin and Ya. B. Zeldovich, *Rev. Mod. Phys.* **61**, 185 (1989).
40. G. F. Carnevale, Y. Pomeau, and W. R. Young, *Phys. Rev. Lett.* **64**, 2913 (1990).
41. L. Frachebourg, *Phys. Rev. Lett.* **82**, 1502 (1999).
42. J. L. Spouge, *Phys. Rev. Lett.* **60**, 871 (1988).
43. K. Gavrilov, *Phys. Rev. E* **56**, 4860 (1997).
44. For a review of the Maxwell model, see M. H. Ernst, *Phys. Reports* **78**, 1 (1981); A. V. Bobylev, *Sov. Sci. Rev. C. Math. Phys.* **7**, 111 (1988).
45. E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **61**, R5 (2000); P. L. Krapivsky and E. Ben-Naim, *J. Phys. A* **35**, L147 (2002); E. Ben-Naim and P. L. Krapivsky, *cond-mat/0202332* and *cond-mat/0203099*.
46. J. A. Carrillo, C. Cercignani, and I. M. Gamba, *Phys. Rev. E* **62**, 7700 (2000).
47. A. Baldassarri, U. M. B. Marconi, and A. Puglisi, *cond-mat/0111066*.
48. P. L. Krapivsky and S. Redner, *Phys. Rev. E* **54**, 3553 (1996).
49. S. N. Majumdar, S. Krishnamurthy, and M. Barma, *Phys. Rev. Lett.* **81**, 3691 (1998).
50. R. Rajesh and S. N. Majumdar, *Phys. Rev. E* **62**, 7735 (2000).

51. Z. Csahok and T. Vicsek, *J. Phys. A* **27**, L591 (1994); O. J. O'Loan, M. R. Evans, and M. E. Cates, *Phys. Rev. E* **58**, 1404 (1998); G. Tripathy and M. Barma, *Phys. Rev. E* **58**, 1911 (1998); J. Krug, *cond-mat/9912411*.
52. J. P. Bouchaud and A. Georges, *Phys. Rep.* **195**, 127 (1990).
53. I. Ispolatov and P. L. Krapivsky, *Phys. Rev. E* **61**, R2163 (2000).

# The Generalized Fundamental Diagram of Traffic and Possible Applications

E. Tomer<sup>1</sup>, L.A. Safonov<sup>1,2</sup>, and S. Havlin<sup>1</sup>

<sup>1</sup> Minerva Center and Department of Physics, Bar-Ilan University,  
52900 Ramat-Gan, Israel

<sup>2</sup> Department of Applied Mathematics and Mechanics, Voronezh State University,  
394693 Voronezh, Russia

**Abstract.** We propose a new optimization strategy based on inducing stop-and-go waves on the main road and controlling their wavelength. Using numerical simulations of a recent stochastic car-following model [11] we show that this strategy yields optimization of traffic flow in systems with a localized periodic inhomogeneity, such as signalized intersections and entry ramps. The optimization process is explained by our finding of a generalized fundamental diagram (GFD) for traffic, namely a flux-density-wavelength relation. Projecting the GFD on the density-flux plane yields a two-dimensional region of stable states, qualitatively similar to that found empirically [7] in synchronized traffic. The empirical finding of the dependence of the wavelength on the average velocity can also be explained using the same approach.

## 1 Introduction

Traffic flow has been a subject of comprehensive study for more than half a century, [1–27] due to its theoretical and practical importance. Recently this field has attracted much interest, especially after new empirical and theoretical studies have shown its clear relation to physical phenomena of current interest, such as phase transitions, critical phenomena, nonlinear dynamics, and chaos (for reviews see e.g. [1–3]).

One of the main open questions in this field regards the validity of the basic concept of the 'fundamental diagram' [4–6] – a functional relation between the flux and the density of cars. This generally recognized relation, used in almost every study in this field, was challenged recently by empirical findings of Kerner [7,8] indicating that such a fundamental diagram does not exist. Instead, stable synchronized traffic states display a two-dimensional region in the density-flux plane. Consequently, the wide scattering of data points representing congested traffic cannot be attributed only to measurement fluctuations, but also to the existence of a range of stable flux values for a given density. The existence of a range of flux values gives rise to a possibility of manipulating the system in order to achieve the highest possible flux.

Other empirical observations show that traffic flow demonstrates complex physical phenomena, both on macroscopic and microscopic scales. Among these experimental findings are: (i) The existence of three phases in traffic flow [9]: Free flow, synchronized flow, and traffic jams; (ii) Metastable free flow in some

intermediate range of densities, and hysteresis in transitions between free and synchronized flow [9]; (iii) Periodic wave motion in stop-and-go traffic, characterized by a typical time period (see [10]), and a 'wavelength' (distance between two nearest narrow jams) [7]. The wavelength is an increasing function of the average velocity [7] for the intermediate range of velocities.

Recently, a stochastic inertial car-following model that may explain (i)-(ii) and other empirical results was proposed [11–13]. The deterministic version of the model shows transitions between three phases with increasing the density: homogeneous free flow, stop-and-go waves, and congested homogeneous flow. Free flow is metastable for some range of densities, leading to hysteresis in the transitions to congested traffic. Moreover, multiple inhomogeneous congested traffic states coexist in the intermediate regime [11]. These stop-and-go waves are periodic for homogeneous systems with periodic boundary conditions, and differ in their wavelength. These states correspond to limit cycles in the phase space. They can be found analytically near the bifurcation points, and can be traced numerically far from this point [12]. Transitions between these states can be induced by noise.

Earlier studies of microscopic car-following models, such as the optimal velocity model (OVM) [14], also report on stable congested states corresponding to limit cycles in phase space. These limit cycles were found to be universal objects [15]. In this model, as well as in other microscopic models such as the Nagel-Schreckenberg (NS) cellular-automata model [16], the flux  $f$  does not depend on the exact configuration of the congested state, and  $f$  is a function only of the density  $\rho$ . In general, traffic flow models predict (or assume) a single-value or double-value function  $f(\rho)$  – namely a fundamental diagram – unlike the experimental findings of a  $2D$  region in the density-flux plane mentioned above.

In this paper we propose a possible explanation to the experimental findings showing that such a fundamental diagram does not exist. Studying the deterministic version of our model [11] we find a  $2D$  region in the density-flux plane that is caused by the existence of a hidden parameter, namely, the wavelength of the stop-and-go states. Therefore, the fundamental diagram  $f(\rho)$  has to be generalized to a density-wavelength-flux relation. We call this relation the *generalized fundamental diagram* (GFD) of traffic. We further analyze the stability of the stop-and-go states, and show that the noise stability threshold is a function of the wavelength – a function which has a single maximum. The wavelength of this most stable state is an increasing function of the average velocity for the intermediate range of velocities – qualitatively similar to that found experimentally, see (iii) above.

Finally, we propose a novel optimization strategy, which aims to optimize traffic by approaching the highest flux values in this  $2D$  region. Our approach is based on inducing stable stop-and-go waves on the main road and controlling their wavelength. It is shown that this strategy yields optimization of traffic flow when implemented in systems with a localized periodic inhomogeneity, such as signalized intersections and entry ramps, in cases of over-saturation. The promising numerical results for optimization are not predicted by common theories. We

explain the optimization process using our finding of a GFD, together with the finding of the existence of noise stability threshold mentioned above. If approved by field experiments, this method would enable to make use of the existing infrastructure to further optimize congested traffic, especially for high densities where congestion cannot be relieved.

The paper is organized as follows. The car-following model is presented in Sec. 2, where the relation to NS-type models is also discussed. In Sec. 3 the GFD is derived and a  $2D$  region in the density-flux plane is found. Sec. 4 discusses the influence of noise, and a qualitative agreement with the experimental findings (iii) above is shown in Sec. 5. The new theoretical results are then applied to propose a novel optimization method in Sec. 6.

## 2 The Model

In microscopic traffic model approaches [11–24], traffic is treated as a flow of interacting particles. In particular, car-following models [11–15,20–23] are usually defined by a set of ordinary differential equations, each of them describing the motion of a single car. In inertial car-following models, these equations, which are usually of second order, describe the interaction between following cars. This interaction is defined by a relation between the acceleration  $a_n$  of the  $n$ 'th car, its headway  $\Delta x_n$ , velocity  $v_n$ , and velocity difference with the car ahead  $\Delta v_n$ , i.e.,  $a_n = a(\Delta x_n, v_n, \Delta v_n)$ . Newton's third law is not satisfied for these type of interactions, since a leading car affects its follower, but not vice versa.

We assume [11] that the acceleration of each car is determined by four factors:

- (a) desire to maintain a safety time gap from the car ahead,
- (b) pre-braking if the car ahead is much slower,
- (c) desire not to exceed significantly the speed limit,
- (d) random noise.

These four assumptions are represented by four corresponding terms in the equations of motion,

$$a_n = a_{(1)}(\Delta x_n, v_n) + a_{(2)}(\Delta v_n, \Delta x_n) + a_{(3)}(v_n) + \eta'. \quad (1)$$

In the following we derive each term in a simple way that still maintain the essential characteristics of human driving described by assumptions (a)-(d). We first consider a single lane homogeneous system with periodic boundary conditions, length  $L$ , having  $N$  vehicles. It is assumed that all drivers have the same parameters (identical drivers). Initial conditions can be homogeneous, random, or periodic.

**(a) Maintaining safety time gap:** To define the first term in the equation of motion (1),  $a_{(1)}(\Delta x_n, v_n)$ , let us assume that all drivers wish to maintain a safety time gap  $T$  from the car ahead when driving in a platoon. This time gap  $T$  is an estimation of the maximal response time of a driver. A time gap of approximately  $T = 2sec$  is usually recommended for safe driving. Maintaining

this time gap, a driver can react safely even to sharp braking of the car ahead, since that by the end of this time  $T$  he starts braking when he is at the same position as was the car ahead when it started braking. Thus, the optimal headway is

$$\Delta x_n^0 = v_n T + D, \tag{2}$$

where  $D$  is the minimal distance between consecutive cars (car length plus bumper to bumper distance).

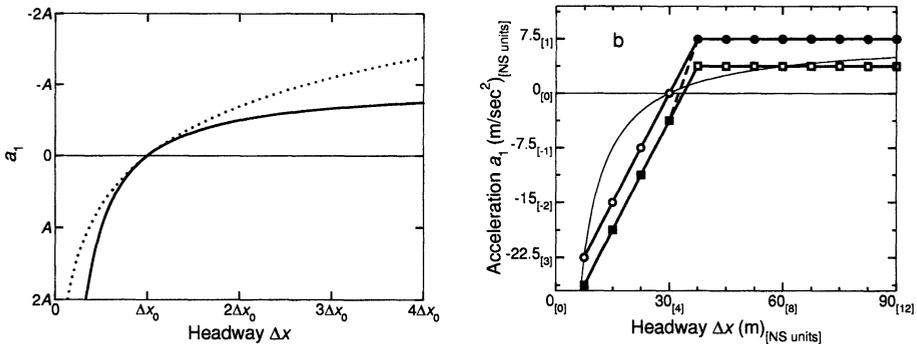
Following the above considerations, we can now define  $a_{(1)}(\Delta x_n)$ . A realistic relation between acceleration and headway should have the following characteristics:

- (i) Zero acceleration for the optimal headway:  $a_{(1)}(\Delta x_n = \Delta x_n^0) = 0$ .
- (ii) Small headway yield sharp braking: for  $\Delta x_n \ll \Delta x_n^0$ ,  $a_{(1)}(\Delta x_n) \rightarrow -\infty$ .
- (iii)  $a_{(1)}$  is an increasing function of  $\Delta x_n$ .

The simplest choice of  $a_{(1)}(\Delta x_n)$  that satisfies (i)-(iii) is

$$a_{(1)} = A \left( 1 - \frac{\Delta x_n^0}{\Delta x_n} \right), \tag{3}$$

where  $A$  is a sensitivity parameter (aggressive drivers can be represented by relatively high values of  $A$ ). In contrast to the optimal velocity model [14], where the acceleration depends on the difference between  $v_n$  and an optimal velocity  $v_0(\Delta x_n)$ , here the acceleration depends on the difference between  $\Delta x_n$  and the optimal headway  $\Delta x_n^0(v_n)$ . The explicit dependence on the headway  $\Delta x_n$  enable us to make the model free of frequent collisions for every choice of parameters values (see also [21,14]).



**Fig. 1. a.** A plot of the function  $a_{(1)}(\Delta x_n)$  given by (3) (solid curve), and of the that given by the logarithmic expression (4) (dotted curve). **b.** The relation between acceleration and headway  $a_{(1)}(\Delta x_n)$  for the deterministic NS model with  $v_{max} = 5$  and  $v_n = 3$  (circles), for the stochastic NS model with the same parameters and  $p = 0.5$  (squares), for the NP model with the same parameters (full symbols, dashed curve), and for our model (3) with the corresponding parameters (thin curve).

As a matter of fact, other possible choices of  $a_{(1)}$  that satisfies (i)-(iii) can also be defined. To find out how our results are affected by the choice of  $a_{(1)}$  we also consider a simple logarithmic relation,

$$a_{(1)} = A \log \left( \frac{\Delta x_n}{\Delta x_n^0} \right). \tag{4}$$

The functions  $a_{(1)}(\Delta x_n)$  given by (3)-(4) are plotted in Fig. 1a. The qualitative (and sometimes even quantitative) characteristics of the stable states of the model are independent of the exact functional choice of  $a_{(1)}$ . In both cases the solutions are fixed points (homogeneous flow) and limit cycles (stop-and-go waves). The range of density values for which homogeneous solution is stable is exactly the same in both cases. This is because the slope  $a'_{(1)}(\Delta x_n^0)$  is the only important property of this function for determining this stability range (see [11]), and it is equal to  $A$  for both choices of  $a_{(1)}$ . In addition, our finding of the generalized fundamental diagram (Sec. 3) is valid for both choices. Therefore, it can be assumed that the results presented below are independent on the exact choice of  $a_{(1)}$  that satisfies conditions (i)-(iii). In the following we use the function given by Eq. (3).

**(b) Pre-braking:** The first term  $a_{(1)}$  in (1) might be sufficient when cars move in a platoon, so that following cars have approximately the same speed. But when a fast car is approaching a much slower car, i.e.  $\Delta v_n < 0$  and  $|\Delta v_n| \sim v_{n+1}$ , this term is not enough (see below). Thus, it is necessary to introduce a second term,  $a_{(2)}(\Delta v_n)$ , in the equation of motion. This term represents the additional negative acceleration caused by this negative velocity difference. If such a term is not included, a fast car approaching a standing car, for example, would start braking only at about time  $T$  before crashing into it – which is, of course, unrealistic.

To determine this term, let us define  $x'_n$  and  $v'_n$  as the position and velocity of the  $n$ 'th car in the system of coordinates of the  $n + 1$ 'th car. Clearly  $x'_n = -\Delta x_n$ , and  $v'_n = -\Delta v_n$ . We discuss here the case of  $v'_n > 0$ , and try to estimate the negative acceleration  $a_{(2)}$  required to reduce  $v'_n$  to zero while car  $n$  is approaching the slower  $n + 1$  car, i.e.  $x'_n$  is reduced to the minimal headway  $D$ . Assuming that this acceleration  $a_{(2)}$  is constant, the time of this process is  $t = \Delta v_n/a_{(2)}$ . Substituting  $t$  in  $-D = -\Delta x_n - \Delta v_n t + \frac{1}{2}a_{(2)}t^2$ , one obtains

$$a_{(2)} = -\frac{(\Delta v_n)^2}{2(\Delta x_n - D)}, \quad \Delta v_n < 0. \tag{5}$$

**(c) Speed limit:** As in real driving, a car in our model can exceed the speed limit. Therefore we denote this speed limit  $v_{per}$  ('permitted velocity'). The corresponding dissipative third term in (1),  $a_{(3)}$ , represents the effective 'repulsive force' acting when the velocity exceeds the permitted velocity. The simplest choice of this a term is

$$a_{(3)} = -k(v_n - v_{per}), \quad v_n > v_{per}, \tag{6}$$

where  $k$  is a constant, representing the coupling to the permitted velocity (relatively high values of  $k$  represent higher 'obedience' to this speed limit).

**(d) Randomness:** The randomness in the driver behavior is usually represented by white uncorrelated noise, denoted by  $\eta'(x, t)$  in (1). For numerical realization of this type of noise, uniformly distributed random numbers are chosen for each car at every iteration. In the following we denote by  $\eta$  the amplitude of this distribution. According to the central limit theorem, the noise becomes Gaussian as the numerical time interval  $\Delta t$  satisfies  $\Delta t \ll T$ . In this case the effective amplitude of the noise is proportional to  $\eta\sqrt{\Delta t}$ .

We can therefore derive a system of equations representing the motion in this model,

$$a_n = A \left( 1 - \frac{\Delta x_n^0}{\Delta x_n} \right) - \frac{Z^2(-\Delta v_n)}{2(\Delta x_n - D)} - kZ(v_n - v_{per}) + \eta', \quad (7)$$

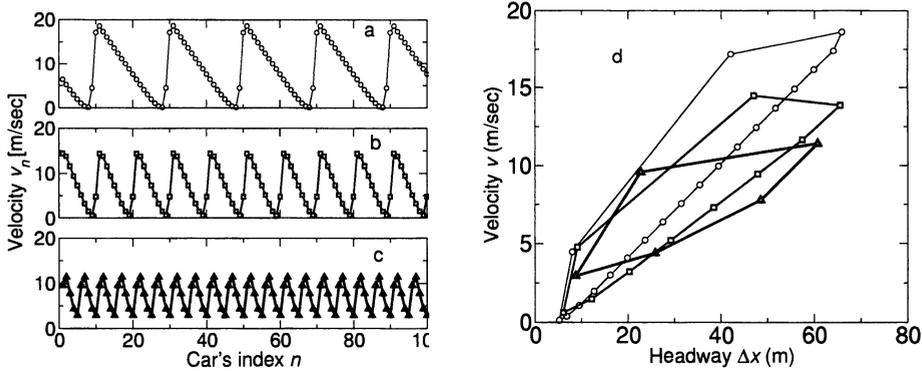
where the function  $Z(x)$  is defined as  $Z(x) = (x + |x|)/2$ . In the following we use this version of the model with parallel updating rule and white noise, unless else is mentioned. The numerical solutions presented below are plotted for the parameters values  $T = 2(s)$ ,  $D = 5(m)$ ,  $A = 3(m/sec^2)$ ,  $v_{per} = 25(m/sec)$ ,  $k = 2(sec^{-1})$ , and a numerical time interval  $\Delta t = 0.1sec$ .

To compare the Nagel-Schreckenberg (NS) model [16] with our model, let us determine  $a_{(1)}(\Delta x_n, v_n)$  and  $a_{(3)}(v_n)$  for the deterministic NS model [17]. Since the speed limit  $v_{per}$  in the NS model is really a maximal velocity, it corresponds to (6) with the limit  $k \rightarrow \infty$ . To find the corresponding  $a_{(1)}$  of the NS model, we should first define an optimal headway for the NS model. This optimal headway is  $\Delta x_n^0 = v_n T + D$  - as in (2) - with the parameter values  $T = 1$  and  $D = 1$ . The reason is that in the deterministic NS model  $a_{(1)}(\Delta x_n) = 0$  for  $\Delta x_n = v_n$ , implying that this is the optimal headway  $\Delta x_n^0$ . To obtain the function  $a_{(1)}(\Delta x_n)$  we recall that in the NS model the acceleration  $a_n$  is equal to 1 for  $\Delta x_n > \Delta x_n^0$ , and that the velocity is reduced to  $\Delta x_n - 1$  for  $\Delta x_n < \Delta x_n^0$ . Therefore

$$a_{(1)}(\Delta x_n) = \begin{cases} 1, & \Delta x_n \geq \Delta x_n^0 \\ \Delta x_n - \Delta x_n^0, & \Delta x_n < \Delta x_n^0 \end{cases} \quad (8)$$

for the deterministic NS model ( $p = 0$ ). The plot of this piecewise linear discrete function is displayed in Fig. 1b in circles, and can be compared to  $a_{(1)}(\Delta x_n)$  of our model (thin curve).

For the stochastic NS model ( $0 < p < 1$ ), a function  $a_{(1)}(\Delta x_n)$  can represent the average behavior of cars. Therefore  $a_{(1)}$  can be found here by subtracting  $p$  from the deterministic  $a_{(1)}$  in Eq. (8), according to the randomness rule in the NS model [16]. Therefore  $a_{(1)}(\Delta x_n)$  remains a piecewise linear function also for the stochastic NS model (Fig. 1b, squares). Note that in both the deterministic and the stochastic NS models,  $a_{(1)}(\Delta x_n)$  becomes a linear function for  $v_n = v_{max}$ . In this case the horizontal part of  $a_{(1)}(\Delta x_n)$  can be replaced with any positive function, due to the influence of the maximal velocity term  $a_{(3)}$ .



**Fig. 2.** Three different stable stop-and-go states of the same density  $\rho = 0.03veh/m$ ,  $A = 1m/sec^2$ , and  $v_{per} = 25m/sec$ . Figs. a-c presents the velocities of all cars of these states, which also yield different values of flux  $f$ . **a.**  $\lambda = 20veh$  ( $f = 0.2618veh/sec$ ), **b.**  $\lambda = 10veh$  ( $f = 0.2160veh/sec$ ), and **c.**  $\lambda = 5veh$  ( $f = 0.2168veh/sec$ ). **d.** Headway-velocity diagrams of the three limit cycles in Figs. a-c:  $\lambda = 20veh$  (circles),  $\lambda = 10veh$  (quares), and  $\lambda = 5veh$  (triangles).

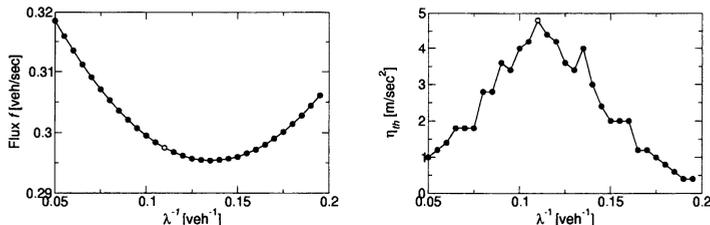
For the stochastic Nagel-Paczuski (NP) model [18], however, random braking can occur only after braking due to insufficient headway. The corresponding function  $a_{(1)}(\Delta x_n)$  is therefore a combination of the deterministic NS function (8) for  $\Delta x_n < \Delta x_n^0$  and that of the stochastic NS model for  $\Delta x_n \geq \Delta x_n^0$  (see Fig. 1b, dashed curve with full symbols). Note, that unlike the NS models [16,17], both the NP and our model demonstrate metastable free flow (see [18,11]). Note also, that unlike the NS models in which  $a_{(1)}(\Delta x_n)$  is linear for stable free flow, this function is not linear for both NP model and our model. This additional nonlinearity seems to be essential for the existence of bistable regime.

### 3 The Generalized Fundamental Diagram

The solutions of the deterministic model (7) for the intermediate density regime are characterized by periodic density waves. Simulations of our model show the existence of multiple stable states which correspond to limit cycles in the phase space. Figs. 2a-c present the velocities of all cars on a road after sufficient relaxation time, for three different initial conditions. It can be seen that the wavelengths  $\lambda$  of these states (in the vehicle space) is different. Note that the flux of these states has also different values (see caption of Fig. 2).

Our numerical simulations also show the existence of other solutions with other wavelengths and flux values than that displayed in Fig. 2a-c. Consequently, depending on initial conditions, different stable wave solutions can emerge with different distances between neighboring humps and different values of flux. This indicates that for intermediate values of densities system (7) has many stable periodic (in, e.g.,  $\Delta x_n, \Delta v_n$  variables) solutions. Hence, in the  $2N$ -dimensional

space of variables  $\Delta x_n, \Delta v_n$  many attractive limit cycles exist. Three examples of these limit cycles are presented in Fig. 2d, for the same states presented in Figs. 2a-c.



**Fig. 3.** a. Typical dependence of the flux on the wavelength. The global density is  $\rho = 0.06 \text{ veh/m}$ ,  $A = 3 \text{ m/sec}^2$ , and  $v_{per} = 25 \text{ m/sec}$ . b. Noise stability threshold of the stable states displayed in the previous Figure, as a function of the wavelength.

As implied above, the flux for the limit cycles is not only a function of the density, but also depends on the wavelength. A typical relation between the flux and the wavelength for some given density is shown in Fig. 3a. As can be seen in this figure, the flux-wavelength relation usually has a single minimum.

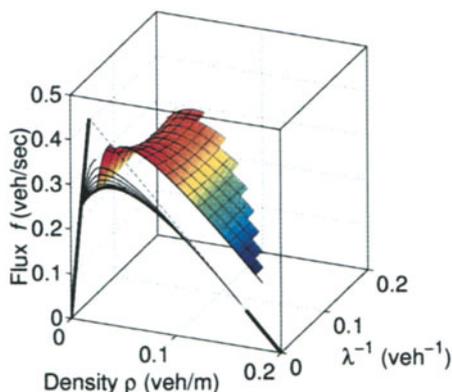
Fig. 3a also indicates that similarly to Kerners' empirical findings, here also exists a range of possible stable flux values for some given value of density. Note that the stable periodic states of the OVM [14] are having the same value of flux for a given density ( $f = f(\rho)$  but  $f \neq f(\lambda)$ , see [15]). The same holds for the deterministic NS model [16], where for all stable states  $f$  depends on  $\rho$  only. For our model, however, the flux is not only a function of the density, but also of the wavelength of stop-and-go traffic waves,

$$f = f(\rho, \lambda). \quad (9)$$

Thus, the fundamental diagram (flux-density relation) has to be generalized into the *generalized fundamental diagram* (GFD) – a flux-density-wavelength relation.

To obtain this flux-density-wavelength relation we perform extensive simulations of the deterministic model (7) on homogeneous systems with periodic boundary conditions, starting from different initial conditions [28]. In these simulations, the density of cars can easily be controlled, due to conservation of mass. However, controlling the wavelength is more difficult. For this purpose it is necessary to start with periodic initial conditions, such as equal initial spacing between cars and harmonic dependence of the initial velocity on the car indexes. But even this consideration may not be enough, since in the basin of attraction of a limit cycle with some wavelength  $\lambda$  does not necessarily include the corresponding harmonic initial states. To overcome this difficulty, we also use as initial conditions the steady state of some other density with the same wavelength, which are rescaled in order to obtain the desired density.

As a result we derive the generalized fundamental diagram (flux-density-wavelength relation) shown in Fig. 4. This figure presents the flux measured



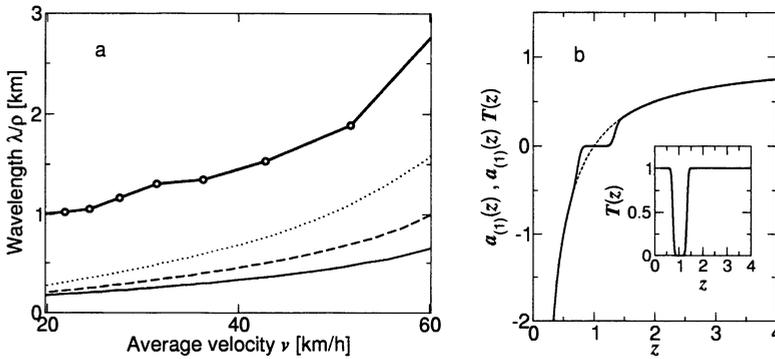
**Fig. 4.** The generalized fundamental diagram (GFD) – flux-density-wavelength relation for the different states of the deterministic model: stable stop-and-go states (surface), stable and unstable homogeneous states (thick and dotted lines, respectively). The wavelength is defined as the average distance (in the vehicles space) between centers of two nearest dense regions in stop-and-go traffic. Some curves with fixed wavelength ( $\lambda^{-1} = 2/60, 3/60, \dots, 12/60 \text{veh}^{-1}$ ) are projected on the density-flux plane (thin curves). The cross-section of the surface for a density constant density  $\rho = 0.06 \text{veh}/m$  is shown in Fig. 3a, demonstrating the typical dependence of the flux on the wavelength.

for the steady states of the deterministic model: stop-and-go waves (surface), homogeneous free and homogeneous congested flow (thick lines). The projection of the surface in Fig. 4 on the density-flux plane (thin curves in Fig. 4) provides a two-dimensional region of stable states, qualitatively similar to that found empirically [7,8] for synchronized flow. Note that this result is obtained even for a deterministic model. Thus, in contrast to the common belief that the flux depends only on the density of cars (fundamental diagram), we obtain a generalized fundamental diagram (Fig. 4) which shows that the flux depends on two variables – density  $\rho$  and wavelength  $\lambda$ .

## 4 The Influence of Noise

In the previous section we have studied the deterministic version of our model, with zero noise term in (7). Nevertheless, real traffic is characterized by the presence of noise, mainly caused by some randomness in the behavior of each single driver, which is our concern in this Section. Our simulations show that even for non-zero noise amplitude, the stop-and-go waves remain stable until some threshold amplitude is reached. But some of the limit cycles are more sensitive to noise than others. For sufficiently large noise amplitude  $\eta$ , the system moves from these metastable states to more stable states. An example for such transition was demonstrated in [11].

To determine the threshold amplitude of noise, extensive simulations have been performed, starting from various initial limit cycles. In these simulations the noise amplitude has been gradually increased, starting from  $\eta = 0$ . Every increase of the noise was followed by sufficiently large transient and evaluation times, during which the average wavelength has been evaluated. The evaluation of the wavelength has been performed directly by measuring all velocities,  $v(n)$ , and counting the number of times this function crosses the average velocity  $v$ . The criterion for supercritical noise (or loss of stability) is the presence of wavelength which is different from the initial wavelength for more than 50% of the evaluation period. Using this technique, the noise stability threshold  $\eta_{th}$  is evaluated for different values of wavelength. A typical relation between  $\eta_{th}$  and  $\lambda$  is shown in Fig. 3b, for  $\rho = 0.06$ . This figure can be compared to Fig. 3a, where the same parameters have been used, and the flux of the same states is shown.



**Fig. 5. a.** The average wavelength of the stop-and-go states (in real space) as a function of the average velocity. Random initial conditions were applied for different values of density, and the  $\lambda$  and  $v$  were measured after sufficiently long transient time, and averaged many iterations of each density. This relation is displayed for the common model (7) with  $A = 3m/sec^2$  (solid lower curve),  $A = 1.5m/sec^2$  (dashed curve),  $A = 0.75m/sec^2$  (dotted curve), and for the threshold version of the model with  $A = 3m/sec^2$  (solid upper curve). **b.** The threshold version of the model. The solid curve corresponds to the modified function  $a_{(1)}$  which includes the threshold in the reaction of the driver, given by Eq. (11),  $A = 1m/sec^2$ . It can be compared to the unmodified expression (dashed curve) with the same value of  $A$ . The threshold function (10) is displayed in the inset.

As can be seen in Fig. 3b, intermediate values of wavelengths are the most stable in the presence of noise. In contrast, cycles with relatively small and relatively large wavelengths are more sensitive to noise, and can be considered as metastable states. The practical meaning of this finding is that for a given level of noise (below the maximal threshold) only the intermediate part of the possible range of wavelength can be found, and the most probable state is the

state with the highest value of threshold. This might be the reason to the fact that in real stop-and-go traffic there exists a typical value of wavelength [7] (see also Sec. 5 below).

From Figs. 3a-b it follows that the most stable state in the presence of noise, denoted by an open circle, has a relatively low value of flux. This fact is one of the motivations for the optimization strategy developed in Sec. 6. The optimization process, which is based on the flux-wavelength relation demonstrated in Fig. 3a, is explained later using our analysis of the different noise thresholds of the different states presented here (Fig. 3b).

## 5 The Wavelength of Stop-and-Go Traffic

Stop-and-go traffic is a well-known congested traffic state characterized by periodic density waves. This state is characterized by the existence of a characteristic time period (approximately 10 minutes, see [10] and references therein). Recently, the existence of a characteristic lengthscale, which is a function of the average velocity, was also measured for stop-and-go traffic [7]. This wavelength, called ' $R_{narrow}$ ', represents the average distance between the narrow jams, and was found to be an increasing function of the average velocity of the synchronized flow downstream.

When comparing the experimental findings to the numerical results of our model, a qualitative similarity is found. Like the empirical results, the average value of the wavelength (starting from random initial conditions) is found to be an increasing function of the average velocity (see Fig. 5a). However, the values of the wavelength we find are considerably lower than that found by [7]. A better quantitative agreement can be achieved using the threshold version of the model described below.

The existence of threshold level of stimulus required for human response is a well-known phenomenon, which, in our case, means that the driver does not accelerate or brake if the headway  $\Delta x_n$  is sufficiently close to the optimal headway  $\Delta x_n^0$ . In other words, drivers usually accelerate (decelerate) only when the difference between the headway and the optimal headway is above (below) some positive (negative) value. The realization of this threshold in our model is by multiplying the acceleration term  $a_{(1)}$  given by Eq. (3) with a threshold function,  $T(\Delta x_n/\Delta x_n^0) = T(z)$ , a function that approaches 0 when  $z = \Delta x_n/\Delta x_n^0 \approx 1$ , and approaches 1 otherwise. We choose to use the function

$$T(z) = \frac{1}{2} \{1 + \operatorname{erf}[\frac{s}{\sqrt{2}}(z - \frac{1}{r})]\} \frac{1}{2} \{1 + \operatorname{erf}[\frac{s}{\sqrt{2}}(r - z)]\}. \tag{10}$$

The values of  $r$  and  $1/r$  determine the upper and lower cutoff values of  $z$ , respectively, and  $s$  is the quality factor of the cutoff. Here we use the parameter values  $r = 0.75$  and  $s = 20$ . The plot of the function  $T(z)$  is displayed in the inset of Fig. 5b. The obtained threshold acceleration term

$$a_{(1)}(\Delta x_n) = A \left(1 - \frac{\Delta x_n^0}{\Delta x_n}\right) T(\Delta x_n/\Delta x_n^0), \tag{11}$$

is displayed in Fig. 5b, for  $A = 1m/sec^2$ .

As a result of this modification, the effective  $A$  can be close to zero (below the threshold) or close to  $A$  (above the threshold). Hence, the values of the wavelength are expected to grow. As can be seen from Fig. 5a, the average wavelength (red curve) is close to that measured by Kerner in real stop-and-go traffic (see [7]).

## 6 Optimization of Congested Traffic by Wavelength Control

Optimization and control of traffic flow [4,29–34] are usually performed using the concept of the fundamental diagram of traffic flow (flux-density relation) [4–6]. However, recent experimental studies of Kerner [7,8] indicate that such a fundamental diagram does not exist, and that stable synchronized traffic states display a  $2D$  region in the density-flux plane. In Sec. 3 we suggest a possible explanation to this finding using the GFD, showing that for a given value of density, there exists a range of possible flux values representing stable states of different wavelengths.

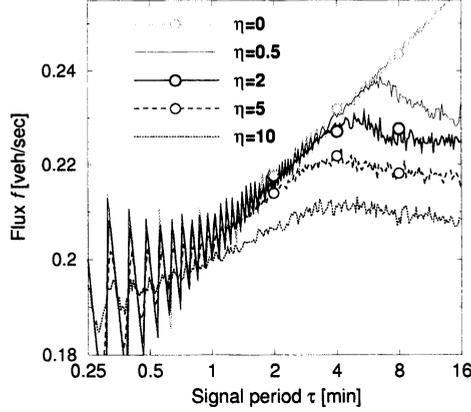
This new theoretical insight will be applied here, in order to achieve the highest possible flux during congestion. For this purpose, a new strategy of optimization is proposed here. The strategy is based on inducing stop-and-go waves on the main road and controlling their wavelength. Using numerical simulations of our stochastic car-following model (7) we show below that this strategy yields optimization of traffic flow when implemented in systems with a localized periodic inhomogeneity.

In the following, two types of systems with a localized periodic inhomogeneity are considered: (a) signalized intersections, and (b) entry ramps with signalized entrances. We focus on cases of over-saturation, i.e., where traffic is congested upstream of the inhomogeneities.

**(a) Signalized intersection:** Signal optimization theories [4,29–31,33,35] are traditionally concerned with optimizing global quantities such as the total delay time of all drivers in the system. Our aim here is different – to optimize the flux in an arbitrary direction of an intersection. In an oversaturated signalized intersection, however, the flux in each direction upstream to the intersection is influenced only by the parameters of the traffic signals of this direction. Thus we can consider a single direction for simplicity (as, e.g., in [29]).

In the studied system, traffic lights are placed at position  $L/2$  on a road with length  $L$  and periodic boundary conditions. The flux  $f$  is measured at the position of the traffic lights. We assume that during a given portion  $P_r$  of the total signal periods  $\tau$ , the intersection is occupied by vehicles coming from the other directions. The rest of the time in each signal period is divided into three parts:  $\tau_g$  and  $\tau_y$  are the durations of the green and yellow lights, and  $\tau_-$  is a given additional safety red light period. Thus  $\tau = P_r\tau + \tau_g + \tau_y + \tau_-$ .

The period of yellow light is realized as follows. When the light is changed from green to yellow, all simulated drivers upstream of the intersection estimate



**Fig. 6.** Relation between signal period and flux for the values of acceleration noise amplitude  $\eta = 0, 0.5, 2, 5, 10\text{m/sec}^2$  (top to bottom). Traffic lights parameters are  $P_r = 1/3, \tau_y = \tau_- = 2\text{sec}$ . The total number of cars in the system is  $N = 400$  and its length is  $L = 10\text{km}$ . While according to (12) almost no influence of  $\tau$  on  $f$  is expected, the numerical simulations of the deterministic model ( $\eta = 0$ , upper curve) show that  $f$  significantly increases with  $\tau$  even for large values of  $\tau$ . Moreover, an optimal signal period is found when the stochastic model is used ( $\eta > 0$  curves). The nine open circles correspond to the nine instances shown in Fig. 7.

the intersection crossing time  $t_n$  by a linear extrapolation of their position. The first car that begins to stop,  $s$ , is the first car that is not able to cross before the light changes to red, i.e.,  $t_s = \min\{t_n | t_n > \tau_y\}$ . For this car,  $\Delta x$  in (7) is replaced with the distance between  $s$  and the traffic lights. The consecutive cars follow  $s$  and stop according to (7). Due to this procedure cars can still cross the intersection during time  $\tau_y$ .

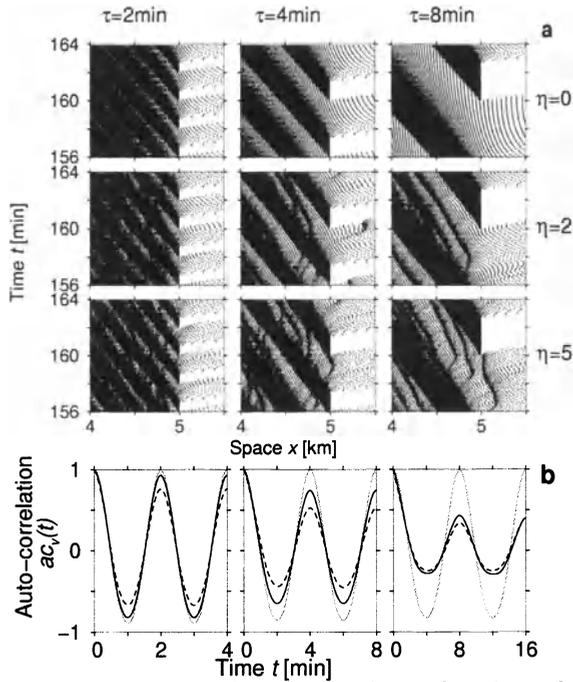
As mentioned above, our aim here is to optimize the flux in a single direction of the intersection, without affecting the parameters of the whole complex system of roads. Thus we consider  $P_r$  to be a given constraint, and explore the relation between the flux  $f$  and the signal period  $\tau$ . According to common optimization theories, the result is expected to be [29]

$$f = f_0 \left(1 - P_r - \frac{\tau_-}{\tau}\right), \tag{12}$$

where  $f_0$  is the constant flux during the green light period. (12) simply expresses the fact that the outflow of a single direction is a product of the actual relative green light period and  $f_0$ . Since usually  $\tau \gg \tau_-$ , the flux should be hardly influenced by  $\tau$ .

However, the results of our simulations are different from (12) for both stochastic and deterministic models. The typical relation between  $f$  and  $\tau$  for a given density shown in Fig. 6 shows a much greater influence of  $\tau$  on  $f$  than that predicted by (12). Apart from trivial flux oscillations which are explained below, Fig. 6 demonstrates a monotonic asymptotic increase of the flux  $f$  as  $\tau$  grows

for the deterministic model  $\eta = 0$ . For the stochastic model  $\eta > 0$ , however, an optimal signal period can be clearly seen. In this case, a crossover is observed from the deterministic monotonically increasing  $f(\tau)$  to a saturated lower value of the flux for large  $\tau$ . Therefore there exists an optimal signal period that yields maximal flux. Such influence of  $\tau$  on  $f$  is not explained by (12).



**Fig. 7.** **a.** Space-time diagrams and **b.** auto-correlation functions of systems with single traffic light with parameters values that are similar to that of the nine instances denoted by circles in Fig. 6. The position of the traffic light is  $x = 5\text{km}$ . The major dense (black) regions moving upstream in Fig. a are caused by the red light, and the distance between such nearest two regions (corresponding to the wavelength) is growing as  $\tau$  grows. The grey curves in Fig. b. correspond to  $\eta = 0$ , the solid curves to  $\eta = 2\text{m/sec}^2$ , and the dashed curves to  $\eta = 5\text{m/sec}^2$ . Comparison to Fig. 6 indicates that the crossover to reduced values of flux in the latter Figure is related to the emergence of small jams, which causes loss of the periodicity (see Fig. a) as well as reduced values of the auto-correlation function,  $ac_v(t = \tau)$  (see Fig. b).

In contrast to [31], where the oscillations observed in  $f(\tau)$  have been related to degree of synchronization of the green lights in the network, here the oscillations are a trivial result of the discrete nature of the flowing media. Since all vehicles have the same parameters, the crossing times of the intersection are identical in each period. This causes discretization in the average number of crossing cars in each period,  $n$ , which becomes a staircase function of  $\tau$ . The flux is given by  $f = n(\tau)/\tau$ , and hence the magnitude of these oscillations decreases with increasing  $\tau$ .

The existence of an optimal signal period can now be explained using our finding of the influence of  $\lambda$  on the flux and on  $\eta_{th}$ . First, it is easy to see that the signal period  $\tau$  controls the wavelength  $\lambda$  of stable stop-and-go waves that are induced by the traffic lights, as  $\lambda = v\tau$ , where  $v$  is the wave velocity [6]. From Fig. 3a it appears that in order to increase the flux,  $\lambda$  – and therefore  $\tau$  – should be increased [37]. For this reason the deterministic model yields a monotonically increasing  $f(\tau)$  (Fig. 6, upper curve). However, for the stochastic model ( $\eta > 0$ ), states for which  $\eta_{th}(\lambda) < \eta$  become unstable according to Fig. 3b – in particular those with relatively high values of  $\lambda$ , which in principle should yield higher values of flux. The increase of the flux with increasing  $\tau$  thus crosses-over to lower values due to the instability of the induced waves. The existence of an optimal  $\tau$  is a result of this effect.

We can therefore see that the non-trivial flux-wavelength relation is the reason for the unexpected behavior of  $f(\tau)$ . The deviations between the theoretical prediction of (12) and the numerical measurements of the deterministic model (Fig. 6) can be explained using the GFD. The crossover in  $f(\tau)$  for the stochastic model, and the existence of an optimal  $\tau$  are related to the loss of the stability of the states with relatively large wavelength due to the noise. (12) can be now modified by assigning  $f_0 = f_0(\tau, \eta)$ .

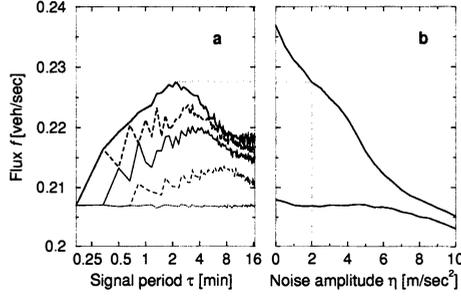
To obtain further support of this explanation let us evaluate a measure for the periodicity of the flow using single vehicle data collected at the intersection. We calculate the auto-correlation function  $ac_v(t)$  [10] of the velocity function  $v(t')$  measured at the intersection,

$$ac_v(t) = \frac{\langle v(t')v(t'+t) \rangle - \langle v(t') \rangle \langle v(t'+t) \rangle}{\langle v(t')^2 \rangle - \langle v(t') \rangle^2}, \tag{13}$$

where a linear interpolation of the discrete function  $v(t)$  is used. The brackets  $\langle \dots \rangle$  indicate averaging over time  $t'$ . Displayed in Fig. 7b are the auto-correlation functions for the nine instances of Fig. 7a, respectively. As can be seen from this Figure,  $ac_v(t = \tau) = 1$  for all the  $\eta = 0$  instances, implying that the flow for these cases is completely periodic, and that the time period of  $v(t)$  is exactly  $\tau$ . A comparison of all instances shown in Fig. 7b to the data in Fig. 6 shows that the flux approaches the deterministic value as long as  $ac_v(t = \tau) \approx 1$ . As  $\eta$  or  $\tau$  are increased, the flow is no longer periodic ( $ac_v(t = \tau) < 1$  in Fig. 7b), small jams emerge (Fig. 7a), and the flux becomes lower (Fig. 6).

**(b) Entry ramp with signalized entrance:** It is a well-know fact that congestion often occurs upstream to entry ramps [9,11,25–27]. A common way to relieve congestions in this case is to control the entrances by traffic signals. The idea is to prevent congestion by homogenizing the flow – by letting cars enter only into low-density regions. Clearly, this method is more efficient for intermediate densities rather than for high densities where congestion cannot be avoided. In the following we examine the possibility of optimization of such congested systems. The traffic light will be used to induce stable stop-and-go waves with optimal wavelength on the main congested road.

The system studied here has periodic boundary conditions and a single entry ramp [34,35]. To make this inhomogeneity periodic, we introduce traffic signals



**Fig. 8.** (a) Relation between signal period on the entry ramp and flux on the main road, for  $P_r = 0.7, 0.5, 0.4, 0.2, 0.0$  (top to bottom), and for noise amplitude  $\eta = 2m/sec^2$ . The total number of cars in the system is  $N = 300$ , the system length is  $L = 10km$ , and the flux is locally measured on the main road at  $100m$  upstream to the on-ramp. Here  $f_{in} = 0.1veh/sec$  and  $f_{max} = 0.333veh/sec$ . (b) A comparison between the optimal flux (upper curve) and the flux without the presence of a traffic light (lower curve), as a function of  $\eta$ .

at the downstream end of the entry ramp, and study its effect on the flux on the main road. The incoming vehicles are allowed to enter the main road during the green light period  $\tau_g$ , and are delayed during the red light period  $P_r\tau$ , so here  $\tau = P_r\tau + \tau_g$ . Unlike the signalized intersection where  $P_r$  was predetermined, here it is one of the optimization parameters, in addition to  $\tau$ . Similar to [25], we introduce an exit ramp at a large distance from the entry ramp, so that the total number of cars in the system is conserved. The entrance and the exit of cars from the ramps are performed in the same manner as in [11]. We focus on cases where the average flux of the incoming vehicles  $f_{in}$  causes congestion on the main road (see [9,26]), but the secondary road is not congested. Since our goal is to optimize the flux on the main road without causing congestion on the secondary road, we set an upper bound for  $P_r$ . This bound is  $P_r \leq 1 - f_{in}/f_{max}$ , where  $f_{max}$  is the maximal possible value of the incoming flux, since cars approach the queue upstream to the traffic light with rate  $f_{in}$ , and this queue is discharged with rate  $f_{max}$  during the green light.

Typical relations between flux and signal period are plotted in Fig. 8a for different values of  $P_r$  and for  $\eta = 2m/sec^2$ . Note that the curve corresponding to an unsignalized entry ramp ( $P_r = 0$ ) is the lowest, implying that the introduction of a traffic light increases the flux on the main road. Moreover, the increase in the flux on the main road is obtained without causing congestion on the secondary road, as we consider only the values of  $P_r$  which are below this congestion threshold. The relative increase in the flux due to the introduction of a traffic light with optimal parameters (Fig. 8b) varies from 1.0% for  $\eta = 10m/sec^2$ , through 10.0% for  $\eta = 2m/sec^2$  and up to 13.9% for  $\eta = 0$ . This increase of the flux  $f$  has even a more significant influence on the growth rate of the congested section of highway upstream of the entry ramp, since this rate is proportional [6] to  $f' - f$ , where  $f'$  is the flux upstream to this region. These results suggest that

even when congestion of real traffic cannot be relieved, the flux can be increased by an optimal signalization of entry ramps to control and stabilize stop-and-go waves.

In addition to the two studied systems, the new optimization approach may also be efficient in traffic control systems that are based on varying speed limits along the road. In these systems, stop-and-go traffic waves can be controlled [36], and therefore can also be optimized using this approach.

To summarize, this work was motivated by the new experimental findings of a two-dimensional representation of synchronized flow in the density-flux plane [7,8]. Using a deterministic car following model [11] we are able to show that the fundamental diagram has to be generalized to include another variable, the wavelength of the stop-and-go waves. The projection of the generalized fundamental diagram (GFD) on the density-flux plane yields a two-dimensional region of stable states, qualitatively similar to that found experimentally for synchronized flow. The different states also differ in their sensitivity to noise. The dependence of the wavelength of the most attractive state on the average velocity is qualitatively similar to that found experimentally. Quantitative agreement can be also approached by including threshold in the driver's reaction. We use these theoretical predictions to propose a novel strategy for traffic optimization, based on inducing stable stop-and-go waves yielding the maximal flux. In general, the interplay between the wavelength-flux relation (Fig. 3a) and the noise stability threshold of the different states (Fig. 3b) determines the optimal wavelength and signal period.

Nevertheless, the encouraging results obtained for the studied systems are mostly based on numerical simulations and analytical studies. A careful field study is thus necessary to verify the model predictions, especially the generalized fundamental diagram, and to test the optimization mechanism.

## References

1. D. Chowdhury, L. Santen, and A. Schadschneider: Phys. Rep. **329**, 199 (2000).
2. D. Helbing: Rev. Mod. Phys. (in the press) (2001); cond-mat/0012229 (2000).
3. D.E. Wolf: Physica A **263**, 438 (1999).
4. A.D. May: *Traffic Flow Fundamentals*, (Prentice Hall, Englewood Cliffs, N.J. 1990).
5. B.D. Greenshields: High. Res. Rec. **14** 468 (1934).
6. M.J. Lighthill and G.B. Whitham: Proc. R. Soc. London Ser. A **229**, 317, (1955).
7. B.S. Kerner: Phys. Rev. Lett. **81**, 3797 (1998).
8. B.S. Kerner: Physics World **12** (8), 25 (1999).
9. B.S. Kerner and H. Rehborn: Phys. Rev. Lett. **79**, 4030 (1997).
10. L. Neubert, L. Santen, A. Schadschneider, and M. Schreckenberg: Phys. Rev. E **60**, 6480 (1999).
11. E. Tomer, L. Safonov, and S. Havlin: Phys. Rev. Lett. **84**, 382 (2000).
12. L.A. Safonov, E. Tomer, V.V. Strygin, and S. Havlin: Physica A **285**, 147 (2000).
13. L.A. Safonov, E. Tomer, V.V. Strygin, Y. Ashkenazy, and S. Havlin: Europhys. Lett. **57**, 151 (2002).

14. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama: *Phys. Rev. E* **51**, 1035 (1995).
15. Y. Sugiyama: 'Dynamical Model for Congestion of Freeway Traffic and its Structural Stability'. In: *Traffic and Granular Flow*, ed. by D.E. Wolf, M. Schreckenberg, and A. Bachem (World Scientific, Singapore 1996) pp 137–149.
16. K. Nagel and M. Schreckenberg, *J. Phys. I (France)* **2**, 2221 (1992).
17. K. Nagel and H.J. Herrmann: *Physica A* **199**, 254 (1993).
18. K. Nagel and M. Paczuski: *Phys. Rev. E* **51**, 2909 (1995).
19. O. Biham, A.A. Middleton and D. Levine: *Phys. Rev. A* **46**, R6124-R6127 (1992).
20. R. Herman and R.W. Rothery: 'Car Following and Steady State Flow'. In: *Proceedings of the 2ns International Symposium on the Theory of Road Traffic Flow, London, 1963*, ed. by J. Almond, (Organization for Economic Cooperation and Development, Paris 1965) pp 1–11.
21. D. Helbing and B. Tilch: *Phys. Rev. E* **58**, 133 (1998).
22. N. Mitarai and H. Nakanishi: *Phys. Rev. Lett.* **85**, 1766 (2000).
23. M. Treiber, A. Hennecke, and D. Helbing: *Phys. Rev. E* **62**, 1805 (2000).
24. S. Cheybani, J. Kertesz, and M. Schreckenberg: *Phys. Rev. E* **63**, 016108 (2001).
25. H.Y. Lee, H.W. Lee, and D. Kim: *Phys. Rev. Lett.* **81**, 1130 (1998).
26. D. Helbing and M. Treiber: *Science* **282**, 2001 (1998).
27. D. Helbing, A. Hennecke, and M. Treiber: *Phys. Rev. Lett.* **82**, 4360 (1999).
28. Interactive simulations of the deterministic model with and without traffic lights can be found at <http://ory.ph.biu.ac.il/2000/traffic/>.
29. D.C. Gazis and R.B. Potts: 'The Over-Saturated Intersection'. In: *Proceedings of the 2ns International Symposium on the Theory of Road Traffic Flow, London, 1963*, ed. by J. Almond, (Organization for Economic Cooperation and Development, Paris 1965) pp 221–237.
30. W.B. Cronje: *Transport. Res. Rec.* **905**, 80 (1983).
31. D. Chowdhury and A. Schadschneider: *Phys. Rev. E* **59**, R1311 (1999).
32. N. H. Gartner, C. Stamatiadis, and P. J. Tarnoff, *Transport. Res. Rec.* **1494**, 98 (1995).
33. T.H. Chang and J.T. Lin, *Transport. Res. B* **34**, 471 (2000).
34. H. Zhang, S.G. Ritchie, and W.W. Recker: *Transport. Res. C* **4**, 51 (1996).
35. M. Treiber and D. Helbing, *preprint* (2001).
36. R. Sollacher and H. Lenz: 'Nonlinear Control of Stop-and-Go Traffic'. In: *Traffic and Granular Flow '99*, ed. by D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf, (Springer, Berlin 2000) pp 315–320.
37. Decreasing  $\lambda$  and  $\tau$  would not yield the same result due to relatively large values of  $\tau_-/\tau$  in Eq. (12).

# Stochastic Resonance Towards Traffic Models

T. Ohira

Sony Computer Science Laboratories 3-14-13 Higashi-gotanda, Shinagawa, Tokyo  
141, Japan

**Abstract.** In this paper, we first give an overview of stochastic resonance, which has gained more research attentions recently. In this phenomena, noise, which is normally considered an obstacle to information processing, is treated to have an constructive effect causing a resonance with external signals. A particular model which causes “resonance” with noise and delay is focused. We see how regular dynamical patterns appear in this context. We then discuss possible directions where such stochastic resonance could be useful. We proceed by considering simple concrete models for computer network traffic and pedestrian traffic.

## 1 Introduction

Many natural and artificial systems are associated with noise or fluctuation. Issues relating to noise have consequently been a major topic in a variety of fields (see, e.g., [1,2]). From the point of view of information processing, noise has been considered an obstacle; however, it was recently found or suggested that noise can actually be an integral part of biological information processing, leading to an active research area named as “Stochastic Resonance” [3–7]. In this paper, we try to connect this phenomena of stochastic resonance to traffic problems. Preliminary investigations in that direction are presented for a computer network traffic model and a pedestrian traffic model.

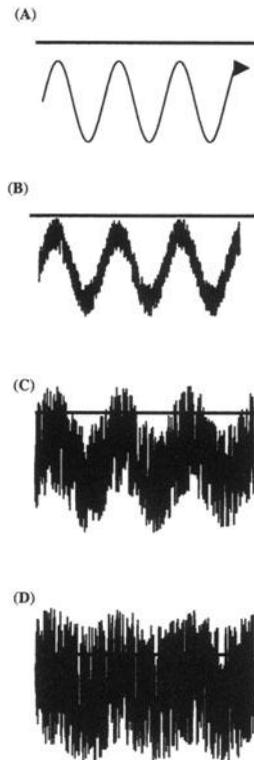
## 2 Stochastic Resonance

In this section, we first give a qualitative description of stochastic resonance. Then, particular models showing resonance with noise and delay, delayed stochastic resonance, are discussed.

### 2.1 Overview

I will give an intuitive view of stochastic resonance. More thorough discussion on the subject is found in references [3–7]. Typically, biological sensing systems have a threshold below which an incoming signal is not detected. In figure 1(A), such situation is schematically shown. Consider the situation where noise with different strength can be added to this signal. Too small amplitude noise does not make a sensing system to detect the signal(1(B)), and too large amplitude

noise can be sensed, but only as a noise and the nature of the signal is destroyed (1(D)). Appropriately tuned noise, however, leads to the detection of the qualitative characteristics of the signal (1(C)). We consider this phenomena as a resonance by tuning the noise, hence the name stochastic resonance. This is only one qualitative picture of stochastic resonance. But, it illustrates the point that noise is used constructively for information processing rather than destructively as is usually the case.



**Fig. 1.** A qualitative mechanism of Stochastic Resonance. The solid line means threshold of detection.

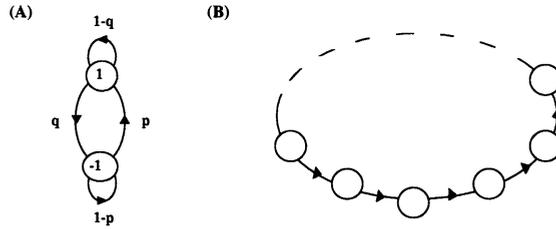
## 2.2 Delayed Stochastic Resonance

Delay is a factor which is also important in traffic problems as well as in other fields [8–13]. Here we give more detailed account for a special case of stochastic resonance. In these models, we do not have an incoming signal. Instead, delay in feedback or interaction is used. We term this phenomena as “delayed stochastic resonance”.

**Single Unit Model** Let us first discuss a very simple single unit model [14], which is defined by three parameters,  $p$ ,  $q$ , and  $\tau$  (Figure 2(A)).  $X(t)$  is the state at time steps  $t$  and the formal definition is given by the conditional probabilities as follows:

$$\begin{aligned}
 P(X(t+1) = +1|X(t-\tau) = -1) &= p, \\
 P(X(t+1) = -1|X(t-\tau) = -1) &= 1-p, \\
 P(X(t+1) = +1|X(t-\tau) = +1) &= 1-q, \\
 P(X(t+1) = -1|X(t-\tau) = +1) &= q.
 \end{aligned}$$

Hence, the transition probability of the model depends on its state at  $\tau$  steps in the past and is a special case of delayed random walks [15,16].



**Fig. 2.** (A) Schematic view of binary model. (B) A ring network

We randomly generate  $X(t)$  for the interval  $t = (-\tau, 0)$ . Simulations are performed in which parameters are varied and  $X(t)$  is recorded up to  $10^6$  steps. From the trajectory  $X(t)$ , we construct a residence time histogram  $h(u)$  for the system to be in the state  $-1$  for  $u$  consecutive steps. Some examples of histograms and  $X(t)$  are shown in Figure 3 ( $q = 1 - q = 0.5$ ,  $\tau = 10$ ). We note that with  $p \ll 0.5$ , as in Figure 3(A), the model has a tendency to switch or spike to the  $X = 1$  state after the time step interval of  $\tau$ . But the spike trains do not last long and result in a small peak in the histogram. For the case of Figure 3(C) where  $p$  is closer to 0.5, we observe less regular transitions and the peak height is again small. With appropriate  $p$  as in Figure 3(B), spikes tend to appear at interval  $\tau$  more frequently, resulting in higher peaks in the histogram. This change of peak height in histograms which reaches maximum (Figure 3(D)) at an appropriate noise level is one way to characterize stochastic resonance. It turns out that this histogram  $h(u)$  can be obtained exactly in the steady state.

$$h(u) = \left(\frac{q}{p+q}\right)^u \left(\frac{p}{p+q}\right)^2 \quad (1 \leq u < \tau), \tag{1}$$

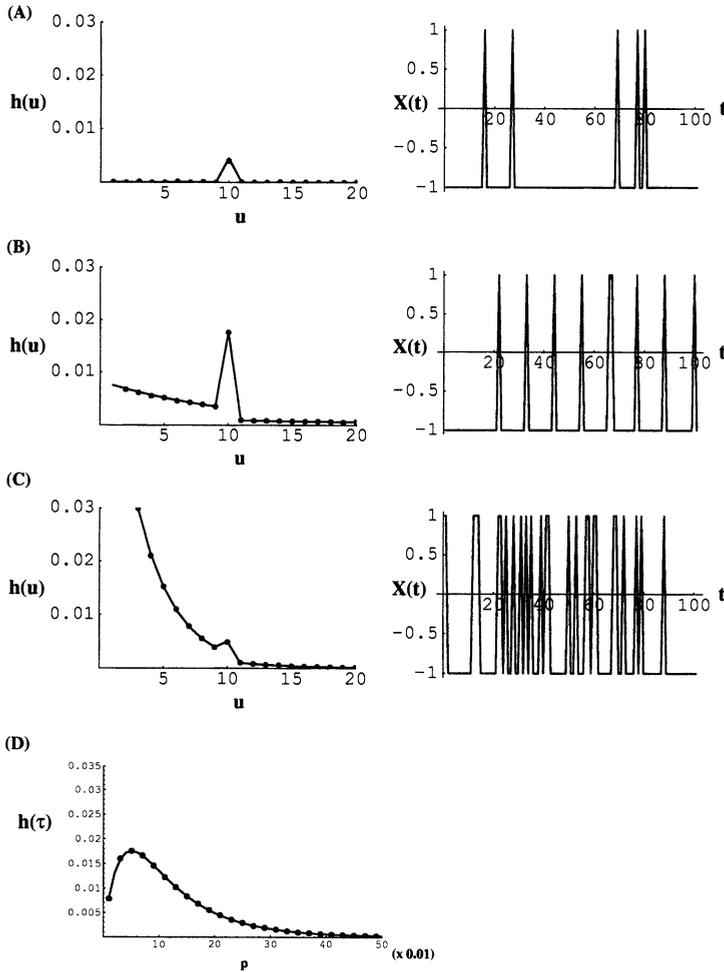
$$h(u) = \left(\frac{q}{p+q}\right)^\tau \left(\frac{p}{p+q}\right) (1-q) \quad (u = \tau), \tag{2}$$

$$h(u) = \left(\frac{q}{p+q}\right)^u (p)^2 \quad (u > \tau). \tag{3}$$

Also, by differentiating equation (2) with respect to  $p$ , we can derive the resonant condition for the peak to reach maximum height as

$$q = p\tau. \tag{4}$$

Choosing an appropriate  $p$  is equivalent to “tuning” noise with other parameters appropriately fixed. In this sense, our model shares a feature of the stochastic resonance.



**Fig. 3.** Residence time histogram and dynamics of  $X(t)$  as we change  $p$ . The values of  $p$  are (A)  $p = 0.005$ , (B)  $p = 0.05$ , (C)  $p = 0.2$ . The solid line in the histogram is from the analytical expression given in equations (1-3). Also, in (D) we show a plot of peak height by varying  $p$ . The solid line is from equation (2). The parameters are  $\tau = 10$ ,  $q = 0.5$ .

**Network Model** The next model we consider is a 1-D ring network of  $M$  binary units (Figure 2(B)). The state  $X_i(t)$  of the  $i$ th unit at time step  $t$  is either  $-1$  or  $1$ . The interaction is assumed to be only one way, and the conditional probability of transition for the  $i$ th unit depends on the state of  $i-1$ th unit at  $\tau$  steps earlier

$$\begin{aligned}
 P(X_i(t+1) = +1 | X_{i-1}(t-\tau) = -1) &= p, \\
 P(X_i(t+1) = -1 | X_{i-1}(t-\tau) = -1) &= 1-p, \\
 P(X_i(t+1) = +1 | X_{i-1}(t-\tau) = +1) &= 1-q, \\
 P(X_i(t+1) = -1 | X_{i-1}(t-\tau) = +1) &= q,
 \end{aligned}
 \tag{5}$$

where  $p$  and  $q$  are probability parameters and the  $M$ th unit is connected to the 1st unit (a ring) (Figure 2(B)).

Typical behavior of a unit in the network with  $M$ ,  $\tau$ , and  $q$  fixed and  $p$  varied is shown in Figure 4 (A) (B) (C). We again see that regular spiking patterns appear with tuned  $p$ . To see this behavior more clearly, we construct a residence time histogram  $h(u)$  for a unit to be in the state  $-1$  for  $u$  consecutive steps as in the single unit case. Some corresponding examples of the histograms are shown in Figure 4, where we fix  $M$ ,  $\tau$ , and  $q$ , but changes  $p$ . The same phenomena as in the single unit model is observed. With  $p$  small, as in Figure 4(D), the model has a tendency to switch or spike to the  $X = 1$  state after the time step interval of  $\tau$ . But the spike trains do not last long and result in a small peak in the histogram. For the case of Figure 4(F) where  $p$  is too large, we observe less regular transitions and the peak height is again small. With appropriate  $p$  as in Figure 4(E), spikes tend to appear at the interval  $s = M(\tau + 1) - 1$  more frequently, resulting in higher peaks in the histogram. This is what we mean by stochastic resonance. As in the single unit case, we can treat this model analytically. As mentioned, the peaks occurs at  $s = M(\tau + 1) - 1$ , and the histogram is given as follows:

$$\begin{aligned}
 h(u) &= \alpha^2 \beta^u & (0 \leq u < s) \\
 &= Q_{++} \beta^s & (u = s) \\
 &= Q_{+-}^2 Q_{--}^{(u-s-1)} \beta^{(2s-u)} & (s+1 \leq u \leq 2s) \\
 &= Q_{+-}^2 Q_{--}^{(u-s-1)} & (2s+1 \leq u),
 \end{aligned}
 \tag{6}$$

where  $\alpha = \frac{p}{p+q}$ ,  $\beta = \frac{q}{p+q}$  and

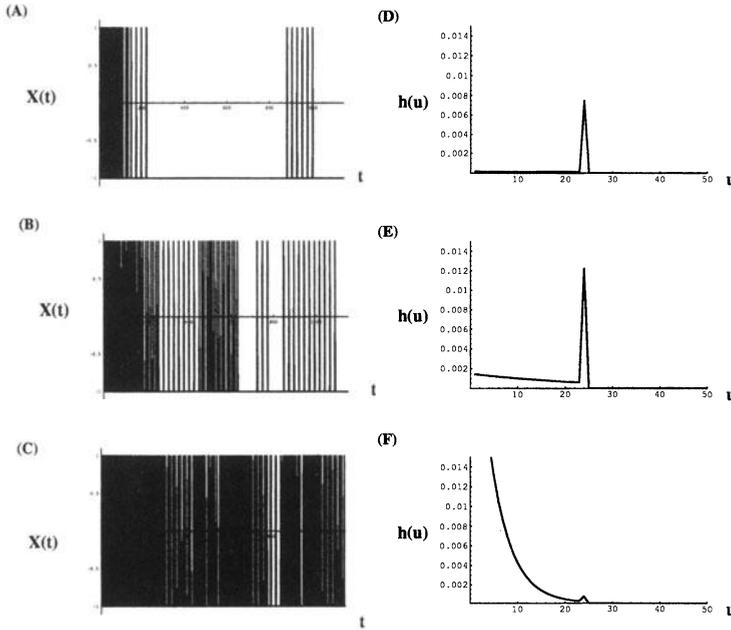
$$\begin{aligned}
 Q_{++} &= \alpha^2 + (1-p-q)^M (\alpha\beta) \\
 Q_{+-} &= \alpha\beta - (1-p-q)^M (\alpha\beta) \\
 Q_{--} &= \beta^2 + (1-p-q)^M (\alpha\beta).
 \end{aligned}
 \tag{7}$$

From this we can again derive the resonance condition as

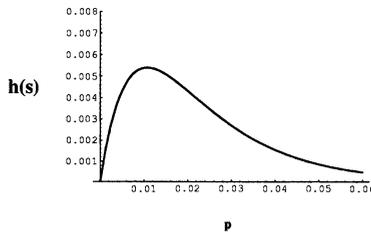
$$q = M(\tau + 1)p.
 \tag{8}$$

The way the peak hight  $h(s)$  changes with  $p$  is shown in Figure 5. The maximum height is given at the resonant condition above.

The main notable point with this network model as well as in the single unit model is, that we do not need to have explicit oscillator to produce a rhythmic dynamical behavior. The delay in the interaction with the tuned stochasticity is sufficient.



**Fig. 4.** Typical dynamics of the model with the corresponding histograms. The parameters are  $M = 4$ ,  $\tau = 5$ ,  $q = 0.04$ , and (A)(D)  $p = 0.0005$ , (B)(E)  $p = 0.0016$ , and (C)(E)  $p = 0.01$ .



**Fig. 5.** The change of the peak height as we change  $p$ . The parameters are  $M = 4$ ,  $\tau = 5$ ,  $q = 0.24$ . Note the maximum height is given at  $p = 0.01$  obeying the resonant condition.

### 3 Traffic Models

We now briefly discuss our preliminary results of stochastic resonance in the context of traffic models.

#### 3.1 Computer Traffic Network Model

The phenomenology of the nature of computer network traffic has commanded much attention recently. Analysis, simulations and experiments based on such concepts as “phase transitions” and “self-similarity” are active research topics (see e.g. [17–19]). This paper also addresses the phase transition nature of computer network traffic [20]. Our main focus, however, is not so much on understanding the nature of traffic itself, rather we propose an autonomously adaptive probabilistic routing strategy to “ease” the phase transition point of the network traffic into a congestion state.

**Model** The network architecture considered in the model consists of nodes placed as a two-dimensional lattice (Figure 6). It is a square with  $N$  nodes (routers) on each side and  $N^2$  nodes as a whole. Packets are generated and destroyed on nodes on the boundary of the lattice (squares in the figure), but not on inner nodes (circles). Inner nodes only forward packets received from neighbor nodes. A node on the boundary generates a packet according to the Poisson arrival with  $\lambda$  and sends it to a destination node selected randomly among the nodes on the boundary (including itself). Each node has a receiving queue of unlimited length through which packets are forwarded to the destination and then destroyed.

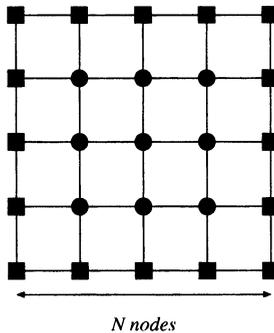


Fig. 6. Network architecture

During each unit time, each node goes through the following process in order to forward packets. It picks a packet from the head of its queue, decides to which node among its neighbors it will forward the packet, and then puts it at

the end of the queue of the selected node. The next node is selected so that the packet is delivered to the destination along the shortest path. If more than one candidate of the next node exist, a strategy is needed to select the recipient. In our simulation we consider two strategies. One of them makes the decision deterministically, which we call “deterministic routing”, and the other does so probabilistically, which we call “probabilistic routing”. With deterministic routing, the node among the candidates to which the least number of packets has been forwarded so far is selected as the next node.

The probabilistic routing strategy we propose and compare with this deterministic routing is given by introducing a particular form of “randomness”. When we have multiple routes  $A$  and  $B$  based on the destination address, we assign the probability to choose a route  $A$  or  $B$  by the following equation:

$$P(A) = \frac{e^{-\beta X_A}}{e^{-\beta X_A} + e^{-\beta X_B}}, \quad (9)$$

$$P(B) = \frac{e^{-\beta X_B}}{e^{-\beta X_A} + e^{-\beta X_B}}, \quad (10)$$

$$1 = P(A) + P(B), \quad (11)$$

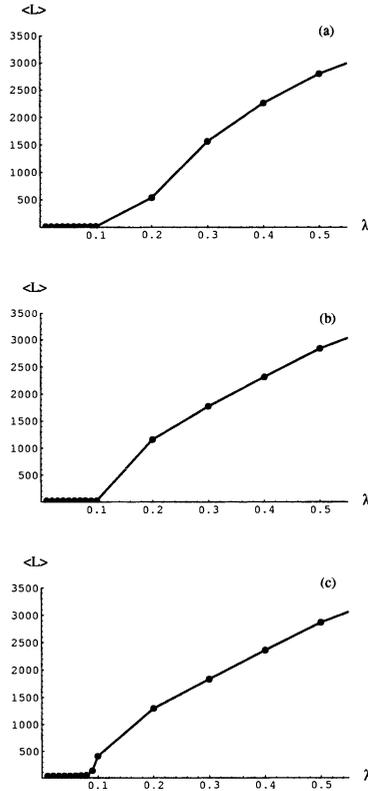
where  $\beta$  is a parameter.  $X_A$  and  $X_B$  are the number of packets the router has already sent in the direction  $A$  and  $B$ .

The characteristics of these routing strategies are “autonomous” and “adaptive”. They are autonomous as their routing choice is made without any reference to or interaction with other routers. And they are adaptive as a routers routing choice changes based on its past choices and hence time.

We now note a couple of points about this probability function. First, we note that the main difference with the deterministic routing is that even if  $X_A > X_B$ , there is some probability of choosing and sending a packet to  $A$ . Second, we can recover the deterministic model by letting  $\beta \rightarrow \infty$ . Also, if we set  $\beta = 0$ , we have a completely random choice of  $A$  and  $B$ , i.e.,  $A$  and  $B$  will be chosen with equal probability of 0.5 regardless of  $X_A$  and  $X_B$ . Hence,  $\beta$  is a control parameter which determines the degree of randomness of choice. This is analogous to the concept of “temperature” in statistical physics [21]. (To be more precise,  $\beta$  corresponds to the inverse temperature.)

This relates to the main motivation for introducing the form of selection probability function given above. In statistical physics, a similar expression arises in “canonical ensembles” and is computationally often realized by the Monte Carlo method [22]. From the experience of phase transitions appearing at some model magnetic systems at a “finite temperature” (i.e., a suitable amount of randomness) with the Monte Carlo method, we can infer that the above type of selection probability function can affect the phase transition point in computer network traffic as well. In the next section, we see that this is indeed the case and with suitable choice of randomness ( $\beta$ ), we can shift the transition point to ease the onset of the congestion phase.

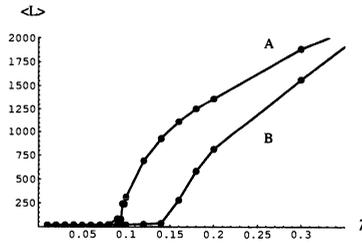
**Simulation results** We quantify network traffic congestion by the average “lifetime” of a packet  $\langle L \rangle$ , which is the average time between the sending and receiving of a packet. (Averages are taken over packets.)



**Fig. 7.** Phase transition behavior with deterministic routing measured in average lifetime of a packet  $\langle L \rangle$  as we vary packet creation rate  $\lambda$ . The system size is varied as  $N =$  (a)15, (b)25, (c)35.

In Figures 7, we show the behavior of  $\langle L \rangle$  as we change the creation rate  $\lambda$  of the packet using the deterministic routing. The simulation is performed with various system sizes of  $N$ , and the system is run up to time step 10,000. The phase transition behavior is clearly observable as  $\lambda$  increases beyond a “critical rate”,  $\lambda_c$ . This transition into the congestion phase is sharper with increasing size as in other physical systems showing phase transitions.

We now compare the deterministic routing with the proposed probabilistic routing. One such example is shown in Figure 8. We can see that the onset of phase transition is moved to higher  $\lambda$  showing that the model network with probabilistic routing can tolerate more packets before going into a congestion phase.

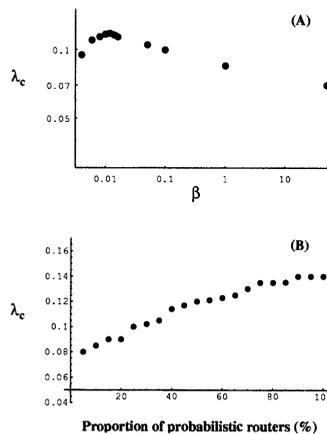


**Fig. 8.** Comparison of the phase transition behavior of the deterministic (A) and the probabilistic ( $\beta = 0.008$ )(B) routing for  $\langle L \rangle$ . The system size is  $N = 20$ .

To examine the effect of randomness for this shift of phase transition point, we plot in Figure 9(A) the phase transition points as a function of  $\beta$ . We see that we need to choose  $\beta$  appropriately ( $\approx 0.01$  in Figure 9(A) ) to have a desired phase point shift.

For individual routers, the deterministic routing appears to be the most balanced way of sending packets to the next router. Use of the probabilistic routing strategy means that this balance is sometimes intentionally upset. The fact that easing of the phase transition point takes place nonetheless means that an emergent collective behavior of routers is playing a crucial role in deciding the congestion nature of the network.

To gain more insight into the collective behavior of the model, we investigate how the phase transition point changes when only a portion of the routers have the probabilistic routing and others operate using the deterministic routing; a representative example is shown in Figure 9(B).



**Fig. 9.** Change of “critical rate”,  $\lambda_c$  as  $\beta$  is varied for the probabilistic routing. The system size is  $N = 20$ . Change of critical rate,  $\lambda_c$  as the proportion of the probabilistic routers is varied in the system. The system size is  $N = 20$  and  $\beta = 0.008$ .

We see that the critical points of phase transition change non-linearly and show saturation as a function of the proportion of probabilistic routers. From a system design point of view, this response shape indicates a possibility of fault tolerance: the effectiveness of the system does not deteriorate appreciably until a certain proportion of routers becoming non-probabilistic. This non-linear shape together with phase transition behavior suggests that the collective behavior of the model is not simply an aggregation of the effect of individual routers. Rather, the interaction among routers, which is indirectly mediated by packets passing through, is playing a role in the collective behavior of the model system.

### 3.2 Pedestrian Model

Among the pedestrian models, which have gained much attention recently, a counter flow model [23] is studied here. Our result is very preliminary, but shows an indication of stochastic resonance of flow in the reverse sense. We mean by “reverse sense” that with tuned “noise” the pedestrian flow shows the worst case of total grid lock.

Pedestrians are randomly placed on a two dimensional rectangular lattice of size  $(W \times L) = (150 \times 400)$ . The boundary condition is periodic on all sides (torus). We consider two sets of  $N = 9000$  pedestrians. One set tries to move to the right and the other to the left on the lattice. We set  $(S_r, S_l)$  as parameters for them to step sideways. At each time step, a pedestrian is chosen to move one step to one of its four neighboring site by the following rules.

- If no one is in front of you in the direction you are heading, he moves one step forward.
- Otherwise, he looks at both his right and left sides.
- If only right (left) side is open and  $S_r > 0$  ( $S_l > 0$ ), he moves to the open site.
- If both sides are open, he moves to the right and left site with the probabilities  $P_r = \frac{S_r}{S_r + S_l}$  and  $P_l = \frac{S_l}{S_r + S_l}$ .
- If neither site is open, he does not move.

With this setting, we performed computer simulations. We obtained the following preliminary results.

- $(S_r, S_l) = (1.0, 0.0)$ . This is the situation where only forward or to the right moves are allowed. We typically observe low flow, but not a grid lock, of pedestrians.
- $(S_r, S_l) = (1.0, 0.3)$ . The left moves are now allowed. However, we typically observe a grid lock situation with no flow.
- $(S_r, S_l) = (1.0, 0.8)$ . The more left moves are taken. Now, we do not see any congestion and they are in free flow.

The above result is an indication that we have a stochastic resonance in the reverse sense: with tuned probability of side stepping, the worst case of no flow situation appeared. More investigations are needed for understanding of this phenomena.

## 4 Discussion

We have considered here a possible connection of stochastic resonance and traffic models. This type of effort is just at the beginning state. We expect, however, that such approach can be useful to investigate congestion problems in various traffic systems.

## References

1. R.L. Stratonovich: *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1963).
2. N.G. van Kampen: *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam 1992).
3. A. Longtin, A. Bulsara, F. Moss: Phys. Rev. Lett. **67**, 656, (1991).
4. K. Wiesenfeld, F. Moss: Nature, **373**, 33 (1995).
5. J.J. Collins, C.C. Chow, T.T. Imhoff: Nature, **376**, 236 (1995).
6. A.R. Bulsara, L. Gammaitoni: Physics Today, **49**, 3, 39 (1996).
7. L. Gammaitoni, P. Hänggi, P. Jung, F. Marchesoni: Rev. Mod. Phys. **70**, 223 (1998).
8. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, Y. Sugiyama: Phys. Rev. E. **51**, 1035 (1995).
9. K. Nakanishi, K. Itoh, Y. Igarashi, M. Bando: Phys. Rev. E. **55**, 6519 (1997).
10. M. Konishi: Neural Computation **3**, 1 (1991).
11. M.C. Mackey, L. Glass: Science **197**, 287 (1977).
12. K.L. Cooke, Z. Grossman: J. Math. Analysis and Applications **86**, 592 (1982).
13. A. Longtin, J. Milton: Biol. Cybern. **61**, 51 (1989).
14. T. Ohira, Y. Sato: Phys. Rev. Lett. **82**, 2811 (1999).
15. T. Ohira, J.G. Milton: Phys. Rev. E. **52**, 3277 (1995).
16. T. Ohira, T. Yamane: Phys. Rev. E. **61**, 1247 (2000).
17. W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson: Communication Review, **23**, 183 (1993).
18. M. Takayasu, H. Takayasu, T. Sato: Physica A, **233**, 824 (1996).
19. M.S. Taqqu, W. Willinger, R. Sherman: ACM/SIGCOMM Computer Communication Review, **27**, 5 (1997).
20. T. Ohira, R. Sawatari: Phys. Rev. E. **58**, 193 (1998).
21. R.K. Pathria: *Statistical Mechanics* (Pergamon Press, Oxford 1972).
22. K. Binder, K: *Monte Carlo Methods in Statistical Physics* (Springer-Verlag, Berlin 1986).
23. M. Muramatsu, T. Irie, T. Nagatani: Physica A, **267**, 487 (1999).

# An Interpretation of a Traffic Engineer on Vehicular Traffic Flow

M. Koshi

Nihon University, EXTEC, UBE Bldg. 9<sup>th</sup> Floor, 3-7-2 Kasumigaseki, Chiyoda-ku, Tokyo 100-0013, Japan

**Abstract.** The present paper focuses on capacity of tunnels and sags on motorways. Free flow becomes congested due to speed reduction caused by sudden change of light condition at tunnel entrances or by unrecognized increase of gradient at sags combined with drivers' behavior that they do not shorten their car-following spacing corresponding to the speed reduction, resulting in reduction of flow rate of successive vehicles. The free flow to congested flow breakdown takes place at the flow rate level of around 3,000 vph per two-lane (one direction). Once congestion queue is formed, the capacity value is determined by drivers' behavior of departing from the queue at its front-end. This departure flow rate is in the range of 2,200 to 2,700 vph per 2-lane and varies depending on time during which the driver has been caught in the queue, light condition (day or night) and some other factors.

## 1 Free Flow and Congested Flow

It is a well-known fact that there are two states of vehicular traffic flow, free flow and congested flow.

Traffic flow takes the form of free flow when there is no influence of capacity restriction of a bottleneck in the downstream section. The characteristics of the flow (flow rate or volume as well as speed) on a road section depends on the inflow from the upstream section and the geometry of the section in question.

Traffic flow takes the form of congested flow in a queue which is formed on the upstream section of a bottleneck. The volume of the congested flow is equal to the capacity of the bottleneck regardless of the traffic demand arriving at the upstream tail of the queue.

Bottlenecks are such sections or spots of road where the capacity is lower than at the upstream sections as well as at the downstream sections. Stop lines (signalized or non-signalized), diverging as well as merging areas, weaving sections, tunnel entrances and sags (road sections of increasing vertical gradient) are typical examples of bottlenecks.

A queue is formed when traffic demand or inflow volume exceeds the capacity of the bottleneck. The queue keeps to grow in length as long as the traffic demand exceeds the bottleneck capacity and decreases in length when the traffic demand becomes lower than the capacity. Congested flow can only be observed in such queues.

As far as the author has learned, any simple homogeneous section of road has never been a bottleneck, i.e., traffic flow never switches by itself from free flow to congested flow just by chance at an arbitrary spot on a stretch of simple homogeneous road section. This is perhaps because in reality it is practically impossible to feed traffic into a simple homogeneous section up to its saturation level of free flow without causing bottleneck phenomenon on the upstream merging areas.

## **2 Capacity of Bottlenecks**

Capacity values of bottlenecks are determined by various aspects of drivers' behavior depending on the types of bottlenecks.

Signalized stop line capacity depends on green light split and saturation flow rate which is departure flow rate from a stopped queue given a "go" signal. Capacity of a "stop" controlled stop line depends on rate of acceptable gaps in the major flow of traffic. Capacity of merging/diverging areas and weaving sections depends on lane-selection, lane-changing and gap-acceptance behavior of drivers in a complex manner.

Capacity values and traffic flow phenomena at tunnel entrances and sags are quite interesting and suggestive for understanding drivers' behavior. Let us take a closer look at them in the following sections.

## **3 Capacity of Free Flow at Tunnel Entrances and Sags**

### **Tunnel Entrances**

Drivers tend to reduce their speed slightly when they enter a tunnel due to the sudden change of light conditions but partly don't reduce their car following spacing because they are not fully conscious about the speed reduction and because they tend to need longer car-following spacing in tunnels than outside of tunnels for the same speed as shown in Fig. 1. This driver behavior brings about reduction of flow rate (or inverse of time headway) of successively following vehicles. When a long platoon passes vehicles in the tail part of the platoon are forced to reduce their speed to a large extent or even come to complete stop due to accumulation of the flow rate reduction of the preceding cars. When traffic volume reaches approximately 3,000 pcu/2-lane the next platoon tends to arrive before the slow moving or even stopped vehicles of the preceding platoon are cleared. This triggers congestion at tunnel entrances.

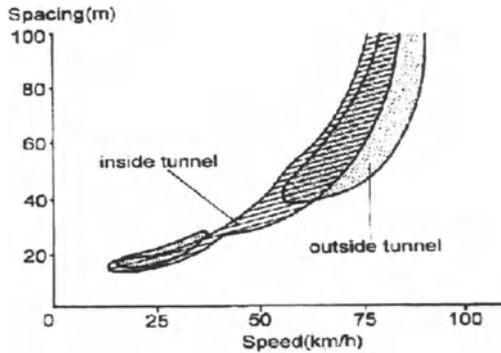


Fig.1 Speed-spacing Relationships Inside and Outside of Tunnel  
Kobotoke Tunnel, Chuo Motorway, Vehicle detector Data (ref. 1)

### Sags

Drivers tend to reduce their speed unconsciously at sags due to gradual increase of gradient when the road alignment is such that the existence of the sag is not easily visible to the drivers. When speed is reduced, the drivers tend not to duly reduce their car-following spacing because they are not fully aware of the speed reduction.

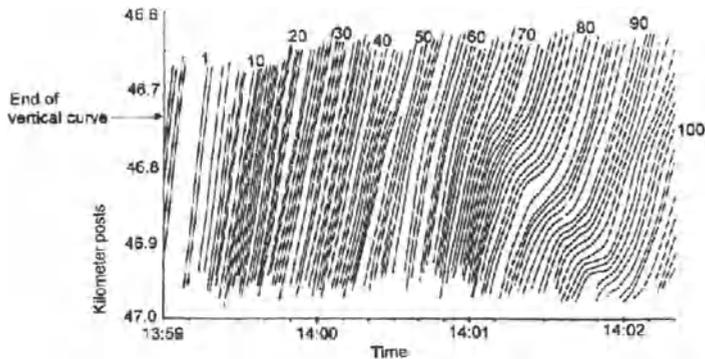


Fig 2 Time-space Trajectories of Vehicles just before the start of Congestion  
Hatano Sag, Tomei Expressway, Aerial Photo Observation, 1991  
(ref.4)

Figure 2 shows the trajectories of vehicles just before the start of congestion at a sag. Vehicles # 35 to # 70 are those who triggered congestion. Fig. 3 shows the typical examples of speed-spacing trajectories of those vehicles. Trajectory type 1 means that the rate of flow increases due to the speed reduction whereas type 2 does not cause any significant change in flow rate and type 3 results in flow rate reduction. Out of 36 vehicles from # 35 to # 70, 27 vehicles (75%) behaved as type 3 whereas 4 (11%) and 5 (14%) did as type 1 and 2, respectively. It is shown in Fig. 3 that accumulation of flow rate reduction brings about break down of free

flow. The breakdown of free flow takes place when the free flow volume exceeds approximately 3,000 pcu/h/2-lane.

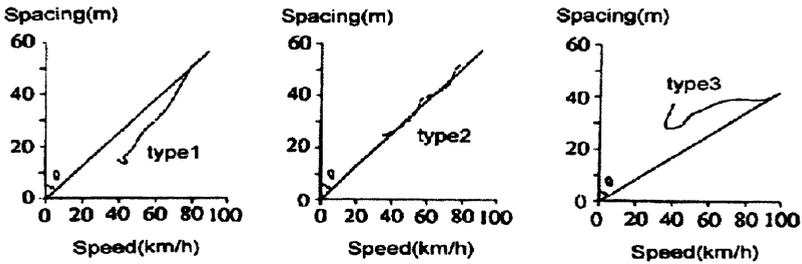


Fig.3 Typical Examples of Speed-spacing Trajectories of Vehicles #35 to #70 in Fig.2 (ref.4)

#### 4 Flow Rate Reduction after Breakdown of Free Flow

Flow rate after breakdown of free flow becomes equal to the rate of departure from the queue at its front end.

Figures 4 and 5 show typical examples of motorway tunnel congestion. Fig. 4 shows speed and volume at Tsuburano Tunnel entrance on August 13, 1988 and Fig. 5 shows the same variables at the same site on the same day but two years later, in 1990.

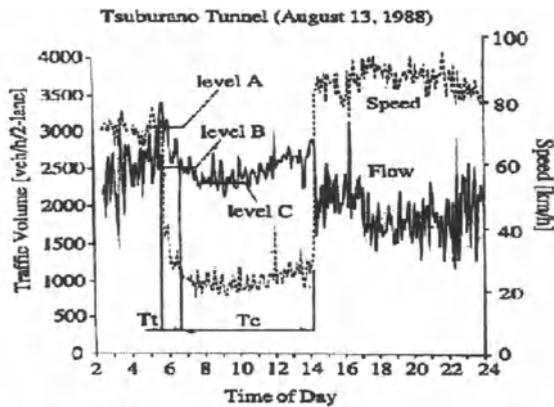


Fig. 4 A Typical Example of Motorway Tunnel Congestion (5-min. Data of Detectors at 200 m Upstream of the Tunnel Entrance). (ref.3)

It can be seen in Fig. 4 that free flow breaks down at volume of level A and the speed drops down to approximately 20 Km/hr taking a transition period of  $T_t$  and then becomes relatively stabilized until the end of congestion during  $T_c$ . The volume also drops down to level B after the transition period  $T_t$  and further down to

level C afterwards. Both speed and volume are still changing, however, even during the relatively stabilized period  $T_c$ . The transition period  $T_t$  was approximately an hour in 1988 congestion whereas it was about two hours in 1990 congestion.

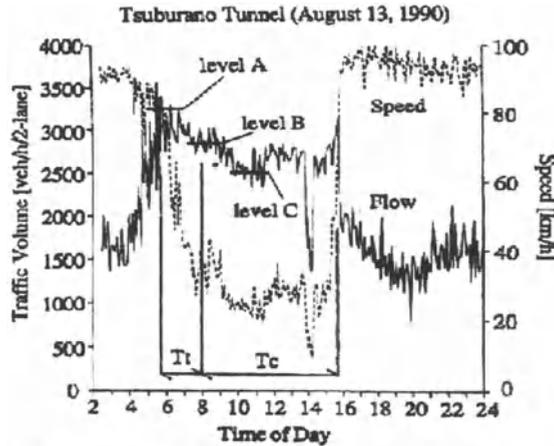
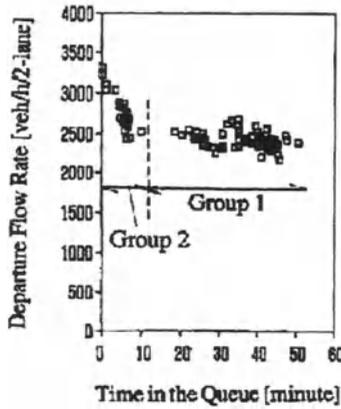


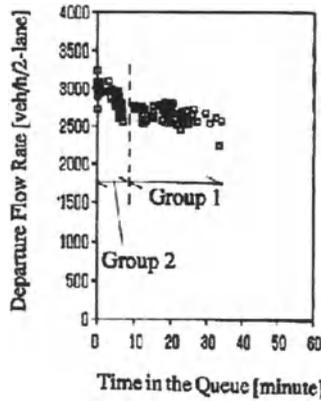
Fig. 5 Another Example of Congestion at the Same Site as in Fig. 1.

Figures 6 and 7 show time during which vehicles have been caught in the queue (TIQ, Time In the Queue) vs. departure flow rate of the vehicles. Figs. 6 and 7 correspond to the congestion shown in Figs. 4 and 5, respectively. Vehicles in Group 1 are those which depart during  $T_c$  in Figs. 4 and 5 and most of those in Group 2 are vehicles which depart during  $T_t$  (transition period) in Figs. 4 and 5 and the rest of vehicles in Groups 2 are those which depart just before disappearance of congestion.

It is found from Figs. 6 and 7 that (1) transition flow from breakdown of free flow to stabilized congestion depends on TIQ of the departing vehicles rather than the absolute time after the breakdown, (2) the transition takes place when TIQ is up to approximately 10 minutes and (3) the departure flow rate still keeps slightly decreasing as TIQ increases even in the region of TIQ is over 10 minutes.



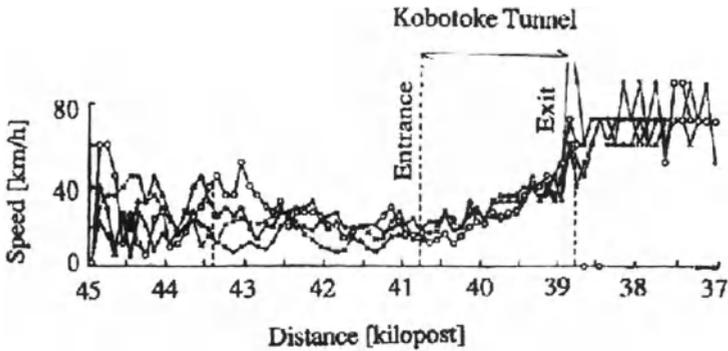
**Fig.6** Departure Flow Rate vs. Time in the Queue for the Congestion Case Shown in Fig. 4  
(15 min. Moving Average at 5 min. Intervals). (ref.3)



**Fig.7** Departure Flow Rate vs. Time in the Queue for the Congestion Case Shown in Fig.5  
(15 min. Moving Average at 5 min. Intervals). (ref.3)

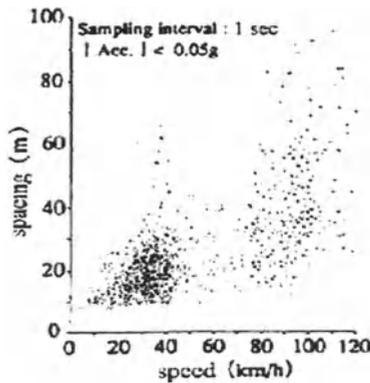
These phenomena come totally from drivers' behavior. Drivers tend to loose more their tension in following cars ahead as they have been caught in the queue longer. Particularly during first ten minutes in the queue the decrease of drivers' tension is much quicker than after 10 minutes of TIQ.

After the transition period  $T_t$  has elapsed, vehicles depart the queue with extremely low rate of acceleration and the drivers do not realize that they are departing the queue and entering free flow condition. Fig. 8 shows an example of speed profile of floating test car runs. It is seen that approximately 2 km is needed to accelerate from 20 km/hr to 60 km/hr.



**Fig.8 Speed Profiles of Floating Runs in Tunnel Entrance Congestion.** (ref.3)

It is a well-known fact that there is a discrepancy in car-following behavior and accordingly in Q-K-V relationships between free flow and congested flow.



**Fig 9 Speed-Spacing Observed in Real Flow**  
Tomoi Motorway, Hadano Sag, Inner Lane (1991)

Figures 9 and 10 show examples of speed-spacing relationship which were observed in real flow and experiments by test cars. In the experiments the test drivers showed continuous speed-spacing relationships through free and congested conditions whereas in real flow drivers tend to take longer spacing in the congested condition than the continuous relations resulting in a discrepancy between free and congested conditions.

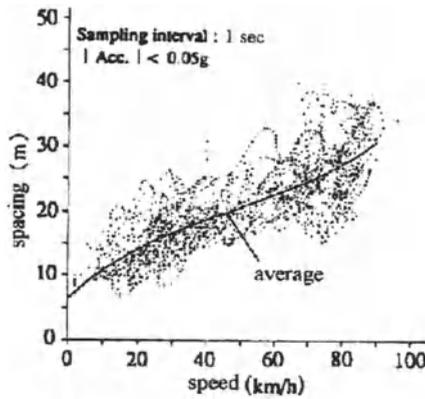


Fig 10 Speed-Spacing in Car Following Experiment  
Chuo Motorway (ref.4)

It is the author’s estimation that ordinary drivers in real flow have basically continuous speed-spacing relationships but as TIQ increases in congested flow they loose their intension to keep following closely to the cars ahead and increase their spacing.

### 5 Effect of Light Condition on Departure Flow Rate

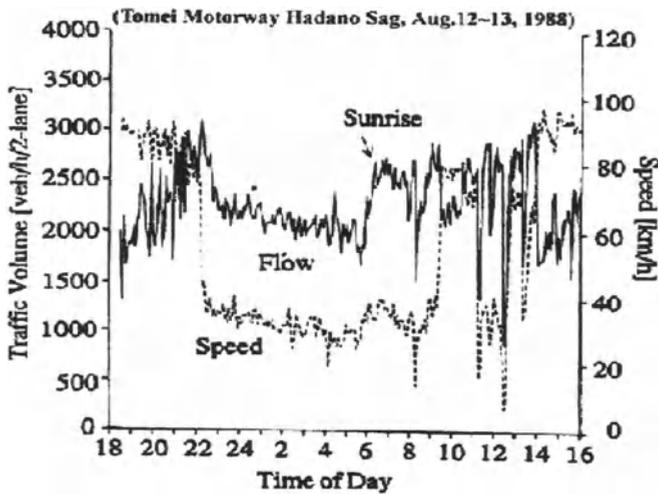
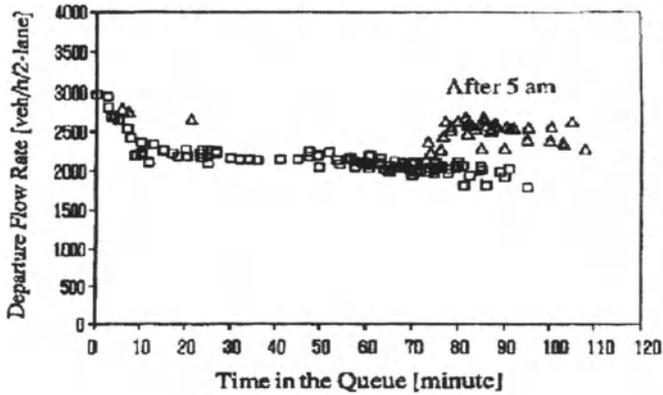


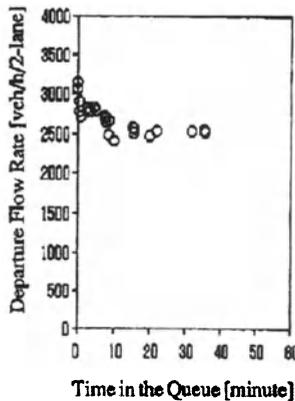
Fig.11 Increase of Departure Flow Rate at Sunrise

Figure 11 shows that departure flow rate from congestion which started at night increases remarkably at sunrise in the morning. Fig. 12 is TIQ vs. the departure

flow rate expression of the same congestion. Fig. 13 shows TIQ vs. departure flow rate of congestion at the same site as in Fig. 11 but only during daytime.



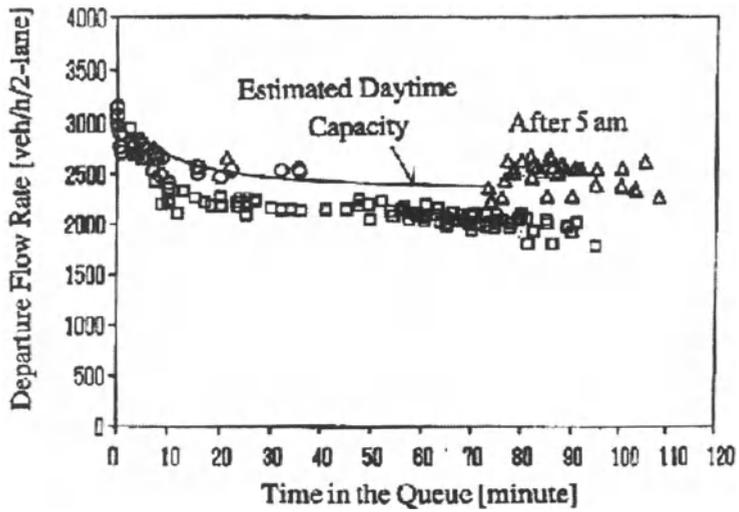
**Fig.12** Departure Flow Rate vs. Time in the Queue at Sunrise for the Congestion Case Shown in Fig.11. (15 min. Moving Average at 5 min. Intervals). (ref.3)



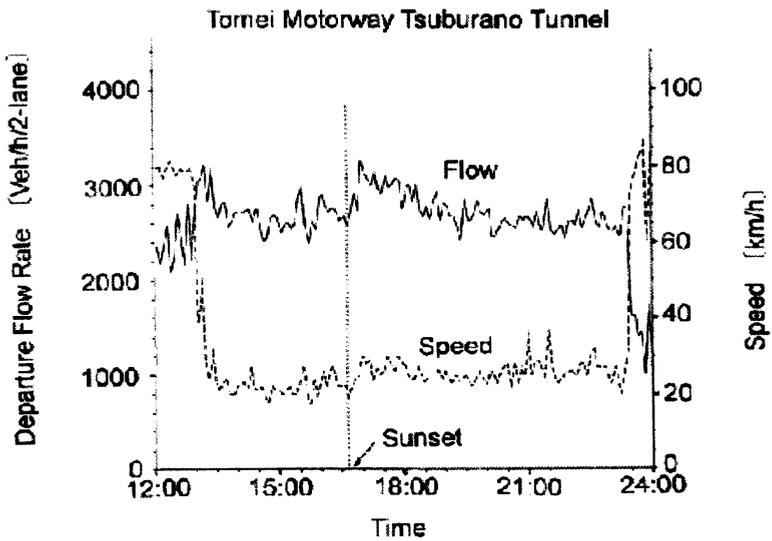
**Fig.13** Departure Flow Rate vs. Time in the Queue during Daytime (At the Same Sites as in Fig.11). (15 min. Moving Average at 5 min. Intervals). (ref.3)

Figure 14 is an overlap of Fig. 12 and Fig. 13. Estimated daytime capacity values are significantly higher than nighttime values as shown in Fig. 14.

Figure 15 shows increase in departure flow rate of tunnel congestion at sunset. Due to the tunnel illumination drivers have better light conditions in the tunnel than in the open section after sunset resulting in higher departure flow rate.



**Fig.14** Comparison of Capacity at Sunrise with the Daytime Capacity (At the Same Sites as in Fig.11). (15 min. Moving Average at 5 min. Intervals). (ref.3)



**Fig.15** Increase of Departure Flow Rate at Sunset

## 6 Synchronization Between Lanes in Congested Conditions

It is a well known fact that vehicular flow tends to oscillate in terms of Q-K-V and the oscillation in the parallel lanes often synchronize with each other.

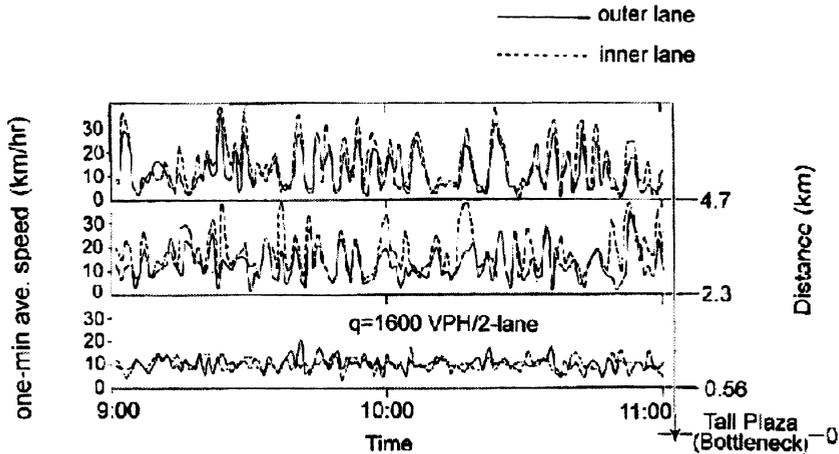


Fig. 16 Speed Profile of Individual Lanes (1min. ave. speed) (ref.2)

The author's hypothesis is that oscillation itself may be generated when the departure flow rate from the front end of the low-speed platoon is larger than the capacity of the bottleneck. The author also assumes that synchronization between lanes may be caused mainly by lane changing from slower lane to faster one.

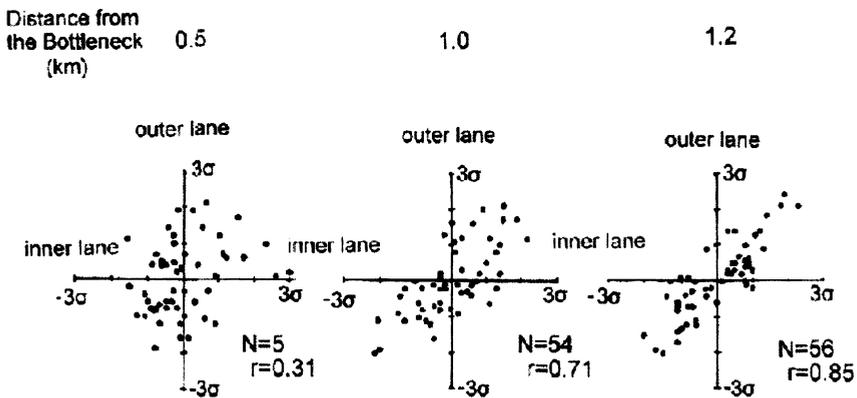


Fig. 17 Speed Correlation Between Lanes (ref.2)

Figure 16 shows speed profile of the two parallel lanes at different spots in a congested flow. It is seen that the amplitude of oscillation is not large at the immediate upstream of the bottleneck.

Figure 17 shows correlation of speeds between the lanes at different spots in a congested flow. It can be seen that synchronization becomes clearer as the upstream distance from the bottleneck increases.

## 7 Concluding Remarks

- (1) The characteristics of vehicular traffic flow is determined totally by drivers' behavior such as speed-spacing of car following and lane selection/changing.
- (2) Dynamic speed-spacing relationships when accelerating particularly determines departure flow rate from a queue or capacity of sag and tunnel bottlenecks and is affected by time during the driver is caught in the queue, light conditions and probably some other factors as well.
- (3) Congested flow is the flow in a queue which is formed in the upstream section of a location-specific bottleneck. Traffic flow does not switch from free flow to congested flow by itself at an arbitrary spot on a stretch of a simple homogeneous road section.

## References

1. M. Koshi, Traffic Phenomena in Expressway Tunnels, IATSS Review, International Association of Traffic and Safety Sciences, Vol. 10, No. 1, 1984.3.
2. M. Koshi et. al., A Study of the Phenomena of Congested Traffic Flows, Journal of Infrastructure Planning and Management, Japan Society of Civil Engineers, No. 306, pp. 59-69, 1981.2.
3. M. Koshi et. al., A Study on Congestion Phenomena at Tunnels and Sags on Motorways, Journal of Infrastructure Planning and Management, Japan Society of Civil Engineers, No. 458/IV-18, pp. 65-71, 1993.1.
4. J. Xing and M. Koshi, A Study on the Bottleneck Phenomena and Car-Following Behavior on Sags of Motorways, Journal of Infrastructure Planning and Management, Japan Society of Civil Engineers, No. 506/IV-26, pp. 45-55, 1995.1.

# Congestion Induced by Bottlenecks in Two-Lane Optimal Velocity Traffic Flow Model

S. Tadaki<sup>1</sup>, M. Kikuchi<sup>2</sup>, K. Nishinari<sup>3</sup>, Y. Sugiyama<sup>4</sup>, and S. Yukawa<sup>5</sup>

<sup>1</sup> Computer and Network Center, Saga University, Saga 840-8502, Japan

<sup>2</sup> Cybermedia Center, Osaka University, Toyonaka 560-0043, Japan

<sup>3</sup> Department of Applied Mathematics and Informatics, Ryukoku University, Ohtsu 520-2194, Japan

<sup>4</sup> Division of Mathematical Science, City College of Mie, Tsu, Mie 514-0112, Japan

<sup>5</sup> Department of Applied Physics, University of Tokyo, Bunkyo 113-8656, Japan

**Abstract.** The optimal velocity (OV) model is one of car-following traffic flow models, which can reproduce some of observed features. The model is discretized to define the coupled map traffic flow model based on optimal velocity functions (CMOV model). Based on the CMOV model, the effect of noises and traffic bottlenecks are discussed. By introducing a simple set of lane-changing rules, the traffic flow on two-lane expressways with a bottleneck is observed.

## 1 Introduction

Physical understanding of the traffic flow on expressways has been improved mainly based on mathematical models and their computer simulations. Most of models describe the traffic flow on one-lane expressways. They seem to succeed to describe some fundamental features of traffic flow phenomena. Most of real expressways, on the contrary, have more than two lanes. To compare simulation results with observational data, traffic flow models for multi-lane expressways should be constructed.

The traffic flow on real two-lane expressways shows some interesting phenomena, which are not observed on one-lane expressways. The first is a *reverse lane usage* phenomenon. Usually drivers are requested to run on the slow lane (the left lane in Japanese expressways) and are allowed to run on the fast (right) lane only for overtaking. The observed data, however, show that the flow on the fast lane exceeds the one on the slow lane [1]. The second is the *synchronized flow*, in which the average velocity on the fast lane is very close to the one on the slow lane during congestion [2,3].

There are mainly three types of traffic flow models. Macroscopic models have bases on hydrodynamic or gas-kinetic descriptions. Microscopic models are cellular automaton models and car-following models. The Optimal Velocity (OV) traffic flow model [4–6], which is a one of car-following models, has succeeded to describe the phase transition from a freely-moving phase to a jam phase by changing the traffic density. We have proposed a new simulation model, Coupled Map traffic flow model based on Optimal Velocity functions (CMOV model) [7–9], by discretizing the original OV model.

In this paper, the CMOV model and the simulation results with the model are briefly summarized. The model is extended to be applicable to two-lane expressways by introducing a simple set of lane-changing rules. We discuss simulation results for two-lane expressways by comparison with real observational data.

## 2 Optimal Velocity Model and Coupled Map Model

The optimal velocity model is one of car-following traffic flow models [4–6]. In car-following traffic flow models, a driver controls his car according to stimuli from the preceding car. The essential idea the OV model introduced is the optimal velocity function  $V_{\text{optimal}}$ . A driver accelerates or decelerates his car to tune the speed to the optimal (safety) velocity  $V_{\text{optimal}}$ , which depends on the headway. In the OV model, the position  $x_n$  of the  $n$ th car obeys the second order differential equation:

$$\frac{d^2x}{dt^2} = \alpha \left[ V_{\text{optimal}}(\Delta x_n) - \frac{dx}{dt} \right], \quad (1)$$

where  $\Delta x_n = x_{n-1} - x_n$  is the headway distance to the preceding car, and  $\alpha$  is the susceptibility. The index  $n$  to distinguish car is omitted hereafter.

The form of the OV function can be chosen arbitrarily under some conditions. For realistic cases, we can use the following form:

$$V_{\text{optimal}}(\Delta x) = \frac{v_{\text{max}}}{2} \left[ \tanh \left( \frac{\Delta x - d}{w} \right) + c \right]. \quad (2)$$

Parameters in Eq. (2) can be selected to fit observed data for the headway-velocity relation (see Table 1) [4,5,10].

**Table 1.** Parameters in the OV function

Parameter	value
$d$	25m
$w$	23.3m
$V_{\text{max}}$	33.6m/sec
$c$	0.913
$\Delta x_{\text{min}}$	7.02m

The OV model can reproduce some observed features for the traffic flow on one-lane expressways. The simple form of the model, moreover, enables us to discuss some analytically exact solutions [11,12]. Because the model is described with a set of differential equations, however, the sequence of cars should be fixed

as an initial condition. We need some modification to apply the OV model to realistic simulations.

We had introduced the Coupled Map traffic flow model based on Optimal Velocity functions (CMOV model) by discretizing the original OV model [7,8]. The aim of the discretization is not to integrate the differential equations numerically but to make a new model applicable to various realistic simulation situations. The CMOV model is defined as follows: when the headway distance  $\Delta x$  is sufficiently large, the position  $x$  and velocity  $v$  of a car are updated as follows:

$$x(t + \Delta t) = x(t) + v(t)\Delta t, \tag{3}$$

$$v(t + \Delta t) = v(t) + \alpha [V_{\text{optimal}}(\Delta x(t)) - v(t)] \Delta t. \tag{4}$$

The time step  $\Delta t$  is a fixed parameter. To avoid rear-end collisions and backward motions of cars, cars are forced to stop if the headway distance  $\Delta x$  is less than the minimum value  $\Delta x_{\text{min}}$ . Namely for  $\Delta x < \Delta x_{\text{min}}$  (3) and (4) are replaced with

$$x(t + \Delta t) = x(t), \tag{5}$$

$$v(t + \Delta t) = 0. \tag{6}$$

These conditions, however, correspond to traffic accidents and are hardly used in simulations.

One of the most important improvements upon the OV model is that the CMOV model allows us to change the sequences of cars. The CMOV model, therefore, enables us to simulate the traffic with open boundary conditions, where new cars are injected from an entrance to the tail of the sequence of cars and ejected from an exit. Moreover we can construct two-lane models with the CMOV model.

### 3 Congestion on One-Lane Expressways

First we observe congested states on one-lane expressways with open boundaries with the CMOV model. The setup for simulation is shown in Fig. 1. The system consists of three segments: the left and right parts are for relaxation, whose lengths are  $L_1$  and  $L_2$ , respectively. The middle part with length  $L_{\text{observe}} = 1\text{km}$  is used for measurements.

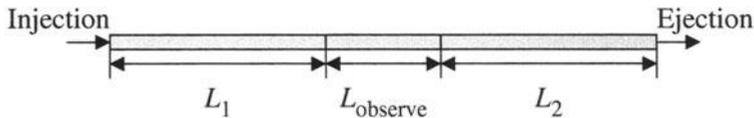
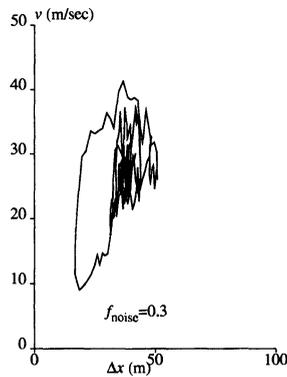


Fig. 1. The setup of an one-lane expressway with open boundaries.

The system has no car at the initial state. From the left side, the upper stream, a car is injected every second with some probability if the distance between the left end and the tail of the sequence of cars is larger than  $\Delta x_{\min}$ . The initial speed of the injected car is zero.

The occurrence of the traffic jam is recognized in the headway-velocity plane. In the original OV model and the CMOV model under a periodic boundary condition, a closed trajectory called a hysteresis loop is observed as an evidence of the jam flow.

If there is no bottleneck in the system, an injected car just catches up the preceding car and maintains the speed. Therefore the congested flow does not occur. In the headway-velocity plane, the trajectory of an injected car does not show a loop structure.



**Fig. 2.** The values of the headway and velocity of a test car for  $f_{\text{noise}} = 0.3$ .

To observe congested states, we introduce a multiplicative noise term into (4).

$$v(t + \Delta t) = [v(t) + \alpha \{V_{\text{optimal}}(\Delta x(t)) - v(t)\} \Delta t] (1 + f_{\text{noise}} \xi), \quad (7)$$

where  $\xi \in [-0.5, 0.5]$  is a uniform random variable and the noise level is given by a parameter  $f_{\text{noise}}$ . The multiplicative form of the noise is a reasonable choice, because the errors are enhanced in the high speed region and suppressed in the low speed region.

Figure 2 shows the trajectory in the headway-velocity plane of a test car for  $f_{\text{noise}} = 0.3$ . The loop structure is clearly shown. The headway shortage created by the noise propagates upstream and induces the congested traffic flow. We call this the *noise induced congested flow*.

The density is observed as the number of cars in the segment  $L_{\text{observe}}$  in Fig. 1 per length. The time average of the density increases as the square of the noise level  $f_{\text{noise}}$  (Fig. 3).

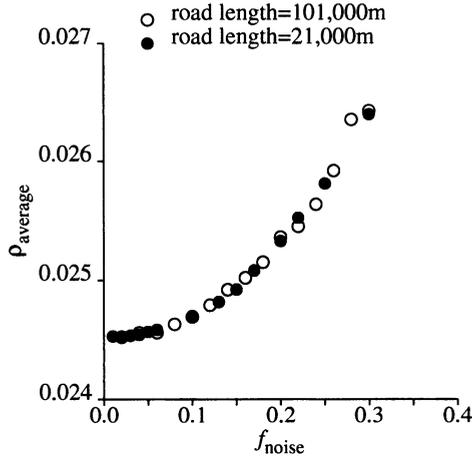


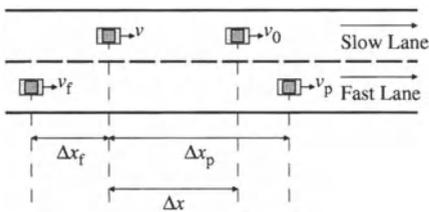
Fig. 3. The average density of cars increases as the square of the noise level.

### 4 Model for Two-Lane Traffic

One of the advantage of the CMOV model is the flexibility in the sequence of cars. Rule-based behavior such as lane-changing can be introduced into the CMOV model. A model of two-lane expressways can be constructed by introducing a set of lane-changing rules into the CMOV model.

Lane-changing rules consist of various elements. They have never been discussed well based with observational data. Moreover, lane-changing behavior depends on traffic regulations and driving customs. So it is difficult to construct a general set of lane-changing rules [13]. Here we introduce a simple and reasonable set of lane-changing rules.

(a) Move to Fast Lane



(b) Move to Slow Lane

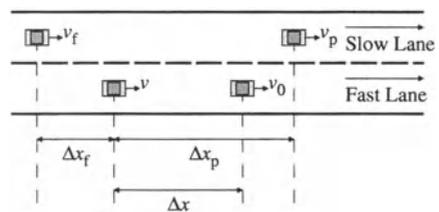


Fig. 4. Lane-changing rules for moving to the fast lane (a) and moving to the slow lane (b).

We introduce a set of asymmetric lane-changing rules to describe the traffic flow on expressways. Each car on the slow lane tries to move to the fast lane for

overtaking if it can not run at the desired speed because of short headway. Each car in the fast lane tries to move back to the slow lane if it can run at the desired speed or faster than the current speed in the slow lane. If the lane-changing is safe, the car really changes the lane with some probabilities. These lane-changing is synchronously preformed for all cars on the slow lane first and on the fast lane second at each time step before the motion described by (3) and (4).

The lane-changing rule for moving to the fast lane is shown in Fig. 4(a). Consider a car running with the speed  $v$  on the slow lane. If the headway  $\Delta x$  is not enough ( $\Delta x < \Delta x_{\text{safe}}$ ), the car wants to move to the fast lane. The lane-changing is allowed if the lane-changing is safe ( $\Delta x_f > D(v)$ ,  $D(v)$  is a function to measure the safe distance from a current speed). If all conditions mentioned above are satisfied, the car changes the lane with a probability  $P_{\text{up}}$ .

The lane-changing rule for moving to the slow lane is the same with those for moving to the fast lane except one additional rule (Fig. 4(b)). The probability for moving to the slow lane is  $P_{\text{down}}$ . The additional rule is introduced because on expressways drivers are requested to run on the slow lane and are allowed to run on the fast lane only for overtaking. If the slow lane is clear ( $\Delta x_p > \Delta x_{\text{clear}}$ ) and the lane-changing is safe, the car moves to the slow lane independently of the headway  $\Delta x$ .

The rule mentioned above contains many parameters. In the simulations discussed here, we use the following values:

$$\Delta x_{\text{safe}} = d + \frac{w}{2}, \quad (8)$$

$$D(v) = d + w \left( \frac{v}{v_{\text{max}}} - \frac{2}{1} \right), \quad (9)$$

$$\Delta x_{\text{clear}} = 5\Delta x_{\text{safe}}, \quad (10)$$

and  $P_{\text{up}} = 0.8$  and  $P_{\text{down}} = 0.5$ .

In addition to the parameters included in the set of lane-changing rules, two parameters are introduced to the model. The first is named an acceleration factor  $f_{\text{overtake}}$ . A driver changing lane to the fast lane wants to overtake the preceding car on the slow lane. He will accelerate beyond his usual desired speed for overtaking. Therefore all cars on the fast lane assumed to change their original desired speed to  $f_{\text{overtake}}v_{\text{max}}$  during running on the fast lane. The second is the variety  $\delta$  of desired speed. The values of desired speed are assumed to distribute uniformly in  $[(1 - \delta/2)v_{\text{max}}, (1 + \delta/2)v_{\text{max}}]$ .

Figure 5 shows the schematic view of the simulated system with two lanes. To observe the congested flow, a road segment with speed limit is installed into the right half of the segment  $L_2$ . In the segment with speed limit, the desired speed  $v_{\text{max}}$  for all cars is reduced to  $rv_{\text{max}}$  ( $r \leq 1$ ). At an observation point, the flow (the number of cars passing the point) and the average speed during 5 minutes are recorded according to the observational data for real expressways.

First we observe the lane-usage characteristic for  $\delta = 0.4$ ,  $f_{\text{overtake}} = 1.2$  and  $r = 0.5$ . The lane-usage characteristic is a ratio of the flow on the fast or slow

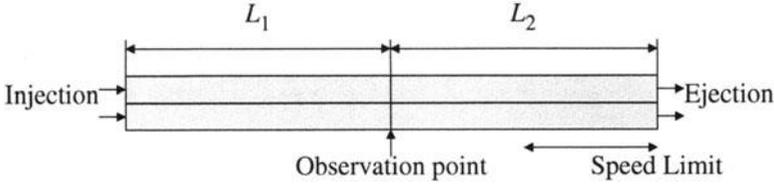


Fig. 5. The setup of a two-lane expressway with open boundaries.

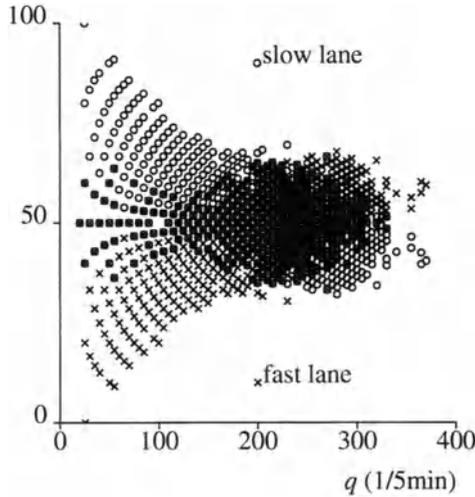


Fig. 6. The lane-usage characteristics for  $\delta = 0.4$ ,  $f_{\text{overtake}} = 1.2$  and  $r = 0.5$ .

lane to the total flow:

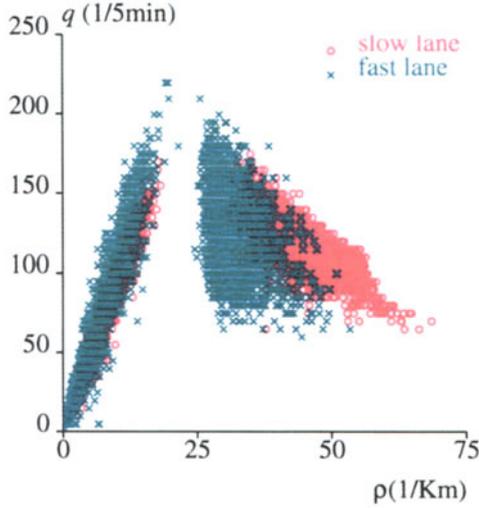
$$r_i = \frac{q_i}{q_{\text{fast}} + q_{\text{slow}}}, \quad i \in \{q_{\text{fast}}, q_{\text{slow}}\}. \tag{11}$$

In the traffic regulation assumed in the lane-changing rules, all drivers are required to run on the slow lane and are allowed to run on the fast lane only for overtaking. The simulation result (Fig. 6) shows that the flow on the fast lane exceeds the one on the slow lane in the high-flow region. The *reverse lane usage* is reproduced.

Figure 7 shows the fundamental diagram, the density-flow relation. The region in the left side of the peak corresponds to the freely moving phase. The right region corresponds to congestion. The flow on the fast lane shows rather sharp peak than on the slow lane.

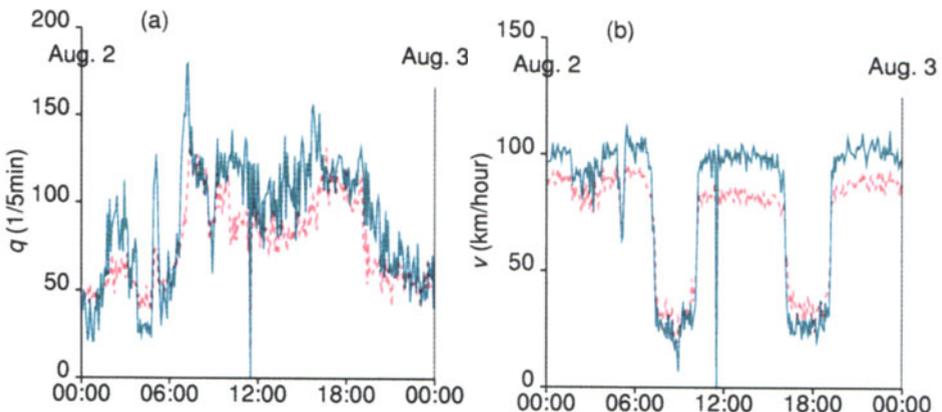
## 5 Comparison with Observed Data

To compare the simulation results with the real traffic flow, we show some examples of observational data. The observational point locates at 170.64 km from



**Fig. 7.** The fundamental diagram for  $\delta = 0.4$ ,  $f_{\text{overtake}} = 1.2$  and  $r = 0.5$ .

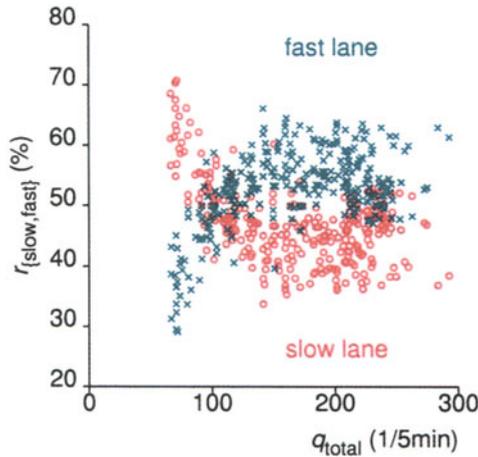
Tokyo on Tomei Expressway linking Nagoya with Tokyo. The traffic flow was observed on lanes bound for Tokyo. There is Nihonzaka Tunnel 2km downstream. The data was taken on 2nd August 1996 by Japan Highway Public Corporation. At the observational point, a pair of induction loop coils is installed and observe the number of car passing the point and the average speed for 5 minutes. The average density is calculated with the flow divided by the speed.



**Fig. 8.** The time sequence of (a) the flow and (b) the speed observed on 2nd Aug. 1996. The solid and broken lines correspond to the values on the fast and slow lanes, respectively.

Figure 8 shows the time sequence of the flow and the speed. The flow (Fig. 8(a)) on the fast lane always exceeds the one on the slow lane. The time sequence of the speed (Fig. 8(b)) shows the clear transition from the freely-moving state to the congested state. On the contrary to the behavior of the speed, any significant changes do not occur in the behavior of the flow at the transition.

The time sequence of the speed during congestion shows strong correlations between the flow on the fast and slow lanes. The speed on the slow lane is larger than the one on the fast lane during congestion against expectations.



**Fig. 9.** The lane-usage characteristic.

As seen in the time sequence of the flow (Fig. 8), the reverse lane usage can be observed in the lane usage characteristic (Fig. 9). The transition from the normal lane usage to the reverse one occurs at the lower total flow than in the simulation (Fig. 6).

## 6 Summary and Discussion

We introduced the coupled map traffic flow model based on optimal velocity functions by discretizing the original optimal velocity model. By temporal discretization of the optimal velocity model, we can introduce a set of lane-changing rules to the coupled map traffic flow model. We employ a set of simple and intuitive lane-changing rules.

The model is applied to an open-boundary two-lane expressway. The simulation can reproduce some macroscopic features, such as the fundamental diagram and the lane usage characteristics. The reverse lane usage, especially, can be reproduced by the simulation. Namely the flow on the fast lane exceeds the one on the slow lane.

One of the key issues for the reverse lane usage is the distribution of the desired speed  $v_{\max}$ . We can observe that cars with higher desired speed prefer to run on the fast lane as expected. It is stressed that the variety of desired speed  $\delta$  and the acceleration  $f_{\text{overtake}}$  on the fast lane is necessary to reproduce the reverse lane usage in the simulation. The importance of the variety of cars was also discussed by Treiber and Helbing [14].

The simulation, however, does not well reproduce some of the observed feature in congestion. The flow on both lanes enter the congested state synchronously in the simulation as observed in real data. Even in congested states, however, the velocity on the fast lane is larger than the one on the slow lane. The acceleration factor  $f_{\text{overtake}}$  affects to keep the difference of the velocity between lanes.

## Acknowledgment

The authors thank the Japan Public Highway Corporation for providing us the observation data. A part of this work is financially supported by Grant-in-aid No. 12650065 and 12680446 from Ministry of Education, Science, Sports and Culture, Japan. It is also supported by Center for Integrated Research in Science and Engineering in Nagoya University, and by Namura Shipbuilding, Co. Ltd.

## References

1. K. Nishinari and M. Hayashi, (Eds.), in: *Traffic Statistics in Tomei Express Way*, (The Mathematical Society of Traffic Flow, Japan, 1999).
2. B.S. Kerner and H. Rehborn, *Phys. Rev. E* **53**, 1297 (1996); *Phys. Rev. E* **53**, 4275 (1996); *Phys. Rev. Lett.* **79**, 4030 (1997).
3. H.Y. Lee, H.W. Lee, and D. Kim, *Phys. Rev. Lett.* **81**, 1130 (1998); *Phys. Rev. E* **59**, 5101 (1999).
4. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Jap. J. Ind. Appl. Math.* **11**, 203 (1994).
5. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995).
6. M. Bando, K. Hasebe, K. Nakanishi, A. Nakayama, A. Shibata, and Y. Sugiyama, *J. Phys. I France* **5**, 1389 (1995).
7. S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Japan* **67**, 2270 (1998).
8. S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Japan* **68**, 3110 (1999).
9. T. Uchida and T. Tadaki, *J. Phys. Soc. Japan*, **70**, 1842 (2001).
10. M. Koshi, M. Iwasaki, and I. Ohkura, *Some Findings and an Overview on Vehicular Flow Characteristics*, in *Proc. 8th Intl. Symp. on Transp. and Traffic Theory*, V.F. Hurdle et al., (Eds.), (1983).
11. Y. Sugiyama and H. Yamada, *Phys. Rev. E* **55**, 7749 (1997).
12. K. Nakanishi, K. Itoh, Y. Igarashi, and M. Bando, *Phys. Rev. E* **55**, 6519 (1997).
13. K. Nagel, D.E. Wolf, P. Wagner, and P. Simon, *Phys. Rev. E* **58**, 1425 (1998).
14. M. Treiber and D. Helbing, *J. Phys. A: Math. Gen.* **32**, L17 (1999).

# Widely Extended Optimal Velocity Model of Traffic Flow and their Linear Stability

K. Hasebe<sup>1</sup>, A. Nakayama<sup>2</sup>, and Y. Sugiyama<sup>3</sup>

<sup>1</sup> Faculty of Business Administration, Aichi University, Miyoshi, Aichi 470-0296, Japan

<sup>2</sup> Gifu Keizai University, Ohgaki, Gifu 503-8550, Japan

<sup>3</sup> Division of Mathematical Science, City College of Mie, Tsu, Mie 514-0112, Japan

**Abstract.** The optimal velocity (OV) model of traffic flow is extended such that the driver receives the information not only from the preceding vehicles but also from many other vehicles including vehicles that follow. We investigated in the linear approximation the damping behaviour of disturbance imposed on the uniform flow.

## 1 Introduction

We extend the original optimal velocity (OV) model [1]

$$\frac{d^2 x_n}{dt^2} = a \left[ V(\Delta x_n) - \frac{dx_n}{dt} \right], \quad (1)$$

to the following form,

$$\frac{d^2 x_n}{dt^2} = a \left[ V(\Delta x_{n+k_+}, \dots, \Delta x_{n+1}, \Delta x_n, \Delta x_{n-1}, \dots, \Delta x_{n-k_-}) - \frac{dx_n}{dt} \right], \quad (2)$$

where  $x_n$  ( $n = 1, 2, \dots$ ) is the position of the  $n$ th vehicle,  $\Delta x_n = x_{n+1} - x_n$  is the headway of  $n$ th vehicle,  $V(\Delta x)$  is the optimal velocity and  $a$  is a constant called sensitivity which has dimension of inverse of time. In (2),  $\Delta x_{n+k_+}, \dots, \Delta x_{n+1}$  are headways of the vehicles ahead of the  $n$ th vehicle and  $\Delta x_{n-1}, \Delta x_{n-2}, \dots, \Delta x_{n-k_-}$  are headways of the vehicles behind the  $n$ th vehicle.

We call forward looking (FL-OV) model [2,3] for  $k_+ > 0$ ,  $k_- = 0$ , and backward looking (BL-OV) model [4] for  $k_+ = 0$ ,  $k_- > 0$ .

All models are assumed to have the same uniform flow solution when they have the same headway, that is

$$V(b, \dots, b) = V(b), \quad (3)$$

for arbitrary  $b$ .

## 2 Linear Analysis

We study the linear stability of uniform solution  $x_n = bn + V(b, b, \dots, b)t$ . Suppose  $y_n$  is the small deviation of  $x_n$ . The linearized equation of  $y_n$  is

$$\ddot{y}_n = a \left[ \sum_{k=-k_-}^{k_+} f_k \Delta y_{n+k} - \dot{y}_n \right] \quad n = 1, 2, 3, \dots, \quad (4)$$

where  $\Delta y_{n+k} = y_{n+k+1} - y_{n+k}$  is the perturbation of the headway,

$$f_k = \frac{\partial}{\partial \Delta x_{n+k}} V(\Delta x_{n+k_+}, \dots, \Delta x_n, \dots, \Delta x_{n-k_-}) |_{\Delta x=b} \quad (5)$$

is the gradient of OV function. The periodic boundary condition is assumed. The solution of (4) is expanded by the modes  $\exp i(n\theta - \omega t)$  with  $\theta = 2\pi j/N$ ,  $j = 0, 1, 2, 3, \dots, N-1$ , where  $N$  is total number of vehicles. The differential equation (4) is reduced to the algebraic equation of  $\omega$

$$-\omega^2 = a \left[ \sum_{k=-k_-}^{k_+} f_k \{ \exp i(k+1)\theta - \exp ik\theta \} + i\omega \right]. \quad (6)$$

The sign of  $\text{Im } \omega$  determines the stability of flow. If  $\text{Im } \omega > 0$  ( $\text{Im } \omega < 0$ ) flow is unstable (stable) and if  $\text{Im } \omega = 0$  flow is marginal. (6) and  $\text{Im } \omega = 0$  gives the relation between  $a$  and  $\theta$ .

$$a(\theta) = \frac{\left[ \sum_{k=-k_-}^{k_+} f_k \{ \sin k\theta - \sin(k+1)\theta \} \right]^2}{\sum_{k=-k_-}^{k_+} f_k \{ \cos k\theta - \cos(k+1)\theta \}}. \quad (7)$$

If  $a(\theta) > a$  ( $a(\theta) < a$ ), the corresponding modes are unstable (stable) and if  $a(\theta) = a$  the corresponding modes are marginal.  $a_c = \{ \max a(\theta) \}$  is called critical sensitivity.

(3) and (5) give the subsidiary condition

$$\sum_{k=-k_-}^{k_+} f_k = V'(b) . \quad (8)$$

It is easy to see the two sets of parameters  $\{a, f_k\}$  and  $\{\lambda a, \lambda f_k\}$  ( $\lambda > 0$ ) are equivalent each other. So instead of (8) we can set

$$\sum f_k = 1. \quad (9)$$

### 3 Set of the Most Stable Parameters

In FL-OV models of  $k_+ = \kappa$ ,  $k_- = 0$  and  $\sum f_k = 1$ , the set of parameters

$$f_\kappa = f_{\kappa-1} = f_{\kappa-2} = \dots = f_0 = \frac{1}{\kappa + 1} \quad (10)$$

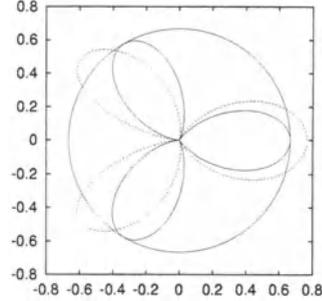
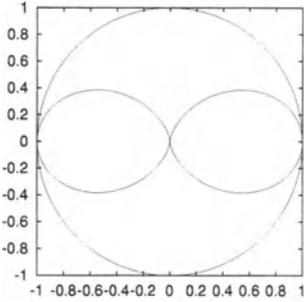
gives the smallest critical sensitivity

$$a_c = 2/(\kappa + 1). \quad (11)$$

In the other words parameters (10) are the solution of minimax problem of  $a(\theta)$ . We call (10) the most stable parameter set. For the parameters (10),  $a(\theta)$  is written as

$$a(\theta) = \frac{1 + \cos(\kappa + 1)\theta}{\kappa + 1} . \quad (12)$$

Figures 1 and 2 show diagrams of (12) for  $\kappa = 1$  ( $a_c = 1$ ) and  $\kappa = 2$  ( $a_c = 2/3$ ). From these figures we can be convinced that (10) minimizes the critical sensitivity.



**Fig. 1.**  $\kappa = 1, f_1 = f_0 = 1/2$ : The circle shows the critical sensitivity  $a_c = 1$ . The instability occurs at both  $\theta = 0$  and  $\theta = \pi$ . The instability at  $\theta = 0$  (long wave mode) is known in the original OV model. That at  $\theta = \pi$  (short wave mode) is new, which is reported first by Sawada [3].

**Fig. 2.**  $\kappa = 2, f_2 = f_1 = f_0 = 1/3$ : The circle shows critical sensitivity  $a_c = 2/3$ . Dashed line shows  $a(\theta)$  with parameters  $f_2 = 0.15, f_1 = 0.50, f_0 = 0.35$  which give a local minimum value of  $a_c = 0.77$ .

In the BL-OV model with  $k_+ = 0, k_- = 1$ , we find

$$a_c = \frac{1}{f_0 - 1/2}, \tag{13}$$

where  $\sum f_k = f_0 + f_{-1} = 1$ .  $a_c$  can have any small value for a large positive value of  $f_0$ . We can not define any parameter set which gives the smallest critical sensitivity.

### 4 Damping Behaviour of Disturbance

We assume disturbance in the uniform flow at  $t = 0$

$$\langle \delta y_n \delta y_m \rangle = \epsilon_1 \delta_{nm}, \quad \langle \delta v_n \delta v_m \rangle = \epsilon_2 \delta_{nm}, \quad \langle \delta y_n \delta v_m \rangle = \epsilon_3 \delta_{nm}, \tag{14}$$

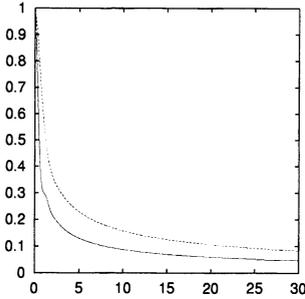
where  $\delta y_n$  and  $\delta v_n$  are disturbances of position and velocity, respectively.  $\langle \dots \rangle$  represents statistical average. For simplicity we discuss the case of  $\epsilon_2 = \epsilon_3 = 0$  hereafter. We examine two test functions

$$D(t) = \sum_n \langle y_n(t)^2 \rangle, \tag{15}$$

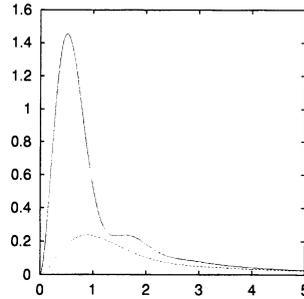
and

$$E(t) = \sum_n \langle v_n(t)^2 \rangle = \sum_n \langle \dot{y}_n(t)^2 \rangle, \tag{16}$$

in the linear approximation. Figs. 3 and 4 show  $D(t)$  and  $E(t)$  for two models: the BL-OV model of  $f_{-1} = -0.5$ ,  $f_0 = 1.5$  and the FL-OV model of  $f_0 = 0.5$ ,  $f_1 = 0.5$ , where we set  $a = 2$ .



**Fig. 3.**  $D(t)$ : Solid and dashed lines correspond to the BL-OV and the FL-OV model respectively.



**Fig. 4.**  $E(t)$ : Solid and dashed lines correspond to the BL-OV and the FL-OV model respectively.

From Fig. 3 and Fig. 4 we find

1. The amplitude  $D(t)$  of the BL-OV model is smaller than that of the FL-OV model.
2. The maximum value of  $E(t)$  of the BL-OV model is larger than that of the FL-OV model, but  $E(t)$  of the BL-OV model damps faster than that of the FL-OV model.

In more complicated models, the backward looking has the same effect as above. To see this, we investigate a hybrid of the FL-OV and the BL-OV models  $k_+ = 1$ ,  $k_- = 1$ . We set  $f_{-1} = -0.5$ ,  $f_0 = 1.0$ ,  $f_1 = 0.5$ . We compare this model to the FL-OV model  $k_+ = 2$ ,  $k_- = 0$  of the most stable parameter set  $f_0 = f_1 = f_2 = 1/3$ . For these two models we set sensitivity  $a = 4$ . Fig. 5 and Fig. 6 represent  $D(t)$  and  $E(t)$  of these models. In Fig. 6, we can see the damping behaviour of the hybrid model is improved. To see global features, we study the time integral of test functions.

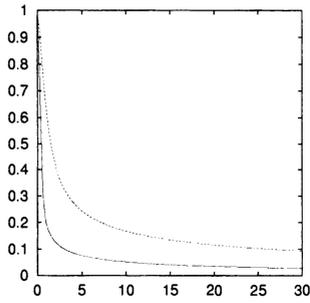
$$ID(t) = \int_0^t D(t')dt', \quad IE(t) = \int_0^t E(t')dt'.$$

Figures 7 and 8 show  $ID(t)$  and  $IE(t)$  for the hybrid model and the FL-OV model of the most stable parameter set.

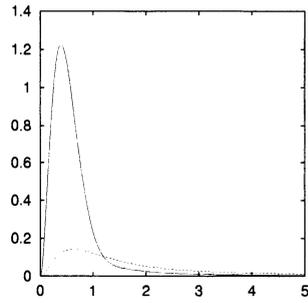
In Fig. 7 we see the damping of the disturbance of position in FL-OV model is very slow.

## 5 Summary and Discussion

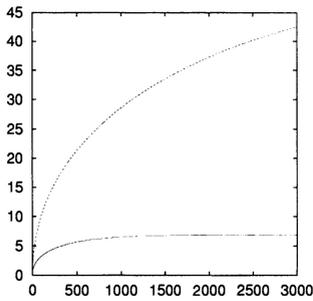
We proposed a wide extension of the OV model, which includes the FL-OV model where the driver obtains the information only from the preceding vehicles,



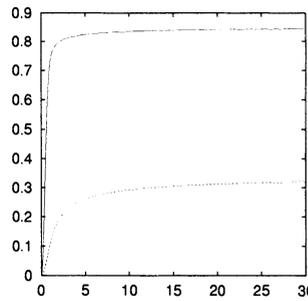
**Fig. 5.** Damping behaviour of  $D(t)$ . Solid and dashed line represent hybrid model and the FL-OV model of the most stable parameter set respectively.



**Fig. 6.** Damping behaviour of  $E(t)$ . Solid and dashed line represent hybrid model and the FL-OV model of the most stable parameter set respectively.



**Fig. 7.** Damping behaviour of  $ID(t)$ . Solid and dashed line represent hybrid model and the FL-OV model of the most stable parameter set respectively.



**Fig. 8.** Damping behaviour of  $IE(t)$ . Solid and dashed line represent hybrid model and the FL-OV model of the most stable parameter set respectively.

the BL-OV model where the driver obtains information from the vehicle right ahead and from the vehicles behind and the hybrid model. We investigated the damping behaviour of the small disturbance imposed on the flow in the linear approximation. It is found that the FL-OV model ( $k_+ = 1$ ) and the BL-OV model ( $k_- = 1$ ) show different damping behaviours. The disturbance of the velocity in the FL-OV model is small at the cost of the large disturbance and the slow damping of position. The disturbance of the position is small and the damping of position is fast in the BL-OV model at the cost of the large disturbance of velocity. We found that the damping behaviours in the hybrid model ( $k_+ = k_- = 1$ ) and the FL-OV model ( $k_+ = 2$ ) have the same property.

## Acknowledgements

This work is partly supported by a Grant-in-Aid for Scientific Research (C) (No.12650065) and (C) (No.13640409) of the Japanese Ministry of Education, Science, Sports and Culture, and by Center for Integrated Research in Science and Engineering in Nagoya University.

## References

1. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, Phys. Rev. E **51**, 1035 (1995); Japan J. Indust. Appl. Math. **11**, 203 (1994); M. Bando, K. Hasebe, K. Nakanishi, A. Nakayama, A. Shibata, and Y. Sugiyama, J. Phys. I France **5**, 1389 (1995); T.S. Komatsu and S.-i. Sasa, Phys. Rev. E **52**, 5574 (1995). Y. Sugiyama and H. Yamada, Phys. Rev. E **55**, 7749 (1997). M. Bando, K. Hasebe, K. Nakanishi, and A. Nakayama, Phys. Rev. E **58**, 5429 (1998); Japan J. Indust. Appl. Math. **17**, 275 (2000).
2. T. Nagatani, Phys. Rev. **E60**, 6395 (1999).
3. S. Sawada, Int. J. Mod. Phys. C **13**, No. 1 (January 2002).
4. K. Hasebe, in: *Proceedings of the 3rd Symposium on Simulation of Traffic Flow*, M. Fukui and Y. Ishibashi, (Eds.) pp.6, (The mathematical society of traffic flow, Japan, 1997)(in Japanese); K. Hasebe, A. Nakayama, and Y. Sugiyama, to appear in Phys. Rev. E.

# Stability of Multi-Lane Traffic Flow

A. Sasoh

Shock Wave Research Center, Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

**Abstract.** Linear and non-linear stabilities of single-lane traffic flow have been fairly well studied; stable, metastable and unstable states can be well defined. However, stability of multi-lane flow is much less understood. In this paper, we study the effects of the relaxation time in car acceleration/deceleration equation, and of the number of lanes on the stability to unsteady disturbance. Multi-lane flow is less sensitive to the strength of the disturbance; once a weak disturbance causes a triggering lane change, that induces successive disturbances and lane changes, eventually resulting in randomly fluctuated flow. The road capacity is influenced mainly by the value of the relaxation time.

## 1 Introduction

The stability of traffic flow is important from the viewpoint of understanding fundamental physical characteristics of the system. The linear stability of single-lane traffic system can be theoretically analyzed and the corresponding criterion can be obtained in a closed form. Bando et al. [1] formulate the linear stability criterion which is applicable to an optimal velocity (OV) model of any form. The state of single-lane traffic is subdivided into 'stable,' 'metastable' and 'unstable' ones. The word 'metastable' implies that the system is stable for weak disturbance, but unstable when its amplitude exceeds a critical value [2,5]. In the metastable states, a peculiar wave propagation behavior can appear; a small disturbance at first propagates downstream almost as an acoustic wave, then goes back to downstream with an increased amplitude as a compressive shock wave ('boomerang effect,' [3,4]. Sasoh [14] studies the effects both of the relaxation time and of the strength of the disturbance on single-lane traffic flow. The critical strength in the metastable state is determined as the function of the relaxation time.

Intensive works on multi-lane traffic [10,12,13,16] have resulted in better understanding its basic characteristics including the existence of the so-called 'synchronized' state [6,7,11]. However, the stability of multi-lane flow, in particular to unsteady disturbance, warrants further investigations to generalize the phenomena. Here, we try to input new insights on the stability of multi-lane traffic flow.

## 2 Multi-Lane Car-Following Model

Here, we conduct a car-following simulation in order to understand the impacts of microscopic processes on macroscopic phenomena. The detail of the methodology of the present study is presented elsewhere [14,15]. We employ right-sided traffic. All quantities are dimensionless. The coordinate in the travel direction and that in the transverse one are designated by  $x$  and  $y$ , respectively. The length and width of car is neglected. The lane width is set to unity. Each car is labeled by a set of sequential numbers  $(i, j)$  ( $i$ ; temporal sequential number in Lane  $j$  counted from the entrance). The velocity components on  $x$  and  $y$  coordinates are designated by  $u$  and  $v$ , respectively.

### 2.1 Equation of Motion

#### 2.1.1 Along Lanes

We use Bando et al.'s OV model [1] with a modification term being added to avoid unrealistic collisions [14].

$$\frac{du}{dt} = \frac{U(\Delta x) - u}{\tau} \quad (1)$$

$$U = \frac{f(\Delta x)}{f(\infty)} \quad (2)$$

$$f(\Delta x) = \tanh(\Delta x - \phi) + \tanh \phi + \alpha \left( \frac{1}{\Delta x^{n_1}} - \frac{1}{\Delta x^{n_2}} \right) \quad (3)$$

$$\Delta x(i, j) \equiv x(i+1, j) - x(i, j) \quad (4)$$

Here,  $t$ ,  $U$  and  $\Delta x$  designate time, effective optimal velocity and headway respectively. In (1) to (3), the car serial numbers  $(i, j)$  is omitted for simplicity. In (1),  $\tau$  acts as the characteristic relaxation time of the car acceleration/deceleration. In the right-hand side of (3), the sum of the first and the second term correspond to the original OV model of Bando et al; the third term is the correction for avoiding unrealistic collision.  $\phi, \alpha, n_1$  and  $n_2$  are set to 2, 0.1, 2 and 4, respectively.

### 2.1.2 Lane Change Rule

A car labeled  $(i, j)$  is motivated to change the lane as necessary reaction to the frontal car motion if the following condition is satisfied.

$$[A] \text{ and } \{[B] \text{ or } [C]\} \quad (5)$$

$$[A] \quad \Delta x(i, j) \leq \Delta x_c \quad (6)$$

$$[B] \quad u(i+1, j) - u(i, j) \leq \Delta u_c < 0 \quad (7)$$

$$[C] \quad u(i+1, j) \leq u_c \quad (8)$$

where  $\Delta x_c$  and  $\Delta u_c$  have respective constant values.

Another criterion for the motivation of lane change (spontaneous lane change) is expressed as

$$\min\{|x(i, j) - x(i', j')|_{\text{lane boundary}}\} \geq \max(\Delta x_s, \Delta x(i, j)) \quad (9)$$

where  $i', j'$  are the serial index number in the lane to which the car is changing.

The lane change can be completed only if the headway in the new lane is kept larger than a critical value,  $\Delta x_l$ , during the whole lane change period.

$$\min\{|x(i, j) - x(i', j')|_{\text{lane boundary}}\} > \Delta x_l, \quad 0 < \Delta x_l < \Delta x_s \quad (10)$$

If either (5) or (9) is satisfied, and if the inequality (10) is also satisfied, the car starts changing the lane. If changes to both lanes are possible, one to transfer is randomly selected with an even probability. If the inequality (10) is not satisfied before the lane change motion starts, the car adjusts its speed in order to wait for the condition to be satisfied:

$$\frac{du}{dt} = \min\{\text{Eq. (1)}, a_0\}, \quad a_0 = \text{const} < 0. \quad (11)$$

The absolute value of the lane change speed is set to constant ( $=v_0$ ). If the condition of the inequality (10) becomes violated after the lane change motion starts, the car goes back to the center of the current lane with the same absolute speed as of  $v_0$ .

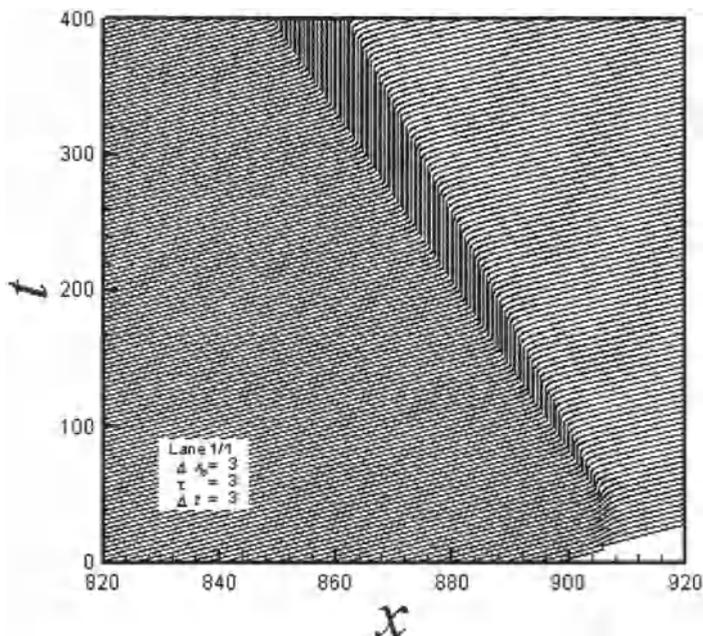


Fig. 1.  $x-t$  diagram of single-lane flow.

### 3. Alleviation of Non-Linear Disturbance

Figures 1 and 2 show  $x-t$  diagrams of single lane and three-lane flows, respectively, under the same condition of  $(\tau, \Delta t) = (3, 1)$ ; a disturbance is generated by the leading car's stop-and-go motion in Lane 1 during  $10 < t < 10 + \Delta t$ . This condition corresponds to a linear instability state in the single-lane flow. In the single-lane flow (Fig. 1), since the disturbance is rather weak, a traffic jam is formed after a delay. The disturbance begins to propagate as an acoustic wave. Then, the amplitude of the wave grows up, and eventually it makes transition to a compression shock wave. Therefore, the trajectory of the wave goes once forward and then back. This behavior is named as the 'boomerang' effect by Helbing [3]. Behind the compression shock, the car velocities vanish. After a while the cars re-start, forming an expansion shock wave. [14,15]. Between these two shocks, the cars experience traffic jam.

Figure 2 shows  $x-t$  diagrams of the three-lane traffic flow. Such a serious traffic jam formation as appearing in the single-lane flow is not observed in the presented time frame.

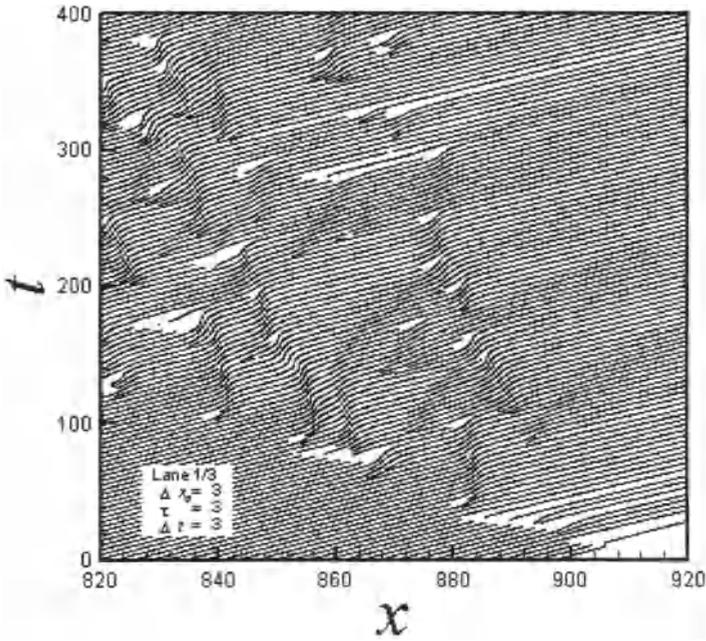


Fig. 2.  $x-t$  diagram of the first lane of three-lane flow.

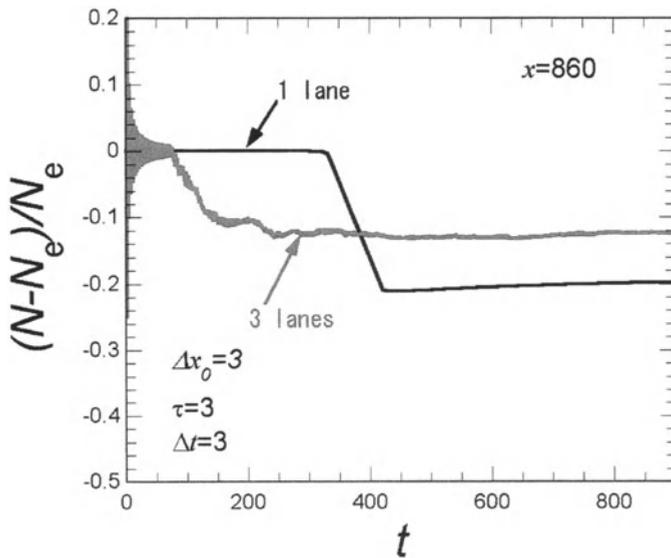


Fig. 3. Time variation of  $(N - N_e) / N_e$ .

In Fig. 3, variations of the decrease in the number of cars  $N$ , which passed the location of  $x=860$ , from the equilibrium value  $N_e$  are plotted. In the single lane case  $N$  starts to drop from  $N_e$  at the moment of the arrival of the compression shock wave. Past the expansion shock wave, the car flux becomes slightly larger than the equilibrium value.

In the three-lane flow  $N$  drops earlier. Yet, the decrease in  $N$  becomes smaller than that of the single-lane flow. Therefore, in this case the instability in the single lane flow is alleviated by adding the extra-lanes.

#### 4 Effect of Relaxation Time

Figure 4 shows the time variation of the number of cars which pass the location of  $x=860$ . Clear dependence of  $N$  on  $\tau$  is seen. For  $\tau=2$ ,  $N$  deviates from  $N_e$  only slightly. With increasing  $\tau$ , the decrease in  $N$  from  $N_e$  becomes increased. On the other hand, the effect of  $\Delta t$  is almost negligible. With  $\Delta t=3$ ,  $N$  starts to drop from  $N_e$  earlier because the flow experiences the stronger disturbance. However, once the flow is fluctuated, the effect of the strength of the triggering disturbance become negligible.

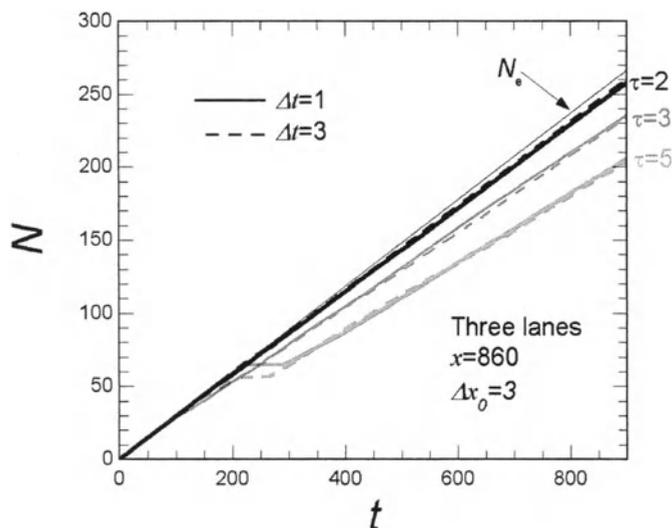


Fig. 4. Time variation of  $N$  for different  $\tau$  and  $\Delta t$ .

## 5 Conclusion

Multi-lane flow is more robust against unsteady-disturbance than the single-lane flow is. Once even a weak disturbance induces lane change, the strength of the triggering disturbance becomes less important. The relaxation time is influential to the effective flow capacity in both cases.

## References

1. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995).
2. M. Herrmann and B.S. Kerner, *Physica A* **255**, 163 (1998).
3. D. Helbing, *Rev. Mod. Phys.* to appear.
4. D. Helbing and M. Treiber, *Science* **282**, 2001 (1998).
5. B.S. Kerner, P. Konhäuser, and M. Schilke, *Phys. Rev. E* **51**, 6243 (1995).
6. B.S. Kerner and H. Rehborn, *Phys. Rev. E* **53**, R4275 (1996).
7. B.S. Kerner and H. Rehborn, *Phys. Rev. Lett.* **79**, 4030 (1997).
8. B.S. Kerner and P. Konhäuser, *Phys. Rev. E* **48**, R2335 (1993).
9. B.S. Kerner and P. Konhäuser, *Phys. Rev. E* **50**, 54 (1994).
10. T. Nagatani, *J. Phys. A: Math. Gen.* **29**, 6531 (1996).
11. B.S. Kerner, *Phys. Rev. Lett.* **81**, 3797 (1998).
12. T. Nagatani, *Physica A* **202**, 449 (1994).
13. M. Rickert, K. Nagel, M. Schreckenberg, and A. Latour, *Physica A* **231**, 534 (1996).
14. A. Sasoh, *J. Phys. Soc. Jpn.* **70**, 3161 (2001).
15. A. Sasoh and T. Ohara, *Proc. Roy. Soc.: Math. Phys. & Eng. Sci.* submitted.
16. V. Shvetsov and D. Helbing, *Phys. Rev. E* **59**, 6328 (1999).
17. M. Treiber, A. Hennecke, and D. Helbing, *Phys. Rev. E* **62**, 1805 (2000).

# **Empirical Traffic Data**

# 8-Figured Hysteresis Loop of OV Model

T. Nishi and D. Takahashi

Department of Mathematical Sciences, Waseda University, Tokyo 169-8555, Japan

**Abstract.** In this paper, we studied about an extended version of optimal velocity (OV) model, which was originally proposed by Bando *et al.* [1] We used a ‘2-step function’ as an OV function and obtained an 8-figured hysteresis loop in phase space. We also succeeded in obtaining the similar hysteresis loop using a smoother OV function.

## 1 Introduction

Optimal Velocity(OV) Model, which was first proposed by Bando *et al.* [1], has been studied by many researchers. OV model is based on the following general formula.

$$\ddot{x}_n = \alpha[V(\Delta x_n) - \dot{x}_n], \quad \Delta x_n = x_{n-1} - x_n, \quad (1)$$

where  $n$  denotes a car number,  $x_n$  a position of  $n$ -th car and  $V(\Delta x_n)$  an OV function depending on a headway  $\Delta x_n$ .

The original OV function is constructed with a hyperbolic tangent function which reflects characteristics of human driving. Sugiyama defined OV function by a step function and derived an exact solution [4].

In this paper, we analyze a traffic flow producing from an OV function using ‘2-step function’ made by superposition of two-step functions. Then we obtain an orbit of hysteresis loop in phase space. Since the figure of orbit is like a shape of character 8, we call this orbit ‘8-figured hysteresis loop’. Such type of solution is not usually observed for OV functions including only one point of inflection.

In Sect. 3, we show numerical results about this hysteresis loop and discuss its stability. In Sect. 4, we show a necessary condition to obtain an exact solution of this loop. In Sect. 5, we modify 2-step function, use a continuous function using hyperbolic tangent function, and examine whether 8-figured hysteresis loop is stable or not.

## 2 2-Step OV Function

We use an OV function named 2-step function defined by

$$V(\Delta x_n) = \begin{cases} 0 & (\Delta x < d_{\min}) \\ v_{\text{mid}} & (d_{\min} \leq \Delta x < d_{\max}) \\ v_{\text{max}} & (d_{\max} \leq \Delta x) \end{cases} \quad (2)$$

Its graph is shown in Fig. 1. In numerical experiments, we use a periodic boundary condition in space. When the car density is near 0.5, we often observe an orbit in phase space as shown in Fig. 2. Since the shape of the orbit is like a character 8, we call it 8-figured hysteresis loop. This loop means that there is two jam phase, one is strong ( $v = 0$ ) and the other is weaker ( $v = v_{mid}$ ). Figure

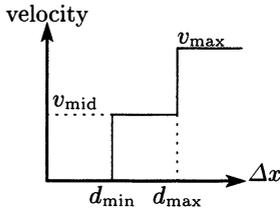


Fig. 1. 2-step function

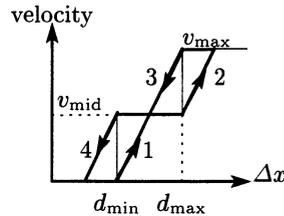


Fig. 2. 8-figured hysteresis loop

3 shows orbits of cars, which trace the hysteresis loop. We can observe three regions. One is the region where cars are not moving at all. We call this region ‘jam region’. Second is the region where cars are running in the maximum speed  $v_{max}$ . We call this region ‘free driving region’. And the last is the region where cars are running in the middle speed  $v_{mid}$ . We call this region ‘slow driving region’.

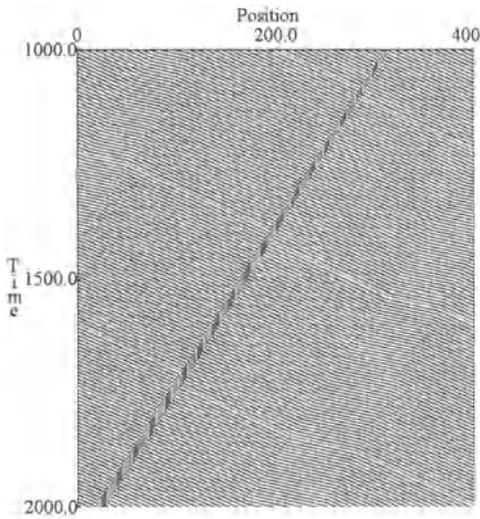


Fig. 3. Orbits of cars

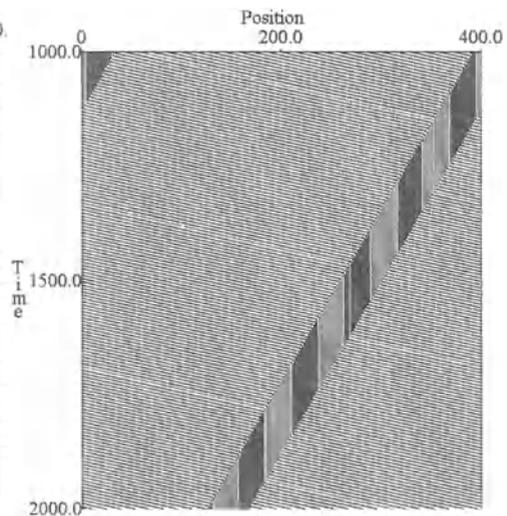


Fig. 4. 1-step function

These three regions are quite stable and coexist for a long time. Slow driving region exists between jam and free driving ones. Each car changes its own region as jam  $\xrightarrow{1}$  slow driving  $\xrightarrow{2}$  free driving  $\xrightarrow{3}$  slow driving  $\xrightarrow{4}$  jam. The number

above the arrow corresponds to Fig. 2. This configuration is stable for a long time though it is not permanent and breaks in the end.

The existence of 8-figured hysteresis loop cannot be observed numerically when we use a monotone increasing OV function with only one point of inflection. Moreover, we cannot observe for 1-step OV function described in [4]. Figure 4 shows orbits of cars for 1-step function to compare with Fig. 3.

### 3 Stability of 8-Figured Hysteresis Loop

Next we discuss the stability of the solution. We search numerically the value range of  $d_{max}, d_{min}$  where 8-figured hysteresis loop appears. The range is shown in Fig. 5.

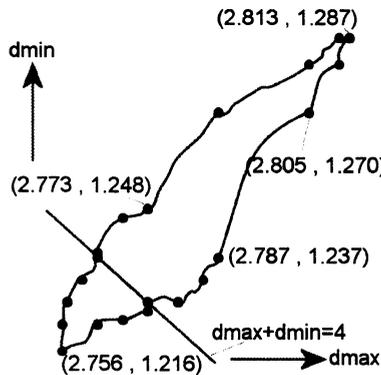


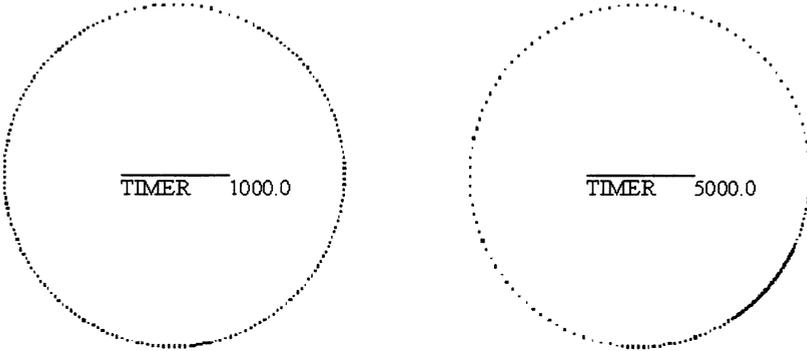
Fig. 5. The value range where 8-figured hysteresis loop appears

We also checked how long the hysteresis loop exists stably for each  $d_{max}, d_{min}$ . Its results are shown in Fig. 6. Since it takes about 400 in time for each car to go around the circuit, the hysteresis loop exists for more than 10 laps.

$d_{min} \backslash d_{max}$	1.222	1.225	1.227	1.232	1.237	1.24	1.245	1.246
2.76	5000	4900	4700	×	×	×	×	×
2.763	5500	5200	5000	4800	4200	×	×	×
2.768	×	5700	5400	5200	4800	4300	3900	3900
2.773	×	6600	6300	6000	5500	4800	4300	4200
2.787	×	×	×	×	7100	5000	4500	4400

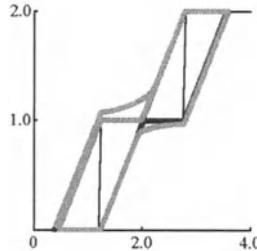
Fig. 6. Time range for hysteresis loop to exist

The reason why the loop breaks is supposed as follows. Figure 7 shows snapshots of car positions when the loop is stable and when begins to break. When hysteresis loop breaks the free driving region ( $v = v_{max}$ ) and the strong jam



**Fig. 7.** Snapshots of car position

region ( $v = 0$ ) become larger, and the weak jam region ( $v = v_{mid}$ ) shrinks. Then only jam region and free driving region are dominated, on the other hand slow driving region disappears, and the hysteresis loop orbit becomes another type of orbit as shown in Fig. 8.



**Fig. 8.** Break of hysteresis loop

#### 4 Necessary Condition for 8-Figured Hysteresis Loop

Next we show the necessary condition to obtain the stable 8-figured hysteresis loop. First we assume that a steady solution exists and it makes a general hysteresis loop shown in Fig. 9. Moreover, let us assume the whole configuration of cars moves at a constant speed  $v_{sys}$  in circuit. Under these assumptions, we can calculate the amount of inflow and outflow at the point corresponding to

$\Delta x_J$ ,  $a$ ,  $b$ ,  $\Delta x_F$ . These amounts must satisfy the following equation to form the steady solution.

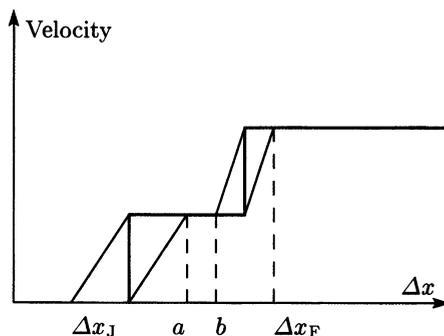


Fig. 9. Steady solution

$$\frac{V(\Delta x_J) - v_{Sys}}{\Delta x_J} = \frac{V(\Delta x_F) - v_{Sys}}{\Delta x_F} = \frac{V(a) - v_{Sys}}{a} = \frac{V(b) - v_{Sys}}{b} \quad (3)$$

Since we have  $V(\Delta x_J) = 0$ ,  $V(\Delta x_F) = v_{max}$ ,  $V(a) = V(b)$ , we obtain the following equations.

$$a = b = \frac{\Delta x_F (V(a) - V(\Delta x_J)) + \Delta x_J (V(\Delta x_F) - V(a))}{V(\Delta x_F) - V(\Delta x_J)} \quad (4)$$

$$v_{Sys} = \frac{\Delta x_F \cdot V(\Delta x_J) - \Delta x_J \cdot V(\Delta x_F)}{\Delta x_F - \Delta x_J} \quad (5)$$

To obtain a steady solution of 8-figured hysteresis loop,  $a$  and  $b$  must be the same. Therefore, the theoretical exact solution of the 8-figured hysteresis loop needs strict condition,  $a = b$ , and we consider the instability of numerical results is due to this condition.

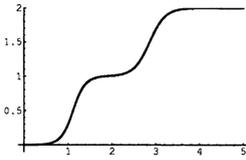
### 5 Continuous 2-Step Function

OV functions in previous sections are of piecewise-linear type. Next we use a continuous OV function and verify the existence of the hysteresis loop numerically. The continuous OV function is defined by hyperbolic tangent functions as follows;

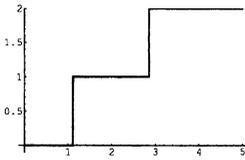
$$V(\Delta x_n) = \frac{1}{2} \left[ \tanh \frac{a \Delta x_n - 2}{\epsilon} + \tanh \frac{b \Delta x_n - 4}{\epsilon} + \tanh \frac{2}{\epsilon} \right] \quad (6)$$

If we take a limit  $\epsilon \rightarrow +0$ , we derive equation (2) from (6). Graphs are shown in Figs. 10 and 11 for different  $\epsilon$  values.

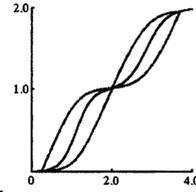
By taking appropriate value for  $a$  and  $b$ , we again obtain 8-figured hysteresis loop for the range of  $\epsilon \leq 0.7$ . An example of the loop is shown in Fig. 12. 8-figured hysteresis loop appears when density is 0.5. Orbits of cars are shown in Fig. 13. This 8-figured hysteresis loop lasts till 7000.0 in time.



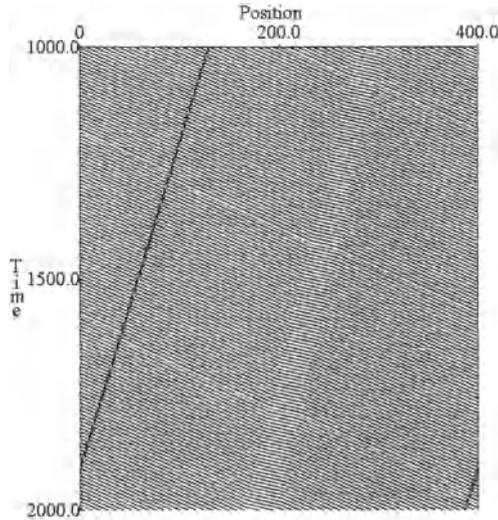
**Fig. 10.** Continuous function with  $\epsilon = 0.5$



**Fig. 11.** Continuous OV function with  $\epsilon = 0.001$



**Fig. 12.** 8-figured hysteresis loop for  $\epsilon = 0.7$



**Fig. 13.** Orbits of cars for continuous 2-step function

## 6 Conclusion

In this paper, we used an OV function with 2-step function and showed an orbit of 8-figured hysteresis loop is stable for a long time in phase space. The orbit may put a new light on the dynamics of traffic flow. We also showed a similar type of loop is stable using a continuous OV function. Though we have not shown above step-type OV function is valid for a real traffic, we can consider the detail of OV function effects a global car configuration.

## References

1. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035-1042 (1995).
2. Y. Sugiyama, *Comp. Phys. Comm.* **121-122**, 399-401 (1999).
3. Y. Sugiyama and H. Yamada, *Proceeding of International Workshop of Traffic and Granular Flow '97*, pp. 301-318 (Springer, Singapore, 1998).
4. Y. Sugiyama and H. Yamada, *Phys. Rev. E* **55**, 7749-7752 (1997).

# Observational Aspects of Japanese Highway Traffic

S. Yukawa<sup>1</sup>, M. Kikuchi<sup>2</sup>, A. Nakayama<sup>3</sup>, K. Nishinari<sup>4</sup>, Y. Sugiyama<sup>5</sup>, and S. Tadaki<sup>6</sup>

<sup>1</sup> Department of Applied Physics, University of Tokyo, Bunkyo 113-8656, Japan

<sup>2</sup> Cybermedia Center, Osaka University, Toyonaka 560-0043, Japan

<sup>3</sup> Gifu Keizai University, Ohgaki 503-8550, Japan

<sup>4</sup> Department of Applied Mathematics and Informatics, Ryukoku University, Ohtsu 520-2194, Japan

<sup>5</sup> Division of Mathematical Science, City College of Mie, Tsu 514-0112, Japan

<sup>6</sup> Computer and Network Center, Saga University, Saga 840-8502, Japan

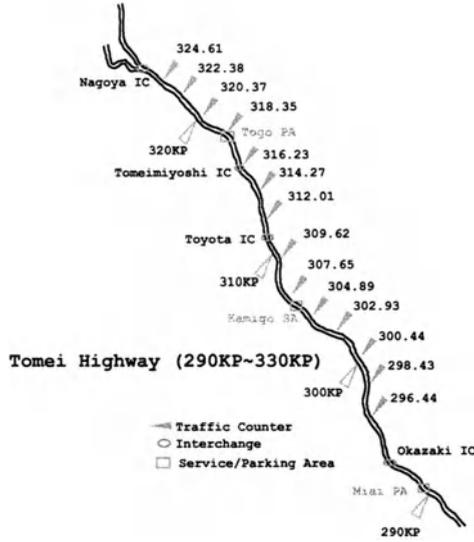
**Abstract.** From mass data of the Japanese highway traffic, we see the behavior of the traffic flow in Japan. Especially, we focus our attention on the multi-lane phenomena. For two-lane traffic, detailed time series of traffic variables are shown. Daily behavior of time series with a traffic jam event is discussed. In addition space-time plots of the jam is also considered. For three-lane traffic, we discuss a role of the middle lane. As a result, it changes according to the traffic volume.

## 1 Introduction

Following the previous article [1], we see observational aspects of Japanese highway traffic. In order to forecast and control behavior of vehicular traffic, development of good simulation models is inevitable. For a general purpose we require a physical microscopic dynamical model of the traffic flow. Its validity should be checked by the consistency with the real observation. In this sense, an observation of the real traffic is crucial for the theoretical consideration. In addition, from a physical point of view, the observation is also important in finding new interesting phenomena.

The authors have a lot of observational data of Japanese highway traffic. The data were observed at about one hundred points through one year on Japanese highways in 1999. They consist of the time series of traffic variables which are 5 minutes averaged flow and velocity per lane [2].

In this article we focus our attention on new features of multi-lane traffic. We will see the observational results of two areas of multi-lane highway. One is the Tomei highway with two lanes near Nagoya. Another one is the Meishin highway with three lanes near Kyoto. For the Tomei highway, we will see detailed time series of traffic variables and space-time plots of traffic jam. For the Meishin highway, we will discuss the role of the middle lane.



**Fig. 1.** Observational area: The Tomei highway from Okazaki IC to Nagoya IC. Gray triangles represent the observational points.

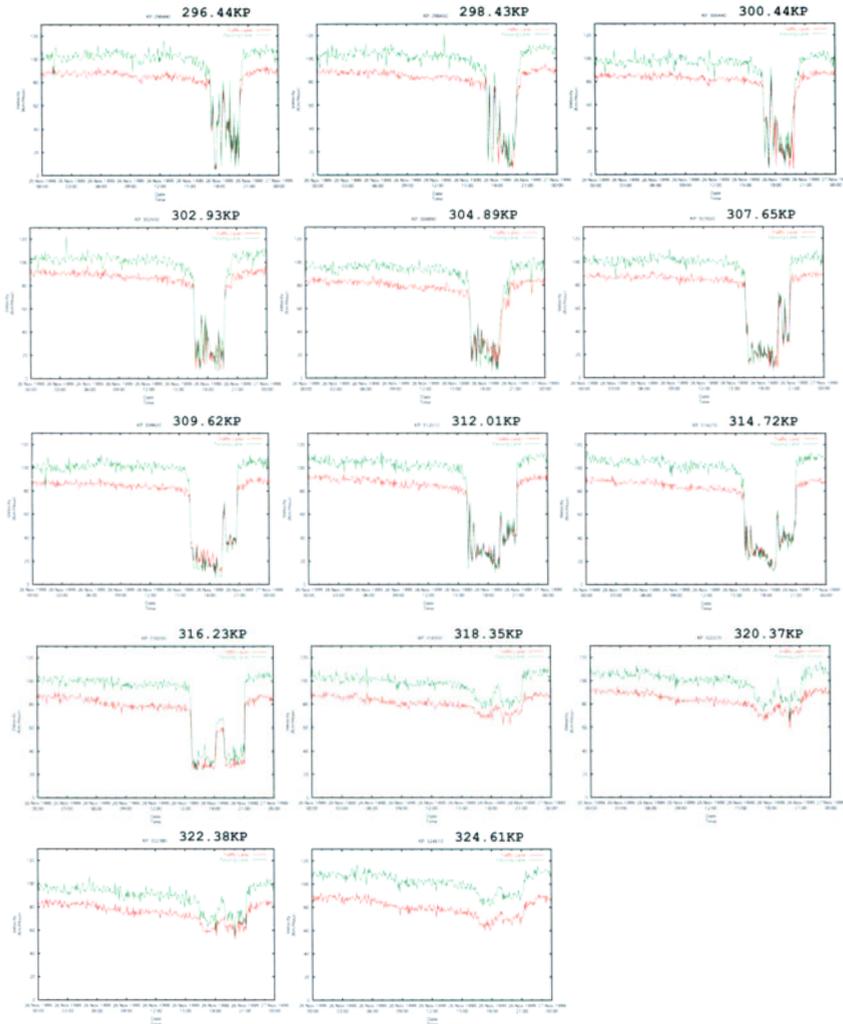
## 2 Two-Lane Traffic

First, we look the results of two-lane traffic. Especially, we see the traffic from Okazaki IC to Nagoya IC area, shown in Fig. 1. In the following results, observations were done for the vehicles which move from the lower right to the upper left in this map. Gray triangles represent the observational points, at where a loop coil detector is built in each lane. They are labeled by “kilo post” which is the distance measured in kilometers from the origin of this highway, that is, Tokyo. In this area some parking or service areas exist. Four interchanges are also jointed to the highway in this area.

It is known that a bottleneck exists near the 316 kilo post. The origin of the bottleneck is composite, but one of the important factors is a sag of the highway. When a traffic jam appears in this area, its front is formed around 316 kilo post frequently and it grows behind there.

### 2.1 Typical Jam Event: Time Series of Velocity

Typical time series of velocity are shown in Fig. 2. Each figure corresponds to time series for one day. The origin of the horizontal axis is the midnight of 26th November 1999. The vertical axis represents the averaged velocity with the unit of km/h. If the traffic flows on the figure from the left to the right and downward are considered as a sequel, then the upper left figure shows the upper stream of the traffic and the lower right one is the lower stream of the traffic. A green line corresponds to the data of passing lane and red one shows the data of traffic

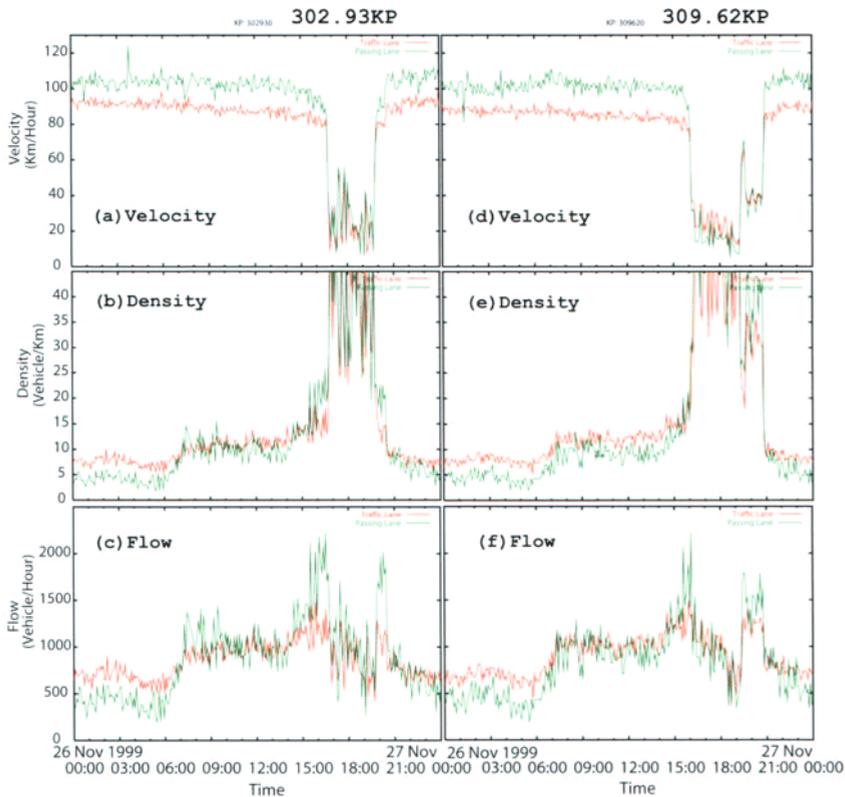


**Fig. 2.** Typical time series of velocity. Traffic flows from the left to the right and the up to the down in the sequel of figure. The data for 26th November 1999 are plotted. A red line corresponds to the data of the traffic lane and a green one corresponds to one of the passing lane.

lane. From, for instance, the figure of the 296.44 kilo post we can see the clear slowdown of velocities. This is a traffic jam. Similar slowdown is also observed in the figures from the 298.43 to the 316.23 kilo posts. But, in the figure of the 318.35 kilo post, such behavior is hard to recognize. This fact reflects that the bottleneck exists near the 316 kilo post and the jam is formed from there to the upper stream.

In addition, the duration of such regions becomes longer closer to the bottleneck. This phenomenon will be clearly seen in a space-time plot of the time series discussed later. The fluctuation of velocity observed in the slowdown period also becomes smaller closer to the bottleneck. This dependence of fluctuation may be caused by the stability of the traffic jam. Near the bottleneck the jam might be stable, but far from the bottleneck the jam is not so stable or even rather unstable. Because of this the fluctuation becomes larger.

## 2.2 Detailed Time Series: Daily Behavior of Two-Lane Traffic



**Fig. 3.** Detailed time series of (a) velocity at the 302.93 kilo post, (b) density at the 302.93 kilo post, (c) flow at the 302.93 kilo post, (d) velocity at the 309.62 kilo post, (e) density at the 309.62 kilo post, and (f) flow at the 309.62 kilo post.

Next let us see the detailed behavior of traffic variables. Typical time series of velocity, density, and flow are shown in Fig. 3. Density was not observed directly. Thus we calculate the density  $k$  from the velocity  $v$  and the flow  $q$  as  $k = q/v$ .

These plots are taken from typical two kilo posts where jam regions appear. The units of velocity, density, and flow are km/h, vehicles/km, and vehicles/h, respectively.

First, we look the time series of velocity (Figs. 3(a) and (d)). As shown in the preceding section, the jam period is clearly observed on both figures. In the case of (d), the velocity of the traffic lane (red line) is higher than the one of the passing lane (green line) in the jam period in contrast to outside of the jam. This reversing of the velocity is sometimes observed in our data. Such phenomenon is dependent on the traffic regulations. In Japan, a vehicle moving in the traffic lane can overtake one in the passing lane legally, staying in its lane. Thus the reversing of the velocity is possible. Outside of a jam the velocity is steady. The velocity in the passing lane is almost larger than the one of the traffic lane. This is natural because large vehicles like trucks must move in the traffic lane with a slower legal velocity than the other vehicles in Japan.

From these figures, we can clearly recognize two stages in the daily behavior of two-lane traffic in which a jam is observed. The first one is a free flow stage which is characterized by a high averaged velocity and its difference between two lanes. The second is a jam period. It can be distinguished by a low velocity and less difference between two lanes. We call this state of the traffic as in *the jam stage*.

An additional structure of the free flow period can be found in the time series of density shown in Figs. 3(b) and (e). In these figures the free flow period can be separated into additional two periods: the first one is occurring from the midnight to the morning of the corresponding day. It can be distinguished by the density difference between the two lanes and by the lower densities than in the other regions. We call the state of the traffic as in *the free flow stage*. Another period is characterized by the almost same densities in the two lanes. In addition, the values of the density is slightly larger than the ones of the free flow period. We call the state of the traffic as in *the middle flow state*. In other words, in the early morning, the role of two lanes is obviously specialized. On the other hand, in the middle flow stage, differentiation of the role of lanes is not clear. The two lanes are qualitatively equivalent.

In the period corresponding to the jam stage of velocity, we find an precursor of jam, especially in Fig. 3(b). Precisely speaking, this precursor is appearing in the free flow region of the time series of velocity. A similar structure is also found just behind the jam region. We call these states of the traffic as in *the pre or post jam state*. These structures are not clearly seen in another time series of the density shown in Fig. 3(e). From other observations, these structures are often found near the not fully developed jam structure. They are seen in the jam structure with large velocity fluctuations.

Corresponding time series of flow is shown in Figs. 3(f) and (g). In the free flow period, similar behavior of the density time series is observed. A remarkable feature is appearing in the precursor of the jam and the structure just behind of the jam period. In these structures the flow becomes larger than in the other stages.

**Table 1.** Characteristics of the daily behavior of two lanes traffic: “diff.” means that a difference between the lanes is clearly recognized.

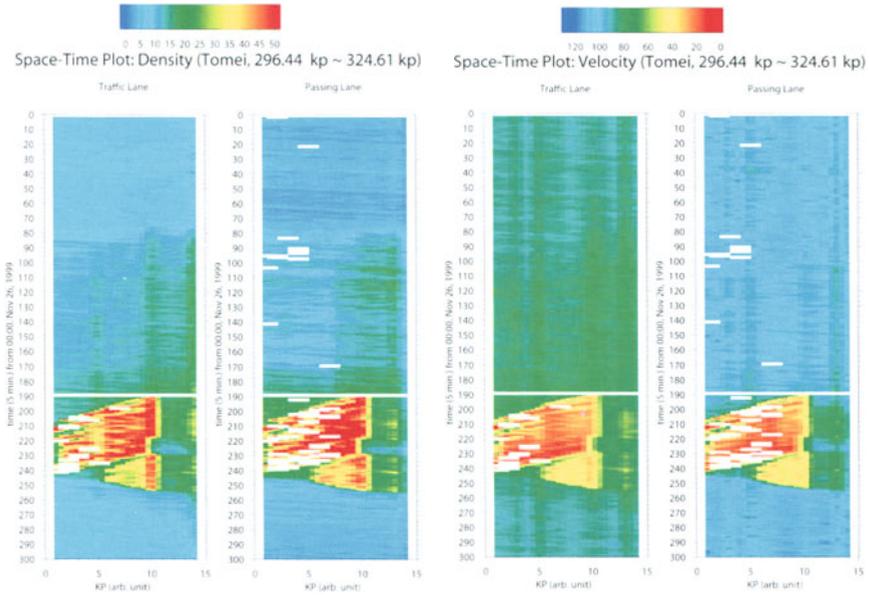
	free flow	middle flow	pre/post jam	jam
velocity	high, diff.	-	-	low, less diff.
density	diff.	less diff.	larger	largest
flow	diff.	less diff.	largest	low

To shortly summarize, daily behavior of two lane traffic can be separated into five stages. They are the free flow, the middle flow, the pre jam, the jam, the post jam stages. These features are characterized as shown in Table 1. For the free flow stage, traffic can distinguished by the high velocity and difference between traffic and passing lanes in velocity, density, and flow. On the other hand, in the middle flow stage, the difference becomes smaller in the density and the flow time series. But, in this stage, there are no notable features in the time series of velocity. The pre and post jam stages are quite similar. In these stages, the flow becomes larger than in the other stages. In the jam stage, the flow and velocity become relatively smaller.

Here we see that the daily behavior can be classified into several stages. It is not clear whether these stages are corresponding to physical phases, since the present situation is dynamically generated and, in addition, it is strongly dependent on the boundary condition of the injection and ejection of the vehicles. If they are phases, we can reproduce the present behavior by a simulation of two lane traffic in circular roads under a fixed density condition [3]. But such behavior has not been reproduced up to date. This is a future challenge of two-lane traffic simulations.

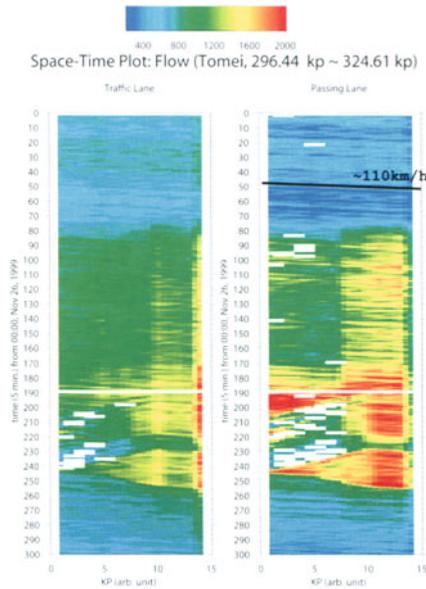
### 2.3 Space-Time Plots

The present data were observed at every two kilo meters. Thus we can get information of space-time behavior of the traffic flow. The resolution of space-time data is about 2 kilometers by 5 minutes. With some interpolation techniques we get the space-time plots of typical jam event shown in Fig. 4. This jam event is same as the one of Fig. 2. The horizontal axis represents the kilo post with a unit of about 2 km. The ticks 1 and 14 correspond to the 296.44 and 324.61 kilo posts, respectively. The vertical axis expresses the time measured from the midnight of 26th November 1999 to 1 a.m. 27th November 1999. The unit of the vertical axis is 5 minutes. A rectangular blank represents lack of data because of errors of the traffic detector. In each figure vehicles move from the upper left to the lower right. We can see bluish stripes with almost same gradients in these figures in the early morning, especially in Fig. 4(c). These stripes express the trajectory of freely moving vehicular bunches. The line shown in Fig. 4(c) corresponds to a trajectory with velocity of about 110 km/h.



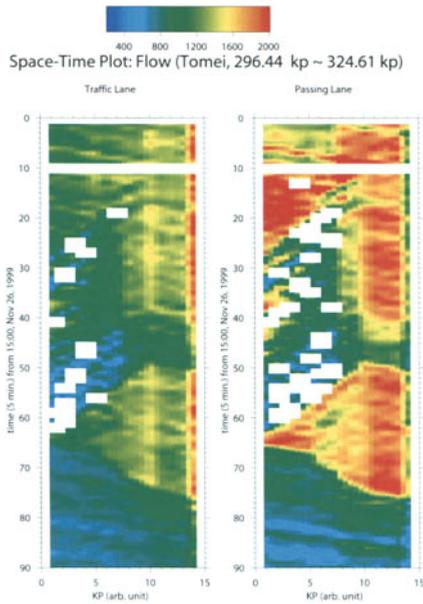
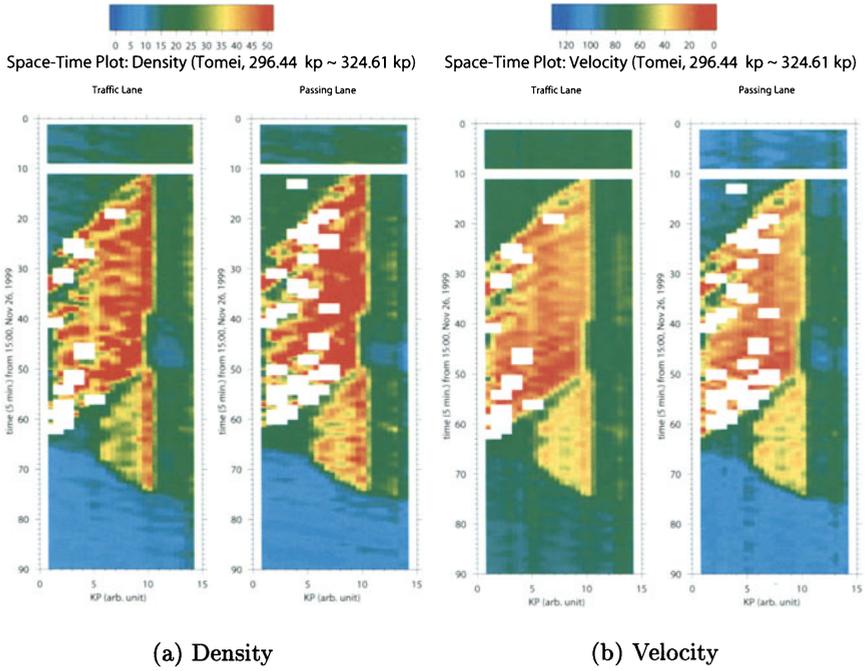
(a) Density

(b) Velocity



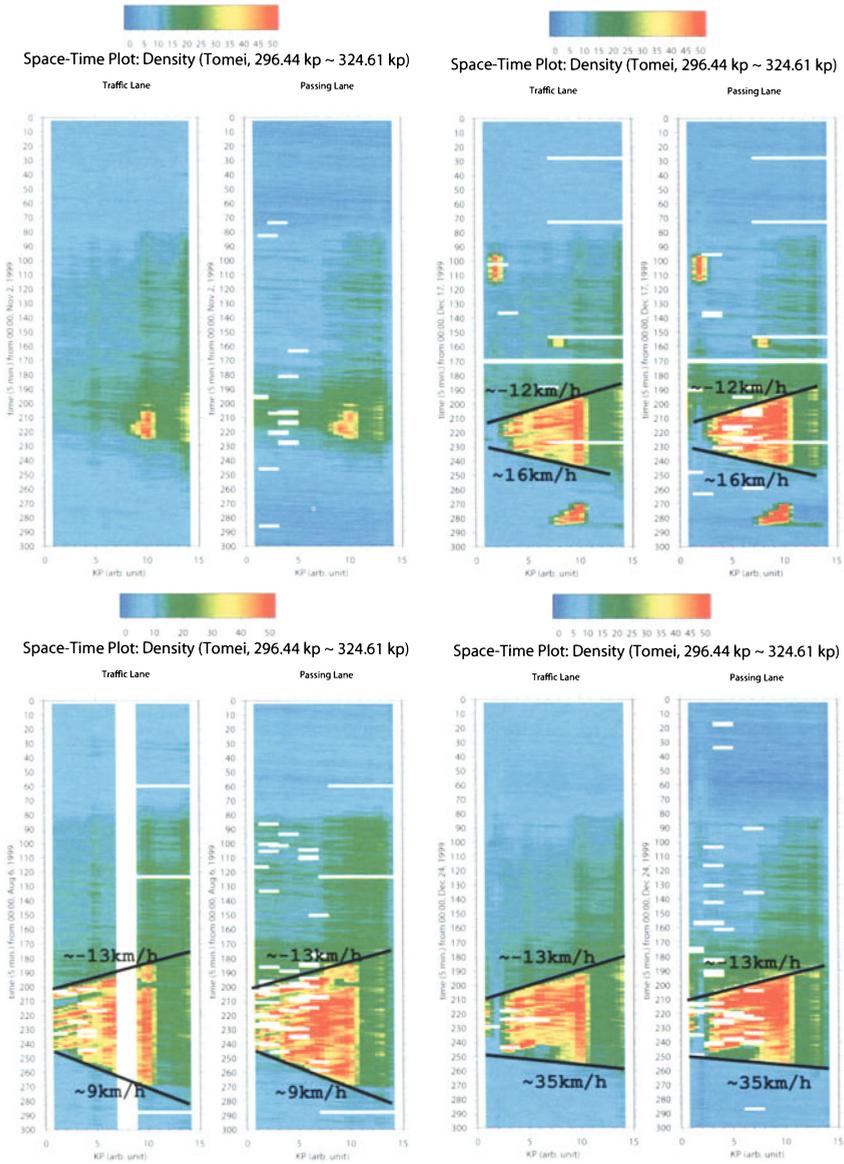
(c) Flow

**Fig. 4.** Space-time plots of density, velocity, and flow on the 26th November 1999 from the 296.44 kilo post to the 324.61 kilo post. Original resolution is about 2 kilo meters by 5 minutes. This figure is plotted with an interpolation technique. A blank represents the error of detector. The units of color scales of density, velocity, and flow are vehicles/km, km/h, and vehicles/h, respectively.



(c) Flow

Fig. 5. Magnification near the jam of Fig. 4.



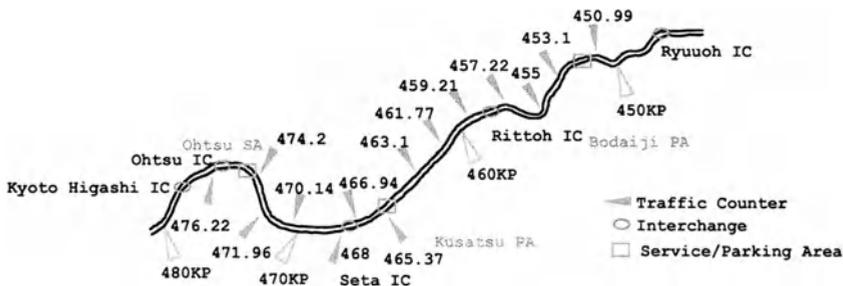
**Fig. 6.** Space time plots of the Tomei Highway: Other jam events observed on 6th August 1999, 2nd November 1999, 17th December 1999, and 24th December 1999.

Look at the afternoon of that day. We can see the red regions in Figs. 4(a) and (b). They are representing the high density and the low velocity, respectively. Therefore, we can regard these regions as traffic jams. We can recognize that the bottleneck exists near 10 or 11 from these figures. These ticks are corresponding to 316.23 and 318.35 kilo posts. Magnifications of the jam structure is shown in Fig. 5. The origin of the time is now taken to be 15 p.m. 26th November 1999. Other conditions of plotting are same as the ones of Fig. 4.

In Fig. 5 we find clear internal structures of the jam pattern. Especially, both in the density and the velocity portraits, it is seen that a congested region of reddish color and an uncongested one of greenish color appear periodically. In addition, such regions propagate the traffic flow with almost the same velocities. This velocity is the same as the one of the growing tail of jam. We can calculate it as  $-12\text{km/h}$  from the figures. In the flow portrait (Fig. 5(c)), the pre and post jam stages are clearly seen as reddish regions. In this picture the traffic jam is not clearly observed. It is only recognized as a low flow region.

We also calculate the velocity of the vanishing tails of the jam as  $16\text{km/h}$  from the figure. This velocity is dependent on the injection condition of the traffic flow. Thus, we expect that the velocities of vanishing tails are different event by event. Actually, we can observe different velocities as shown in Fig. 6. On the other hand, the velocities of the growing tails are almost the same, because they are dependent on the properties of the bottleneck, not the injection conditions of the traffic.

### 3 Three-Lane Traffic



**Fig. 7.** Observational area: Meishin highway from Ryuuoh IC to Kyoto Higashi IC. From Rittoh IC to Seta IC the highway has an additional one lane. Thus from the 459.21 kilo post to the 466.94 kilo post we can observe the behavior of three-lane traffic.

Next, we see results for three-lane traffic. The observational area is shown in Fig. 7. In this area the highway, the Meishin highway, has a three-lane part. From

Rittoh IC to Seta IC, an additional one lane exists. In this region, we can observe the behavior of the three lanes traffic. In the following results, observations were done of the traffic flowing from the right (450.99 kilo post) to the left (476.22 kilo post). It is already known that two bottlenecks exist near the 453.1 kilo post and the the downstream of the 476.22 kilo post.

First, we look at space-time plots of traffic variables in Fig. 8. The daily behavior of the traffic flow on the 20th July 1999 is shown. Conditions of plotting are the same as ones of Fig. 4. Congested regions are observed near the two bottlenecks. The region near the downstream of the 476.22 kilo post is a jam, because characteristic internal structures of jam are recognized. This jam even extends into the three lanes area. The region near the 453.1 kilo post is also jammed.

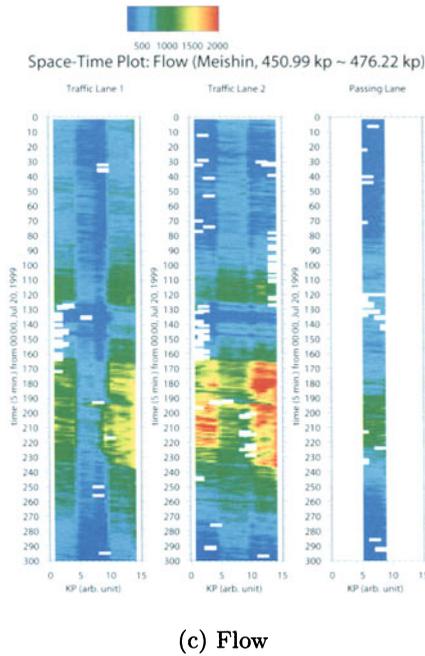
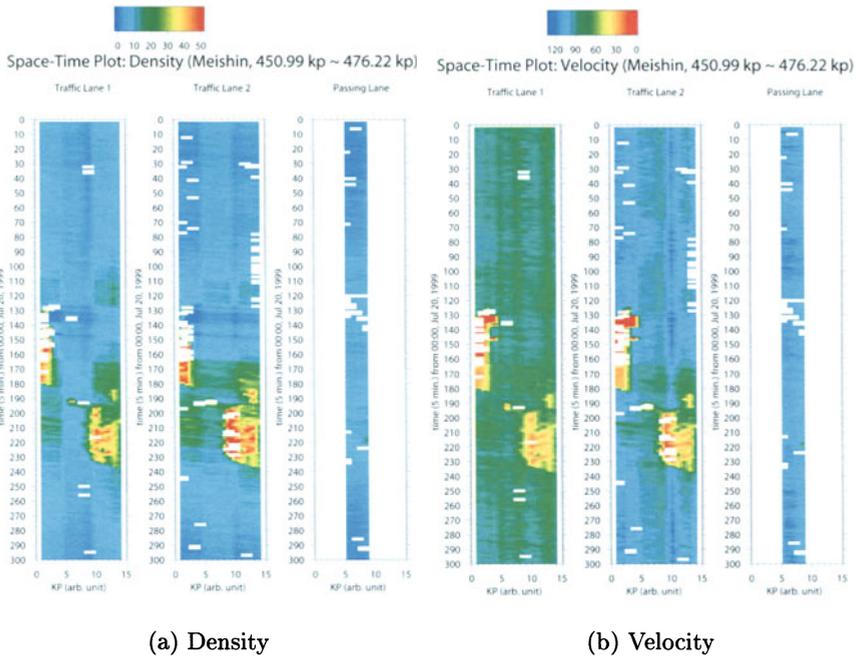
Next, we see detailed time series. Especially, we focus our attention on the three-lane behavior of the traffic flow. In Fig. 9, daily behavior of three-lane traffic is shown. The top figure corresponds to one of the densities, the middle one to one of the flows, and the bottom is corresponds to the velocity. This observation was done at the 461.77 kilo post on the 20th July 1999. The red, green, and blue lines correspond to the outer, middle, and inner lanes, respectively. In Japan the outer and middle lanes are traffic lanes. And the most inner lane is a passing lane.

Let us consider the role of the additional lane from these figures. At first, we can see the several spikes in the density and velocity time series. The spikes near 16:00 represent an actual traffic jam. Another spike near 11:30 is not a jam, because the flow is actually zero at the same time.

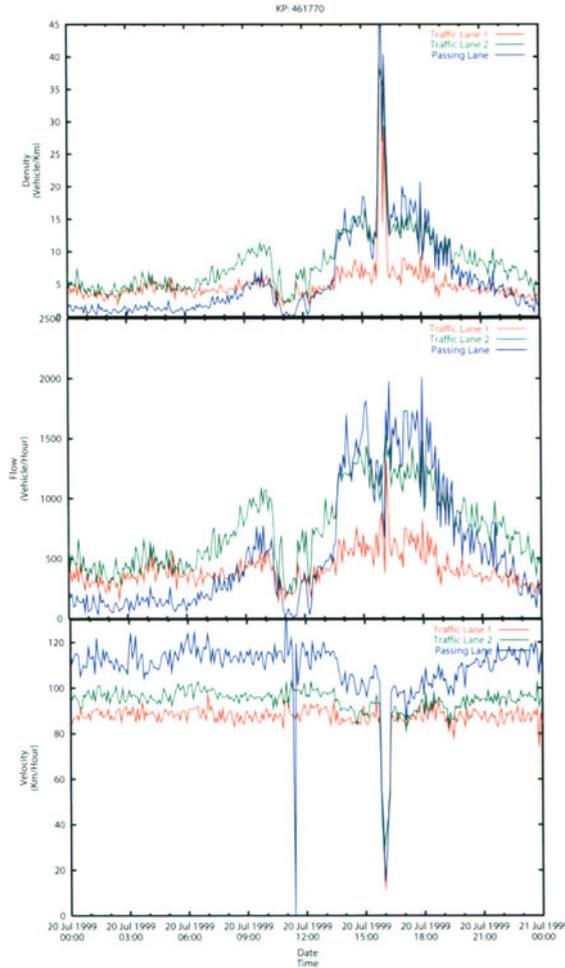
In the region from midnight to the early morning, which region corresponds to the free flow stage of the daily behavior of two-lane traffic, we can observe that the densities of the outer and middle lanes are almost the same. For the flow, its situation is also the same. In the velocity time series, on the other hand, such behavior cannot be recognized. These facts show that the middle lane and the outer lane play the same quantitative role in the stage of low traffic volume.

Even though the quantitative behavior of these traffic states is similar, the tendency of the traffic variables density and flow is not similar between the middle lane and the traffic lane. Rather the behavior of the passing lane is similar to the middle lane in the free flow stage. The density and the flow of the passing lane are increasing from 6:00 to 10:00 in the same manner as the ones of the middle lane.

In contrast, for the regions corresponding to the pre and post jam regions, the qualitative behavior of the variables are similar between the middle lane and the passing lane. In addition, even the quantitative behavior of these two lanes is also the same. In those regions, the outer lane is different from the other two lanes. It is remarkable that the velocity of the outer lane is not so down in these regions, in contrast to the middle and passing lane. At the same time, the density is not so high. It shows that the outer lane is less sensitive to the change of the traffic volume.



**Fig. 8.** Space-time plot of the Meishin Highway: (a) density, (b) velocity, and (c) flow behavior is shown. Observations were done on the 20th July 1999. The observational region is from the 450.99 kilo post to the the 476.22 kilo post. From the 459.21 kilo post (tick 5) to the 466.94 kilo post (tick 9), an additional one lane exists.



**Fig. 9.** Detailed time series of density, flow, and velocity at the 461.77 kilo post. The observation was done on the 20th July 1999. The red, green, and blue lines are corresponding to the outer, middle, and inner lanes.

We summarize the role of the middle lane as follows: In the free flow, the middle lane is the same as the traffic lane quantitatively. It is also the same as the passing lane qualitatively. In the pre and post jam stages, the middle lane is similar to the passing lane both in qualitative and quantitative aspects. From the traffic volume point of view, we can say the followings: For lower traffic volume, the middle lane is the same as the traffic lane quantitatively. But, for higher traffic volume, the role of the middle lane becomes the one of the passing lane. Thus, the role of the middle lane is dependent on the traffic volume.

## 4 Summary

In this article, we see the observational aspects of Japanese highway traffic. Especially, we looked two observational areas in detail: For a two-lane part of the Tomei highway, detailed time series of daily behavior of two-lane traffic has been discussed. As a result, we found five stages of traffic state in daily behavior. But it is unclear whether these stages are real phases of traffic flow. In addition, space-time plots of jam events and jams themselves are shown. From these figures, internal structures of traffic jam are clearly observed.

For three-lane traffic, we have discussed the role of the middle lane from detailed time series of the traffic variables. As a consequence, we find that the character of the middle lane is variable. At the lower traffic volume, the middle lane plays a similar role to the traffic lane quantitatively. On the other hand, qualitative behavior of the middle lane, such as hourly incremental trends, is quite resembling the one of the passing lane. For the higher traffic volume, the middle lane behaves like the passing lane both from quantitative and qualitative points of view.

Here we have discussed only observational results. There is no theoretical interpretation nor computational study. We should develop such investigations based on mass traffic data analysis. Simulations of multi-lane traffic is now in progress by the authors.

## Acknowledgment

The authors thank Japan Highway Public Corporation for providing mass data of traffic flow.

This work is partly supported by a Grant-in-Aid for Scientific Research (C) (No.12650065), (C) (No.12680446), and (C) (No.13640409) of the Japanese Ministry of Education, Science, Sports and Culture, and by Center for Integrated Research in Science and Engineering in Nagoya University.

## References

1. M. Kikuchi, A. Nakayama, K. Nishinari, Y. Sugiyama, S.-i. Tadaki, and S. Yukawa: "Long Term Traffic Data from Japanese Highway" (in this volume).
2. Observational details are found in the previous article [1].
3. The authors try modeling and simulations for two lane traffic. See, for instance, S.-i. Tadaki, M. Kikuchi, K. Nishinari, Y. Sugiyama, and S. Yukawa: "Congestion Induced by Bottlenecks in Two-lane Optimal Velocity Traffic Flow Model" (in this volume).

# Long-Term Traffic Data from Japanese Expressway

M. Kikuchi<sup>1</sup>, A. Nakayama<sup>2</sup>, K. Nishinari<sup>3</sup>, Y. Sugiyama<sup>4</sup>, S. Tadaki<sup>5</sup>, and S. Yukawa<sup>6</sup>

<sup>1</sup> Cybermedia Center, Osaka University, Toyonaka 560-0043, Japan

<sup>2</sup> Gifu Keizai University, Ohgaki 503-3550, Japan

<sup>3</sup> Department of Applied Mathematics and Informatics, Ryukoku University, Ohtsu 520-2194, Japan

<sup>4</sup> Division of Mathematical Science, City College of Mie, Tsu 514-0112, Japan

<sup>5</sup> Computer and Network Center, Saga University, Saga 340-3502, Japan

<sup>6</sup> Department of Applied Physics, University of Tokyo, Bunkyo 115-3656, Japan

## 1 Introduction

Recent development of physics-based mathematical modelling of highway traffic has brought us to a new stage of research. For example, reproduction of overall shape of the  $q$ - $k$  diagram, which itself has once been a big challenge, is now one of the least requirements that any *good* model should achieve. The interest for physicists has been shifted to deeper understanding of traffic behavior such as details of the instability near the capacity and spatio-temporal structures of traffic jams. Understanding them is, without doubt, important also for traffic controls in future.

Now that physics-based models have matured so that their practical use for real-time traffic simulations is within reach. What we really need is vast traffic data taken at several different situations on highways. Importance of the observed data at this stage is two fold: First, in order to make models more suitable for practical use by, for example, tuning some adjustable parameters or taking new effects into account, numerical simulations should be compared with real traffic flows in detail. Second but not least, in order to distinguish phenomena universal to traffic flow in general and ones particular to some situations or regulations, which differ from country to country and from place to place, we need to compare traffic data taken at several different places and under several different situations such as bottlenecks of different types. Especially, in the case of real multi-lane traffic, traffic behavior would be affected strongly by the details of the regulations and the types of bottlenecks.

So far, traffic data have been analyzed mainly from the viewpoints of traffic engineering. Traffic data analyzed and reported by physicists are still in a limited number from only a few countries, such as Germany and Korea [1]. We need much more. In this paper we introduce our data analysis project. In the following, we present some traffic data taken at several points on highways in Japan. The data are provided by the courtesy of the Japan Highway Public Corporation. A part of data has already been published as a booklet [2]. Here we focus our attention

to spatial variations of long-time properties such as the  $q$ - $k$  diagrams and power spectra of fluctuations in flow and velocity. A detailed analysis is still in progress so that the main purpose of this paper is just to show some examples of the data.

## 2 Setting

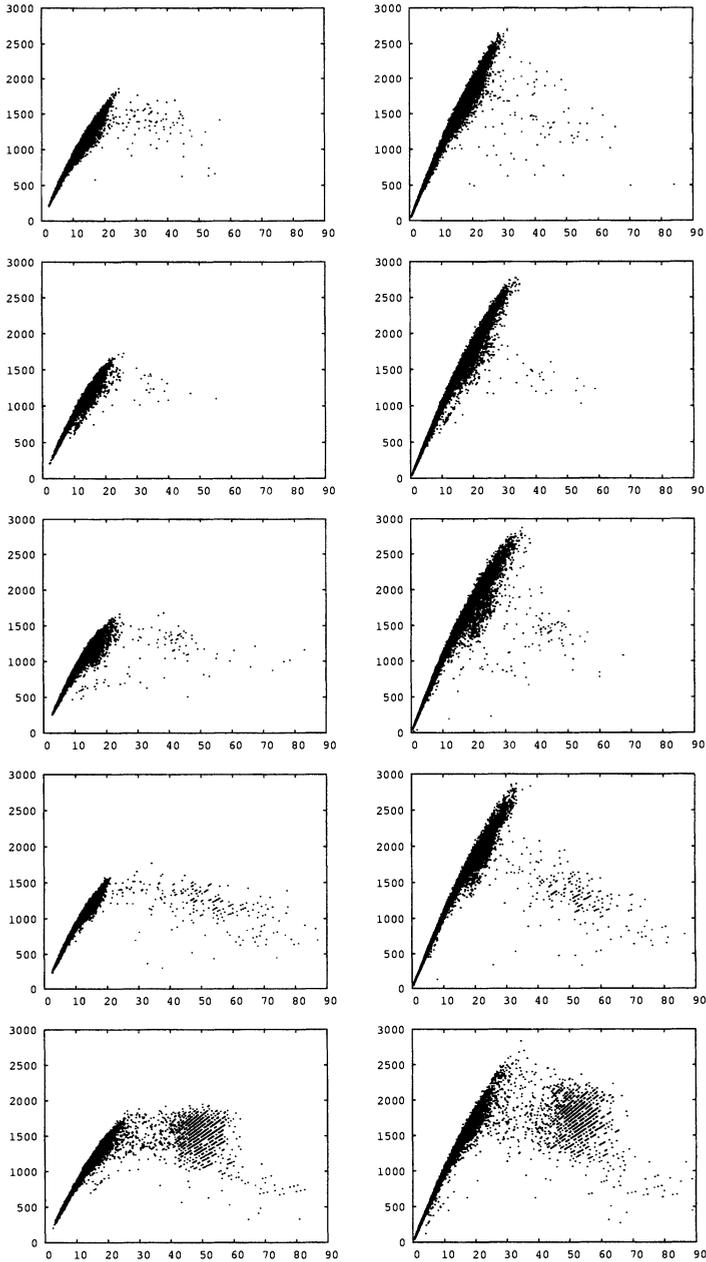
A Pair of loop coils, which are separated by 7 meters, are installed at every 2 kilometers of every highway in Japan by the Japan Highway Public Corporation. It records passing of vehicles over the loop coils and measures their velocity. Normally, the number of vehicles and their average velocity in every 5 minutes are recorded and archived. Thus, *raw* data we can get are time serieses of 5 minutes averages of flow and velocity at each location.

Currently, we have got data for one year taken in 1999 at about 100 different locations in mid-western Japan, some of which are consecutive along the highway so that we can analyze spatial variations of the traffic flow. Among what we show in the following is data only of one month, i.e. from Aug. 1999 taken at Tomei expressway bound for Tokyo near the Nagoya interchange. The data shown were taken at 5 consecutive locations at 316.23, 318.35, 320.37, 322.38, and 324.61 km measured from Tokyo. There are three potential traffic bottlenecks in that region: the Togo–Miyoshi interchange, the Togo parking area, and the Nagoya interchange located near the 316 km point, 318 km point, and 325 km point from Tokyo, respectively. In mid-August we have heavy traffic conjestions here and there in Japan, because many people move for spending the summer vacation. Occurence of heavy conjestion is actually seen in the traffic data in this region as we will see in the following.

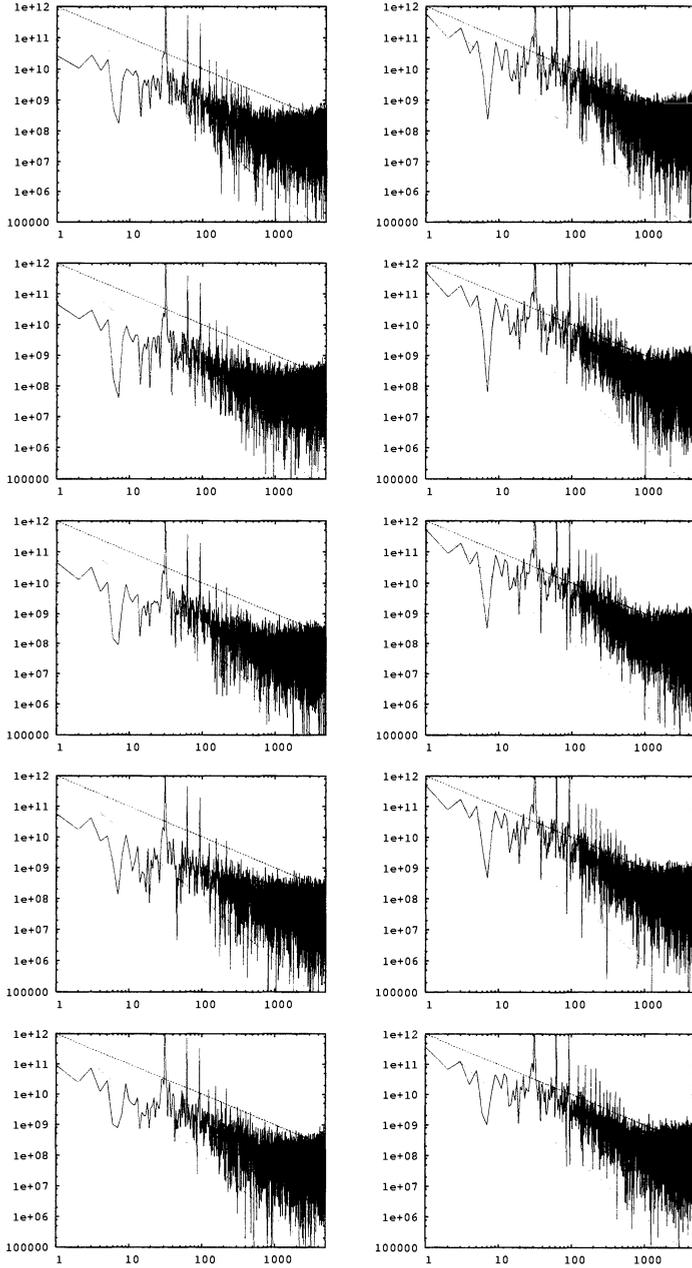
As for the traffic regulations, the legal maximum velocity on highways in Japan is normally 100km/h. In case of two-lane traffic, the outer lane is the driving lane and the inner lane is used for passing.

## 3 $q$ - $k$ Diagrams and Power Spectra

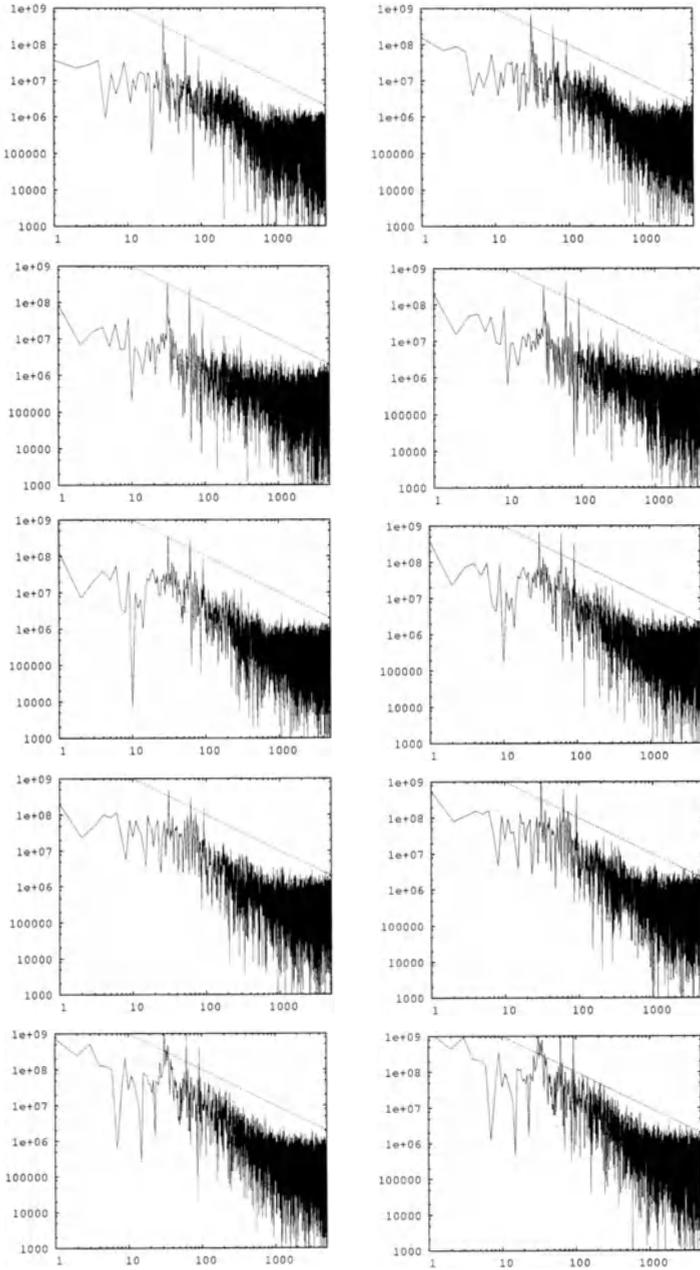
Figure 1 shows the  $q$ - $k$  diagrams both for the driving lane and for the passing lane at the 5 consecutive points mentioned above. It should be noted that the density  $k$  is not a quantity measured directly by the loop coils but it is calculated from the flow and the average velocity. Effects of heavy conjestions are clearly seen. Moreover, the spatial development of the conjestion can be read by comparing the  $q$ - $k$  diagrams for consecutive points. In the present case, the effect of the heavy conjestion is observed just downstream of the Nagoya interchange and gradually weakened along the highway bound for Tokyo. The difference between the driving lane and the passing lane is particularly noticeable; the instability near the maximum capacity is much clearer in the passing lane. The spatial variation of the  $q$ - $k$  diagrams like this contains rich information compared with a  $q$ - $k$  diagram taken at one location alone. It will be useful especially in improving mathematical models.



**Fig. 1.**  $q$ - $k$  diagrams for 5 consecutive points near Nagoya bound for Tokyo. From top to bottom the diagrams represent the data taken at 316.23, 318.35, 320.37, 322.38, 324.61 km points from Tokyo. The left column is for the driving lane and the right for the passing lane. The abscissa represents the density  $k$  in unit of [vehcles/km], and the ordinate the flow  $q$  in unit of [vehicle/hour].



**Fig. 2.** Power spectra of fluctuations in flow recorded in 5 minutes taken at the same points as Fig.1. Left column is for the driving lane and right for the passing lane. The abscissa represents frequency  $f$  in unit of  $1/5$ min. Slopes which indicate  $1/f$  and  $1/f^2$  are also shown as references.



**Fig. 3.** Power spectra of the fluctuations in the average velocity in 5 minutes taken at the same points as Fig.1. The left column shows data of the driving lane and the right of the passing lane. The abscissa represents the frequency  $f$  in unit of  $1/5\text{min}$ . The slopes which indicate  $1/f$  and  $1/f^2$  are also shown as references.

Next, we show power spectra of the fluctuations. Power law, or so-called  $1/f$  fluctuation, in highway traffic was first reported in 70's by Musha and Higuchi [3]. After that, however, power-law fluctuations have been discussed mostly by theoretical or numerical studies [4] and so far only a report of the observation by Wagner and Peinke has been an exception as far as we know [5]. Since the work by Musha and Higuchi was based on only a few hour's observation, much longer observations have been demanded. Figure 2 plots the power spectra of the fluctuations in the flow. Data for the same period and the same 5 points as above are shown. The power spectra become white at high frequencies. On the other hand, clear power-law behaviors are seen in both lanes in a lower frequency range corresponding to a length of about half an hour to about half a day. The range varies slightly from place to place. In spite of the long-time (one month) averages that were taken, power spectra are still scattering heavily so that the only thing we can say at present about the power law is that its power lies between -1 and -2. We plot two lines having slopes  $1/f$  and  $1/f^2$  just for references.

Figure 3 shows the power spectra of the fluctuations in the average velocity for the same 5 points as before. We can see power-law behavior again. But the frequency range where the power-law decay is observed seems to differ more dramatically from place to place. Interestingly, power-law behavior is most clearly seen at 326.14 km point, which is closest to Nagoya interchange. Detailed analysis on this power-law behavior is left for future work.

So far, we have shown some examples of long-term traffic behavior. Intensive data analysis has just been started so that results will be published in a near future. Some aspects of short-time traffic behaviors observed in the traffic data we have are described in the next paper in this proceedings.

## Acknowledgement

We would like to thank the Japan Highway Public Corporation for providing us vast traffic data. This work is supported by Center for Integrated Research in Science and Engineering (CIRSE) in Nagoya University.

## References

1. Extensive references are found in: D. Helbing, *Rev. Mod. Phys.* **73**, 1067 (2001).
2. K. Nishinari and M. Hayashi (Eds.), *Traffic Statistics in Tomei Express Way*, (The Mathematical Society of Traffic Flow, Japan, 1999).
3. T. Musha and H. Higuchi, *Jpn. J. Appl. Phys.* **15**, 1271 (1976); T. Musha and H. Higuchi, *Jpn. J. Appl. Phys.* **17**, 811 (1978).
4. For example, K. Nagel and H. Herrmann, *Physica A* **199**, 254 (1993); K. Nagel and M. Paczuski, *Phys. Rev. E* **51**, 2909 (1995); S. Yukawa and M. Kikuchi, *J. Phys. Soc. Jpn.* **65**, 916 (1995); S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa, *J. Phys. Soc. Jpn.* **67**, 2270 (1998).
5. P. Wagner and J. Peinke, *Z. Naturforsch.* **52a**, 600 (1997).

# **Telematics and Drivers' Behaviour**

# The Local Occupation Probability Method for Evaluating Traffic Flows

Y. Ishibashi<sup>1</sup> and M. Fukui<sup>2</sup>

<sup>1</sup> School of Communications, Aichi Shukutoku University, Nagakute-cho, Aichi Prefecture 480-1197, Japan

<sup>2</sup> Nakanihon Automotive College, Sakahogi-cho, Gifu 505-0077, Japan

**Abstract.** The local occupation probability method is shown to be useful to analyze the traffic flow in the cases that there are bottlenecks on roads. Concrete examples of the results obtained by making use of this method are presented for the effects of a single bottleneck on a one-dimensional road and for a crossing of two one-dimensional roads whereas the maximum velocity is an integer  $m$  in general ( $m > 0$ ).

## 1 Introduction

The rule-184 model of Wolfram has been often used in the cellular automaton (CA) studies of traffic flows, where each car is supposed to move forward by one cell, when possible, in a unit time [1]. This model has been extended to the cases, where cars can move forward by the maximum  $m$  cells, when possible. This extended model is sometimes called the Fukui-Ishibashi model [2].

To evaluate the flow in these model, Wang et al. introduced the probability  $p_i$  that the  $i$  site is occupied by cars or, in other words, the fraction of the time when the  $i$  site is occupied [3,4]. They could successfully obtain the flows in various cases by evaluating Boolean correlation functions.

The purpose of the present paper is to apply this method, which we call, for the time being, the *local occupation probability method*, to the bottleneck and the crossroad problems.

## 2 Bottleneck Problems in the Multi-Unit Velocity Cases

Let us define the problem. Suppose cars on a one-lane circuit can move forward by the maximum of  $m$  cells when possible. The phrase "when possible" means that the cars may move only  $d$  cells if  $d$ , the empty space in front, is smaller than  $m$ . On the circuit there is a gate, which is opened with the probability  $r$  (see Fig. 1 for the location of the gate and the way of numbering cells).

Let us denote the density of the cars, i.e., the ratio of the number of cars and the total number of sites on the circuit, as  $p$ . If  $p$  is very small, there will be no effect

of the gate and the flow will be  $mp$ . On the other hand, if the density is very high, again there will be no effect of the gate, because then the jam appears here and there, irrespective of the operation (opening and closing) of the gate, and there the flow is given as  $1-p$ . Therefore, the essential problem in this case is to evaluate the flow for the intermediate density.

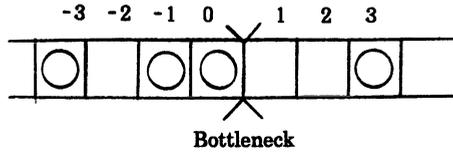


Fig. 1. The bottleneck between sites 0 and 1.

We consider the one-dimensional cyclic road, whose length is  $L$ , and the gate is located between the sites 0 and 1. Suppose that the maximum velocity is  $m$  and the probability of opening the gate is  $r$ . In front of the gate the jam is formed, and so the probabilities that the sites  $i$  ( $i < 0$ ) are occupied are all the same;

$$p_{-1} = \dots = p_{-i} = p_0, \tag{1}$$

where  $J$ , indicating the length of the jam, depends upon the overall car density.

After passing through the gate, the cars run freely and there will appear a sequence of occupation probabilities of the period  $m$ . By simple considerations we find the relations between  $p_0$  and  $p_i$  as [5],

$$p_i = p_0 r^{m-i+1}, \quad 1 < i < m. \tag{2}$$

Among these quantities, the conservation of the probability must hold, i.e.,

$$p_0 + \dots + p_m = 1. \tag{3}$$

From (2) and (3) it is seen that

$$p_0 = 1 / (1 + r + \dots + r^m) \tag{4}$$

The flow  $f$  is given as the product of the average density after the gate and the maximum velocity, i.e., in the present case it is equal to

$$f_b = 1 - p_0 = r(1 - r^m) / (1 - r^{m+1}) \tag{5}$$

(b: bottleneck).

Thus, the lower critical density for the jam caused by the bottleneck is obtained from

$$f_b = mp, \tag{6}$$

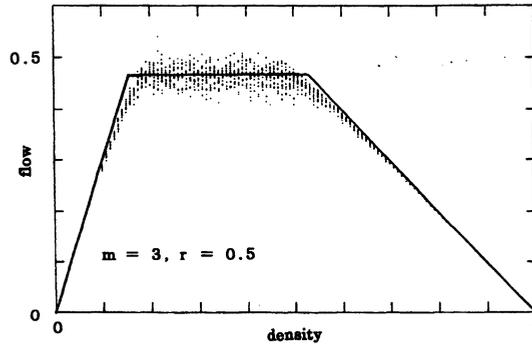
and the upper one follows from

$$f_b = 1 - p. \tag{7}$$

For  $m=1$  the length of the jammed part  $J$  is obtained from the relation that

$$Jp_0 + (L - J)p_1 = Lp . \tag{8}$$

It is easy to see that at the lower critical density the jam length is zero as it should, and at the upper one the jam extends to all the length  $L$ . These are all consistent with the previous findings [6]. The flow is shown in Fig. 2 as the function of the density for the cases  $m=3$  and  $r=0.5$ .



**Fig. 2.** The fundamental figure for  $m = 3$  and  $r = 0.5$ . The solid line shows  $3p$ ,  $r(1 - r^3) / (1 - r^4)$ ,  $1 - p$  from the low density side.

### 3 The Crossroads

The crossroad can be regarded as a kind of a bottleneck, where the eastbound cars disturb the flow in the northbound cars and vice versa. The difference from the case studied in the previous section is that in the ordinary bottleneck the disturbance is probabilistic, while on the crossroad the disturbance will be periodical. In spite of this difference, we can also apply the local occupation probability method to the crossroad problems.

The model is shown in Fig. 3, where the eastbound and the northbound cars cross at the crossing point denoted with 0. The local occupation probabilities for the eastbound and the northbound cars are expressed with  $p_i$  and  $q_i$ , respectively, where  $i$  indicates the site.

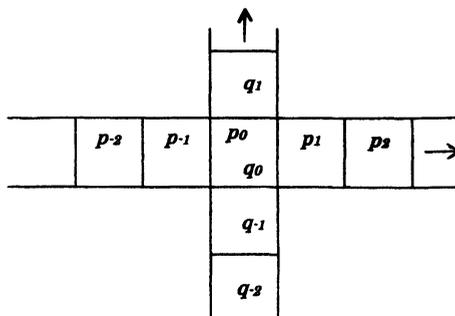


Fig. 3. Two crossing roads.  $p_i$  and  $q_i$  denote the local occupation probabilities.

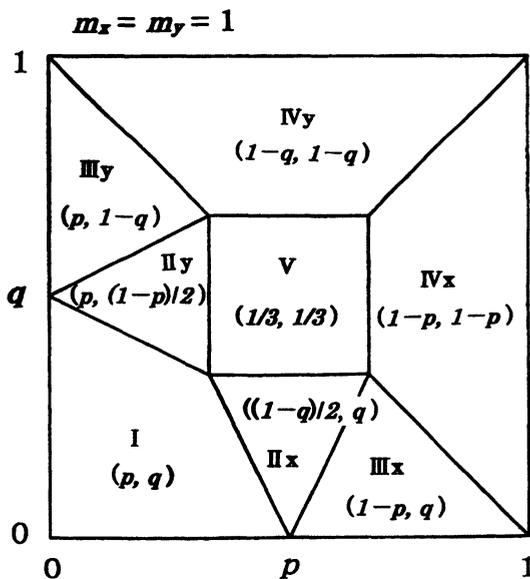


Fig. 4. The phase diagram for the case  $m_x = m_y = 1$ . The flows  $f_x$  and  $f_y$  are presented as  $(f_x, f_y)$ .

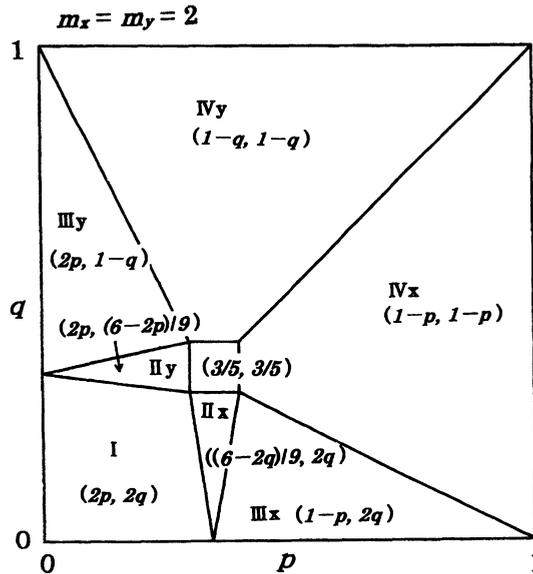


Fig. 5. The phase diagram for the case  $m_x = m_y = 2$ . The flows  $f_x$  and  $f_y$  are presented as  $(f_x, f_y)$ .

The purpose here is to show the phase diagrams of the flow as functions of the eastbound and the northbound car densities,  $p$  and  $q$ , for the cases that the maximum velocities of the eastbound and the northbound cars,  $m_x$  and  $m_y$ , respectively, are equal, i.e.,  $m_x = m_y = m$ . It is known that in such cases a phase diagram consists of eight regions, though they can be classified into five by considering the symmetry (Fig. 4). In the region I the cars in both direction form the free flow, in the region II the cars with low overall density disturb the traffic on the other direction, and in the region III the jam is formed mostly by other cars in the same direction. In the region IV even the flow in the direction of a low car density cannot be free, and the region V is very special in the sense that the flows in both directions are the same.

The concrete expressions of the flow in each region are indicated in the phase diagram for  $m=1$  (Fig. 4) and  $m=2$  (Fig. 5), though regarding the concrete mathematical procedures the readers are referred to Ref. 7. It is seen that as  $m$  increases the region V shifts to the lower density side.

#### 4 Concluding Remarks

In this paper we have shown that the local occupation probability method is a useful method to analyze the traffic flow in the cases that there are bottlenecks on roads. The obtained results for concrete examples are presented for the effect of a

single bottleneck on one-dimensional roads and that of the crossing of two one-dimensional roads when the maximum velocity is an integer  $m$  in general ( $m > 0$ ).

It is needless to say that this method can be used in analyses, regarding the bottleneck effect, of the effect of multiple bottlenecks on roads and, regarding the crossroad problems, of the effect of the crossing point on the phase diagrams in the case of an asymmetric velocities, i.e.,  $m_x > m_y$ . Studies on these problems are now under progress and will be published elsewhere shortly.

## References

1. S. Wolfram, *Rev. Mod. Phys.* **55**, 601 (1983).
2. M. Fukui and Y. Ishibashi, *J. Phys. Soc. Jpn.* **65**, 1868 (1996).
3. B.H. Wang, L. Wang, P.M. Hui, and B. Hu, *Phys. Rev. E* **58**, 2876(1998).
4. B.H. Wang, L. Wang, P.M. Hui, and B. Hu, *Physica B* **279**, 237 (2000).
5. Y. Ishibashi and M. Fukui, *J. Phys. Soc. Jpn.* **70**, 1237 (2001).
6. S. Yukawa, M. Kikuchi, and S. Tadaki: *J. Phys. Soc. Jpn.* **63**, 3609 (1994).
7. Y. Ishibashi and M. Fukui, *J. Phys. Soc. Jpn.* **70**, 2793 (2001).

# Cooperative Driving: Taking Telematics to the Next Level

R.G. Herrtwich and G. Nöcker

DaimlerChrysler AG, Telematics Research, Alt-Moabit 96a, 10559 Berlin and  
HPC E224, 70546 Stuttgart, Germany

**Abstract.** While existing telematics services largely focus on improving mobility and comfort, an upcoming generation of car-centric services will target vehicle and driver safety. Future vehicles will serve as sensors and actuators – and they will be connected by communication links. In new services, information communicated from outside the vehicle or obtained from an adaptive map database will be used as an additional sensor for driver assistance – and potentially at one point for vehicle control. Utilizing higher positioning accuracy and reduced communication latency, telematics will enable new safety applications ranging from alerting the driver about an imminent danger to active interference into the vehicle controls. Examples of such services are hazard warning or collision avoidance. Most of those applications require a sense of the environment of the vehicle as well as information about the traffic ahead including cross-traffic. Telematics extends on-board sensors, permits to communicate intentions, and facilitates road courtesy. It opens the path to an application domain which we call cooperative driving. Traffic adaptive behavior and harmonious driving can solve the conflict of safety and efficiency in dense traffic. Cooperative driving is the ultimate behavior in traffic. It means to gather information and share it with other drivers, to adapt to the surrounding traffic and environment conditions, to interact with other traffic participants in a solution-oriented way, and to obey traffic rules and regulations. The most prominent goals of cooperative driving are to increase traffic safety and traffic flow.

## 1 The Vision

In the 50's and 60's visions came up reaching from automated road traffic to “flying cars”. These visions were inspired by the progress made in electronics and computer technology. Not all these dreams have come true as a view on nowadays traffic shows, but some of them have come closer to reality than ever. Communication units and chip set for GPS (Global Positioning System) have become small, cheap, and are available off the shelf. Portable phones and map-based vehicle guidance systems have become very popular. On the vehicle side, first assistance systems such as Adaptive Cruise Control (ACC) were developed and entered the market in the last two years. Well equipped with environment sensors and drive-by-wire functions, these cars can act as sensors and actuators in traffic. Now the point has come where connectivity of vehicles can be foreseen as a standard vehicle feature and substantial vehicle functions can be improved by or based on

communication and telematics. Autonomous driver assistance systems for automatic cruising, lane keeping, and lane changing up to collision mitigation or avoidance will be melted together with navigation and communication. New functions of cooperative driving will arise, which will lead to more safety and efficiency on the roads of the future. Thus, the vision of vehicle-supported cooperative traffic is no longer an unreachable goal.

Accident statistics [1] from the German Ministry of Transportation show that 87.4% of accidents with injured persons on German roads in 1999 were caused by drivers. 26.7% were caused by driving too fast and too close (with respect to the situation where the accident occurred). 25% of the accidents were caused by right of way violation or entry, exit and turning mistakes. Further 11.5% happened because of wrong lane usage, or wrong overtaking or passing by. So more than 63% of the accidents could have been avoided or mitigated by proper and earlier reaction of the drivers. Cooperative Driving will assist the driver by advance information and maneuvering aid.

Traffic safety and traffic efficiency seem to be two contrary goals in traffic. Where safety asks for large headway and low speed, highly efficient dense traffic means fast and close driving. Traffic research especially in the field of traffic phases (Fig. 1) led to a new understanding of traffic dynamics and enabled new methods of forecasting congestion patterns [2,3,4,5,6,7,8,9].

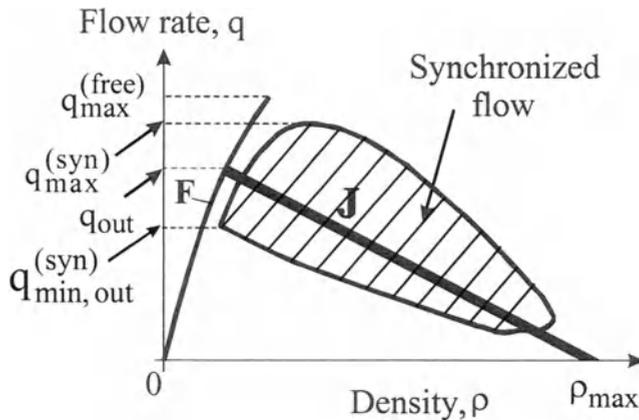


Fig. 1. Sketch from B. S. Kerner showing the synchronized flow as a third phase in traffic.

New results in traffic research show a large potential for improvements in traffic flow. In [10] it is figured out that undisturbed fast and dense traffic flow is about 1.5 times higher than the outflow of a traffic jam. Stabilizing flow and damping of disturbances in dense traffic by harmonious driving can improve traffic flow, because there are still reserves, which are wasted today through unharmonious driving.

A study from the German Bundesanstalt für Straßenwesen (BASt) showed that about 85% of all road accidents are based on misjudgment, distraction and sleepiness of the drivers. Road accidents caused an economic loss of about DM 68 bil-

lion in 1998 in Germany. In a ADAC study traffic jams on the German Autobahn are caused in 33% by accidents, in 31% by construction sites, and in 32% by too much traffic. The economic loss is estimated at DM 200 billion per year. Cooperative driving will not only serve the driver, but will lead to a considerable economic benefit by traffic safety and improved traffic flow.

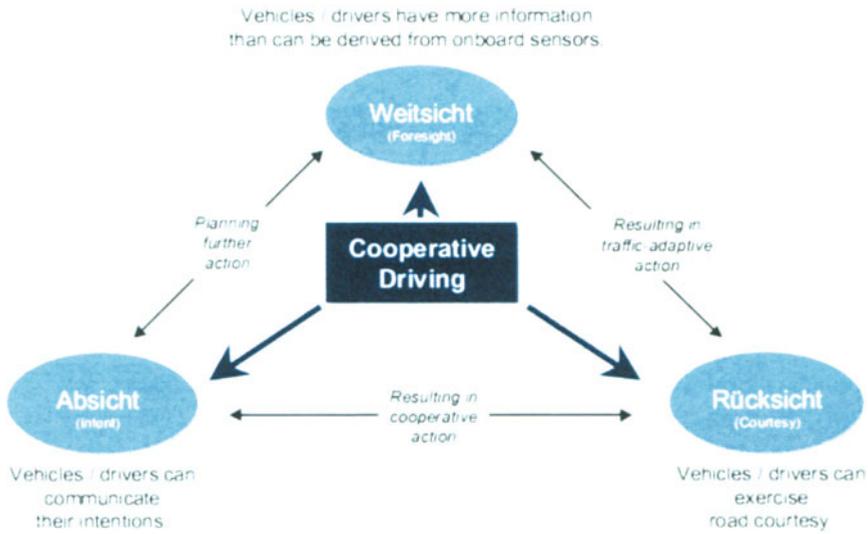
The idea of cooperative driving is not all that new and earlier research programs already addressed this theme. Our work started in 1988 in the PROMETHEUS Project (Subproject Copdrive) and was presented as part of the Common European Demonstrator CED 4 in Munich. Cooperative driving maneuvers such as lane changes were carried out with manually operated radio location/orientation and communication. GPS was not available at that time. During the course of the project, the focus changed towards registration and communication of warning messages. In the PROMOTE Project Chauffeur heavy trucks were coupled via distance sensors and vehicle-to-vehicle communication. The telematic coupling enables stable following control with short headway. In Japan, the AHSRA Project investigates coordinated driving through infrastructure-based communication and inter-vehicle communication, where we have successfully participated in the SmartCruise21 Demo in Tsukuba in November 2000.

## 2 Introducing Cooperative Driving

Cooperative driving is the ultimate behavior in traffic. It means to gather information and share it with other drivers, to adapt to the surrounding traffic and environment conditions, to interact with other traffic participants in a solution-oriented way, and to obey traffic rules and regulations. We distinguish three levels of cooperative driving: (1) cooperative information, where drivers know everything going on around them, even things they cannot sense themselves; (2) cooperative traffic flow, where cars synchronize their drive for higher traffic efficiency, better fuel economy, and collision avoidance; (3) cooperative maneuvering, where more complex driving situations, e.g., at intersections, are resolved.

Driving cooperatively means to gather information about weather and traffic through onboard sensors, refine it by algorithms, and share it with others. Drivers and assistance systems have to understand the physics of traffic. They must interact with others in a solution-oriented way and obey traffic rules and regulations. Cooperative vehicle systems need a holistic view of the environment, and have to adapt to the surroundings. They have to behave well in given traffic situations, and to understand their options. Our aim is to implement vehicles that have a sense of traffic and can, thus, either support their drivers in behaving cooperatively and traffic-adaptive or act cooperatively themselves.

The benefits of Cooperative Driving (Fig. 2) are shown by the following triangle of foresight, intent and courtesy.



**Fig. 2.** Benefits of Cooperative Driving.

Telematics adds to the information of the driver and the onboard sensors of a vehicle by extending the information horizon of the vehicle through communication. This foresight (Fig. 3) enables drivers and assistance systems to react earlier and adapt to the traffic around them.



Information Horizon	Simple sensors	Complex sensors	Telematics
Traffic Information	little	local	remote / predictive
System Quality	Worse than the driver	As good as the driver	Better than the driver

**Fig. 3.** Foresight.

In dense traffic and complex maneuvering situations early communication of intentions leaves other traffic participants more time to react by shifting the starting point of the reaction forward.

To exercise road courtesy is the law (e.g., German Road Traffic Regulations, StVO§1 Basic Rules): (1) The participation in road traffic requires permanent caution and respect for each other. (2) Every road user needs to act such that no other road user is jeopardized, harmed, or more than absolutely necessary restricted or troubled.

Cooperative driving has the potential to improve traffic climate more than any other measure.

### **3 Technical Elements of Cooperative Driving**

Key elements of cooperative driving are connected vehicles, located vehicles, and service centers. Cooperative vehicles must be connected to other vehicles and to service centers as well as to infrastructure information systems.

Therefore, communication is the most crucial component of cooperative driving. Connected vehicles have a big advantage compared with isolated vehicles. While isolated vehicles can only sense the obvious and control their own action, connected vehicles have non-sensible information from far ahead and about drivers' intentions at their disposal. This enables negotiation and coordination of actions.

A connected vehicle serves as a remote sensor and remote actuator. But remote information - at least for the time being - is not a substitute for on-board sensors. The most attractive use of telematics is to complement other sensors, either for calibration or for redundancy. The connected vehicle as a remote sensor makes vehicle-generated data about drivers' intentions, vehicle status and motion, and about the environment available to other vehicles and service centers. Safety-relevant information can be communicated directly to other vehicles in the vicinity. Traffic and weather-related information will be collected as "floating car data" in service centers and be distributed from there - potentially a new business. Based on the available information, the connected vehicle works as a remote actuator where recommendations are carried out by the driver, when he is in the loop, or by the vehicle controller, when an emergency situation leaves no room for manual intervention. This enables warning and information systems well beyond today's driver assistance systems.

The second key element of cooperative driving is the located vehicle. The connected vehicle needs to know its position relative to infrastructure and relative to other vehicles. Advanced maps and high-precision position sensing are the basic techniques. Positioning provides the reference frame for cooperative driving. As loosely-located vehicles have only a general sense of the environment, drivers or other sensors have to provide fine tuning and adjustments. Precisely-located vehicles have precise lane information and location of other vehicles and obstacles and can, thus, execute maneuvers on a very high level of confidence.

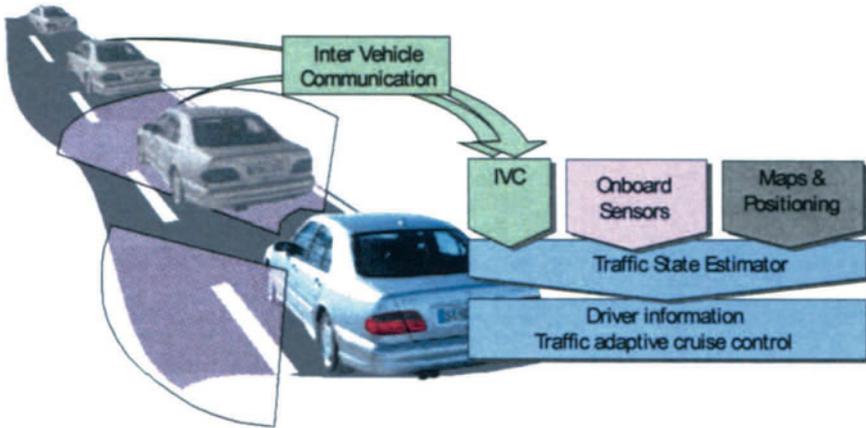
The third key element is the service center. Telematics services can deduce location information from vehicles, and provide location-based traffic and position-aware safety information to vehicles. Service centers connect vehicles to other vehicles over time and space: Without a critical mass of equipped vehicles in a given area, information will be lost and ‚forgotten‘ over time. While distributed information is attached to time, i.e., to vehicles in the vicinity and action affects cars in the vicinity, central information is attached to a certain space rather than just time and will potentially affect cars at any range. A service center is a repository for historical information and data deduced from many vehicles.

#### **4 Applications of Cooperative Driving**

The first level of cooperative driving is cooperative or advance information, where drivers know everything going on around them, even things they cannot sense themselves, potentially ahead of time, and can condition themselves to react more quickly and in a better and safer way. In this level, systems for optimal driving strategy are developed, with respect to traffic and weather situations such as low speed ahead because of overtaking trucks or reasonable speed reduction because of icy roads, snow, or fog.

The second level is cooperative or harmonious traffic flow. This means to avoid detrimental behavior, where cars synchronize their drive for higher traffic efficiency, better fuel economy, and collision avoidance. The development of systems for this level is closely connected to the results of traffic theory and research. Driver assistance systems must behave in traffic like experienced drivers. Human drivers normally show a traffic-dependent behavior. They observe the actual traffic situation around them and adapt their driving strategy to traffic flow and density and to changes of both. They adjust their following behavior and their choice of lane. Therefore, driver behavior in traffic has to be investigated in respect to the surrounding traffic and must be transformed in rules and control strategies.

The quality of virtual drivers depends strongly on the quality of their traffic information. By means of onboard sensors, information about speed and distance of vehicles on neighboring lanes is available. By use of this information, driving of assistance systems can be adapted to the local traffic situation. A forecast that extends the visibility range of the driver, is reached by inter-vehicle communication. This forecast enables safe and comfortable driving with low emissions and less fuel consumption. A system called onboard traffic monitor uses sensor and communication information combined will calculate the local traffic situation an the situation ahead with the help of an onboard traffic state estimator. The adaptive vehicle (Fig. 4) becomes an important basic system for the future and is essential for harmonious driving.

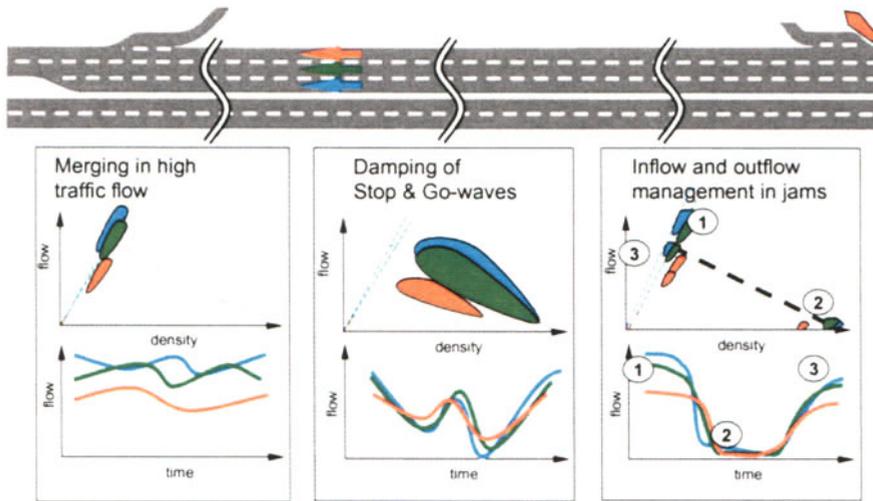


**Fig. 4.** Adaptive Vehicle.

If communication and driver assistance systems are used to improve traffic flow, one must come from macroscopic considerations to microscopic conclusions. The analysis of single vehicle data, especially of speeds and time gaps, is necessary in order to get a better understanding of following behavior and clustering effects.

Cooperative and harmonious traffic flow (Fig. 5) offers new chances for the stabilization of high traffic flows and for the absorption of critical, local disturbances. Through the use of sensors, connectivity, and positioning and through their intelligent combination, an improvement of comfort, traffic performance, and safety can be achieved in dense traffic. Research work in the field of traffic-adaptive driving started in 2000 [11] and is now carried out within the INVENT Program, funded by the German Ministry of Education and Research. First systems are aiming for improved merging procedures at bottlenecks and for the damping of stop&go waves. Starting with information systems, this work will end in adaptive adjustment of intelligent cruise control systems.

Experiments in real traffic would be dangerous and too expensive, so the development of cooperative and adaptive systems has to be done in the first stage in virtual reality. Traffic simulation is essential for a better understanding of the effects caused by new systems. Because traffic is an active multi-particle system, traffic-adaptive systems will change macroscopic properties of traffic by their modified microscopic behavior. Research in this field will be an important task for the future.



**Fig. 5.** Harmonious Traffic flow by cooperative driving.

The third and highest level of cooperative driving is cooperative maneuvering - which brings all of the above together. Here, more complex driving situations, e.g., at intersections, are resolved and drivers/cars negotiate their driving maneuvers for higher road safety and road efficiency. Examples for cooperative maneuvering are manifold. Assistance systems for interweaving of traffic streams on highways at on- and off-ramps and at lane drops are investigated. Safe lane change must be possible even if the sight is limited. Finding the right timeslot when making a left turn and showing that one is waiting for a long time already and that help would be appreciated needs communication of intentions. Cooperative systems must know the traffic rules which affect the right of way, zipping at bottlenecks or passing at obstacles. Cooperative vehicles will communicate to the succeeding vehicles that one intends to pass or turn or that it is safe to pass and provide information about oncoming traffic. Assistance in crossing of streets and safe crossing at complex intersections is the most complex goal.

## 5 Some Metrics

Most important influence factors for establishing cooperative driving are communication latency and frequency, positioning accuracy, and required penetration.

Looking at the communication latency, we find volatile information about obstacle positioning and vehicle maneuvering which must be directly communicated within milliseconds. Safety-relevant information on the nearby area should also be communicated fast and directly. Dynamics of viscid information about road and weather conditions as well as traffic situation lays in the area of minutes and can

be communicated by service centers. Information that is more stagnant such as road layout and vehicle dynamics can be updated at a monthly rate.

Control applications for collision avoidance and path optimizing need a very high positioning accuracy of about 0.1m. Alarm applications need accuracy in the meter range and information applications require only a rough positioning accuracy of about 10m. The next development steps of the GPS Constellation will provide the framework for alarm systems in the 2006-2008 time frame, and for control application in the 2012-2015 time frame.

Communication latency vs. positioning accuracy defines the Position-Aware Safety Systems (PASS) application space (Fig. 6).

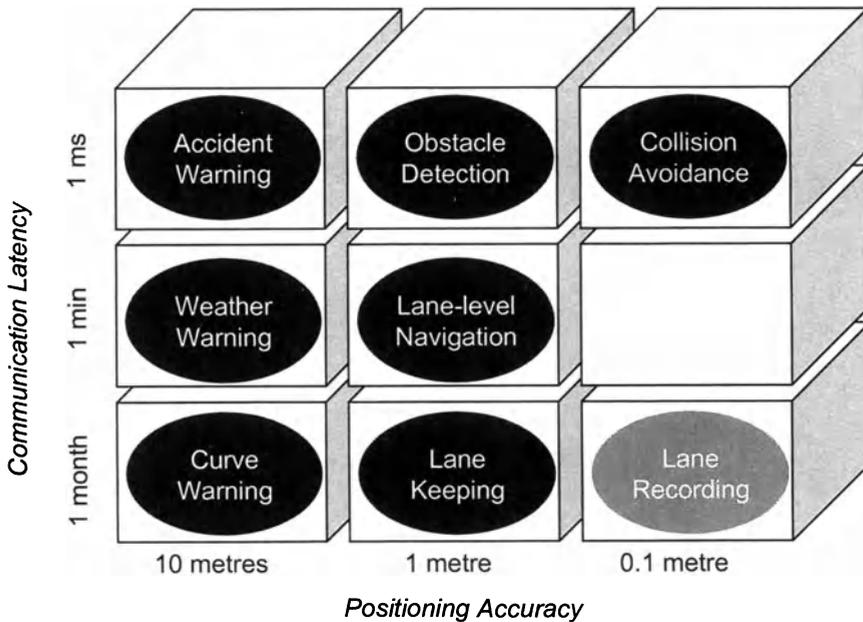


Fig. 6. Position Aware Safety Systems.

While advance information systems can be established with only some vehicles, especially when information centers are used, harmonious driving needs many vehicles to show an effect on traffic flow. The necessary penetration rate for cooperative maneuvering is very high. Most vehicles must be equipped with communication when maneuvering shall increase safety on the road. It is obvious that the demand for standardization increases with the penetration level. Because of complexity and availability reasons we will start with advance information and end with cooperative maneuvering.

## 6 Conclusion

Research work has started in the area of cooperative driving, but fundamental challenges remain. Communication systems, world models, and assistance systems have to be further developed. Drivers need predictable and liable systems as “little helpers” and they need always to know who is responsible for what their cars do. Taking care of the human-machine interface is essential for successful systems in the market. And a legal framework for new systems with such a high impact on traffic has to be generated. So, we do not have efficient and accident-free traffic soon - but we come a lot closer!

## References

1. Verkehr in Zahlen 2000: Hrsg. Bundesverkehrsministerium, Deutscher Verkehrs-Verlag GmbH, Hamburg, 2000.
2. B.S. Kerner: Experimental Features of Self-organization in Traffic Flow, *Physical Review Letters*, Vol. 81, 3797, 1998.
3. B.S. Kerner: A Theory of Congested Traffic Flow, *Proceedings of the Third International Symposium on Highway Capacity*, R. Rysgaard (Ed.), Vol. 2, pp. 621-641, 1998.
4. B.S. Kerner: Congested Traffic Flow: Observations and Theory, *Transportation Research Record*, Vol. 1678, 160-167, 1999.
5. B.S. Kerner: The Physics of Traffic, *Physics World*, 25-30, No. 8, 1999.
6. B.S. Kerner: Theory of Breakdown Phenomenon at Highway Bottlenecks, *Transportation Research Record*, Vol. 1710, 136-144, 2000.
7. B.S. Kerner: Complexity of Synchronized Flow and Related Problems for Basic Assumptions of Traffic Flow Theories, *Networks and Spatial Economics*, Vol. 1, 35, 2001.
8. B.S. Kerner: Tracing and Forecasting of Congested Patterns for Highway Traffic Management, *IEEE IST Conference Proceedings*, pp. 88-93, 2001.
9. B.S. Kerner, R. G. Herrtwich: Traffic Flow Forecasting, *Automatisierungstechnik*, Vol. 49, 2001.
10. B.S. Kerner, H. Rehborn: Experimental Properties of Complexity in Traffic Flow, *Physical Review*, Vol. 53, R4257-R4260, 1996.
11. G. Nöcker, D. Hermann, A. Hiller: Telematics Based Traffic Organisation, *Proceedings of the 7th World Congress on Intelligent Transport Systems*, Torino, Italy 2000.

# Modeling of Vehicular Behavior from Road Traffic Engineering Perspectives

H. Ozaki

Department of Civil and Environmental Engineering, Toyo University, Kawagoe, 350-8585, Japan

**Abstract.** This paper studies modeling of vehicular behavior from the view of road traffic engineering. Bottleneck phenomena at roadways are described by several observed facts. Simulation studies are presented with modeling of car-following behavior and human factors to show reproduction of the bottleneck phenomena and their improvements by ITS applications.

## 1 Introduction

Road traffic engineering needs modeling. It needs more knowledge of road traffic flow to realize preferable road designs and traffic controls with less real trial-and-error basis. Reliable models that are acquired through analyses can be applied to next sites.

Requirements of models from the road traffic engineering are two-fold. Models should be accurate enough to describe real world to be applied for engineering purposes. They should also be easy to handle with observable parameters to support decision-makings on selection of alternative plans.

This paper discusses modeling of vehicular behavior. Special interest is given to the capacity bottleneck phenomena at homogeneous roadway sections.

## 2 Capacity Bottlenecks on Roadways

Capacity bottlenecks phenomena appear where traffic demand exceeds roadway capacity. These phenomena may reasonably occur at intersections or merging sections where several streams of traffic sum up above the roadway capacity. The other natural examples would be lane-drop sections or lane closures where roadway capacity locally decreases compared with the upstream sections.

However, there also exist very unreasonable bottlenecks at homogeneous roadway sections. The typical sites are vertical alignment sags and long tunnels.

There are two phases observed at these bottleneck sites. The sites trigger the breakdown of flow or occurrence of congestion. They also are characterized as where the heads of congestion queue persist.

Fig. 1 illustrates vehicles' trajectories on the space-time plane. It is collected at Hadano sag section on Tomei Expressway when the breakdown of flow takes place. Sag sections tend to reduce the speed of vehicles by the increase of gradient. Facing the deceleration of lead vehicles, some following vehicles try to absorb disturbance of flow by reducing their headways, but others do not resulting in the occurrence of decelerating wave which propagates upstream.

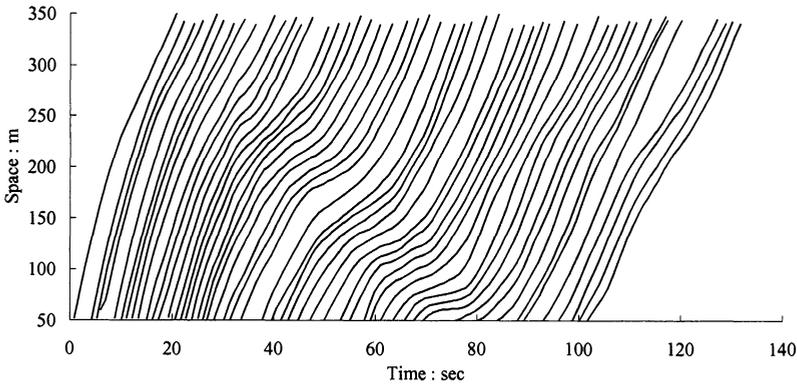


Fig. 1. Breakdown of flow observed at Hadano sag section, (Xing 1992).

Once breakdown of traffic flow occurs, very low capacity is observed during congestion. According to the observation, capacity falls as low as 2,500 to 2,800 vehicles per hour (vph) per two-lane at Japanese expressway. Moreover, capacity tends to decrease, as queue grows longer.

Another point is low acceleration of vehicles from the head of congestion queue. Several kilometers are needed for speed recovery.

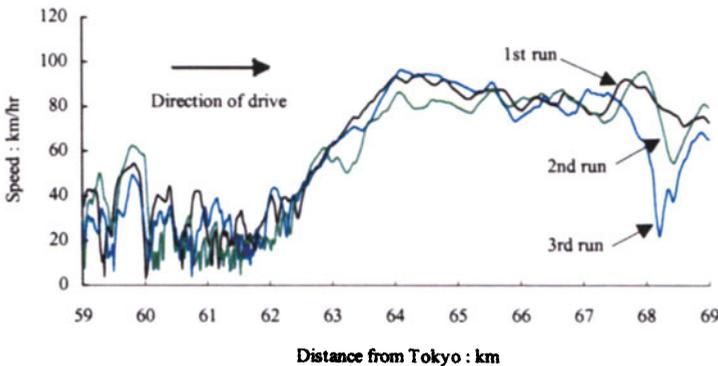


Fig. 2. Speed profiles collected by a test vehicle at Nakano sag section.

Fig. 2 compares three speed profiles collected by a test vehicle along ten kilometers where Nakano sag section locates at 62 kilo-post (kp) on Chuo Expressway. The series of speed data of three runs are collected about one hour time interval during congestion. It shows how the head of congestion queue and speed recovery profile are stable.

Fig. 3 shows a clearance-speed trajectory of the test vehicle along the ten kilometers stretch observed by one of the three runs. During stop and go operation along the first three kilometers, speed of the vehicle is bounded as high as 40 to 50 km/hr levels. Going through the bottleneck sag, the vehicle speed reaches 80 to 100 km/hr after slow acceleration.

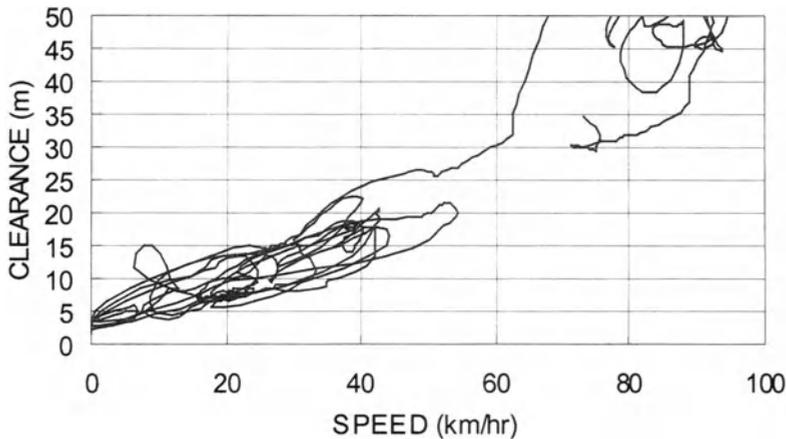


Fig. 3. Clearance-speed trajectory of the test vehicle (First run).

Fig. 4 is a schematic view of vertical alignment of Hadano sag, which locates at 47 kp on Tomei Expressway. Vertical gradients change from  $-2.6\%$  to  $-0.3\%$  and 500 meters of gradient curve is inserted for 2.3% increase, which should be mild enough for the road design standard. There are three vehicle detectors installed along the roadway. They locate at 47.5 kp, 46.6 kp, and 45.9 kp.

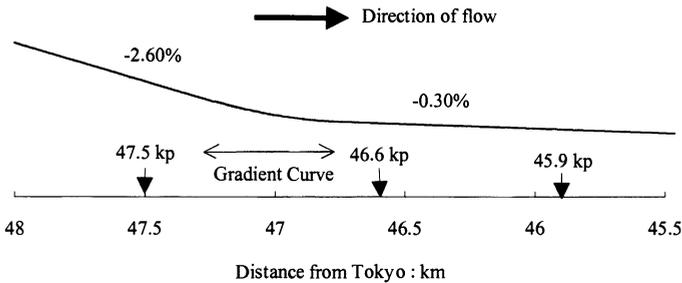


Fig. 4. A schematic view of vertical alignment of Hadano sag.

Fig. 5 shows the four-hour time series of 1-minute volumes (two lanes for one direction) collected at the three detectors. Before 14 o'clock, 1-minute volumes fluctuate, which is characterized as random arrival of traffic flow. However, after about 14 o'clock when the congestion starts, 1-minute volume at every detector becomes very stable. Since there are neither entrances nor exits along the sag, volumes at the three detectors appear reasonably the same.

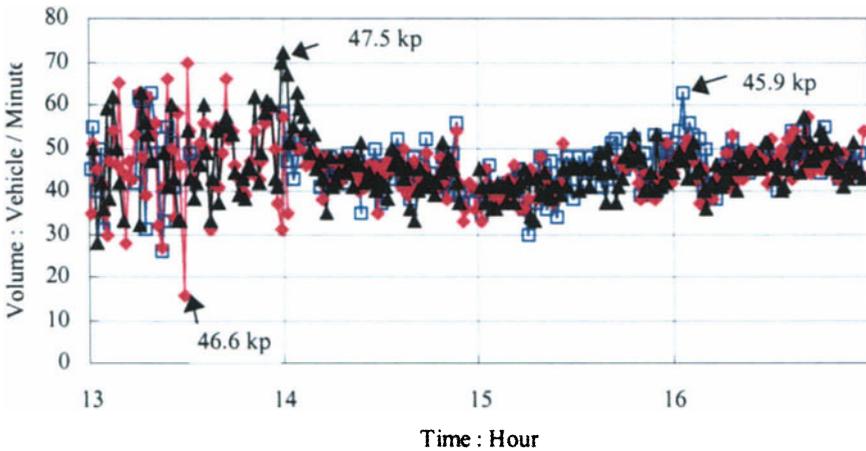


Fig. 5. 1-minute volumes at three detectors on Hadano sag.

Figs. 6(a)-(c) illustrate 1-minute lane-by-lane volumes at the three sites. Fig. 6(a) is observed at 47.5 kp, the entrance of vertical gradient curve. Fig. 6(b) is at 46.6 kp, the exit of vertical curve, and Fig. 6(c) is observed at 45.9 kp, further downstream. Comparison of the three figures, one may find the following.

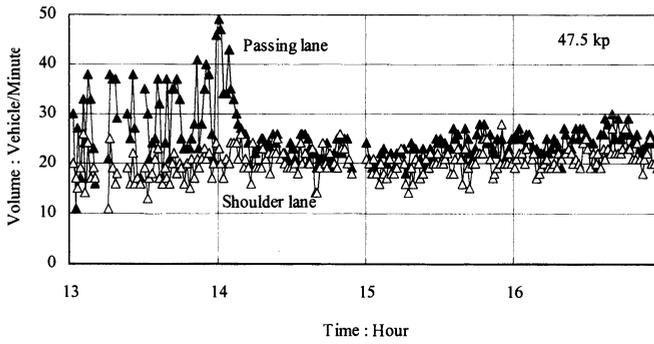


Fig. 6(a). 1-minute lane-by-lane volumes observed at 47.5 kp.

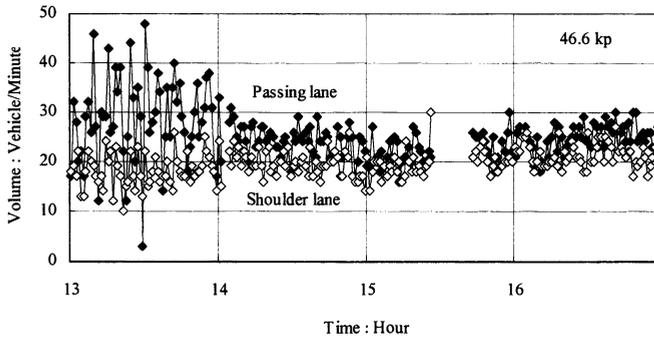


Fig. 6(b). 1-minute lane-by-lane volumes observed at 46.6 kp.

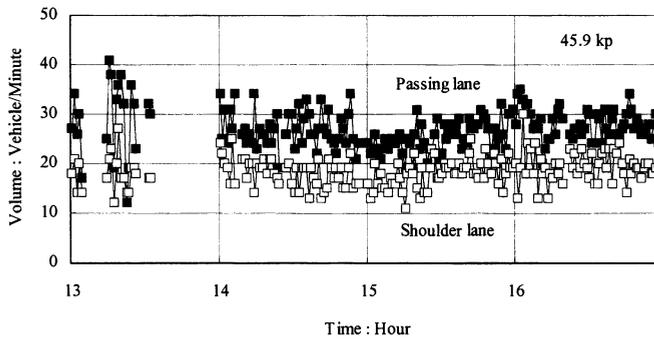


Fig. 6(c). 1-minute lane-by-lane volumes observed at 45.9 kp.

- Before congestion occurs, more vehicles use the passing lane. The flow rates on the passing lane fluctuate and are sometimes very high. On the contrary, usage of the shoulder lane is less and rather stable.

- Very high rates of one-minute flow are temporally observed on the passing lane at 46.6 kp (the exit of the sag) before the start of the congestion.

- At around 14 o'clock, a very dense vehicular platoon arrives at 47.5 kp, which is estimated to cause the breakdown of flow at the sag.

- After congestion develops, rates of flow on both lanes become similar and stable. Lane change actions gradually take place at downstream of the sag.

Fig.7 compares 1-minute average speeds at the three sites. It shows that after congestion occurs, speed is very stable and site dependent.

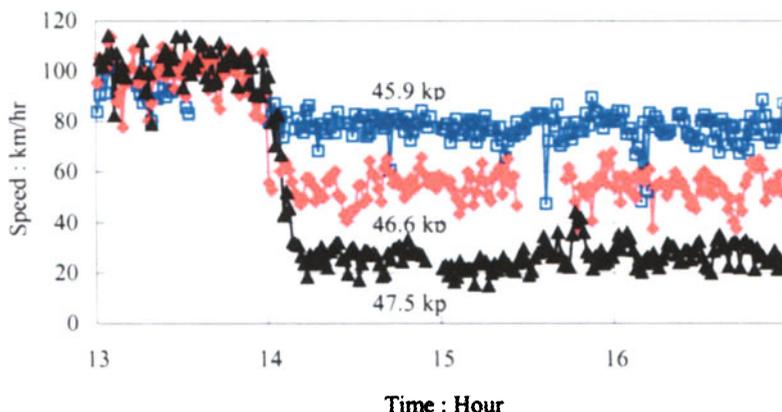


Fig. 7. 1-minute average speed at the three sites.

Following are the summaries of several facts learned from various observations on Japanese expressways.

- Sags with mild enough gradient change tend to become capacity bottlenecks. However, congestion hardly occurs at sags where change of gradient can be easily perceived.

- There is slight difference in queue discharge flow among lanes, which implies that lane changes are less around the bottlenecks.

- Capacity gradually decreases after captured for long period. It is estimated that the period of time caught in the queue has some effects on the capacity reduction at the bottleneck.

- Good lighting atmosphere improves capacity, which is observed at bottleneck tunnels after sunset. It is also observed that at bottlenecks sags capacity becomes larger after sunrise.

- Another interesting phenomena is observed at clearance works after accidents. Accidents may produce bottleneck also at an opposing direction. This is considered to be related by the drivers' "looking aside at-site".

All those observations indicate that there are some human factors related to the bottleneck phenomena at homogeneous roadway sections. Since drivers' percep-

tion to identify vertical gradient change is not reliable at mild enough sag sections, they might drive as they would do on level sections. There may be some delay in finding gradient increase.

If we consider the drivers' perception to estimate speed or spacing inside tunnels, lighting condition should affect miss-estimation. It would also be reasonable to estimate that drivers' actions to follow lead vehicles fall into unintentional delays to react by looking aside at accident sites, and some may lose intention to keep proper headway to their lead vehicles after captured into congestion for long time period.

### 3 Modeling of the Phenomena

This section describes the hypothesis of driver's factors that effect bottleneck phenomena at homogeneous roadway sections. Then it is applied with a proposed car-following model to simulate traffic flow around vertical alignment sag.

Car-following behavior has been extensively studied since 1950's. It has been proposed a notable car-following model, called "GM model", which describes the acceleration of the follower as relational to relative speed [1].

$$\dot{V}_{F1}(t + T_1) = \alpha \cdot \frac{[V_F(t)]^m}{[S(t)]^l} \cdot [V_L(t) - V_F(t)] \tag{1}$$

where

- $V_F$  = speed of the follower,
- $V_L$  = speed of the leader,
- $S$  = spacing between the follower and the leader,
- $T_1$  = reaction time lag to relative speed, and
- $\alpha, l, m$  are parameters.

Koshi et al. pointed out that desired speed or spacing should be considered on actual car-following behavior and they are more effective at congested situations around bottleneck sections [2].

Considering the so many previous studies, Ozaki proposed an elaborated car-following model for level roadway sections [3]. It has three terms to describe the followers' acceleration as (2).

The three terms imply response to relative speed, response to deviation from the target spacing, and adjustment to follow the behavior of the lead vehicle, respectively. In effect, (2) appears that two terms are added to the model (1).

$$\dot{V}_F(t) = \dot{V}_{F1}(t) + \dot{V}_{F2}(t) + \dot{V}_{F3}(t) \tag{2}$$

$$\dot{V}_{F2}(t + T_2) = \beta \cdot \frac{1}{[S(t)]^n} \cdot \{S(t) - f[V_F(t)]\} \tag{3}$$

$$\dot{V}_{F3}(t+T_3) = \delta \cdot \frac{1}{[S(t)]^h} \cdot \dot{V}_L(t) \quad (4)$$

where,

- $f(V_F)$  = function of target spacing when the speed of follower is  $V_F$ ,
- $T_2$  = reaction time lag for response to deviation from the target spacing,
- $T_3$  = reaction time lag for response to lead vehicle, and
- $\beta, \delta, n, h$  are parameters.

Parameters of the model are calibrated with observed data collected on congested situations on the level roadway.

### 3.1 Influence of Vertical Gradient

Considering the facts observed at roadways, it is hypothesized that drivers delay their adjustment to mild enough vertical gradient change. This hypothesis is modeled by that drivers are able to adjust the vertical gradient  $\theta$  where they drove at time  $T_{grad}$  before. It is modeled that optional deceleration term (5) is added to the elaborated car-following model (2) at time  $t$  as:

$$-g [\sin\theta(t) - \sin\theta(t - T_{grad})] \quad (5)$$

where,

- $g$  = acceleration of gravity,
- $\theta(t)$  = vertical gradient at the point vehicle passes at time  $t$  and
- $T_{grad}$  = delay in adjusting to vertical gradient.

Along the constant gradient section, this term is not effective. At sag section, since the gradient gradually increases along the gradient curve, this term reduces the output of acceleration by the car-following model (2).

### 3.2 Simulation Study -1

The purpose of the simulation study is to confirm whether the modeling of hypothesis induces persistent capacity bottleneck at sags.

The simulation starts with the initial vehicle that enters a sag section which models Hadano sag on Tomei Expressway with constant 30 km/hr. Then it starts acceleration to the assigned final target speed. The positions of the start of the acceleration are controlled by three cases; entrance and exit of the gradient curve, and center of the two positions. The final target speed is either 60, 90, 120 km/hr. All the rest followers are assumed identical; they follow their leaders according to the elaborated car-following model with delayed compensation of the gradient curve increase by five seconds.

Input conditions of the simulations are summarized as Table 1.

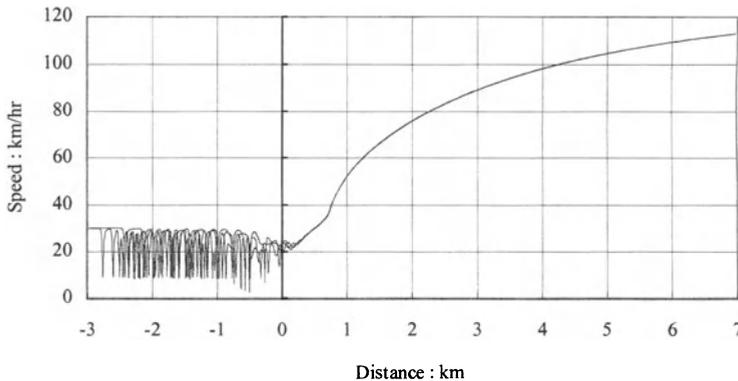
**Table 1.** Input conditions of simulation study –1.

Items	Input conditions
Roadway	1-lane with no overtaking
A sag	2.3% gradient increase 500m gradient curve
Arrival flow	uniform, 1500 vph, 30 km/hr
Initial vehicle	accelerates from the sag $V_{1Target}$ (target final speed) = 60, 90, 120 km/hr $X_{aStart}$ (position of acceleration start)= Entrance, Middle, Exit of gradient curve
Followers	the elaborated car-following model
Compensation lag	$T_{grad}$ = 5 seconds

**3.3 Simulation Result -1**

Trials of simulation result in that the head of congestion queue and queue discharge flow become stable. It is also confirmed that vehicles uniformly start acceleration from the sag.

The results are found to be independent on the behavior of initial vehicle. Fig. 8 shows a comparison of the speed profiles of 500th followers. Three profiles, although they may not be distinguished, are simulated under target speed of the initial vehicle is either 60, 90, or 120 km/hr.



**Fig. 8.** Comparison of the speed profiles of 500th followers.

## 4 Application of ITS

Since the bottleneck phenomena at homogeneous roadway section are related with human factors, services of intelligent transport systems (ITS) may assist to alleviate the problems. One of the possibilities would be assistance of drivers' operation by intelligent cruise-assist vehicles.

Cruise-assist vehicles can maintain target speed or target time headway to keep flow rate at car-following operation by precise measurements at real time basis for less reaction time lags. It may also be achieved by inter-vehicle communication if information of lead vehicles is provided to followers at real time. It would attain a good support for human limitation on perception and operation [3].

### 4.1 Simulation Study -2

This section presents a study on the introduction of cruise assist vehicles to alleviate sag bottleneck phenomena [6].

Simulations at a sag section are tried with random participation of cruise assist vehicles among unequipped normal vehicles by human operation.

Cruise-assist vehicles are assumed to try to follow the leader by 1.6-second target headway. The simulations aim to study the possible occurrence of the congestion. The arrival vehicles form platoons of 80 vehicles that enter the sag section with constant speed of 80 km/hr.

Table 2 is the summary of input conditions of the simulations.

**Table 2.** Input conditions of simulation study -2.

Items	Input conditions
Roadway	1-lane with no overtaking
A sag	2.3% gradient increase 500m gradient curve
Arrival flow	2000 vph average, 80 km/hr 25 sets of platoon (80 vehicles) + gap
Speed limit	80 km/hr
CA vehicles	target headway = 1.6 sec target speed = 80 km/hr penetration rate = 0,20,40,60,80,100%
Normal vehicles	the elaborated car-following model
Compensation lag	$T_{grad} = 5$ seconds

### 4.2 Simulation Results -2

The trials of simulations resulted that a sag bottleneck still appears under low penetration rate of cruise-assist vehicles. However, total breakdown of flow is alleviated by the growth of penetration rate. Figs. 9(a) and (b) show the speed profiles of a vehicle along the way. With low penetration rate of 20%, there still per-

sists capacity bottleneck at the sag, however with higher penetration rate of 40%, only the slight reduction of speed is observed.

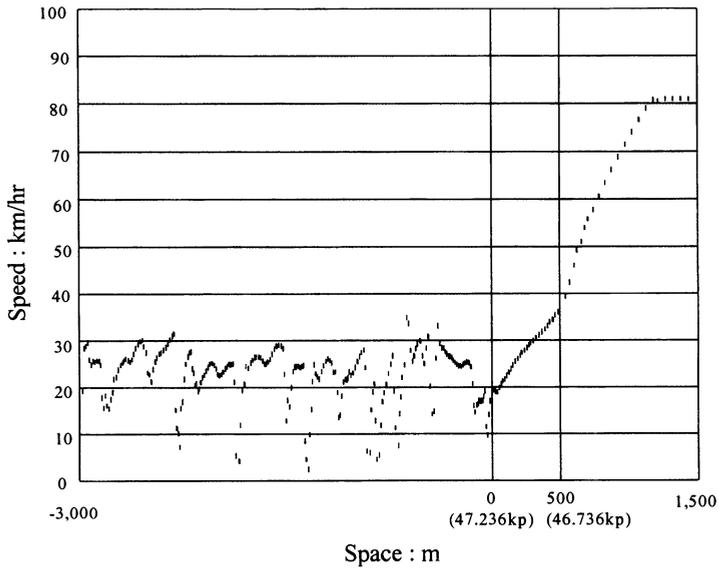


Fig. 9(a). Speed profile of a sampled vehicle with penetration rate of 20%.

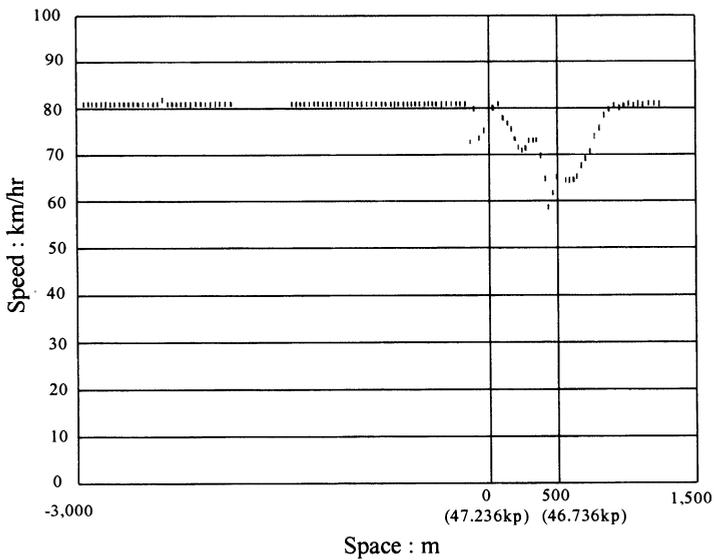


Fig. 9(b). Speed profile of a sampled vehicle with penetration rate of 40%.

## 5 Conclusion

In this paper, purposes and some example applications on the modeling of road traffic flow is reported relating to the capacity bottleneck phenomena. It is concluded as follows:

1) Road traffic engineering needs usable models with several requirements and purposes.

2) Congested flow is generated by reasons at downstream bottlenecks where demand exceeds capacity.

3) Sags or tunnels can be exceptional bottlenecks on homogeneous roadway sections, where human factors have effects.

4) Bottleneck phenomena at sags are regarded due to the misperception & delayed reaction to vertical gradient increase, and are simulated by car-following models with additional deceleration of delay on gradient compensation.

5) Medium rate of penetration of cruise-assist vehicles hopefully has good possibility to alleviate sag bottlenecks

Further research on traffic flow should be explored both from theoretical and practical point of views.

## References

1. D.C. Gazis, R. Herman, and R.W. Rothery, Nonlinear follow-the-leader models of traffic flow, *Oper. Res.* **9**(4), 545-567 (1961).
2. M. Koshi, M. Kuwahara, and H. Akahane, Capacity of sags and tunnels on Japanese motorways, *ITE Journal* **62**(5), 1-22 (1992).
3. H. Ozaki, Reaction and anticipation in the car-following behavior, in: C.F. Daganzo (Ed.), *Transportation and Traffic Theory*, pp 349-366, (Elsevier, 1993).
4. H. Ozaki, Assistance of drivers to mitigate highway capacity problems, in: *Proc. of the Second World Congress on Intelligent Transport Systems*, Yokohama, Vertis, IV, pp. 1727-1731 (1995).
5. J. Xing, A study on bottleneck phenomenon and car-following behaviour on motorways, (Ph. D. Thesis, University of Tokyo, 1992).
6. T. Yokota, M. Kuwahara, and H. Ozaki, A study of AHS effects on traffic flow at bottlenecks, in: *Proc. of the Fifth World Congress on Intelligent Transport Systems*, Seoul, Vertis, (CD-ROM, 1998).

# ITS and the Revolution of Automobiles

Y. Noguchi

Toyota Motor Corporation

**Abstract.** In this paper we discuss, how cars, roads and information are going to change oneself and consequently, what kind of revolutions the society would receive. Further we show how Toyota is going to cope now and in the future with the rapidly changing world.

## 1 General Discussion

First, the market size related to Intelligent Transport Systems (ITS) in Japan is shown in Fig. 1. As shown in the bar graph, the service market is expected to expand substantially from year 2000. The market will double its size every 5 years, and in 2005, 330 thousands new jobs are supposed to be created, and furthermore in 2015, 1 million 70 thousands new jobs will be offered.

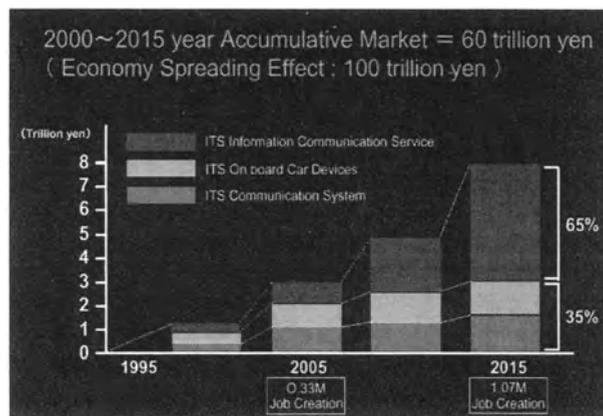


Fig. 1. Market Size of ITS Related Information; Communication Area of Japan.

Notable fact is that the market size of services surpasses the size of the system market and the in-vehicle devices market. As a whole, the accumulated market size until 2015 will be 60 trillion yen. And as much as 100 trillion yen economy ripple effect is expected.

The first stage of automobile performance history was the improvement of the fundamental performances such as driving, steering and stopping. And it progressed in the improvement of comfort. Recently, emphasis is put on safety and

environment aspects. And hereafter, the age comes when information performance of automobile is required. In other words, the car changes its role from the status symbol to an indispensable moving space. Eventually, the day will come when everything we do at home can be done in the car.

ITS (Intelligent Transport System) is the next generation of transport systems as a whole. It is expected to solve the current transport problems brought by automobiles. The cars and roads will be equipped with information and communication systems. ITS is expected to solve problems like traffic accidents, traffic jams and environmental pollution with the help of intelligent cars and intelligent roads.

The following three aspects are considered to be the main features of ITS.

First is about the social and public aspect, the enhance of convenience and comfort by solving the various problems caused by automobiles.

Second is about the product aspect, providing more competitive vehicles with higher functionality.

Third is about the new business aspect, the exploration of new businesses by developing systems of high-level usage.

In the 21<sup>st</sup> century information performance in addition to safety, environmental friendliness and driving performance, is very important. Information performance will not be solved or improved by the car itself. The concerted relationship between cars and road infrastructure is essential.

Figure 2 shows 5 major areas where Toyota is involved in the ITS business.

The first field is Car Intelligence (center of the chart). Here intelligence is given to the car itself.

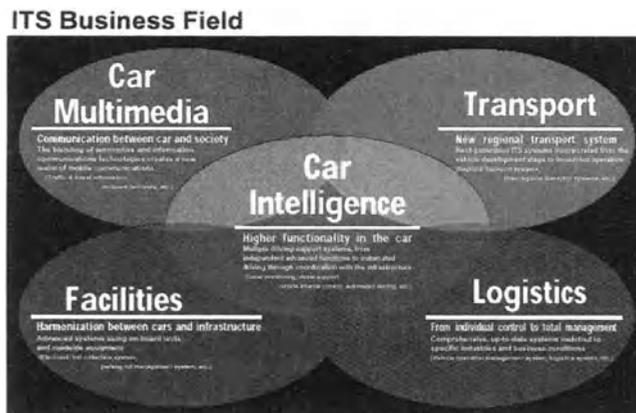


Fig. 2.

The second field is Car Multi-media (upper left of the chart), showing the business development based on car navigation systems.

The third area is about Facilities (lower left part), dealing with ETC and its expanded services.

The fourth field is about Logistics (lower right part), aiming to a more effective transportation of goods and vehicle operation management.

The last part is about Transport (upper right part), dealing with new regional public transport systems.

## 2 The Car Intelligence

The development steps giving more functions to the car are shown in Fig. 3. The vertical axis shows the level of traffic management while the horizontal axis shows the level of social infrastructure. Higher level of traffic management and higher level of social infrastructure will make the car itself more intelligent or lead to autonomous and more intellectually powered car. In other words, the car itself changes from the current status to a car that gives information to the driver, assists the driver, and guides itself depending on the surrounding traffic condition.

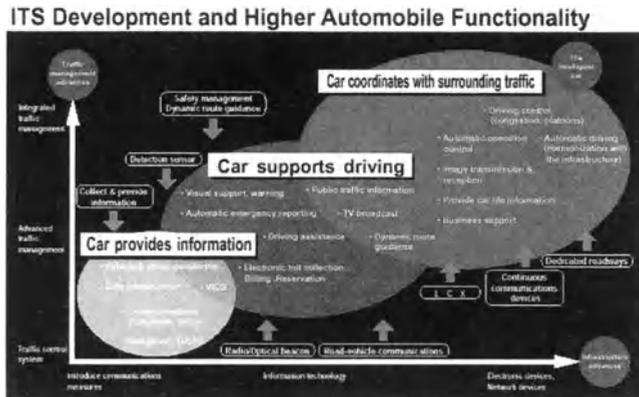


Fig. 3.

One of the most notable products of Car Intelligence that was already introduced in the market is Radar Cruise Control launched in 1996. When a car that is running with preset speed (determined by cruise control), is going to overtake a slower moving car, the cruise control unit calculates the distance between the cars by radar and decelerates the car speed to follow the car in front with constant distance. If the car in front leaves the driving lane of the car, the following vehicle accelerates to the original speed. This is called a clever intelligent auto cruise control. It is well accepted by the customers and more than 50,000 units were sold up to now.

The product called “Back Guide Monitor” was launched last January. This system not only shows the backward view, but also tells the driver how to steer trying to park the car. It is very convenient to those who are not good at parking cars.

### 3 Car Multimedia

The car navigation sales recently increased dramatically in Japan. Last year, alone 1 million 950 thousand units were sold, and accumulated sales have reached 7.8 million pieces. This is the world's highest penetration rate.

The accumulated sold number of Vehicle Information and Communication Systems (VICS) that provides road and traffic information to the car navigation systems, reached 2.8 million units. Now each car navigation out of two is equipped with VICS. It has become very popular and familiar.

The system image of Monet (Mobile Network) is shown in Fig. 4. Since 1997 various information for car drivers have been provided on demand and in real time. Latest information is provided responding to the user's request on the monitor screen or by voice through a digital mobile phone.

"Live view" of the roads, "weather forecast", "restaurant guide", "e-mail" and "parking guide" are typical current services of Monet. Internet is also accessible. The basic contents were substantially enriched compared with the original one at the beginning.

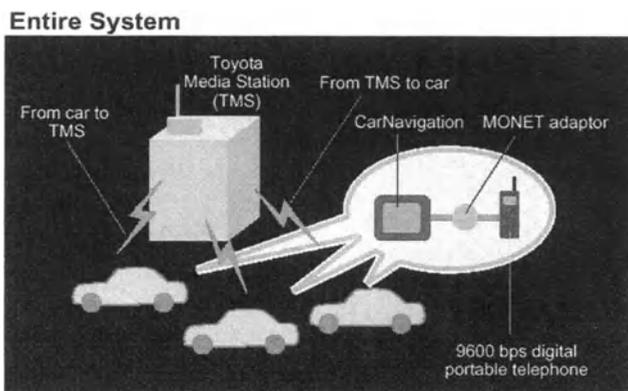


Fig. 4. Monet (Mobile Network).

Emergency Reporting Systems or so called Mayday Systems provide information like the location of the car or the driver ID in case of the explosion of the air-bag or by pressing a button automatically to the rescue center that then informs the police or fire department. This is done immediately after traffic accidents or cases like that.

In several overseas markets, this kind of service is already in operation by organizations. In Japan, related governmental agencies and private companies are preparing to introduce similar services, and the Japan Mayday Service Co. Ltd was established and launched its service last September. Toyota participates in this company as a major equity owner.

## 4 Facilities

In this field, the infrastructure plays a major role. ETC or Electronics Toll Collection used on the expressways is a typical example. Fig. 5 shows the image of ETC systems used on highways. Automatic toll collection of the traffic fee is executed by communication between tollgates and vehicles. Actually more than 35% of the high way congestions take place at the toll plazas. The system is expected to reduce traffic jam, exhaust emission and other pollution around the tollgates substantially.

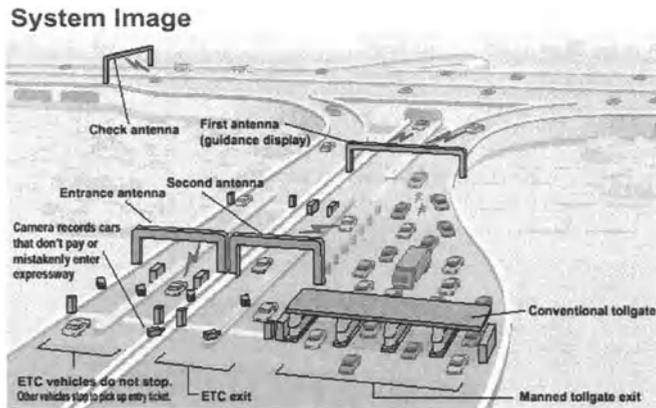
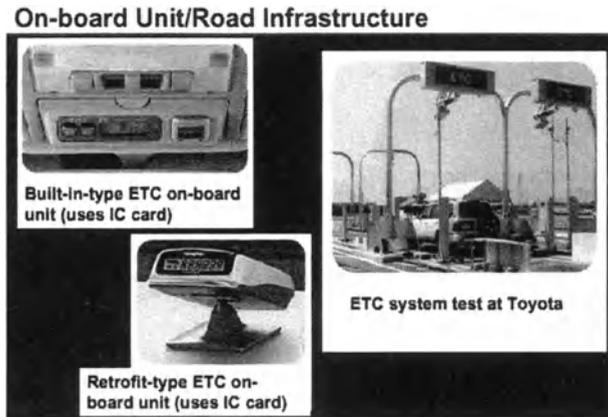


Fig. 5. ETC (Electronic Toll Collection System).

ETC is a rather new system in Japan. The system has just started. But some 20 countries mainly in Europe and North America already use this system. Also in Asian countries as Malaysia, China and Singapore have already started the service. But the norms of the system have not been standardized yet. For example, there are many frequencies for the communication and furthermore there are active and passive methods even at the same frequency. Besides, there are tag method and IC card method as onboard equipment as well. Consequently, it is required to install specific onboard equipment for each to pass the different gates smoothly. A unification of standards for a common usage has now become the urgent task that has to be solved. ISO is investigating the standardization of the norms, and it is expected that 5.8 GHz frequency would be the international standard.

The introduction of ETC in Japan will proceed and cover 75 percent of the traffic on the highway in 2003. The time will come soon when almost all toll plazas will be equipped with ETC.

Toyota's on-board units and roadside systems are shown in Fig. 6. Two types of on-board units are provided, a built-in-type and a retrofit-type. The infrastructure has been developed as well, resulting in total systems with high performance and high reliability.



**Fig. 6.** ETC (Electronic Toll Collection System).

## 5 Logistics

Transportation efficiency of commercial vehicle is decreasing year by year while the amount of traffic jams is increasing. And consequently, transportation costs become a big burden for the society. On the other hand the demand or volume of transportation is expected to grow continuously. So, it is considered as one of the world's urgent problems to be solved. To find a revolutionary solution is considered to be our responsibility.

Our taxi operation control system is shown in Fig. 7. It combines communication and car navigation. It is named TIME-t, which stands for Toyota Intelligent Mobility Enhancement for taxi, and was launched 3 years ago.

At the taxi dispatch center, the system perceives the location and operation condition of each vehicle and delivers the most appropriate car to the customer, or the center leads the vehicle to the most appropriate road so that the customer will not have to wait too long. Consequently, it provides merits to customers, taxi companies and taxi drivers.

The followings effects occur after introducing our system at a Taxi Cab Company in Toyota city.

- Increase of efficiency of reception and cab delivery operations.
  - Time saving : 35%
  - Efficiency improvement : 2.5 times  
(35 deliveries to 90 deliveries per hour)
- Decrease of man power in the delivery center  
(17 persons to 9 persons)
- Increase of safety and customer service.
  - Substantial decrease of accidents, and decrease of waste of time.

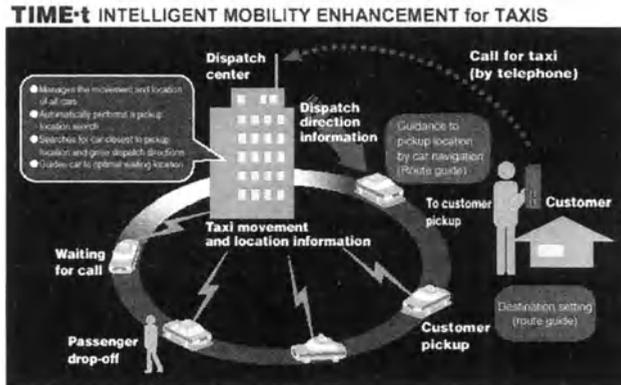


Fig. 7. High Performance Vehicle Motion Management System.

The system shown at Fig. 8 is called TIME-d (TIME for delivery). The purpose of this system is to increase the efficiency of the operation circle by utilizing communication. The operation circle includes all points between merchandise delivery and fee collection. This system was introduced to increase the efficiency of filling up merchandises in automatic venders 3 years ago.



Fig. 8. Intelligent Mobility Enhancement for Delivery.

By TIME-d, the driver is informed about the inventory condition of the vender machine by PHS, and all he has to do is filling up, when he reaches the automatic vender. Sales data are automatically sent and collected and further more even a production plan can be created.

The following effects raised by a local Coca Cola Bottler who is actually using the system.

- Improvement of operation : + 17%
- Increase of vender machine units per driver : + 27%
- Increase of sales by better stock management.

Another similar system called TIME-w is provided by Toyota. Here “w” means welfare. It connects welfare facilities and delivery vehicles by cellular communi-

cation, and makes it possible to send and return persons to and from facilities smoothly, safely and comfortably.

At the facility, operators can locate the vehicle and receive information about the condition of the persons in the vehicle at real time. So they are able to instruct immediate changes of routes or give appropriate treatments according to condition changes of the persons. The system is actually being used at a welfare facility for older people in Toyota city.

TIME-b is also another similar system. Here “b” stands for bus. To support the community buses, this intelligent bus location system was developed and now is in operation. At the bus stop, the customer can easily see where the next bus is and how long it takes until its arrival. Moreover the system responds to the customer’s phone inquiry at the bus stop and tells him the information by synthetic voice, so that he does not have to wait long even in hot summer, cold winter or in heavy rain.

Figure 9 shows a system distributing the built cars. Communication devices are positioned in the carrier trucks. The operation control center and the distribution center are networked. By this system, dynamic control of both, carrier trucks and loaded vehicles is realized. In this way, the system provides efficient, high level distribution. It is only an example. Toyota provides similar systems to various industries and businesses, and contributes in an efficient, sophisticated distribution.

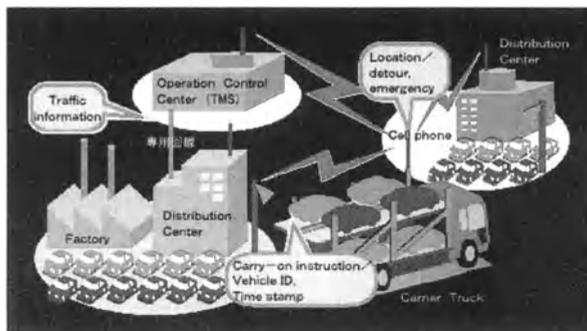


Fig. 9. Distribution System.

## 6 Transport

New transport systems by utilizing ITS have been developed. Today many new transport systems that should be suitable for the next generation are investigated actively in the world. “Intelligent Multi-mode Transit System (IMTS)” and “EV Commuter Shared Use System” have been developed as new types of public transport systems for the purpose of improving the entire transport systems.

The system image of IMTS is shown in Fig. 10. It includes the merits of trains and buses. It is designed for economic and flexible public transport systems for

medium range and medium volume traffic. In this system, the vehicles automatically drive in platoons on the dedicated road in the main-lane like trains on the rail, while on ordinary roads the driver manually drives each vehicle. It is a completely new concept traffic system, used in two ways.

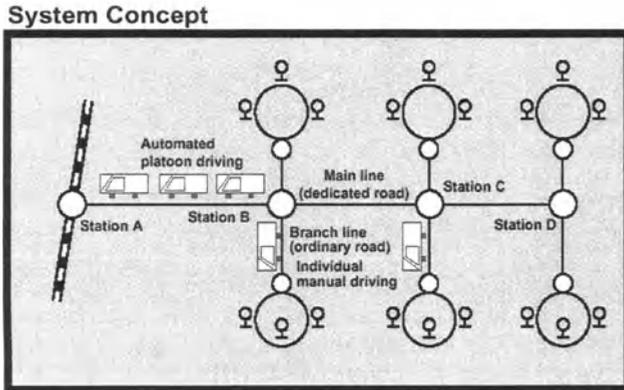


Fig. 10. Intelligent Multimode Transit System (MTS).

Wide range usage such as access to the airports or other towns and inter city transport in rural areas is expected to be effective. For example, it is considered useful by traffic measures in the area of the Expo 2005 that will be held in the Aich Prefecture.

IMTS has a cost advantage over other new traffic systems. Construction costs of other new traffic systems are said to be 4 to 5 billion yen per kilometer. On the other hand, those of IMTS are estimated to be 1 to 2 billion yen depending on the situation. In addition, ordinary mass production buses can be used for IMTS. So, concerning the vehicle itself, IMTS is half to one third less expensive than the usage of trains. Also, by changing the unit number of convoys in busy periods and idling periods such as in the morning or evening, IMTS can save operation cost without reducing the operation frequency and convenience.

The vehicle goes along the road lane automatically, sensing the magnetic nails embedded in the road. Driving information of the vehicle ahead is transmitted to the following vehicles by vehicle-to-vehicle communication technologies. Also, relative speed and distance between the vehicles are calculated. By these systems, the entire convoy goes smoothly without clashing. Stop, departure or separating the vehicles in a convoy is operated by the direction of the control center and communication between the road and the vehicles.

Various experiments have been conducted in our R&D facilities in Higashi-Fuji since 1998. IMTS proceeded to actual use in April 2001. In Awaji Island IMTS is now in operation. Like these examples, IMTS has been developed based on Toyota's technologies that have been acquired in the course of pursuing the development of the intelligent car.

Figure 11 shows the shared personal commuter system called "Crayon". It utilizes a small electric vehicle for a personal commuter. The operation experiment

started using two passengers EV called “e-com” which was exhibited at the Tokyo Motor Show in 1997. The system was introduced to leave free cars in the specific area such as sight seeing places, business zones and so on. Also, a shared use of e-com has been introduced in a suburban area of Tokyo as a project of the Ministry of Industry and Trade (currently Ministry of Economy, Trade and Industry) since January 2001.

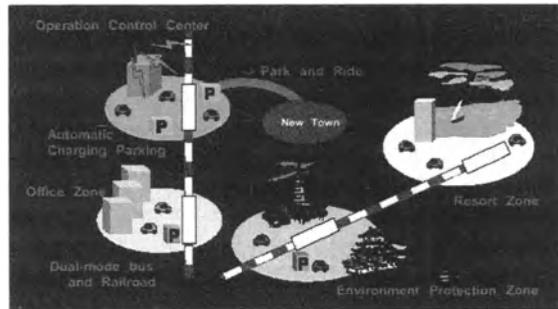


Fig. 11. Crayon – Image of Use.

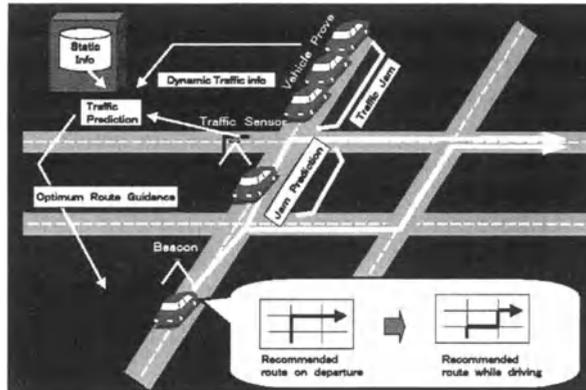
## 7 Mitigation of Traffic Congestion

As referred earlier, the efficiency of transportation is decreasing substantially because of traffic congestions and transportation costs have become a big burden to society. A Solution for the problem especially in the urban area is an urgent matter. In order to contribute much more to the mitigation of traffic congestions, the three key technologies, which have been developed by Toyota, should be pointed out.

The first one is the technology of traffic forecast. This means, more informative information for the drivers;

- 1) Will the congestion get worse or better?
- 2) How long will it take to get through the congestion?

In order to accomplish these requirements, it is important to improve static and dynamic information and develop a simulation technology based on these information (Fig. 12). To do this, a closer public/private cooperation and deregulation such as openness for various information are indispensable.



**Fig. 12.** Traffic Prediction Method.

The second one is the so-called network navigation. This concept consists of the optimum route guidance executed by infrastructure side such as an information service provider. Such a provider always has to have updated map database, many static traffic related information, and gather plenty of dynamic information from various sources. Then a forecast of traffic phenomena is possible. This must result in better solutions for flow disperse and congestion mitigation. The private sector should devote the technical development and the public sector should make information to be open to society.

The third one is TDM (Traffic Demand Management) including capable information. For the better traffic flow dispersion, not only space but also the time-axis consideration is needed. To do this, a seamless information service at home, at the office and at kiosks as well as in the automobiles is essential. And obviously its information should be accurate, fresh and reliable. It makes traffic demand shift on the time axis, and helps inter-modal transportation as a whole.

In addition, it is considered that the application of the above mentioned technology for commercial vehicles is much more effective to improve traffic, because the logistics of commercial vehicles is one of the causes for urban traffic congestion, while the demand of the goods transportation will surely increase. In addition in the urban area in particular, the concept "inter-modality" would be the best candidate to provide smart solutions of traffic related problems.

## 8 Effects Brought by ITS

Various efforts which were referred earlier and positive attitude of both, government offices and private sectors for the development of ITS will have fruitful result in 2010. The achievement of the projects shown in Fig. 13 and ITS will spread widely in the society and furthermore will change society substantially.

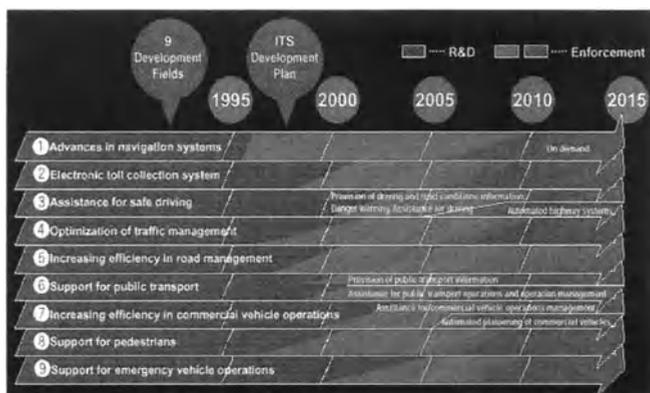


Fig. 13. Comprehensive Plan for the Promotion of ITS.

What will ITS bring to the society? Four major areas, according to the VERTIS which is now ITS-Japan and if various themes of ITS penetrate and spread in the society, the following situations will be realized.

First, in the area of safety: Mortal traffic accidents will be reduced to half of the current level in 30 years.

Second, in the area of smooth traffic: Traffic jams will be reduced to one fifth of current level in 20 years.

Third, in the area of environment: Fuel economy of vehicles will be improved by 15 percent, CO<sub>2</sub> emission will be reduced by 15 percent, and also NO<sub>x</sub> in the city zone is expected to be reduced by about 30 percent in 30 years.

Fourth in the area of new business: It is expected that new industries will create about 60 trillion yen in the next 20 years.

## 9 Closing

Although there are many problems to be solved, the 21<sup>st</sup> century is believed to be the one of dreams where the problems are solved one by one. These problems, international or domestic, government matters or private matters should be solved for the purpose of realizing a truly affluent traffic society.

# Modelling the Impact of ACC-Systems on the Traffic Flow at Macroscopic Modelling Level

C. Demir

DaimlerChrysler AG, Calwer Straße, 71059 Sindelfingen, Germany<sup>1</sup>

**Abstract.** Within the area of transport telematics, systems like adaptive cruise control systems (ACC-Systems), early warning systems, etc. are developed or improved. Recently ACC-Systems are launched into the market, so the penetration rate will increase in the future. ACC-Systems were researched very detailed by using microscopic models or by field tests with a limited number of equipped vehicles. The expected impact of this systems on the traffic flow was shown by microscopic analysis for different penetration rates. Because macroscopic models are able to simulate the traffic within real time, these models are used within model based traffic management systems instead of microscopic models. But macroscopic models cannot model the effect of systems with vehicles using adaptive cruise control, which have an direct effect on microscopic variables. What we are missing now are methods for modelling the impacts of such systems on the traffic flow on the level of macroscopic modelling to be able to consider these impact within traffic management systems. In this contribution an existing macroscopic traffic flow model will be extended with suitable macroscopic terms to be able to consider the impact of these systems on the traffic flow.

## 1 Introduction

Adaptive Cruise Control (ACC) concepts have been widely developed during the past decade and are launched into the market by only a few car manufacturers. The main objective of these concepts are the enhancement of driving comfort. That means from the view of the driver, a ACC-System has to keep automatically a desired distance to the car in front. To give a safety feeling to the driver the distance should depend on the speed. To reach this aim the driver has to give a desired time headway  $\delta_0$ , which will be used to calculate a speed dependent distance to the car in front.

A ACC-Radar device measures the distance to the car in front and calculates the time headway  $\delta = dx v^{-1}$ . Then the ACC-System adapts by acceleration or deceleration the determined time headway  $\delta$  to the given time headway  $\delta_0$ . But maximum acceleration and deceleration are, because of comfort aspects, limited to  $+2 \text{ ms}^{-2}$  and  $-3 \text{ ms}^{-2}$ . Therefore the ACC-System is at current time only useable for

---

<sup>1</sup> The contribution was done at Technical University Hamburg-Harburg, Denicke Str. 17, 21073 Hamburg

free motorway traffic with  $v > 40 \text{ kmh}^{-1}$ . Because detailed analysis of the ACC-System is not intention of this paper the calculation of relative speed and relative distance to the car in front etc. are not given in detail.

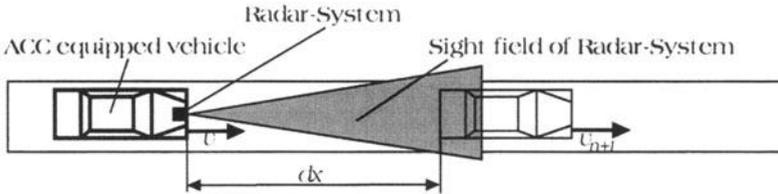


Fig. 1.1. Concept of the ACC-System.

The ACC-System harmonises the distance between vehicles and as an indirect effect the safety within the traffic is increases. But harmonising the distances between the vehicles means also the changing of driver behaviour within the traffic flow. It means further that traffic flow will be influenced if several vehicles are equipped with such a system. That means in another words, that the ACC-System has a direct impact on the driver's behaviour such as the driving distance, the acceleration,... (and other microscopic indicators) and each change in the driver's behaviour or in the microscopic indicators causes changes in the behaviour of the vehicle collectives such as the traffic density, the traffic volume,... (and other macroscopic indicators).

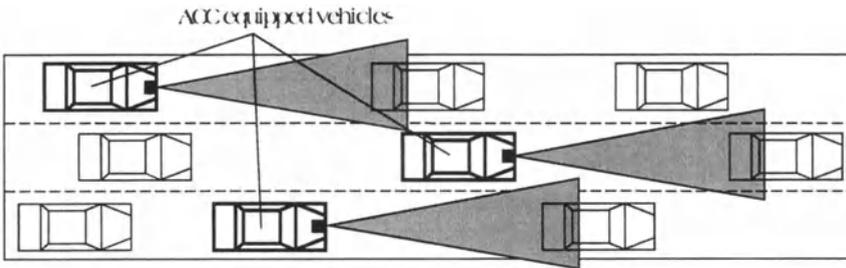


Fig. 1.2. Influencing the traffic flow by several ACC equipped vehicles.

The two parameters  $\delta_0$  (the given time headway) and  $\varepsilon$  (penetration rate: percentage of equipped vehicles within the whole traffic) reflect the impact of the ACC-System on the traffic flow.

In previous studies with the aid of different traffic simulation models, the effects of the ACC on driving behaviour, acceptance and safety were studied detailed. In these studies also the effects on the traffic flow at high penetration rates were researched detailed by simulations. However less attention has been paid considering the effects of these systems within future traffic management systems for the case of increasing penetration rates.

Because macroscopic models can simulate the traffic flow within real time, these systems are used within model based traffic management systems like in

model based dynamic speed limitation systems on motorways. Microscopic models are not applicable within such on-line control systems, because they require high computing resources and the computing speed depends on the number of vehicles within the simulation network. Because there exists no macroscopic model, which can simulate the impact of ACC-Systems on the traffic flow, the goal of this contribution is to extract the impact of ACC-Systems on the traffic flow and to describe the impact at the macroscopic modelling level.

The adaptive cruise control systems however cannot be modelled directly on the level of macroscopic modelling, because these systems have a direct effect on microscopic variables like distance behaviour, vehicle speed, vehicle acceleration / deceleration, which cannot be modelled direct within macroscopic models. But changes within microscopic variables causes changes in the fundamental diagram, which is the base of most macroscopic models. The caused changes in the fundamental diagram were shown in previous studies.

In this contribution an existing macroscopic traffic flow model (SIMONE – Simulation of MOtorway NEtworks) will be extended with suitable macroscopic terms to be able to consider the impact of these systems on the traffic flow on macroscopic level. This model is used currently within traffic management systems at Rhein-Main-Area and at the northern of Munich in Germany.

Because data are needed to analyse the macroscopic impacts of ACC-Systems and because there is no possibility to get data with high penetration rates from field sites, a microscopic model (MIKROSIM) was used to generate such data by microscopic simulations. Because the used microscopic model has to provide the possibility of realistic (detailed) reproduction of acceleration / deceleration process, a Wiedemann based microscopic model is used here.

## 2 Used Traffic Flow Models

### 2.1 Microscopic Traffic Flow Model (MIKROSIM)

MIKROSIM was developed at the department of automation and control theory at the University Hamburg-Harburg for simulating inner-urban street networks. The model is based on the linear car following model of Wiedemann. The model is based on the assumptions, that

- the process of traffic is given as a sequence of situations,
- each driver in a given situation has several action alternatives, which can create different subsequent situations,
- each driver selects under considering his intention an alternative course of action,
- The model of Wiedemann is limited to single-lane motorways without track switching possibility. This model was extended by lane changing possibilities for roadways with n-lanes, so that traffic can be modelled more realistically.

## 2.2 Macroscopic Traffic Flow Model (SIMONE)

SIMONE was developed at the department of automation and control theory at the University Hamburg-Harburg for fast simulations of motorway networks. The model is based on the non-linear model of Payne. On the basis of the continuity equation of traffic, Cremer developed an empirical relationship for the temporal course of the average speed on motorways (2).

For modelling a motorway, the selected motorway section is divided into several sections, the so-called segments, with a length of several hundred meters. Within these segments traffic conditions are regarded as homogeneous. With the aid of the average traffic density, average speed and traffic volume the traffic on the respective segment is characterised. For the characterisation of the traffic on the total regarded motorway section, these variables are evaluated for all segments within this motorway section.

For the description of the temporal course of traffic in a segment two dynamic relations in time-discrete form are used. The first one is the continuity equation of traffic

$$c_i(n+1) = c_i(k) + \frac{T}{\Delta_i} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)]. \quad (1)$$

This equation means, that the number of vehicles in the regarded segment changes during the time interval  $kT \leq t \leq (k+1)T$  only with the number of on-coming and off going vehicles from the ramps ( $T$  is the simulation time step and  $k$  is the index for the current time step). This is the continuity equation in time-discrete form.

The second dynamic equation is an empirical relationship for the temporal course of the average speed

$$v_i(k+1) = v_i(k) + \frac{T}{\tau} [V(c_i) - v_i]_{(k)} + \frac{T}{\Delta_i} [v_i(v_{i-1} - v_i)]_{(k)} + \frac{\nu T}{\Delta_i \tau} \left[ \frac{c_i - c_{i+1}}{c_i + \kappa} \right]_{(k)} \quad (2)$$

in the segment  $i$  with the constants

- $\tau$  Relaxation constant [s] indicates the time, which a driver collective requires, in order to adapt its driving behaviour to the surrounding traffic density,
- $\nu$  Sensitivity factor [veh.<sup>2</sup>/km] indicates, the sensitivity of the model to react to the foreseen traffic density,
- $\kappa$  Reducing factor [veh./km] reduces the strong influence of too low values of traffic density,
- $V(c_i)$  Speed-density relationship.

The traffic volume is calculated as a weighted average value at the end of the segment by using of the average density and average speed in the neighboured segments

$$q_i(k) = \alpha v_i(k) c_i(k) + (1 - \alpha) v_{i+1}(k) c_{i+1}(k), \quad (3)$$

whereby  $\alpha$  ( $0 \leq \alpha \leq 1$ ) is a further model parameter.

In nearly all macroscopic models, so also in SIMONE, the assumption is made, that in the stationary state of traffic the average speed is a function of the density – speed-density characteristic or VC-characteristic:

$$V(c) = v_f u_2 \left[ 1 - \left( \frac{c}{c_{max}} \right)^{l(3-2u_2)} \right]^m \tag{4}$$

with the model parameters

- $v_f$  average free speed at free traffic flow, at  $c \approx 0$ ,
- $c_{max}$  maximal average traffic density at congestion,
- $l, m$  Parameters, which restrict the form of VC-characteristic,
- $u_2$  Restriction of the free average speed at speed limitations.

In the literature different forms of the VC-characteristics or the fundamental diagrams are known. Thereby it is differentiated between single and multi-phase diagrams. The traffic simulation model SIMONE exhibits a single-phase fundamental diagram.

### 3 Macroscopic Modelling of the Impact of ACC-Concepts

#### 3.1 Limits of Analytical Approaches

For this the following differential equation system should be considered (equation of the ACC-System)

$$m_{car} \frac{d}{dt} v(t) = K_2 \frac{d}{dt} dx(t) - K_1 dx(t) + K_1 v(t) \mathcal{G}_0, \tag{5}$$

$$\frac{d}{dt} dx(t) = v_{n+1}(t) - v(t), \tag{6}$$

- with  $m_{car}$  mass of vehicle,
- $v(t)$  vehicle speed,
- $v_{n+1}(t)$  speed of vehicle in front,
- $K_1, K_2$  damping factors of ACC-System

for  $n$  vehicle with the conditions

- 100 % of vehicles are equipped with ACC-Systems
- speed range of ACC-System is  $[0, v_{max}]$
- there are no lane changing manoeuvres.

The stationary state (balance state for  $t \rightarrow \infty$ ) of the differential equation system is then given by (all derivatives are zero)

$$dx = v(t) \delta_0 \text{ and } v(t) = v_{n+1} \tag{7}$$

Traffic density is reverse proportional to the distance between the vehicles

$$c = \frac{1}{l_{car} + dx} = \frac{1}{l_{car} + v\vartheta_0} \tag{8}$$

Converting this equation to  $v$  it follows

$$v = \frac{1}{c\vartheta_0} - \frac{l_{car}}{\vartheta_0} \tag{9}$$

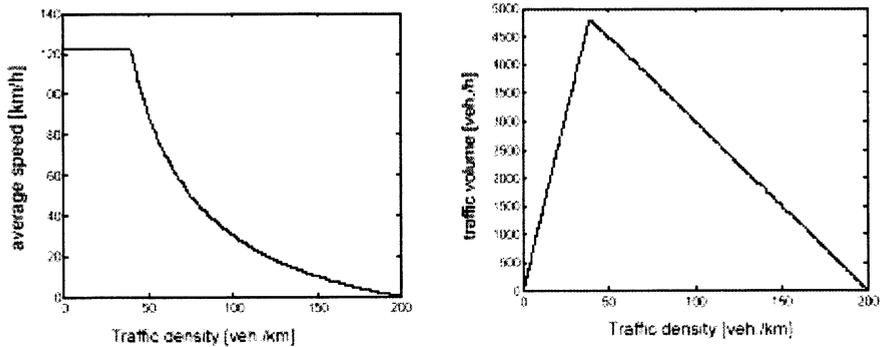
To avoid non realistic values for  $v$  at very low traffic densities the following limitation for  $v$  will be introduced

$$v = \begin{cases} v_f & \text{for } v > v_f \\ \frac{1}{c\vartheta_0} - \frac{l_{car}}{\vartheta_0} & \text{for } v \leq v_f \end{cases} \tag{10}$$

For the traffic volume the following equation is still valid

$$q = vc \tag{11}$$

Both, average speed and traffic volume are given in the following figures for a density range of [0, 200 veh./km] and for  $\delta_0 = 0.6$  sec. (for one lane).



**Fig. 3.1.** The impacts of ACC-Systems on the VC-diagramm and on the fundamental diagramm.

The analytical modelling approach is limited because

- it can only considered traffic scenarios with 100 % penetration rate
- lane change manoeuvres cannot be considered
- real system behaviour of ACC (only useable for  $[v_0, v_{max}]$  with  $v_0 \gg 0$ ) cannot be considered
- real traffic goes into congested state before reaching the stationary state of controller (ACC-System)

Because it is not possible to derive the impacts of ACC concepts on the traffic flow by using analytical methods, a simulation based modelling approach is used here.

### 3.2 Simulation Based Modelling Approach

Because analytical approaches are not suitable here, traffic data, including ACC impacts, are needed to analyze the impact of ACC on macroscopic variables (traffic density, average speed, traffic volume). But ACC was recently introduced, so there are no traffic data available, which includes high penetration rates, which is needed for impact analysis. Therefore the mentioned microscopic model is used as real traffic environment, to generate the required traffic data.

For this the algorithms (5) and (6) of the ACC-System are integrated into the microscopic model. Appropriate traffic scenarios are defined, which include transitions (free flow  $\rightarrow$  congested flow, congested flow  $\rightarrow$  free flow) and simulated with different combinations of  $(\delta_0, \epsilon)$ . Macroscopic variables/indicators will be established and the dependencies of the macroscopic indicators from  $\delta_0$  and  $\epsilon$  will be determined. The approach is shown in the following flow diagram.

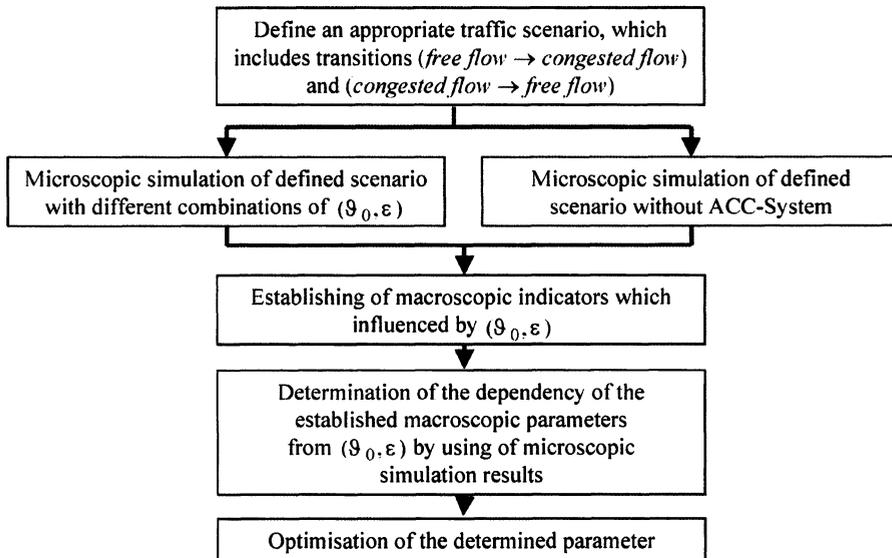


Fig. 3.2. Simulation based modelling approach.

### 3.3 Generating of Traffic Data with ACC-Systems by Microscopic Simulation

Because ACC systems were developed only for motorway traffic, these systems are switched off automatically for the inner-city street network. This happens in that way, that the systems are switched off at speeds less than 40 km/h. That means, that these systems are also not applicable within congestion's on motorways. In the case of switching off, the driver is warned optically or acoustically, in order to take over the driving manoeuvre.

In order to prevent that the ACC controller is switched on and off permanently, if the regarded vehicle oscillates around this velocity, here a hysteresis is introduced. For example the ACC controller is switched on, if the speed is larger than  $v_0+1$  [km/h] and switched off, if the speed is smaller than  $v_0-1$  [km/h].

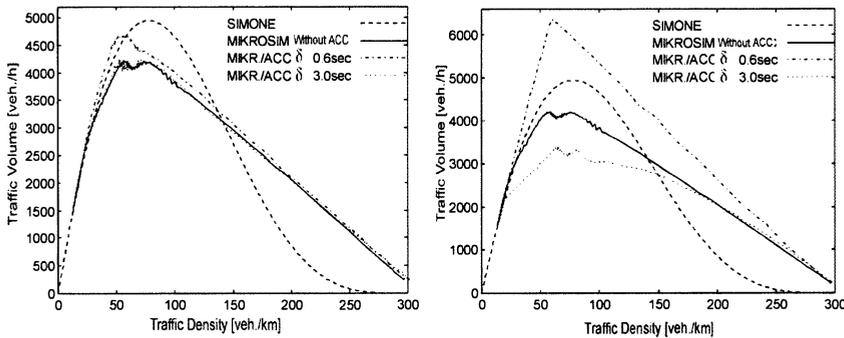
In the following, simulation results with different equipment degrees  $\varepsilon$  and time headways  $\delta_0$  will be generated, to get a wide base for the further analysis.

Penetration rate of vehicles $\varepsilon$ [%]	10	20	40	60	80	100		
Time headway's $\delta_0$ [sec.]	0,6	0,8	1,2	1,5	1,8	2,0	2,5	3,0

**Table 3.1.** ACC parameter for the simulation scenarios.

At the selection of the simulation scenarios great importance was attached to the dynamic processes (free flow, congested flow, transition from free to congested or from congested to free flow) within the traffic flow, so that the fundamental diagram could be generated as completely as possible.

In the Fig. 3.3 it is determined, that low equipment degrees do not have substantial effects on the dynamics of the traffic flow or on the relations between the macroscopic traffic variables, independent of the time headways of the ACC controller. At higher equipment degrees and small time headways the rate density characteristic is in so far modified that at higher traffic densities higher average speeds are possible. This becomes apparent in the fundamental diagram (Fig. 3.3) as an increase of the maximum traffic volume of the measurement point or as an increase of the road capacity. At larger time headways the average speeds are less than that at the scenario without ACC controller. This becomes again apparent in the fundamental diagram as a reduction of the road capacity (Fig. 3.3).



**Fig. 3.3.** Fundamental diagram at lower and higher penetration rates (10%, 100%).

The efficiency of the road is thus increased by the adjustment of smaller time headways or smaller following distances and reduced by the adjustment of larger time headways. The ACC controller caused increases or decreases of the road efficiency leads to the situation, that the congestions dissolve faster or more slowly than on normal conditions without ACC controllers.

The analysis of the microscopic simulations has given the following results:

- The ACC-System has no impact on the average speed within the free traffic flow
- The ACC-System has no impact on the maximum traffic density within congestions
- The ACC-System changes directly the average following distances between the vehicles and therefore indirectly the traffic volume
- The ACC-Parameters (time headway and penetration rate) cause not only a linear but also a non-linear change of the traffic density

The basis of modelling traffic flow on the macroscopic level is the fundamental diagram. This describes the collective driver behaviour of the individual driver-vehicle-units. That means, that each modification in the behaviour of the driver-vehicle-units becomes apparent in a modified fundamental diagram. Since the ACC systems imply a direct modification of the distance behaviour of the individual driver-vehicle-units, this can be considered macroscopically by using a new or an adapted fundamental diagram. The distance behaviour of the individual driver-vehicle-units is shaped thereby by the time headway and the modification degrees of the collective distance behaviour is shaped by the penetration rate of all vehicles. A very low penetration rate of (1%), e.g., compared with a penetration rate of (80%), would cause no substantial modification of the collective distance behaviour or also the fundamental diagram.

The VC characteristic is certainly described by the average traffic density, maximum traffic density within congestion and by the average speed for free flow. Microscopic investigations showed that the ACC systems have neither with low nor with high equipment degrees effects on the average speed for free flow vehicles. Also the maximum traffic density within a congestion is not influenced by using the ACC controller, because the system is switched of automatically in stop and go traffic and also within congestions. The following distances between the vehicles are reduced or increased by the application of the ACC controller, depending on the adjustment of the time headways. The average following distance of the vehicles is reverse proportional to the average traffic density, so that the traffic density for smaller time headways is increased and reduced for larger time headways. This effect is considered here by the introduction of the parameter  $\mu_1$ . The adjusted time headways cause in connection with the penetration rate of the ACC controller not only a linear, but also a non-linear modification of the traffic density. Experimental investigations showed that the non-linear correlation between the traffic density, maximum traffic density and the average speed can be described exponential. Therefore a second parameter  $\mu_2$  is introduced here into the analytical description of the fundamental diagram to describe the non-linear influence of the ACC controller on the quotient between traffic density and maximum traffic density.

$$V(c) = v_f \left[ 1 - \left( \frac{\mu_1 c}{c_{\max}} \right)^{\mu_2} \right]^m \quad (12)$$

### 3.4 Parameter Optimization

The parameters  $\mu_1$  and  $\mu_2$  are optimised for each ACC parameter set, which is given in the Tab. 3.1, with the aid of microscopic simulation results. As quality functional here the mean square error was selected

$$\mathfrak{S}(\mu_1, \mu_2) = \frac{1}{S} \frac{1}{N} \sum_{j=0}^{S-1} \sum_{k=0}^{K-1} ((v_j(k) - \hat{v}_j(k))^2 + \sigma(q_j(k) - \hat{q}_j(k))^2), \tag{13}$$

in order to reduce the deviations between the macroscopic simulation results and the real traffic data - produced with MIKROSIM - to a minimum. The parameter thereby means:

- $v_j(k)$  real measured (within microscopic model) local speeds at the  $k$ -th time step and  $j$ -th segment,
- $\hat{v}_j(k)$  within macroscopic model estimated local speeds at the  $k$ -th time step and  $j$ -th segment,
- $q_j(k)$  real measured (within microscopic model) local traffic volumes at the  $k$ -th time step and between  $j$ -th and  $(j+1)$ -th segment,
- $\hat{q}_j(k)$  within macroscopic model estimated local traffic volumes at the  $k$ -th time step and between  $j$ -th and  $(j+1)$ -th segment
- $\sigma$  Balancing factor between different error ranges of values for average speed and traffic volume,
- $K$  Number of simulation time steps,
- $S$  Number of segments within the considered road section.

The balance factor was determined by pre-simulations and analysis of real data for the parameter  $\nu$  and  $q$  as  $\sigma=0.0008$  by Cremer and is also used within the current optimisation.

After the optimisation of the parameters  $\mu_1$  and  $\mu_2$  for each individual ACC parameter set  $\varepsilon$  and  $\delta_\rho$ , these parameters can be represented as functions  $\mu_1 = f(\delta_\rho, \varepsilon)$  and  $\mu_2 = f(\delta_\rho, \varepsilon)$  of the ACC parameters.

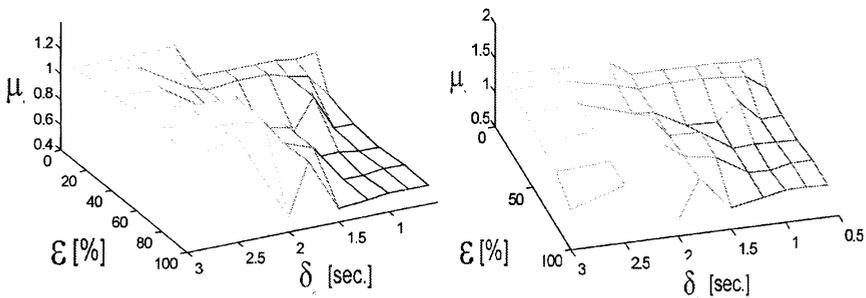


Fig. 3.4. 3 dimensional graph of the parameter sets of  $\mu_1$  and  $\mu_2$ .

The analytic form of these functions can be finally determined by two-dimensional interpolation and can be represented as follows

$$\mu_1(\vartheta, \varepsilon) = P_1(\vartheta, \varepsilon) = \sum_{k=0}^5 \sum_{i=0}^8 L_k^\varepsilon L_i^\vartheta \mu_1^{ki} \quad (14)$$

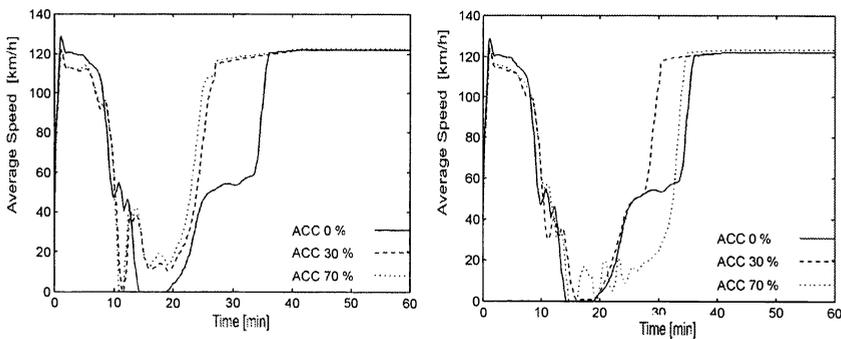
with  $\mu_1^{ki}$  from the Tab. 3.1 and

$$\mu_2(\vartheta, \varepsilon) = P_2(\vartheta, \varepsilon) = \sum_{k=0}^5 \sum_{i=0}^8 L_k^\varepsilon L_i^\vartheta \mu_2^{ki} \quad (15)$$

with  $\mu_2^{ki}$  from the Tab. 3.1.

### 3.5 Simulation Results

After the determined polynomials  $\mu_1(\delta_0, \varepsilon)$  and  $\mu_2(\delta_0, \varepsilon)$  were integrated into the macroscopic model SIMONE, simulations with the ACC parameter  $\delta_0=0.6\text{sec.}$ ,  $\delta_0=2.0\text{sec.}$ ,  $\varepsilon=30\%$  and  $\varepsilon=70\%$  were evaluated as an example scenario. The results are given as follows



**Fig. 3.5.** Simulation results for  $\delta_0=0.6\text{sec.}$  and  $\delta_0=2.0\text{sec.}$

At smaller time headways an improvement of the traffic conditions - faster dissolution of congestions - is already obtained with small to middle penetration rates. A further increase of the penetration rate does not give a substantial gain

At larger time headways, with low penetration rates, less improvement of the traffic conditions occurs. This can be explained by the fact that on the average the kept time headways - of the normal driver - are not substantially influenced by low penetration rates and the vehicles equipped with ACC controller have less influence on the harmonisation of traffic flow. With high penetration rates however the average value of the time headways increases, so that, due to the slower reaction of the vehicles among themselves the dissolving of traffic jams lasts longer.

## 4 Conclusion

ACC-Systems will come up more and more with new cars into the market, so the impact on the traffic flow will increase with increasing penetration rates (increasing number of ACC-equipped vehicles within the traffic). These impacts are analysed very detailed with different penetration rates by using microscopic models.

Because microscopic models (with detailed reproduction of the acceleration/deceleration process) are not appropriate to be used for model based on-line traffic control systems (e.g. variable speed adaptation on motorways), macroscopic models are used for such systems. But there exists no macroscopic model, which is able to model the impact of ACC-Systems on the traffic flow.

In this contribution an existing macroscopic model was extended to consider indirectly the impacts of ACC concepts on the macroscopic level of traffic flow modelling. The simulation results of the extended macroscopic model show that modelling the impact of ACC-Systems on the traffic flow could be done without great deviation to simulations at the microscopic level.

Simulation results also show, that lower time headways increase the traffic flow rate and higher time headways decreases the traffic flow rate. Until a certain penetration rate the traffic flow rate is increasing and higher penetration rates give no more benefit.

Optimisation capabilities exist, e.g., within the validation process of the extended model with data sets (including ACC-Systems) from field sites.

## References

1. S. Becker, M. Bork, H.T. Dorissen, G. Geduld, O. Hoffmann, K. Naab, G. Nöcker, P. Rieth, and J. Sonntag, Summary of Experience with Autonomous Intelligent Cruise Control (AICC), Part 1: Study Objectives and Methods. Proc. Of the 1<sup>st</sup> World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems, pp. 1828-1835 (1994).
2. R.E. Chandler, R. Herman, and E.W. Montroll, Traffic Dynamics: Studies in Car Following. Operation Research, pp. 165-184 (1958).
3. M. Cremer, Der Verkehrsfluß auf Schnellstraßen: Modelle, Überwachung und Regelung, Springer Verlag, Berlin, Heidelberg, New York, 1979.
4. C. Demir, M. Cremer, S. Donikian, and S. Espie, Investigating the Impact of AICC Concepts on Traffic Flow Quality. 5<sup>th</sup> World Congress on Intelligent Transport Systems, Seoul, 1998.
5. S. Donikian, S. Espie, M. Parent, and G. Rousseau, Simulation Studies on the Impact of AICC Studies. 5<sup>th</sup> World Congress on Intelligent Transport Systems, Seoul, 1998.
6. A. Hochstädter, Regelgesetze für die Führung von Fahrzeugen in Konvois und deren Erprobung in einer mikroskopischen Simulationsumgebung, Schriftenreihe der Automatisierungstechnik, Band 16, Doktorarbeit, Technische Universität Hamburg-Harburg, Arbeitsgruppe Automatisierungstechnik, 1998.

7. J.H. Hogema and W.H. Janssen, Effects of Intelligent Cruise Control on Driving Behaviour – A Simulator Study, 3<sup>rd</sup> World Congress on Intelligent Transport Systems, 1996.
8. S.L.S. Jakoby, J.S. Kowalik, and J.T. Pizzo, Iterative Methods for non-linear optimization problems, Englewood-Cliffs, New Jersey, Prentice Hall, 1972.
9. B.S. Kerner, The Physics of Traffic, Physics World, Volume 12, Bristol, 1999.
10. M. Koshi, M. Iwasaki, and I. Ohkura, Some findings and an overview on vehicular flow characteristics, Proc. 8<sup>th</sup> Int. Symp. On Transportation and Traffic Theory, ed. Hurdle, Hauer, Stewart, Toronto, 1981.
11. D. Lange and R. Struif, Fahrverhalten im Bereich von Streckenbeeinflussungsanlagen, Straßenverkehrstechnik 4/97, Bonn.
12. J. Ludmann and M. Weilkes, Investigation of Intelligent traffic systems by means of simulation, 4<sup>th</sup> World Congress on Intelligent Transport Systems, Berlin, 1997.
13. M. McDonald, G. Marsden, C. Demir, S. Gayda, J. Vatn, and S. Espie, Deployment of Inter-Urban ATT Test Scenarios (DIATS) – A European Progress Report. SAE Future Transportation Technology Conference, USA, 1998.
14. M. McDonald and G. Marsden, Deployment of Inter-Urban ATT Test Scenarios (DIATS) – Technical Progress Review. 5<sup>th</sup> World Congress on Intelligent Transport Systems, Seoul, 1998.
15. M. McDonald and J. Wu, The Integrated Impacts of Autonomous Intelligent Cruise Control on Motorway Traffic Flow, Proceedings of the ISC 97 Conference, Boston, USA, 1997.
16. K. Nagel and M. Schreckenberg, K. Physique I 2, 2221 (1992).
17. J.A. Nelder and R. Mead, A Simplex Method for Function Minimization. Computer Journal, 7, S. 308-313 (1964).
18. G. Sala and P. Fabio, Strategies for Intelligent Cruise Control (ICC) application, 3<sup>rd</sup> World Congress on Intelligent Transport Systems, 1996.
19. R. Sauer and Szabo, I+II+III+IV. Mathematische Hilfsmittel des Ingenieurs. Band 139-142, Springer Verlag, Berlin, Heidelberg, New York, 1968.
20. P. Spallucci and W. Törnig, Numerische Mathematik für Ingenieure und Physiker. Band 2, Numerische Methoden der Analysis. Springer Verlag, Berlin, Heidelberg, New York, 1990.
21. G. Steierwald and R. Lapierre, Verkehrsleittechnik für den Straßenverkehr, Band I, Grundlagen und Technologien der Verkehrsleittechnik. Springer Verlag, Berlin, Heidelberg, New York, 1987.
22. B. van Arem, J.H. Hogema, and S.A. Smulders, The Impact of Autonomous Intelligent Cruise Control on traffic flow, 3<sup>rd</sup> World Congress on Intelligent Transport Systems, 1996.
23. R. Wiedemann, Simulation des Straßenverkehrsflusses, Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, Heft 8, 1974.
24. T. Yokota, S. Ueda, and S. Murata, Evaluation of AHS (Automated Highway Systems) Effects on Mean Speed by Static Method. 5<sup>th</sup> World Congress on Intelligent Transport Systems, Seoul, 1998.
25. K. Yoshimoto and T. Kokubo, The effect of automatic driving on traffic flow, 4<sup>th</sup> World Congress on Intelligent Transport Systems, Berlin, 1997.
26. H. Zackor and S. Herkt, Verbesserung des Steuerungsmodells zur Wechselwegweisung auf Autobahnen in Rhein/Main-Gebiet. Berat. Ing. Steierwald/Schönharting, i.A. Herg. Landesamt für Straßenbau, Wiesbaden, 1980.

# A Cellular Automaton Model for Dynamic Route Choice Behavior in Urban Roads

M. Fukui, T. Yosida, and H. Oikawa

Nakanihon Automotive College, Sakahogi-cho, Gifu-ken 505-0077, Japan

**Abstract.** Traffic flow of cars making trips from an origin (O) to a destination (D) on two-dimensional two-way roads is studied using a cellular automaton model. The crossings are at every  $C_r$  site in the square lattice road and drivers turn at the crossing in order to arrive at their destination through the shortest route from the present position. As for driver, two kinds of drivers are defined. When they can't advance through the shortest route by the congestion on the road, (1) the dynamic driver immediately changes the advancing direction at the crossing and (2) the static driver does not change the direction to keep the shortest route, but finally changes the route after waiting for  $T_s$  time-steps. The traffic flow is simulated by the CA model of these road systems, and effects of the route-choice behavior of the drivers on the occurrence of a traffic jam are investigated. It has been found that the static drivers become the seeds for the global jam even if they are less and the supply of traffic information to the drivers can avoid the occurrence of the global jam by the encouragement of detours around the congestion.

## 1 Introduction

For traffic flow on a two-dimensional city road, Biham *et al.* [1] have proposed a one-lane CA model with two kinds of cars advancing perpendicularly to each other. Their model has a phase transition of the first order type between the free moving phase and the jam phase. Cesta *et al.* [2] have extended the model by the introduction of stochastic turns of the cars. Honda *et al.* [3] have proposed a bi-directional traffic model on two-dimensional two-lane city roads with four kinds of cars able to turn stochastically to three directions and found a sort of continuous transition. In the present study, we propose a deterministic traffic model with cars making a trip in the two-dimensional area of a two-lane road network and apply two kinds of drivers with different route-choice behavior, in order to investigate the effects of the route-choice and the supply of traffic information on the traffic flow.

## 2 Model

Two-way urban roads are expressed by a two-dimensional two-lane square lattice. The crossings are at every  $C_r$  sites in the  $W \times W$  square lattice. Two cars advancing to the opposite direction can be contained in a cell. A crossing

can contain two cars as well. The driver begins a trip from an origin (O), i.e. an arbitrary site in the two-dimensional lattice to a destination (D) which is  $\Delta Dx$  and  $\Delta Dy$  ( $-Wx \leq \Delta Dx \leq Wx$  and  $-Wy \leq \Delta Dy \leq Wy$ ) apart from the origin. The cars advance obeying the rule-184 CA model (Fig. 1). The drivers turn at a crossing in order to arrive at the destination through the shortest route from the present position. When deviations of the present position of the car from the destination are denoted as  $\Delta x$  and  $\Delta y$  for x- and y-directions, respectively, the turning rule of the car at the crossing is defined as follows.

When  $\Delta x > \Delta y$ , the driver chooses routes in order to decrease the deviations by the order of ①x-direction, ②y-direction, and ③-y-direction or ④stays there if he can't go ahead (Fig. 2). As for driver, two kinds of drivers are defined in the system;

(1) dynamic driver ( $Nd$ : number of the driver, waiting time at the crossing:  $Ts=0$ ) who immediately changes advancing direction at the crossing when he can't advance through the shortest route because of congested road.

(2) static driver ( $Ns$ : number of the driver, waiting time : $Ts>0$ ) who does not change the direction to keep the shortest route, but finally changes the route by order of ①-③ after waiting for  $Ts$  time-steps. The static driver behaves in the same manner as the dynamic driver when  $Ts = 0$ . On the roads, cars are driven by the dynamic driver or/and the static one. The ratio of the number of dynamic drivers to the total number of drivers is defined as  $Dr=Nd/(Nd+Ns)$ .

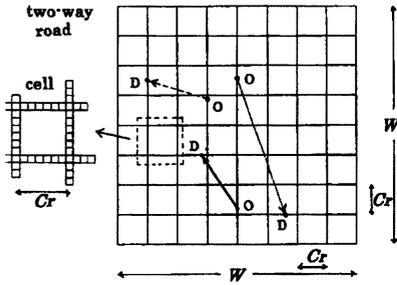


Fig. 1. CA urban roads.

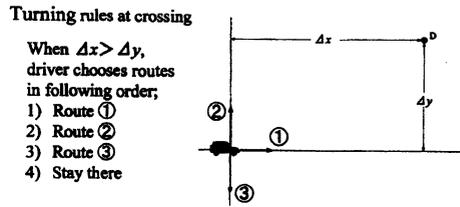


Fig. 2. Turning rule at the crossing.

### 3 Simulation and Results

The traffic flow is simulated in the two-dimensional lattice of  $W=36$  with periodic boundary condition and the origin O and destination D of the trip are randomly set on the road within  $Wx=Wy=36$ . Cars start from the origin

and go to the destination through the square lattice road. When a car arrives at the destination, the car disappears from the road and a new car appears in an arbitrary site. Thus the number of cars is conserved.

First, the car density dependency of the the mean velocity of the cars for various waiting times  $T_s(=0-6)$  in a road system, where the distance between the crossings  $C_r$  is unity, is simulated and shown Fig. 3. In the low-density region of the car, the dynamic car ( $T_s=0$ ) moves freely with velocity=1 by quick turning at the crossing. Whereas, the velocity of static cars waiting for non-zero time-steps ( $T_s>0$ ) decreases linearly as the density increases. When the density is over a critical density, the mean velocity decrease abruptly and the traffic flows go into congested states. For dynamic car with  $T_s=0$ , the local jam appears here and there in the road system and the mean velocity is not zero, whereas the congested state becomes a global jam with a small velocity for cars with  $T_s>0$ .

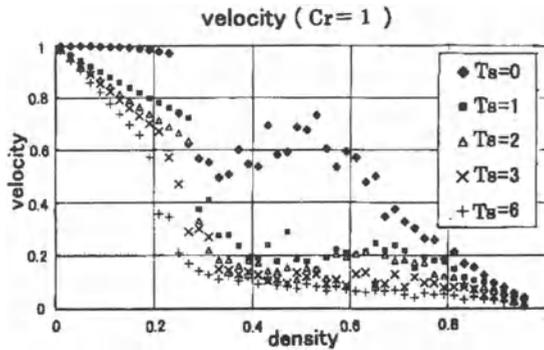
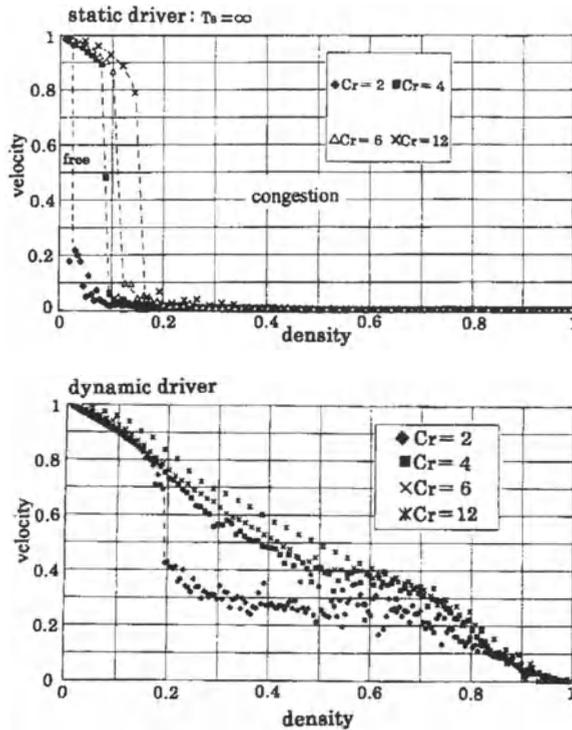


Fig. 3. Mean velocities of cars of various waiting times  $T_s(=0-6)$ .

Next, the effects of the various distances  $C_r$  between the crossings are studied. When all drivers are static ones ( $T_s=\infty$ ) and their density increases, the traffic flow transits to a global jam and the mean velocity of the cars abruptly goes to zero like a first-order phase transition (Fig. 4(a)). The transition density becomes lower as  $C_r$  becomes smaller. When all cars are the dynamic ones, the velocity for  $C_r>3$  decreases gradually and the road system does not go down to the global jam (Fig. 4(b)).

Next, when both types of drivers are driving the cars, velocities and traffic flows of both drivers have been studied. They are shown in Fig. 5 for various ratios  $D_r$  of the dynamic driver and  $C_r=4$ . The velocity of the dynamic driver is larger than that of the static driver in every  $D_r$  (Fig. 5(a)). When the drivers are all dynamic, they advance keeping away from traffic jam and the global jam is prevented. If static drivers ( $T_s=\infty$ ) appear in the drivers



**Fig. 4.** Velocities of (a) static cars and (b) dynamic cars on the urban road with various distances ( $Cr=2,4,6,12$ ) between the crossings.

pool, they produce a deadlock on the road and the traffic jam appears. The static drivers induce the global jam even if they are less, because they become the seeds for the jam.

The relative traveling time and the number of cars of both drivers arriving at the destination in the case of  $T_s=4$ ,  $D_r=0.5$ , and  $Cr=1$  are shown in Fig. 6. The relative traveling time is defined in comparison with the time needed by a car for the journey in the free flow. In the low-density region under the jam transition, the relative traveling time of the dynamic driver is smaller than that of the static one (Fig. 6(a)) and the arrival number at the destination is also larger for the dynamic car than for the static ones (Fig. 6(b)). The results indicate that the dynamic driver can successfully arrive at the destination by avoiding the congestion. In the high-density region above the jamming transition, no differences are found for both the traveling time and the arrival number of both drivers.

When the ratio  $D_r$  of dynamic driver varies at a fixed total density  $\rho=0.2$  with  $Cr=1$  and  $T_s=4$ , the variations of velocities and relative traveling times

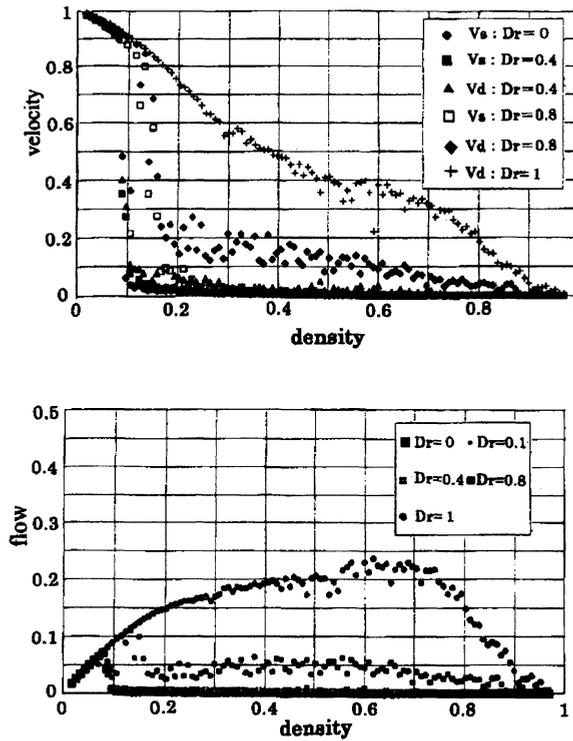


Fig. 5. Individual velocities and total flows of both drivers for various ratios  $Dr$  and  $Cr=4$ .

are shown in Fig. 7. Here, the total density  $\rho=0.2$  is fixed at a density below the jamming transition if  $Dr=0.5$ , as indicated in Fig. 6. The results show the fact that a jam does not occur and the traffic flow can increase in the road system, if the ratio of the dynamic drivers  $Dr$  increases, whereas the jam does occur if the ratio of the dynamic drivers  $Dr$  decreases. This indicates that the supply of traffic information to the drivers may be effective in a traffic flow management.

We have simulated the traffic flow in the urban roads expressed by a CA model with the route-choice behaviors, where cars make trips from an origin to a destination in the two-dimensional two-way roads. Two-types of drivers: a static and a dynamic driver, behaving differently when turning at the crossing on the road. (1) the dynamic driver who immediately changes its advancing direction at the crossing when he can't advance through the shortest route because of a congested road. (2) the static driver who does not change the direction to keep the shortest route, but finally changes the

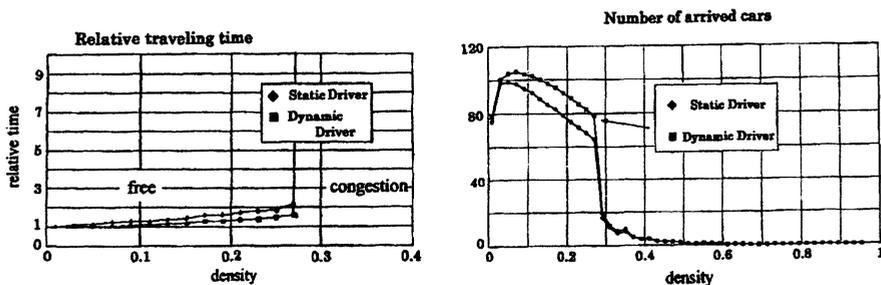


Fig. 6. Relative traveling times and numbers of cars arrived at the destination of both drivers in case of  $T_s=4$ ,  $D_r=0.5$  and  $C_r=1$ .

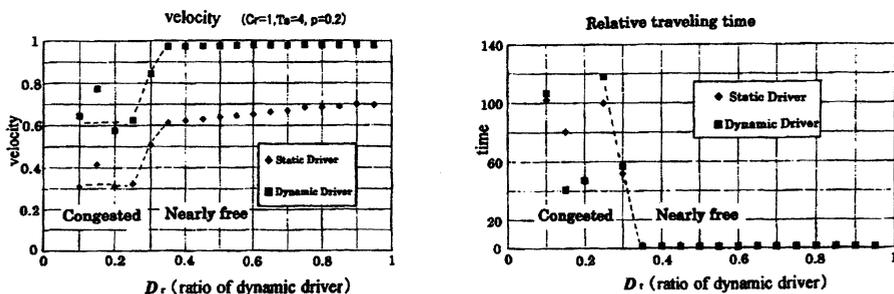


Fig. 7. Velocities and relative traveling time for various ratios  $D_r$  at a fixed total density  $\rho=0.2$ ,  $C_r=1$  and  $T_s=4$ .

route after waiting for  $T_s$  time-steps. Drivers turn at a crossing in order to arrive at the destination through the shortest route from the present position. Simulations on the velocity of the cars, the relative traveling time, and the number of arriving cars have been performed in the various ratios of two kinds of drivers and densities of cars. the effect of the route-choice behaviors of the drivers on the traffic is investigated. It has been found that the traffic flow is improved, when the ratio of the dynamic driver increases. The static driver becomes the seed for the global jam even if he is less and the supply of traffic information to the drivers can avoid the occurrence of the global jam by encouraging of detours around the congestion.

## References

1. O. Biham, A.A. Middleton, and D. Levine, Phys. Rev. A **46**, R6124 (1992).
2. J.A. Cuesta, F.C. Martinez, J.M. Molera, and A. Sanchez, Phys. Rev. E **48**, R4175 (1993).
3. Y. Honda and T. Horiguchi, J. Phys. Soc. Jpn. **69**, 3744 (2000).

# Experimental Investigation of Day-to-Day Route Choice Behaviour

R. Selten<sup>1</sup>, M. Schreckenberg<sup>2</sup>, T. Pitz<sup>1</sup>, T. Chmura<sup>1</sup>, and J. Wahle<sup>3</sup>

<sup>1</sup> Laboratory of Experimental Economics, Adenauerallee 24-42, 53113 Bonn, Germany

<sup>2</sup> Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany

<sup>3</sup> TraffGo GmbH, Grabenstr. 132, 47057 Duisburg, Germany

**Abstract.** The paper reports laboratory experiments on a *day-to-day route choice game* with two routes. Subjects had to choose between a main road  $M$  and a side road  $S$ . The capacity was greater for the main road. 18 subjects participated in each session. In equilibrium the number of subjects is 12 on  $M$  and 6 on  $S$ . Two treatments with 6 sessions each were run at the computerized Laboratory of Experimental Economics at Bonn University using RatImage [1]. Feedback was given in treatment I only about own travel time and in treatment II on travel time for  $M$  and  $S$ . Money payoffs increase with decreasing time. The main results are as follows: Though mean numbers on  $M$  and  $S$  are very near to the equilibrium, fluctuations persist until the end of the sessions in both treatments. The Fluctuations are smaller under treatment II. The effect is small but significant. The total number of changes is significantly greater in treatment I. Two response modes can be observed: A *direct* response mode reacts with more changes for bad payoffs whereas a *contrary* response mode shows opposite reactions. Subjects' road changes and payoffs are negatively correlated in all sessions.

## 1 Introduction

Understanding individual travel behaviour is essential for the design of Advanced Traveler Information Systems (ATIS), which provide real-time travel information, like link travel times [2,3]. However, the response of road users to information is still an open question [4-6]. Moreover it is not clear whether more information is beneficial [4]. Drivers confronted with too much information may become **oversaturated** in the sense that information processing becomes to difficult and users develop simple heuristics to solve the problem [7].

Drivers may also **overreact** to information and thereby cause additional fluctuations. Thus, the behaviour of the drivers has to be incorporated in the forecast [4,5,8]. ATIS can reduce fluctuations only if behavioural effects are correctly taken into account.

The Literature reports a number of experiments on route choice behaviour [5,6]. Here we focus on the route choice in a generic two route scenario, which already has been investigated in the literature [9,10]. However our aim is to present

experiments with a large number of periods and with sufficiently many independent observations for meaningful applications of non-parametric significance tests.

If one wants to investigate results of day-to-day route choice which can be transferred to more realistic environments, it is necessary to explore individual behaviour in an interactive experimental set-up. Does behaviour converge to equilibrium? Does more feedback reduce fluctuations? What is the structure of individual responses to recent experiences? Our experimental study tries to throw light on these questions.

## 2 Experimental Set-Up

Subjects are told that in each of 200 periods they have to make a choice between a main road  $M$  and a side road  $S$  for travelling from A to B. They were told that  $M$  is faster if  $M$  and  $S$  are chosen by the same number of people. No further information was given to the subjects. The number of subjects in each session was 18. The time  $t_M$  and  $t_S$  depends on the numbers  $n_M$  and  $n_S$  of participants choosing  $M$  and  $S$ , respectively:

$$\begin{aligned}t_M &= 6 + 2n_M \\t_S &= 12 + 3n_S.\end{aligned}$$

The period payoff was  $40 - t$  with  $t = t_M$  if  $M$  was chosen and  $t = t_S$  if  $S$  was chosen. The total payoff of a subject was the sum of all 200 period payoffs converted to money payoffs in DM with a fixed exchange rate of 0.015 DM for each experimental money unit. One session took roughly one and a half hours. All pure equilibria of the game are characterized by  $n_M = 12$  and  $n_S = 6$ .

Two treatments have been investigated. In treatment I the subjects received feedback only about the time on the road chosen after each period. In treatment II feedback was provided about both times. Six sessions were run with treatment I and six with treatment II.

## 3 Equilibrium Predictions and Observed Behaviour

Figure 1 shows the number of participants on the side road  $S$  as a function of time for a typical session of treatment II.

It can be seen that there is no convergence to the theoretical equilibrium. There are substantial fluctuations until the end of the session. The same is true for all sessions of both treatments. The overall average of numbers of participants on  $S$  is very near to the equilibrium prediction. The standard deviation of the session average from 6 is never greater than 0.17 and it is only 0.07 on the average. The fluctuations can be measured by the standard deviation of the number of participants choosing  $S$  per period. This standard deviation is between 1.53 and 1.94. In

view of these numbers one can speak of substantial fluctuations in each of the 12 sessions.

The fluctuations are a little larger under treatment I than under treatment II. The effect is significant. The null-hypothesis is rejected by a Wilcoxon-Mann-Whitney-Test on the significance level of 4.65 % (one sided).

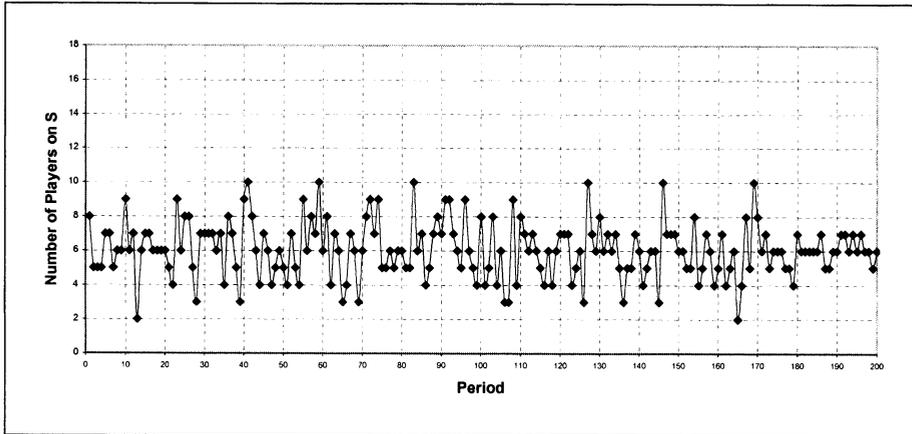


Fig. 1. Number of participants on  $S$  [a typical session of treatment II].

The game underlying the experiment has many pure strategy equilibrium points. In all of them the number of participants on the side road is 6, but the set of players who choose  $S$  can be any set of 6 players. The multiplicity of pure strategy equilibria poses a coordination problem which may be one of the reasons for non-convergence and the persistence of fluctuations. Feedback on both travel times vs. feedback on only own travel time has a beneficial effect by the reduction of fluctuations, but this effect is relatively small.

The fluctuations are connected to the total number of road changes within one session. The Spearman-rank-correlation between the total number of road changes and the standard deviation of the number of participants per period on  $S$  is 0.795. This is significant on the level of 1 % (one sided). The total number of road changes under treatment I is greater than under treatment II. A Wilcoxon-Mann-Whitney-Test rejects the null-hypothesis only on a very weak significance level of 8.98 % (one sided).

Under treatment I subjects who mainly choose only one of the roads feel the need to travel on the other road from time to time in order to get information on both roads. Under treatment II there is no necessity for such information gathering. This seems to be the reason for the greater number of changes and maybe also for the stronger fluctuations under treatment I. Table 1 presents an overview of the data for Experiment I and II.

		number of players on S		number of road changes		Yule coefficients Q	
		mean	std. dev.	mean	std. dev.	mean	std. dev.
Experiment I	session I 01	6.02	1.814	5.08	2.298	0.214	0.654
	session I 02	5.91	1.691	3.87	1.865	0.373	0.592
	session I 03	6.01	1.849	5.16	1.934	0.277	0.525
	session I 04	5.85	1.748	5.19	1.931	0.191	0.603
	session I 05	6.10	1.810	5.28	2.391	0.313	0.584
	session I 06	6.03	1.793	4.35	2.083	0.332	0.542
	treatment I	5.98	1.784	4.82	2.084	0.283	0.585
Experiment II	session II 01	5.98	1.640	3.99	2.001	0.365	0.591
	session II 02	6.05	1.584	3.68	2.039	0.374	0.536
	session II 03	5.99	1.530	3.67	2.091	0.308	0.552
	session II 04	6.10	1.935	5.19	2.320	0.271	0.584
	session II 05	6.06	1.631	4.67	2.480	0.246	0.738
	session II 06	6.17	1.692	4.44	2.044	0.122	0.557
	treatment II	6.06	1.669	4.27	2.163	0.281	0.597

Tab. 1. Overview of the data for Experiment I and II.

## 4 Response Mode

A participant who had a bad payoff on the road chosen may change his road in order to travel where it is less crowded. We call this the *direct* response mode. A road change is the more probable the worse the payoff is.

The *direct* response mode is the prevailing one but there is also a *contrarian* response mode. Under the contrarian response mode a road change is more likely the better the payoff is. The contrarian participant expects that a high payoff will attract many others and that therefore the road chosen will be crowded in the next period.

	change	stay
payoff < 10	$c_-$	$s_-$
payoff > 10	$c_+$	$s_+$

Tab 2. 2x2 table for the computation of Yule coefficients.

The equilibrium payoff is 10. Payoffs perceived as bad tend to be below 10 and payoffs perceived as good tend to be above 10. Accordingly we classified the response of a subject as direct if the road is changed after a payoff smaller than 10 or not changed after a payoff greater than 10. An opposite response is classified as contrarian. Table 2 shows the numbers of times in which a subject changes roads

( $c_-$  for a payoff below 10 and  $c_+$  for a payoff above 10), or stays at the same road ( $s_-$  for payoff below 10 and  $s_+$  for a payoff above 10).

For each subject such a 2x2 table has been determined and a Yule coefficient  $Q$  has been computed as follows.

$$Q = \frac{c_- \cdot s_+ - c_+ \cdot s_-}{c_- \cdot s_+ + c_+ \cdot s_-}$$

The Yule coefficient has a range from  $-1$  to  $+1$ . In our case a high Yule coefficient reflects a tendency towards direct responses and a low one a tendency towards contrarian responses.

In each of four sessions, one of them in Treatment I and three in treatment II, there was one player for whom no Yule coefficient could be determined since these four subjects never change roads. These subjects are not considered in the evaluation of Yule coefficients.

Evidence for the importance of both response modes can be found in the distributions of Yule coefficients within a session. If the two response modes were not present in behaviour one would expect distributions of Yule coefficients concentrated around 0. However the number of subjects with extreme Yule coefficients below  $-0.5$  or above  $+0.5$  tends to be greater than the number of subjects with Yule coefficients in the middle range between  $-0.5$  and  $+0.5$ . A Wilcoxon one sample test supports this alternative hypothesis by rejecting the null-hypothesis that none of both numbers tends to be greater than the other, on the significance level of 1 % (two sided).

If one classifies subjects with Yule coefficients above  $+0.5$  as direct responders and subjects with Yule coefficients below  $-0.5$  as contrarian responders, then one receives 44 % direct responders, 14% contrarian responders and 42% unclassified subjects.

## 5 Payoffs and Road Changes

In all sessions the number of road changes of a subject is negatively correlated with the subject's payoff. Even if subjects change roads in order to get higher payoffs, they do not succeed in doing this on the average. This suggests that it is difficult to use the information provided by the feedback to one's advantage.

## 6 Conclusion

The study has shown that the mean numbers on both roads tend to be very near to the equilibrium. Nevertheless, fluctuations persist until the end of the sessions in both treatments. This is of particular interest in view of the fact that the experiments run over 200 periods which is unusually long and should be enough to show a tendency of convergence to equilibrium, if there is one.

Feedback on both road times significantly reduces fluctuations in treatment II compared to treatment I. However the effect is small. There is a significant rank correlation between the total number of road changes and the size of fluctuations. In treatment I road changes may serve the purpose of information gathering. This motivation has no basis in treatment II. However road changes may also be attempts to improve payoffs. The finding of a negative correlation between a subject's payoff and number of road changes suggests that on the average such attempts are not successful.

Two response modes can be found in the data, a direct one in which road changes follow bad payoffs and a contrarian one in which road changes follow good payoffs. One can understand these response modes as due to different views of the causal structure of the situation. If one expects that the road which is crowded today is likely to be crowded tomorrow one will be in the direct response mode but if one thinks that many people will change to the other road because it was crowded today one has reason to be in the contrarian response mode. We have presented statistical evidence for the importance of the two response modes.

## Acknowledgement

We are grateful to Sebastian Kube who wrote the program for running the experiments.

## References

1. K. Abbink and A. Sadrieh, RatImage – Research Assistance Toolbox for Computer-Aided Human Behavior Experiments, SFB Discussion Paper B-325, University of Bonn (1995).
2. J. Adler and V. Blue, *Transp. Res. C* **6**, 157 (1998).
3. W. Barfield and T.A. Dingus, *Human Factors in Intelligent Transportation Systems* (Lawrence Erlbaum Associates Inc., Mahwah, New Jersey, 1998).
4. M. Ben-Akiva, A. de Palma, and I. Kaysi, *Transp. Res. A* **25**, 251 (1991).
5. P. Bonsall, *Transportation* **19**, 1-23 (1992).
6. H.S. Mahmassani and Y.H. Liu, *Transp. Res. C* **7**, 91-107 (1997).
7. G. Gigerenzer, P.M. Todd, and ABC Research Group (eds.), *Simple heuristics that make us smart* (Oxford University Press, 1999).
8. J. Wahle, A. Bazzan, F. Klügl, and M. Schreckenberg, *Physica A* **287**, 669-681 (2000).
9. Y. Iida, T. Akiyama, and T. Uchida, *Transp. Res. B* **26**, 17-32 (1992).
10. R. Hall, *Transpn. Res. C* **4**, 289-306 (1996).

# Optimal Traffic States in a Cellular Automaton Model for City Traffic

R. Barlovic<sup>1</sup>, E. Brockfeld<sup>2</sup>, A. Schadschneider<sup>3</sup>, and M. Schreckenberg<sup>1</sup>

<sup>1</sup> Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg,  
Lotharstr. 1, 47048 Duisburg, Germany

<sup>2</sup> German Aerospace Center (DLR), 12489 Berlin, Germany

<sup>3</sup> Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany

**Abstract.** The impact of global traffic light control strategies for city networks is analyzed in a recently proposed cellular automaton model. The model combines basic ideas of the Biham-Middleton-Levine model for city traffic and the Nagel-Schreckenberg model for highway traffic. The city network has a simple square lattice geometry. All streets and intersections are treated equally, i.e., there are no dominant streets.

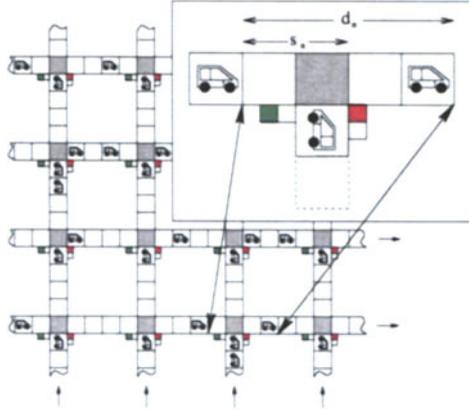
## 1 Introduction

Nowadays mobility is regarded as one of the most significant ingredients of a modern society. Unfortunately, the capacity of the existing street networks is often exceeded. In urban networks the flow is controlled by traffic lights and traffic engineers are often forced to question if the capacity of the network is exploited by the chosen control strategy. One possible method to answer such questions could be the use of vehicular traffic models in control systems as well as in the planning and design of transportation networks. For almost half a century there were strong attempts to develop a theoretical framework of traffic science (for an overview see [1–3]).

In this paper we analyze the impact of global traffic light control strategies, in particular synchronized traffic lights, traffic lights with random offset, and with a defined offset in a recently proposed cellular automaton (CA) model for city traffic. Chowdhury and Schadschneider [5,6] combine basic ideas from the Biham-Middleton-Levine (BML) [7] model of city traffic and the Nagel-Schreckenberg (NaSch) [8] model of highway traffic. In order to take into account the more detailed dynamics of the NaSch model, they extended the BML model by inserting finite streets between the cells. On the streets vehicles drive in accordance to the NaSch rules.

## 2 Model

As one can see from Fig. 1, the network of streets build a  $N \times N$  square lattice. All intersections are assumed to be equitable, i.e., there are no main roads in the network where the traffic lights have a higher priority. In accordance with the BML model streets parallel to the  $x$ -axis allow only single-lane east-bound



**Fig. 1.** Snapshot of the underlying lattice of the model. In this case the number of intersections in the quadratic network is set to  $N \times N = 16$ . The length of the streets between two intersections is chosen to  $D - 1 = 4$ .

traffic while the ones parallel to the  $y$ -axis manage the north-bound traffic. The separation between any two successive intersections on every street consists of  $D - 1$  cells. The traffic lights are chosen to switch after a fixed time period  $T$ . The length of the time periods for the green lights does not depend on the direction and thus the “green light” periods are equal to the “red light” periods. In addition to the ChSch model [5], we improved the traffic lights by assigning an offset parameter to every one (see [9] for an detailed explanation). This modification can be used for example to shift the switch of two successive traffic lights in a way that a “green wave” can be established in the complete network.

In analogy to the NaSch model the speed  $v$  of the vehicles can take one of the  $v_{max} + 1$  integer values in the range  $v = 0, 1, \dots, v_{max}$ . The dynamics of vehicles on the streets is given by the maximum velocity  $v_{max}$  and the randomization parameter  $p$  of the NaSch model which is responsible for the movement. The state of the network at time  $t + 1$  can be obtained from that at time  $t$  by applying the following rules to all cars at the same time (parallel dynamics):

- Step 1: *Acceleration*:  $v_n \rightarrow \min(v_n + 1, v_{max})$
- Step 2: *Braking due to other vehicles or traffic light state*:
  - Case 1: The traffic light is red in front of the  $n$ -th vehicle:
 
$$v_n \rightarrow \min(v_n, d_n - 1, s_n - 1)$$
  - Case 2: The traffic light is green in front of the  $n$ -th vehicle:
 

If the next two cells directly behind  
the intersection are occupied

$$v_n \rightarrow \min(v_n, d_n - 1, s_n - 1)$$

else  $v_n \rightarrow \min(v_n, d_n - 1)$
- Step 3: *Randomization with probability  $p$* :  $v_n \rightarrow \max(v_n - 1, 0)$
- Step 4: *Movement*:  $x_n \rightarrow x_n + v_n$

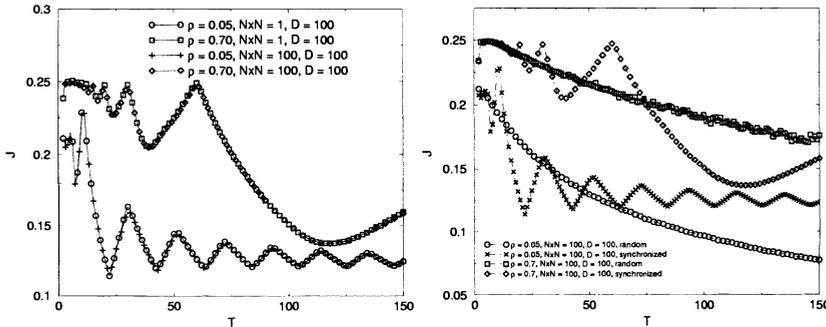
Here  $x_n$  denotes the position of the  $n$ -th car and  $d_n = x_{n+1} - x_n$  the distance to the next car ahead (see Fig. 1). The distance to the next traffic light ahead

is given by  $s_n$ . The length of a single cell is set to  $7.5\text{ m}$  in accordance to the NaSch model. The maximal velocity of the cars is set to  $v_{max} = 5$  throughout this paper. Since this should correspond to a typical speed limit of  $50\text{ km/h}$  in cities, one time-step approximately corresponds to  $2\text{ sec}$  in real time. Note, that Case 2 of Step 2 is modified slightly in comparison to [6]. Due to this modification, a driver will only occupy an intersection if it is assured that he can leave it again. This is done to avoid the transition to a completely blocked state (gridlock) that is undesirable when exploring high densities.

### 3 Strategies

In order to improve the overall traffic conditions in the considered model different global traffic strategies are investigated. At this point it has to be taken into account that there are no dominant streets. This makes the optimization much more difficult and implies that the green and red phases for each direction should have the same length. For a main road intersection with several minor roads the total flow usually can be improved easily by optimizing the flow on the main road.

#### 3.1 Synchronized Traffic Lights



**Fig. 2. Left:** The mean flow strongly oscillates in the case of synchronized traffic lights. A small mini network segment shows the same dynamics like a large network. **Right:** A random shift in the switching leads to a more flexible strategy, e.g., without oscillations. The mean flow is remarkable higher in comparison to synchronized traffic lights.

The typical dependence between the time periods of the traffic lights and the mean flow in the system is shown in Fig. 2 (left) for synchronized traffic lights. For low densities one finds a strongly oscillating curve with maxima and minima at regular distances whereby the optimal traffic states are determined by the travel times between the intersections. The traffic light cycle time corresponding to the

maximum system flow is equal to  $T_{max} = D/2v_{free}$ .<sup>1</sup> Similar oscillations can be even found at very high densities. It is further interesting that if considering a “mini network” with only one single intersection the same results are obtained as for large networks. In [9] it is shown that a simple phenomenological approach based on the description of the dynamics in such a “mini network” is capable to explain the impact of the cycle times in very good agreement with the numerical results even for large networks. Moreover it is very interesting that although the vehicle movement is stochastic (NaSch model) and the mean density on the streets in the network fluctuates, there is no local conglomeration of vehicles in the network leading to remarkable deviations in the flow in comparison to the idealized “mini network” where the density on the streets is fixed (see Fig. 2 (left)). It seems that the signalized intersections of the model interact with the density fluctuations in a way that the vehicles are equally distributed in the network. Also for high densities one can find a strong dependence of the mean flow in the system for the chosen cycle times (see Fig. 2). Obviously for high densities this dependence is not caused by free flowing vehicles, but determined by the movement of jams. The fraction of time the green light is not blocked by a jam controls the overall flow.

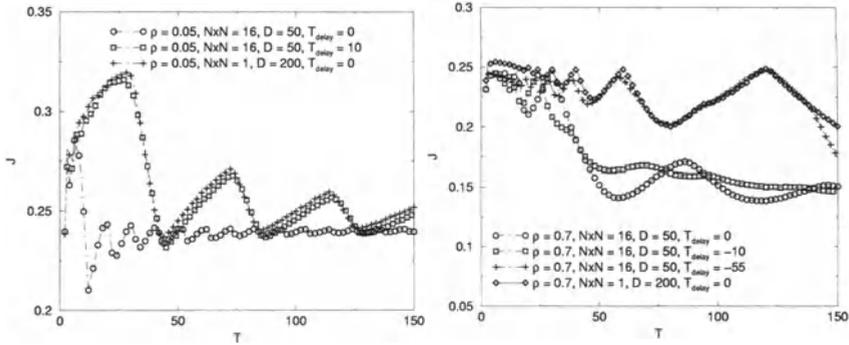
### 3.2 Random Offset Strategy

In this section we want to point out that switching successive traffic lights with a random shift can lead to a more flexible strategy, e.g., without oscillations. Moreover, in contrast to a system with synchronized traffic lights a random shift between the intersections can lead to a remarkable higher global system flow. To give an insight into the effects induced by the random offset we depicted the throughput in the network in dependence of the cycle times in Fig. 2 (right). Obviously the strong oscillations found in the curves corresponding to the synchronized strategy are suppressed by the randomness in the switching. Thus the random offset strategy leads to a smoothed curve which is very useful when adjusting the optimal cycle times in a network. For free-flow densities the random offset strategy outperforms the synchronized strategy for relatively low cycle times because unfavorable states (states with minimal global flow) are avoided by the randomness. However, at high densities the oscillations are suppressed in a similar manner as for the low density case. Hence, as for low densities, this strategy gives an improved flexibility when adjusting optimal cycle times in the network and an improved overall flow in a wide area.

### 3.3 Green Wave Strategy

For the “green wave” strategy the individual offset parameter that is assigned to every intersection is used to implement a certain time delay between the traffic light phases of two successive intersections. This delay can be constituted in the whole network [9]. To quantify the improvement obtained by the “green wave”

<sup>1</sup>  $v_{free} = v_{max} - p$  is the velocity of free-flowing vehicles.



**Fig. 3.** The global flow is plotted for the “green wave” strategy and compared to a system with synchronized strategy. The left diagram shows the free-flow case of the system while in the right diagram the influence of the green wave strategy in the high density state is plotted.

strategy the overall network flow is plotted against the cycle time (see Fig. 3) and compared with the synchronized strategy. The left diagram corresponds to the free-flow case of the system. Obviously, the green wave strategy with a properly chosen offset parameter shows reasonable improvements over the strategy with synchronized traffic lights. The optimal “green wave strategy” is to adjust the time delay such that the first vehicle trespassing an intersection will arrive at the next one exactly when it switches to green. The corresponding optimal delay time is given by  $T_{delay} = D/v_{free}$ . In this way, the “green wave” strategy is capable to pipe all the vehicles through the streets as if there is only one intersection left in the system. This agreement to a system with only one intersection but an equal total street length is demonstrated in Fig. 3 for low as well as high densities. By definition no “green wave” can be established at high densities, but a suitable offset in the switching between successive traffic lights can lead anyhow to an improved flow. Obviously the dynamics for high densities is governed by the motion of large jams that move oppositely to the driving direction. The optimal system state would be reached if a jam moves backward from one intersection to the one before and blocks it while the traffic light is red anyway so that afterwards moving vehicles (outflow of the jam) can take advantage of the green phase as much as possible. For high densities the optimal delay time is equal to  $T_{delay} = D/v_{jam}$ .<sup>2</sup>

### 4 Summary and Discussion

We have analyzed the ChSch model which combines basic ideas from the Biham-Middleton-Levine (BML) model of city traffic and the Nagel-Schreckenberg (Na-Sch) model of highway traffic. In our investigation we focused on global traffic

<sup>2</sup>  $v_{jam}$  is the velocity of backwards moving jams.

light control strategies and tried to find optimal model parameters in order to maximize the network flow. For this purpose we started with the original formulation of the ChSch model where the traffic lights are switched synchronously. It is shown that the global throughput of the network strongly depends on the cycle times, i.e., one finds strong oscillations in the global flow in dependence of the cycle times for low as well as for high densities. In order to allow a more flexible traffic light control the ChSch model was enhanced by an additional model parameter. This new parameter is assigned to every intersection representing a time offset, so that the traffic lights are not enforced to switch simultaneously anymore. Consequently to avoid the strong oscillations we analyzed a network where traffic lights are switched at random. It is shown that the strong oscillations are completely suppressed by randomness. Thus the random offset strategy can be very useful if a control strategy is required which is not very sensitive to the adjustment of the cycle times. This is in particular the case if the density on the streets is strongly fluctuating. The random offset strategy outperforms the standard ChSch model with synchronized traffic lights at low densities for small cycle times and at high densities for all cycle times. An explanation for the profit at high densities is the fact that some parts of the network are completely jammed while in other parts of the network the cars can move nearly undisturbed. This additional gain due to the inhomogeneous allocation of vehicles indicates that an autonomous traffic light control based on local decisions could be more effective than the analyzed global schemes. Moreover a two dimensional “green wave” is implemented with the help of the offset parameter. The “green wave” gives much improvement to the flow in comparison to the synchronized strategy and the random offset at low densities and has even an incisive impact on the throughput at high densities. Although the “green wave” strategy is capable to give a strong improvement, the dependence between flow and the cycle time found in the original ChSch model remains.

## References

1. D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf (Eds.), *Traffic and Granular Flow '99*, (Springer, 2000).
2. D. Helbing, *Verkehrsdynamik: Neue Physikalische Modellierungskonzepte*, (in German), (Springer, 1997).
3. D. Chowdhury, L. Santen, and A. Schadschneider, *Phys. Rep.* **329**, 199 (2000); *Curr. Sci.* **77**, 411 (1999) and *Comp. Science & Techn.* **2**(5), 80 (2000).
4. S. Wolfram, *Theory and applications of cellular automata*, (World Scientific, 1986).
5. D. Chowdhury and A. Schadschneider, *Phys. Rev. E* **59**, R1311-1314 (1999).
6. A. Schadschneider, D. Chowdhury, E. Brockfeld, K. Klauack, L. Santen, and J. Zittartz, in: *Traffic and Granular Flow '99*, (Springer, 2000).
7. O. Biham, A.A. Middleton, and D. Levine, *Phys. Rev. A* **46**, 6124 (1992).
8. K. Nagel and M. Schreckenberg, *J. Physique I* **2**, 2221 (1992).
9. E. Brockfeld, R. Barlovic, A. Schadschneider, and M. Schreckenberg, *Phys. Rev. E* **64**, 056132 (2001).

# Traffic Flow Analysis Based on Multiagent

H. Tsunashima, Y. Nishi, T. Honjyo, and T. Sakai

Nihon University, College of Industrial Technology, 1-2-1 Izumi-cho, Narashino-shi, Chiba 275-8575, Japan

## 1 Introduction

In recent years, in line with the greater use of automobiles, traffic congestions have frequently occurred on city streets and on highways due to the effects of increases in traffic volume, the greater complication of the road system, the operational mistakes of drivers, etc. This has been taken up as a serious traffic problem. In order to discuss methods of solving this problem, research and developmental work on traffic flow simulators by computer is being carried out. However, the dynamics of vehicles and their driver's characteristics, which are important factors for any traffic flow analysis, have not been considered sufficiently enough in the past. In this study, we propose, as an intermediate model between macroscopic models and microscopic ones, an analytical technique based on multiagent, which allows for the consideration of the driver characteristics, vehicle dynamics in complex traffic flow as shown in Fig.1.

## 2 Modeling of Traffic Condition by Multiagent

### 2.1 Agent-Oriented

The artificial life system is a study simulating biological activities in computers. In this study, the agent is considered as a kind of artificial life system. Its property is defined as follows: "An agent selects and executes its actions autonomously in accordance with its accumulated rules as knowledge relative to its own environment (region); moreover, in case two or more agents are present within the same region, each agent exists independently of any other, and there exist no global rule that dictate the actions of all the agents; any global behavior obtained as a result is of a spontaneous nature." By using the new technique, it is possible to give different rules to each vehicle (personality) and determine its actions autonomously under the knowledge accumulated in accordance with the environment surrounding the vehicle.

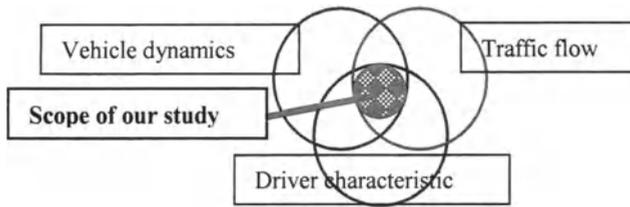


Fig.1. Scope of our study.

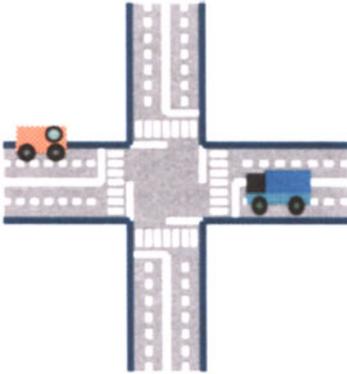
## 2.2 Application to the Road Traffic System

The concept of a road traffic system expressed by multiagent is shown in Fig. 2. The traffic flow system defined as multiagent is classified into road agent and vehicle environment. The road agent shows road forms. The vehicle agent is classified into the driver and the vehicle characteristic. The vehicle characteristics are composed of a weight, a length, braking and acceleration performance, which are defined for each vehicles. The driver agent has a field of vision in front and in the rear and, if other vehicle agents enter its field of vision, it can obtain information concerning them. The relationship between the driver agent, the vehicle agent, and the road agent is shown in Fig. 3.

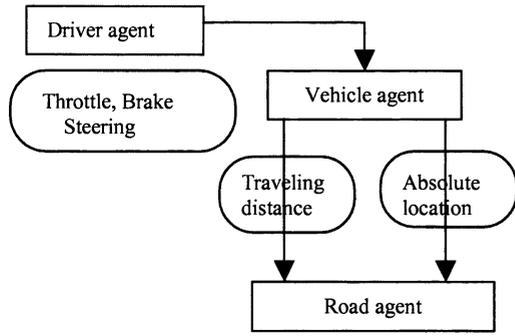
## 3 Design of Agents

### 3.1 Vehicle Agent

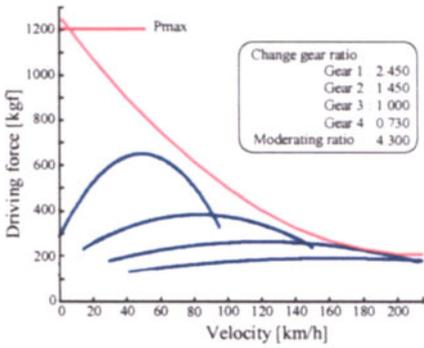
The vehicle agent has a total of nine characteristics: total vehicle weight, total vehicle length, vehicle width, vehicle height, cross section, air drag coefficient, engine performance, braking capability, and control delay. In order to use a traffic flow simulator to express a traffic flow, it is necessary to take into consideration the acceleration characteristics and the deceleration characteristics of all the vehicles on the scene. In particular, the acceleration and deceleration characteristics of a large vehicle are different from those of an ordinary car, and they are anticipated to exert a strong influence on the traffic flow. This proposed simulator carried out the modeling of the engine characteristics of both a normal vehicle and a large-size Vehicle. The driving force characteristics of the normal and large-size vehicles thus modeled are indicated in Fig. 4 and Fig. 5. It should be pointed out here that although the driving characteristics in these figures are altered by gear ratios, this modeling operation used simplified models that would show the characteristics in continuous lines. In this modeling, the driving velocity and maximum driving force of each vehicle are defined in terms of the relationship between the rotational speed and torque of the engine.



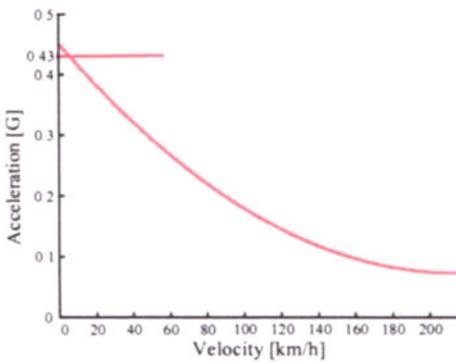
**Fig.2.** A traffic flow expression by agents.



**Fig.3.** Relation of agents.

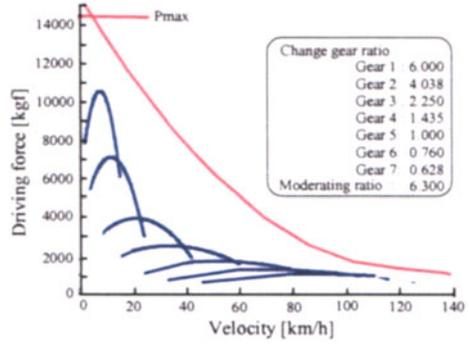


(a) Driving force curve.

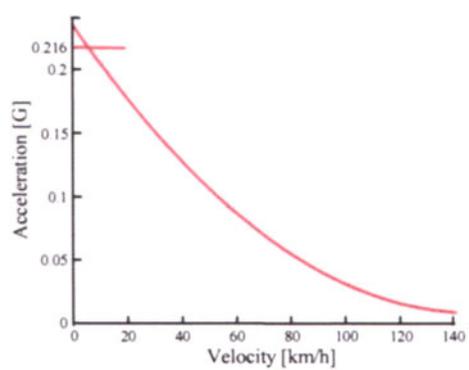


(b) Acceleration.

**Fig.4.** Driving characteristics for normal vehicle.



(a) Driving force curve.



(b) Acceleration.

**Fig.5.** Driving characteristics for large-sized vehicle.

The vehicle agent passes the output of the driver agent through a low-pass filter and then calculates, on the basis of the operational load, the driving force and the braking force outputted according to the velocity at that point in time. In addition, the vehicle agent calculates the acceleration (or deceleration) from the running resistance and driving force (or braking force) and determines the velocity. It then hands over the velocity information thus obtained to the road agent (refer to the lower side of Fig. 6).

### **3.2 Driver Agent**

The driver agent has a total of seven characteristics, namely the current distance to a vehicle ahead, the current driving velocity, the distance to the target vehicle, the front view range, the rear view range, the control delay. In the actual driving of a vehicle in traffic, the distance to a vehicle ahead varies depending on the traffic regulation in force. Here, the distance to the target vehicle was given as a function of velocity determined from actual measurement values.

The relation between the vehicle agent and the driver agent is illustrated in Fig. 6. The driver agent checks to see whether a vehicle ahead is present in its field of vision or not, and depending on the presence or absence of such a vehicle, it makes an appropriate selection from among possible actions, namely control of speed, control of the distance to the vehicle immediately ahead, and lane change, and executes the selected action. The changing of lane is carried out after it passes a comprehensive judgment on relations between the velocities of the own vehicle, the vehicle immediately ahead, the vehicle immediately behind, target velocity, distances to other vehicles, etc. The acceleration or deceleration relative to the target velocity is controlled by means of PID control, and consideration is here given to the reaction delay of the driver. The operational load (acceleration operation load and braking load) outputted by the driver agent is handed over to the vehicle agent which calculates the acceleration force and deceleration force.

## **4 Influence on the Traffic Flow of the Driver's Characteristic**

### **4.1 Outline of Simulation**

The purpose of this simulation is to examine whether difference in the front view range of the driver influences traffic flow. The conditions for this simulation are listed in Fig. 7 and Table 1. Vehicle #0 is a vehicle at rest. Vehicle #2 and #3 were given the characteristics of a large-size vehicle, while all the other vehicles were provided with those of an ordinary vehicle. An investigation was carried out for two cases: front view range of 100m and 300m.

### 4.2 Result of Simulation

The mean velocity of the traffic flow as a whole is shown in Fig. 8. It is clear from Fig. 8 that the effect of this deceleration affected vehicles in the back to such an extent that the mean velocity of the flow rate as a whole was severely influenced. In addition, it was made clear that by securing a wider front view range, this effect can be made milder.

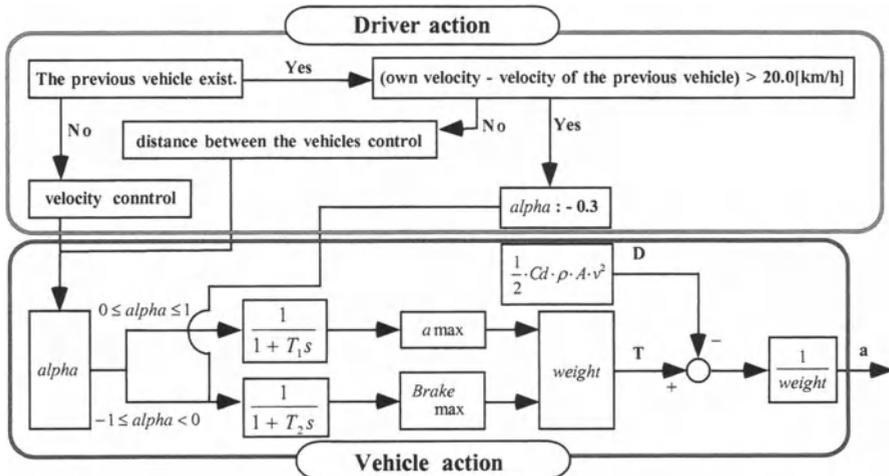


Fig.6. Control of vehicle agent.

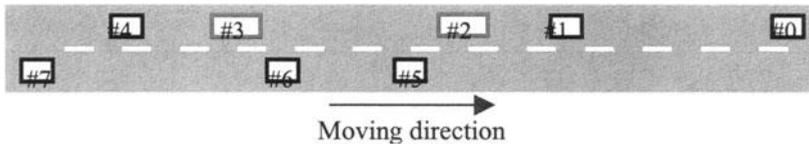


Fig.7. Initial location of the vehicles.

Table 1. Simulation condition.

	Headway [m]	Velocity [m/s]	Target velocity [m/s]	Maximum velocity [m/s]	Front View range [m]	Rear View range [m]
#0	---	0.0	0.0	0.0	---	---
#1	500	22.2	22.2	27.8	100, 300	50
#2	80					
#3	200					
#4	100		25.0	33.3		
#5	---		27.8	27.8		
#6	160		27.8	33.3		
#7	220		27.8	33.3		

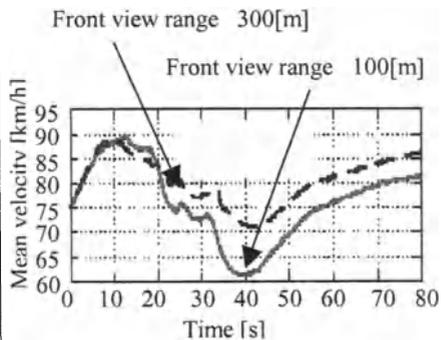


Fig.8. Mean velocity of traffic flow.

## 5 Application to Urban Traffic

Our simulator has analyzed traffic flow on the straight lane. However, it is necessary that the traffic flow simulator can analyze traffic flow in urban areas in order to treat various traffic conditions. Consequently, the road agent should be modified. The vehicle agent and cross agent are defined as new agent and are combined with the road agent. The relation of these agents is shown in Fig. 9.

### 5.1 Remodeling of Road Agent

The road agent shown in Fig. 9 consists of three areas (the outflow area, the moving area and the inflow area). The outflow area and the inflow area have their own state index. The outflow area changes its state index into zero when the vehicle agent enters the outflow area. In this way, the outflow area shifts the vehicle agent to the cross agent. The inflow area changes its state index into zero when the cross agent requests the entering of the vehicle agent into the inflow area. In this way, the inflow area obtains the vehicle agent from the cross agent. The behavior of the moving area is the same as the one of the existing road agent.

### 5.2 Remodeling of Vehicle Agent

The vehicle agent is modified under the consideration of the new road agent. This means that the vehicle agent can enter into the road agent or cross agent when these state indexes are zero.

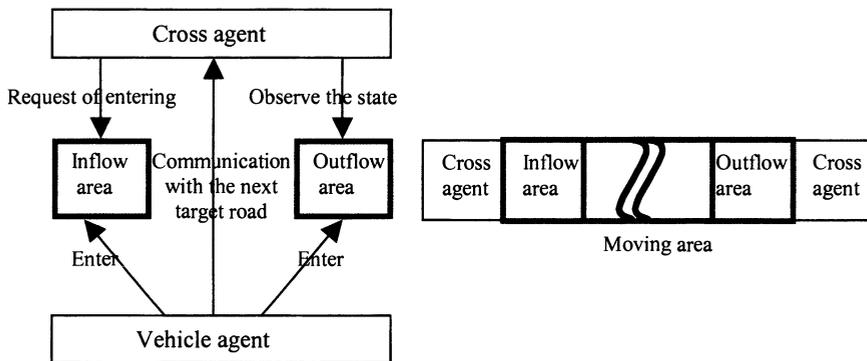


Fig.9. Action of agents.

### 5.3 Modeling of Crossing Agent

The cross agent has its own state index. The cross agent changes its state index to zero when the state index of the outflow area is set to zero. In this way, the cross agent allows the vehicle agent to exit from the outflow area. And then, the cross agent communicates with the vehicle agent located in the next target road, requests the entering of the vehicle agent into the inflow area.

## 6 Conclusions

In this paper, an agent-oriented method of expressing traffic environments and the modeling of driver characteristics and vehicle characteristics were shown, and the basic features of the newly developed traffic flow simulator were examined. As a result, the following facts were made clear:

- (1) It was possible to express the characteristics of each vehicle and its driver in a traffic flow by means of agents.
- (2) By carrying out simulations, it was made clear that the proposed technique makes it possible to grasp the driving conditions of each vehicle and thereby to evaluate its safety in an actual traffic environment.
- (3) It was shown that the driving conditions of each vehicle are substantially affected by its front view range and, in turn, affect the mean velocity of the traffic flow as a whole.

## References

1. Y. Nishi, Structural Analysis based on Multiple Agents, Proceedings of JSME Annual Meeting, JSME 1999 (in Japanese).
2. Nishizoe et al., An Attempt at Structural Analysis Using Multiagents, Transactions of the Japan Society of Mechanical Engineers, No.99-1, 79-80.
3. Tsunashima et al., Traffic Flow Analysis By Multiagent, 5<sup>th</sup> International Symposium on Vehicle Control, 505-509 (2000).

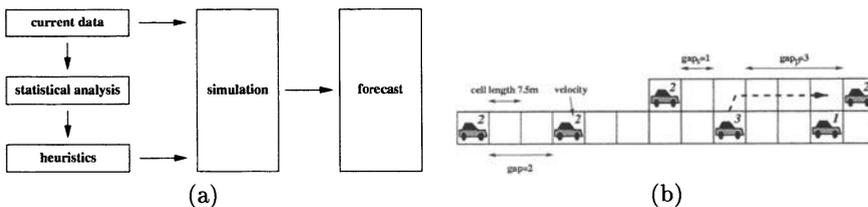
# Traffic Forecast Using a Combination of On-Line Simulation and Traffic Data

R. Chrobok, A. Pottmeier, J. Wahle, and M. Schreckenberg

Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg,  
Lotharstr. 1, 47048 Duisburg, Germany

**Abstract.** In this contribution an approach to traffic forecast using a micro-simulator is presented. In order to provide network-wide information about the current traffic state a cellular automaton traffic flow model is combined with measured data. The framework is applied to the freeway network of North Rhine-Westphalia (NRW), where data from about 3,500 inductive loops are available and provided on-line minute by minute.

## 1 How Do We Forecast?



**Fig. 1.** (a): Schematic sketch of the forecast procedure. The input to the simulator are current and historical data, which are provided in form of heuristics. Using both forecast is carried out. (b): Part of a road in a cellular automaton model.

There are many different ways for predicting traffic conditions. In general, the prognosis horizon is the first and most important parameter, since it determines, which procedure proves as the most effective forecast method. A second important detail is the input data, i.e., the number and the location of the sources of the data. Different approaches have been proposed in the past. Neural networks often are used for predicting traffic flow, speed data or travel times up to 15 minutes. To forecast traffic jams, spatial correlations can be used taking into account the dynamics of a moving jam [1]. For a useful long-term prediction the current traffic data lose their weight and it is more important to use experience about the past, so-called heuristics, in form of a statistical data base consisting of traffic time series [2].

Coupled with current and statistical data the use of on-line simulations supplies the possibility of both: short- and long-term forecasts of a whole network.

To receive the current traffic state data are coupled directly with the simulation. To obtain a first prognosis the simulation has to be performed faster than real-time. Thereby, the main problem is to consider the traffic demand at the boundaries of the network in the simulation. Therefore, the collected data is analyzed statistically to consider traffic states in form of heuristics that are useful for predictions. So, with regard to traffic forecasts the simulation tool serves as a connector of the current with the historical data (Fig. 1 (a)).

## 2 Simulation Model

### 2.1 Principle of On-line Simulation

The starting point for each traffic forecast should be information about the current traffic state in the whole network. Unfortunately, in most urban areas and freeway networks data are only available from separate points in the network. The common sources of information are inductive loops or video detection devices which provide data permanently. However, all information provided is local. Note that an exception to this rule are floating car data (FCD).

Our approach to provide network-wide information is to connect the local data with a micro-simulator. Thus, information can be derived for those regions which are not covered by measurements. The local traffic counts are combined with the network structure under consideration of realistic traffic flow dynamics. This is the basic idea of on-line simulations: *Local traffic counts serve as input for traffic flow simulations to provide network-wide information* [3].

### 2.2 Basic Model

Since the simulator is based on on-line data the model employed should be efficient. Due to their design cellular automaton models are very effective in large-scale network simulations [3–5].

Models which reproduce the dynamic phases of traffic are still under debate. Thus the original cellular automaton by Nagel and Schreckenberg [6] is used in a first version of this simulator. For completeness, the definition of the model is briefly reviewed. The road is subdivided in cells with a length of  $\Delta x = \rho_{\text{jam}}^{-1} = 7.5$  m/veh, with the density of jammed cars  $\rho_{\text{jam}} \approx 133$  veh/km (Fig. 1 (b)). Each cell is either empty or occupied by only one vehicle with an integer speed  $v_i \in \{0, \dots, v_{\text{max}}\}$ , with  $v_{\text{max}}$  the maximum speed. The motion of the vehicles is described by the following rules (*parallel dynamics*):

- R1** Acceleration:  $v_i \rightarrow \min(v_i + 1, v_{\text{max}})$ ,
- R2** Deceleration to avoid accidents:  $v_i' \rightarrow \min(v_i, \text{gap})$ ,
- R3** Randomization: with a certain probability  $p$   
do  $v_i'' \rightarrow \max(v_i' - 1, 0)$ ,
- R4** Movement:  $x_i \rightarrow x_i + v_i''$ .

The variable *gap* denotes the number of empty cells in front of the vehicle at cell  $i$ . A time-step corresponds to  $\Delta t \approx 1$  sec, the typical time a driver needs to react.

The first two rules (**R1**, **R2**) describe a somehow optimal driving strategy, the driver accelerates if the vehicle has not reached the maximum speed  $v_{\max}$  and brakes to avoid accidents, which are explicitly excluded. This can be summed up as follows: *drive as fast as you can and stop if you have to!* However, drivers do not react in this optimal way: they vary their driving behavior without any obvious reasons, reflected by the *braking noise p* (**R3**). It mimics the complex interactions between the vehicles and is also responsible for spontaneous formation of jams.

### 3 Network

A crucial point in the design of every simulator is the representation of the road network. Like in other simulators the network consists of basic elements: links and nodes. A node is either a connection between two links or a sink/source at the boundary or an on/off-ramp. At the sources vehicles are added with regard to input data. On the sinks the cars are simply removed every time step. Links are directional elements representing the roads of the network that connect nodes.

Combining these elements mentioned above it is possible to build complex structures of freeway networks like junctions or intersections. The data used for the network stems from the *NW-SIB*, a GIS system provided by the state of North Rhine-Westphalia. Table 1 shows the design parameters.

**Table 1.** The network of North-Rhine Westphalia

Area	34.000 km <sup>2</sup>
Inhabitants	18.000.000
On- and off-ramps	830
Intersections	67
On-line loop detectors	3.500
Other loop detectors	200
Overall length	2.500 km

### 4 Incorporation of Historical Data

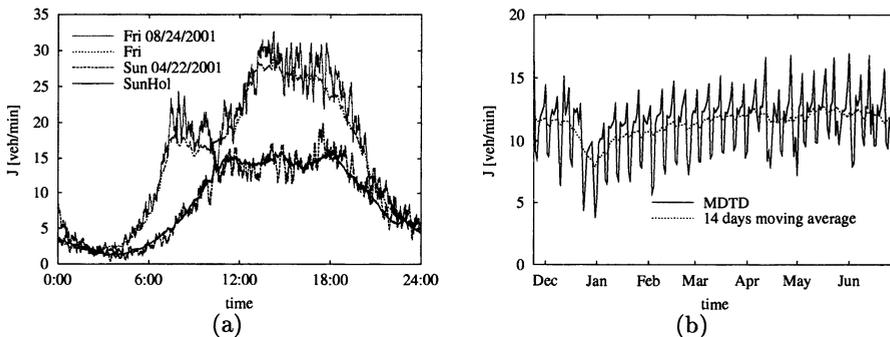
The simulations describe the dynamics in a network but are lacking information about the boundaries. Especially for forecasts, reasonable data has to be incorporated. Therefore, heuristics will be used to predict the flow of the sources.

In order to develop heuristics for traffic forecast, i.e., experience about recurrent events, historical data have to be analyzed. Therefore, it is useful to classify certain days and events in categories. Two different characteristics can be distinguished: daily and seasonal. Seasonal differences appear, e.g., due to school holidays. Additionally, there are daily differences: on working days a sharp morning peak is found which is absent on Sundays or holidays. Besides there are special unpredictable events.

In order to classify days, the daily traffic demand, i.e., the flow of vehicles vs. time, has to be investigated. Therefore, the flows per minute  $J_n(t)$  at a certain time  $t$  are accumulated. Then the data are summed over all classified days  $N_{\text{days}}(t)$  and this result is divided by  $N_{\text{days}}(t)$ :

$$J_{\text{dem}}(t) = \frac{\sum_{n=1}^{N_{\text{days}}(t)} J_n(t)}{N_{\text{days}}(t)}. \tag{1}$$

One advantage of this procedure is the opportunity to analyze even days with an incomplete set of data. The resulting traffic time series are subdivided in eight classes, taking into account the different demand during the weekdays. Figure 2 (a) shows the daily traffic demand of two example classes. On Friday you can see a morning peak and a broaden afternoon peak that are missing on Sundays.



**Fig. 2.** Traffic flow  $J$  vs. time. The data are stemming from an inductive loop on the BAB 1 at kilometer 278.1. (a) The data of two special days and the corresponding classified days. It is quite obvious that the data is useful for long-term predictions. (b) Periodic fluctuations can be seen on the mean daily traffic demand (MDTD). The high peaks belong to Fridays, the low dips to the weekends. Christmas and New Year’s eve cause the two lowest dips in December. The 14 days moving average shows the seasonal fluctuations.

In general the highest number of vehicles during one day is measured on Fridays, a little bit more than on Mondays. Tuesdays, Wednesdays and Thursdays are very similar, and the lowest number of cars is measured on Saturdays,

Sundays and holidays. These facts determine the periodic fluctuations in Figure 2 (b). Also the two very low dips due to Christmas and New Year’s eve can be seen.

The 14 days average shows the seasonal differences: a region of low traffic in the second half of December followed by an arising traffic flow during the first half of the year with a small dip in April due to school holidays and the maximum in May. All these results are in good agreement with the observations made considering urban traffic in [2].

## 5 Forecast Considering Special Events

Similar to holidays or long vacations, there are sometimes special events which influence traffic patterns drastically. Because heuristics lack the influence of the current situation whereas short-term-predictions do not take into account the experience and the knowledge about previous events, both methods are combined in the following.

As short-term forecast method the smoothing averaged traffic flow of the last recent minutes  $J_c(t_0)$  is used. For all prognosis horizons  $\Delta\tau$  the predicted traffic demand  $J_{\text{pred}}(t_p)$  at the time  $t_p$  is the sum of the average demand  $J_{\text{dem}}(t_p)$  of the sample class and the difference of  $J_c(t_0)$  and  $J_{\text{dem}}(t_p)$  weighted with  $k$  (2).  $t_0$  is the point in time when the prognosis is made.

$$J_{\text{pred}}(t_p) = J_{\text{dem}}(t_p) + k \cdot \Delta J(t_0), \tag{2}$$

with

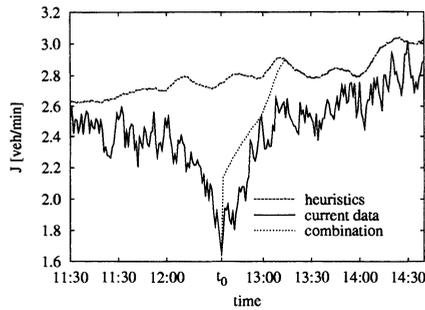
$$\begin{aligned} \Delta J(t_0) &= J_c(t_0) - J_{\text{dem}}(t_0), \\ k &= \begin{cases} \eta \left(1 - \frac{\Delta\tau}{\Delta\tau_{\text{max}}}\right), & \text{if } 0 < \Delta\tau \leq \Delta\tau_{\text{max}} \\ 0 & \text{if } \Delta\tau > \Delta\tau_{\text{max}} \end{cases} \\ \Delta\tau &= t_p - t_0. \end{aligned}$$

Obviously, for  $\Delta\tau > \Delta\tau_{\text{max}}$  the heuristics is used as forecast. The factor  $\eta$  is a coefficient determining the relevance of the current and historical data respectively. Reasonable is  $0 < \eta < 1$ ; for  $\eta = 0$  only the sample graph is used, for  $\eta = 1$  the prognosis curve starts at the current value.

In order to test this model for extreme values, data from the day of the solar eclipse has been chosen to demonstrate the quality of the model. Results are depicted in Fig. 3. There is a large anomaly of the traffic flow, which can not be predicted by heuristics. However, the combination of the short-term-forecast with the heuristics improves the quality of the prediction obviously.

## 6 Summary

In this paper a framework for performing forecasts with on-line simulations has been presented. The heart of the simulator is a microscopic traffic flow model



**Fig. 3.** The irruption of the traffic flow during the solar eclipse as an example for the combination of short-term- and long-term-forecasting. The current data lose their influence at a prognosis horizon  $\Delta\tau_{\max}$ . There it merges with the curve of the heuristics. The combination of both methods offer an improved prediction.

which is supplemented by the network structure and real-world traffic data stemming from about 3,500 loop detectors of the freeway network of North Rhine-Westphalia.

To get heuristics that are useful for forecasts a statistical analysis has been presented considering daily and seasonal differences. Taking special events into account a method is proposed, that combines current with historical data to supply both: short and long term prediction.

## References

1. B.S. Kerner, H. Rehborn, and M. Aleksic, 'Forecasting of traffic congestion', in: *Traffic and Granular Flow '99*, D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf, (Eds.), pp. 339–344, (Springer, Heidelberg, 2000).
2. R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg, 'Three categories of traffic data: Historical, current, and predictive', in: *Proceedings of the 9th IFAC Symposium Control in Transportation Systems 2000*, E. Schnieder and U. Becker, (Eds.), pp. 250–255, (IFAC, Braunschweig, 2000).
3. J. Esser and M. Schreckenberg, 'Microscopic simulation of urban traffic based on cellular automata', *Int. J. of Mod. Phys. C* **8**, 1025 (1997).
4. K. Nagel, J. Esser, and M. Rickert, 'Large-scale traffic simulations for transport planning', in: *Ann. Rev. of Comp. Phys. VII*, D. Stauffer, (Ed.), pp. 151–202, (World Scientific, Singapore, 2000).
5. M. Schreckenberg, L. Neubert, and J. Wahle, 'Simulation of traffic in large road networks', *Future Generation Computer Systems* **17**, 649–657 (2001).
6. K. Nagel and M. Schreckenberg, 'A cellular automaton model for freeway traffic', *J. Physique I* **2**, 2221–2229 (1992).
7. J. Wahle, O. Annen, C. Schuster, L. Neubert, and M. Schreckenberg, 'A dynamic route guidance system based on real traffic data', *Euro. J. Op. Res.* **131**, 302–308 (2001).

# On-Ramp Control

D. Huang

Department of Physics, Chung Yuan Christian University, Chung-li, Taiwan

**Abstract.** We study the phase transition on a highway induced by the fluctuations of on-ramp flow. The highway traffic is provided by a hydrodynamical model. We analyze the characteristics of perturbations to induce congestion near an on-ramp. The phase boundary is obtained. A scaling relation is revealed. We also analyze the time evolution of the local density profile. Conventional control mechanisms to regulate the on-ramp flow are examined. A new control scheme is proposed to suppress the congestion.

## 1 Introduction

The study of traffic flow near on-ramps of a highway system has revealed a rich spectrum of phenomena and attracted research interests from physicists recently [1–5]. On the section of a highway without ramps, the well formed traffic jams always propagate in the upstream direction while near an on-ramp a new type of congestion appears. The traffic jams seem to be confined. They move back and forth around the ramp with the same structure. Or the congestion may extend with time, but always have one end fixed at the ramp. Or they may even stay motionless. The current theoretical research is then focused on the characteristics of this new type of congestion [6–10]. To physicists, the highway traffic is basically a one-dimensional many-body system with strong correlations among vehicles. The system is driven far from equilibrium, where the steady states are characterized by non-vanishing flows. The congestion is taken as a non-trivial phase transition resulting from the instability of the system. In the paper, we study the hysteretic phase transitions induced by the fluctuations of on-ramp flow. The highway traffic is simulated by a well calibrated hydrodynamic model [11–13]. The resultant phase diagram is presented in the next section. The characteristic of the fluctuations is analyzed. In section 3, we study the time evolution of density profile in the emergence of congestion. Various control schemes aimed to suppress the congestion are discussed in the final section.

## 2 Phase Diagram

In the hydrodynamic model of highway traffic, the system is described by the following partial differential equations,

$$\rho \left( \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \right) = \frac{\rho}{\tau} [V(\rho) - v] - c_0^2 \frac{\partial \rho}{\partial x} + \mu \frac{\partial^2 v}{\partial x^2}, \quad (1)$$

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = q_i(t) \phi(x). \quad (2)$$

These two equations are analogous to the Navier-Stokes equation and the continuity equation, respectively. The highway traffic is described by the local density  $\rho(x, t)$  and the local velocity  $v(x, t)$ . The traffic flow is understood as  $q(x, t) = \rho(x, t) v(x, t)$ . The intrinsic properties on the main highway are prescribed by the safe-velocity functional  $V(\rho)$  and three parameters:  $\tau$ ,  $c_0$ ,  $\mu$ , which are constants related to the effects of relaxation, anticipation, and dampening, respectively [11–13]. The on-ramp flow is denoted by  $q_i(t)$  with a spatial distribution  $\phi(x)$  localized at  $x = 0$ . In this work [14], we perform extensive numerical simulations over a system of 60 km highway with an on-ramp right in the middle (position  $x = 0$ ).

When the vehicle density is small, the highway traffic is a homogeneous flow. In the upstream ( $x \ll 0$ ), the safe velocity is achieved and the flow  $q(x, t)$  is a constant denoted by  $q_{up}$ . The on-ramp flow  $q_{rmp}$  provides a small transition layer near the ramp, which transforms the upstream flow into the downstream flow. In the downstream ( $x \gg 0$ ), the safe velocity is also achieved and the flow is simply  $(q_{up} + q_{rmp})$ . The system is then specified by two parameters: the upstream flow  $q_{up}$  and the on-ramp flow  $q_{rmp}$ .

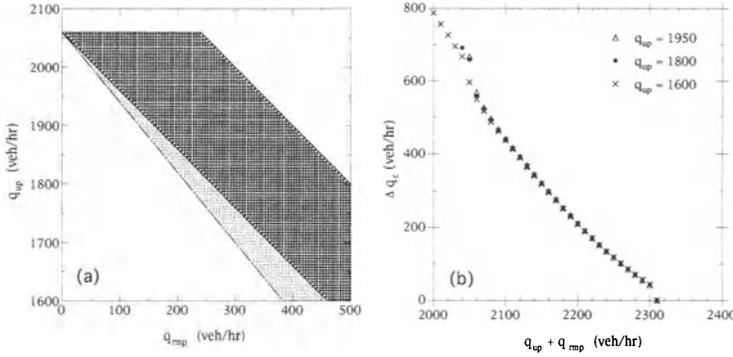
When the vehicle density increases, the congestion becomes one of the stable solutions. The transitions between the homogeneous flow and the congestion can be triggered by fluctuations of the on-ramp flow. We study the criterion of perturbations to induce such transitions. The system is prepared to start with a homogeneous flow specified by two parameters:  $q_{up}$  and  $q_{rmp}$ . To simulate the inflow of a large number of vehicles within a short period, the perturbations are prescribed in the following form,

$$q_i(t) = \begin{cases} q_{rmp} + \Delta q & t_0 < t < t_0 + \Delta t \\ q_{rmp} & \text{otherwise.} \end{cases} \quad (3)$$

The perturbations are characterized by two parameters:  $\Delta q$  and  $\Delta t$ , which specify the height and duration of the extra on-ramp flow, respectively. The initial time  $t_0$  is irrelevant as we start from a stationary state.

The induced phase transitions from the homogeneous flow to the congestion can be observed within a certain range of  $(q_{up}, q_{rmp})$ . The phase diagram is shown in Fig. 1(a). When  $q_{up} + 1.2 q_{rmp} < 2060$  vehicles per hour (veh/hr), the homogeneous flow is the only stable solution. The perturbations always decay with time. When  $q_{up} > 2060$  veh/hr, the congestion will not be confined to the ramp. The effects of an on-ramp become insignificant. When  $q_{up} + q_{rmp} > 2300$  veh/hr, the homogeneous flow becomes unstable. The congestion can be induced by a very small perturbation. Within these boundaries, the transitions can be induced when the perturbations are large enough, i.e., with a given  $\Delta t$ , a low bound of  $\Delta q$  is required to trigger the transition. This critical value is denoted by  $\Delta q_c$ , which depends on  $q_{up}$ ,  $q_{rmp}$ , and  $\Delta t$ . We further observe that within the region  $2060 \text{ veh/hr} < q_{up} + q_{rmp} < 2300 \text{ veh/hr}$  and  $q_{up} < 2060 \text{ veh/hr}$ , the critical value  $\Delta q_c$  is scaled with the downstream flow  $(q_{up} + q_{rmp})$ . i.e.,  $\Delta q_c$  depends only on  $\Delta t$  and the combination  $(q_{up} + q_{rmp})$ . The results are shown in Fig. 1(b). The critical height  $\Delta q_c$  decreases with the increase of both the

downstream flow ( $q_{up} + q_{rmp}$ ) and the duration  $\Delta t$ . It is also observed that  $\Delta q_c$  scales approximately with  $\sqrt{\Delta t}$  at a fixed ( $q_{up} + q_{rmp}$ ). Thus the quantity ( $\Delta q_c \sqrt{\Delta t}$ ) scales approximately with ( $q_{up} + q_{rmp}$ ).



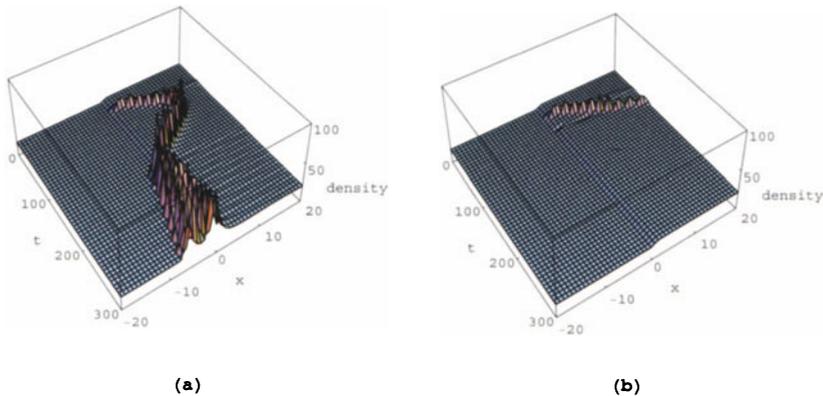
**Fig. 1.** (a) Phase diagram in the  $(q_{up}, q_{rmp})$  plane. The scaling region is shown by black dots; the non-scaling region is shown by the grey dots. (b) Critical extra flow  $\Delta q_c$  as a function of downstream flow  $(q_{up} + q_{rmp})$  for various values of  $q_{up}$ . The parameter  $\Delta t = 5$  min.

Outside the scaling region, the phase transitions can still be observed within the region of  $q_{up} + q_{rmp} > 2060$  veh/hr and  $q_{up} + 1.2 q_{rmp} < 2060$  veh/hr. Such small region can also be expressed as  $q_{rmp} < 2060 - q_{up} < 1.2 q_{rmp}$ . In such cases, the critical value  $\Delta q_c$  depends on  $q_{up}$  and  $q_{rmp}$  separately. When  $q_{up}$  approaches 2060 veh/hr, the phase transitions can only be triggered with a small range of  $q_{rmp}$ . As  $q_{up}$  decreases, a much wider range of  $q_{rmp}$  is supported. As the downstream flow is less than in the scaling region, the congestion can only be induced by a  $\Delta q_c$  larger than in the scaling region, which can be taken as the manifestation of boundary-induced phase transitions [17]. In contrast, the instability in the scaling region is mainly determined by the downstream flow.

### 3 Density Profile

Next, we consider the time evolution of density profiles in the emergence of congestion. The congestion near the on-ramp is caused by the inflow of a large amount of vehicles within a short period. With naive expectation, the appearance of traffic jams is attributed to the incapability of the main highway in dissipating the extra flow from the on-ramp. The extra on-ramp flow and the congestion are then expected to appear simultaneously. However, this naive expectation is valid only within the small range of non-scaling region, i.e.,  $q_{up} + q_{rmp} < 2060$  veh/hr and  $q_{up} + 1.2 q_{rmp} > 2060$  veh/hr. The lower bound of the extra flow  $\Delta q_c$  is quite large in these cases. When  $q_{up}$  is large, the congestion appears to be stationary; when  $q_{up}$  is small, the congestion expands to the upstream as time evolves.

In the scaling region, i.e.,  $2060 < q_{up} + q_{rmp} < 2300$  veh/hr, the congestion can be induced by a much smaller  $\Delta q_c$ . In such cases, quite contrary to the naive expectation, the traffic jams appear much later in time than that of the extra inflow. Delay of more than one hour is often observed. The typical results are shown in Fig. 2(a). Right after the extra inflow, a small traffic jam appears and propagates away from the ramp to the downstream. The traffic jam enlarges slowly and the speed of propagation is decreasing. At a certain time, the traffic jam turns around and propagates upstream, i.e., moving back to the on-ramp. The traffic jam enlarges quickly and a huge congestion is induced when it moves across the on-ramp. When the extra flow is only slightly larger than  $\Delta q_c$ , the small traffic jam can propagate a long distance before its turning around. The appearance of the congestion is then delayed to a much later time. In contrast, when the extra flow is less than  $\Delta q_c$ , the phase transition will not be induced. We can still observe the small traffic jam moving away from the ramp downstream. In such cases, the traffic jam propagates with a slowly increasing speed but decreasing amplitude. And it just disappears in a later time. The  $q_{up}$ -dependence is similar to the cases of the non-scaling region. When  $q_{up}$  is large, the congestion is observed to oscillate back and forth around the ramp; when  $q_{up}$  is small, the congestion expands upstream as time evolves.



**Fig. 2.** (a) The typical density profile  $\rho(x, t)$  in the scaling region. The parameters:  $q_{up} = 1800$  veh/hr,  $q_{rmp} = 310$  veh/hr,  $\Delta q = 420$  veh/hr,  $\Delta t = 5$  min, and  $t_0 = 30$  min. (b) The on-ramp is closed for 1 minute at  $t = 60$  min.

In the non-scaling region, a well formed traffic jam can not be sustained in the main highway, neither upstream nor downstream. The congestion can only emerge right at the ramp. In the scaling region, the traffic jams still can not be sustained upstream while they can be developed downstream. The congestion at the ramp can be related to a well formed traffic jam downstream. Thus the same phase transition can also be triggered by a traffic jam caused by fluctuations

downstream. As the traffic flow further increases, the traffic jams can also be developed upstream when  $q_{up} > 2060$  veh/hr. Then the congestion will no longer be confined to the vicinity of the on-ramp. The traffic jams developed downstream always move upstream. They just move across the on-ramp smoothly without inducing significant effects. The congestion caused by the extra inflow from on-ramp will soon move away from the ramp upstream.

## 4 Discussions

In this paper, we study the emergence of congestion induced by fluctuations of the inflow through the on-ramp. The congestion is not caused by random fluctuations. A definite type of perturbations is required to trigger the transitions. An extra flow within a short time is necessary, which is characterized by two parameters:  $\Delta q$  and  $\Delta t$ . To induce the congestion, the extra flow must be larger than the critical value  $\Delta q_c$ , which then depends on  $q_{up}$ ,  $q_{rmp}$ , and  $\Delta t$ . It is interesting to note that the criterion for a phase transition is not directly associated with the number of vehicles in the extra flow, which is equal to  $(\Delta q_c \Delta t)$ . We observe that the quantity  $(\Delta q_c \sqrt{\Delta t})$  scales with the downstream flow  $(q_{up} + q_{rmp})$ . At the same downstream flow, a larger number of vehicles is required when the perturbation has a longer duration. The scaling also implies that the same extra flow will induce the congestion at a smaller  $q_{rmp}$  when  $q_{up}$  is larger.

The congestion is induced, not spontaneously emerged. Within a wide range of flows, both the free flow and the congestion are stable. Thus appropriate control mechanisms can be employed to suppress the traffic jams. From the phase diagram of Fig. 1(a), the highway is free of congestion when  $q_{up} + 1.2 q_{rmp} < 2060$  veh/hr. With a larger upstream flow  $q_{up}$ , the on-ramp flow  $q_{rmp}$  should be restricted to a smaller value. The conventional on-ramp control is carried out by setting up a traffic light at the ramp before the intersection with the main highway. The traffic light switches periodically to red and green. By allocating different times to red and green phases, the average value of the on-ramp flow can be easily restricted. With the operation of the traffic light, the phase diagram is the same as shown in Fig. 1(a) with the parameter  $q_{rmp}$  denotes the average value of the on-ramp flow. In the non-scaling region, the phase transitions must be triggered with a very large value of  $\Delta q_c$ . Mathematically, it is correct, but impractical in reality. If we consider only the phase transitions in the scaling region, a softer restriction of  $q_{up} + q_{rmp} < 2060$  veh/hr will guarantee the highway to be free from the congestion. The operation of the traffic light should depend on the upstream flow. With the increase of  $q_{up}$ , one should decrease  $q_{rmp}$  by allocating more time to the red phase. To increase the capacity of a highway, one should find the way to limit the maximum value of the inflow fluctuations. As the downstream flow increases, the critical value  $\Delta q_c$  decreases. The congestion is easier to induce and the highway becomes more unstable. On the other hand, if one could limit the value of  $\Delta q_c$ , the restriction can be pushed toward the limit of  $q_{up} + q_{rmp} < 2300$  veh/hr.

The above control scheme based on the phase diagram is a conservative one. The highway traffic is maintained at the status where the congestion is impossible to emerge. The traffic flow is kept under a certain upper bound. Thus the capacity of a highway is limited. From the study of the density profile, we propose a new control scheme to eliminate the congestion more effectively. The small traffic jam propagating downstream with a decreasing speed can be taken as an early symptom of the later emerging congestion. By closing the on-ramp for a short period, say 30 seconds, the small traffic jam can be totally eliminated. Thus the congestion will not appear. The typical results are shown in Fig. 2(b). In this scheme, the timing is important. When the traffic jam is still developing and propagates downstream, it can be eliminated by closing the on-ramp shortly. When the traffic jam is well developed and starts to propagate upstream, it can only be eliminated by closing the ramp for a much longer period, say 10 minutes. Once the congestion appeared at the ramp, closing the ramp will no longer be an option. Usually, it will take more than half an hour to dissipate the congestion at this stage. In this scheme, closing the on-ramp temporarily provides an effective way to suppress the formation of congestion. The on-ramp is open until the symptom of congestion appeared. Then the ramp is closed for a short period. Thus the traffic flow on the highway can be supported at a much larger value. With the advance of Intelligent Transportation Systems (ITS), the highway traffic will be monitored more carefully. This makes the scheme possible. Whenever a small traffic jam propagating downstream with a decreasing speed is detected, closing the on-ramp temporarily will keep the highway free from congestion in a later time.

## References

1. D. Helbing and M. Treiber, *Science* **282**, 2001 (1998).
2. B.S. Kerner and H. Rehborn, *Phys. Rev. Lett.* **79**, 4030 (1997).
3. H.Y. Lee, H.W. Lee, and D. Kim, *Phys. Rev. Lett.* **81**, 1130 (1998).
4. D. Helbing and M. Treiber, *Phys. Rev. Lett.* **81**, 3042 (1998).
5. B.S. Kerner, *Phys. Rev. Lett.* **81**, 3797 (1998).
6. E. Tomer, L. Safonov, and S. Havlin, *Phys. Rev. Lett.* **84**, 382 (2000).
7. L. Safonov, E. Tomer, V.V. Strygin, and S. Havlin, *Physica* **A285**, 147 (2000).
8. N. Mitarai and H. Nakanishi, *Phys. Rev. Lett.* **85**, 1766 (2000).
9. P. Nelson, *Phys. Rev.* **E61**, R6052 (2000).
10. T. Nagatani, *Physica* **A280**, 602 (2000).
11. B.S. Kerner and P. Konhäuser, *Phys. Rev.* **E48**, R2335 (1993).
12. B.S. Kerner and P. Konhäuser, *Phys. Rev.* **E50**, 54 (1994).
13. B.S. Kerner, S.L. Klenov, and P. Konhäuser, *Phys. Rev.* **E56**, 4200 (1997).
14. The following parameters are adopted[15,16]:  $\tau = 0.5$  min,  $c_0 = 54$  km/hr,  $\mu = 600$  veh km/hr,  $V(\rho) = V_0 (1 - \rho/\rho_0)/[1 + 100 (\rho/\rho_0)^4]$ ,  $\phi(x) = \exp[-x^2/(2\sigma^2)]/(\sqrt{2\pi}\sigma)$ , where  $V_0 = 120$  km/hr and  $\rho_0 = 140$  veh/km are the maximum values for velocity and density, respectively, and  $\sigma = 60$  m denotes the length of the on-ramp.
15. H.Y. Lee, H.W. Lee, and D. Kim, *Phys. Rev.* **E59**, 5101 (1999).
16. H.Y. Lee, H.W. Lee, and D. Kim, *Phys. Rev.* **E62**, 4737 (2000).
17. M.C. Chou and D.W. Huang, *Phys. Rev.* **E63**, 056106 (2001).

# The Influence of Tollbooths on Highway Traffic

D. Huang and W. Huang

Department of Physics, Chung Yuan Christian University, Chung-li, Taiwan

**Abstract.** We study the effects of tollbooths on the traffic flow. Various types of toll collection are examined, which can be characterized either by a waiting time or a reduced speed. The phase separation results in a saturated flow, which is observed as a plateau region in the fundamental diagram. The effects of lane expansion near the tollbooth are examined. The full capacity of a highway can be restored. Besides the numerical results, we also obtain analytical expressions for various quantities.

## 1 Introduction

Recently, traffic problems have attracted considerable attention from physicists [1,2]. The highway traffic is basically a many-body system of strongly interacting vehicles. The emergence of congestion can be taken as a phase transition in a one-dimensional nonequilibrium system [3]. From our daily experience, traffic jams occur more often near a tollbooth than in other parts of the highway. In this paper, we investigate how the presence of tollbooths affects traffic flow. The tollbooths on a highway serve two purposes: to collect tolls and to regulate the flow. Besides to increase revenue for governments, road pricing is also considered to be one of the most efficient approaches to reduce congestion and has been investigated currently by both economists and transportation researchers [4]. In the following, we will focus on the effects in regulating the traffic flow.

In this paper, we study the traffic flow and the phase transition in the presence of tollbooths. We simulate the traffic flow on a single lane highway by using the Nagel-Schreckenberg model [5]. The cellular automaton approach has succeeded in reproducing realistic features of highway traffic. The model will be briefly reviewed in the next section. Different types of toll collection will be implemented. The effects of the conventional lane expansion near the tollbooth will be studied in Section 3. The conclusions are summarized in the last section.

## 2 Model

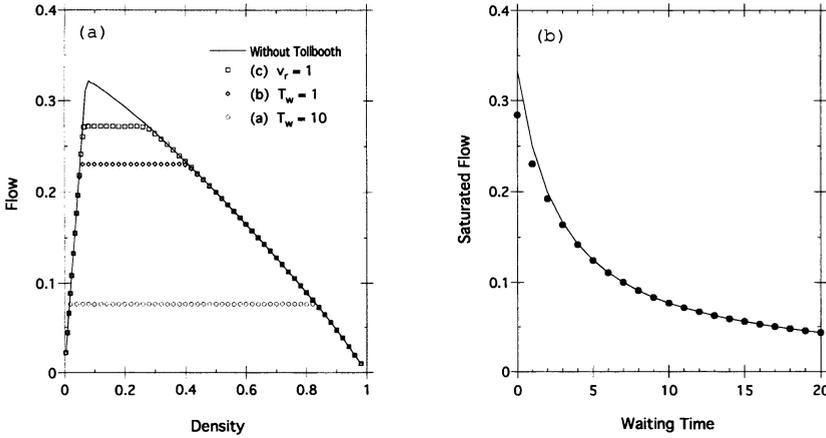
In the basic model, the one-dimensional highway is divided into  $L$  sites. Each site can be either empty or occupied by a vehicle with an integer speed  $v \in \{0, 1, \dots, v_{max}\}$  where  $v_{max}$  is the speed limit. With periodic boundary conditions, the number of vehicles is conserved. At each time step, the configuration of  $N$  vehicles is updated by the following four rules, which are applied in parallel to all vehicles. The first rule is the acceleration. If the speed of a vehicle is lower

than  $v_{max}$ , the speed is advanced by one. The second rule is the slowing down due to other vehicles. If a vehicle has  $d$  empty sites in front of it and a speed larger than  $d$ , the speed is reduced to  $d$ . The third rule is the randomization, which introduces a noise to simulate stochastic driving behavior. The speed of a moving vehicle ( $v \geq 1$ ) is decreased by one with a braking probability  $p$ . In the fourth rule, the position of a vehicle is shifted by its speed  $v$ . Iterations over these simple rules already give realistic results. The model contains three parameters: the maximum speed  $v_{max}$ , the braking probability  $p$ , and the average density  $\rho = N/L$ . To simulate the real traffic on highway, the distance between two sites is set to 7.5 meter; the update time-step is 1 second. The speed limit  $v_{max} = 5$  corresponds to 135 km/hr. The braking probability  $p = 0.5$  provides the reasonable fluctuations. The single lane highway has the maximum flow  $f_{max} \approx 0.32$  at the density  $\rho \approx 0.08$ , which corresponds to a maximum flow of 1200 vehicle/hr at a density of 10 vehicle/km.

A selected site is designated as the tollbooth, on which each vehicle has to stop and pay the toll. We study three different types of conventional toll-collection. In case (a), a driver pays the toll by cash and has to wait for change and the receipt. It takes about 10 seconds for each vehicle to wait on the tollbooth. In case (b), a driver just hands over the prepaid ticket. Without waiting for change and the receipt, the process can be finished by 1 second. In case (c), with the electrical toll system (ETS), the vehicle is not required to stop but can pass the tollbooth with a reduced speed of about 30 km/hr. In the simulations, a waiting time of  $T_w = 10$  is assigned to case (a) whenever a vehicle occupies the site of the tollbooth. Similarly, in case (b), a waiting time of  $T_w = 1$  is assigned. In case (c), a reduced speed of  $v_r = 1$  is enforced when the vehicle passes the tollbooth. The numerical results are shown in Fig. 1(a). The effect of a tollbooth can be easily observed as the plateau in the fundamental diagram. The three different types of toll collection give qualitatively the same result. The presence of a tollbooth sets the traffic flow to a saturated value in the middle range of the density. Case (a) gives the lowest value of the saturated flow and case (c) the highest one. When the traffic flow is less than the saturated flow, either in the very low density or the very high density region, it follows the fundamental diagram without a tollbooth.

When the density is low, vehicles move freely and independently. The presence of a tollbooth will not reduce the traffic flow. When the density increases over a critical value  $\rho_{c1}$ , traffic jams begin to emerge. A phase separation is observed. The drivers experience the congestion before the tollbooth; after passing the tollbooth, vehicles move freely. As the density continues to increase, the traffic flow remains in the saturated value  $f_s$ . The vehicle queue before the tollbooth increases accordingly. The phase separation disappears when the density further increases over the critical value  $\rho_{c2}$ . The traffic jams become dominant. The traffic flow decreases with the increase of density. The effects of a tollbooth become negligible again.

Within the plateau region, the behavior of the local densities can be well described by the Maxwell construction of the first order phase transition. The



**Fig. 1.** (a) The fundamental diagram for various types of toll collection. Parameter  $L = 2000$ . (b) Saturated flow  $f_s$  as a function of waiting time  $T_w$ . The analytical results from Eq. (1) are shown by the solid line.

following analytical expressions can be obtained,

$$f_s = \frac{1 - p}{(T_w + 1)(1 - p) + 1}, \tag{1}$$

$$\rho_{c1} = \frac{(1 - p)(v_{max} - p)}{(T_w + 1)(1 - p) + 1}, \tag{2}$$

$$\rho_{c2} = \frac{(T_w + 1)(1 - p)}{(T_w + 1)(1 - p) + 1}. \tag{3}$$

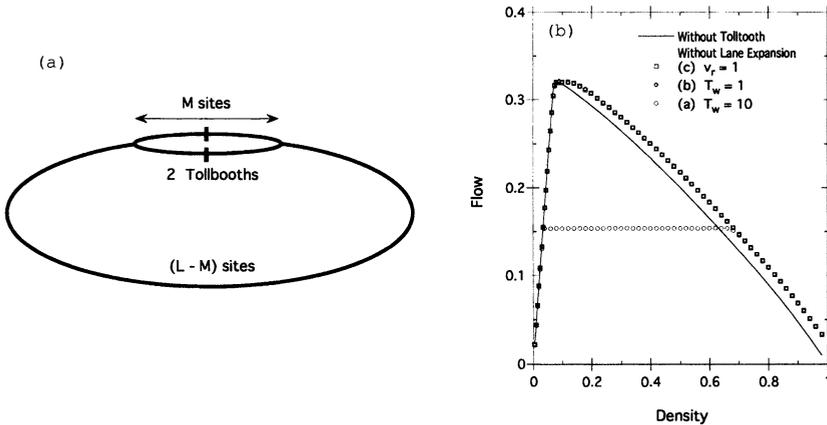
The numerical results can be well described, see Fig 1(b). When  $T_w$  is small, the overestimate of  $f_s$  is related to the overestimate of traffic flow in the mean-field approximation of. The length of vehicle queues before the tollbooth can also be calculated analytically. The presence of a tollbooth introduces a first order phase transition when the density reaches a critical value. Long vehicle queues appear before the tollbooth. As the phase transition is a thermodynamical property of the bulk, the appearance of traffic jams before the tollbooth also reflects the property of the whole lattice. As shown in the above equation, the solution  $L_q$  is always proportional to  $L$ . Thus the length of a vehicle queue is not measured by the number of sites, but by the fraction of lattice. With a given density above the critical one, the vehicle queue occupies a fixed fraction of the lattice, i.e., on a larger lattice appears a longer vehicle queue.

In cases where vehicles are not required to stop on the tollbooth, as in case (c), a reduced speed of  $v_r \geq 2$  has only negligible effects on the fundamental diagram. It is interesting to observed that the case of  $v_r = 1$  is very similar to the case of  $T_w = 0$ . However, the saturated flow in the former is slightly less than that of the latter.

As a single tollbooth already saturates the traffic flow, the presence of more tollbooths has only negligible effects on the fundamental diagram. The distribution of the tollbooths is irrelevant, provided the distance between two tollbooths is not too close, say less than 10 sites. The traffic flow is controlled by the density. Beyond the critical density, the total length of a vehicle queues is fixed, with fixed density and the lattice size. However, the length of a vehicle queue before each tollbooth is mainly determined by the initial configuration.

### 3 Lane Expansion

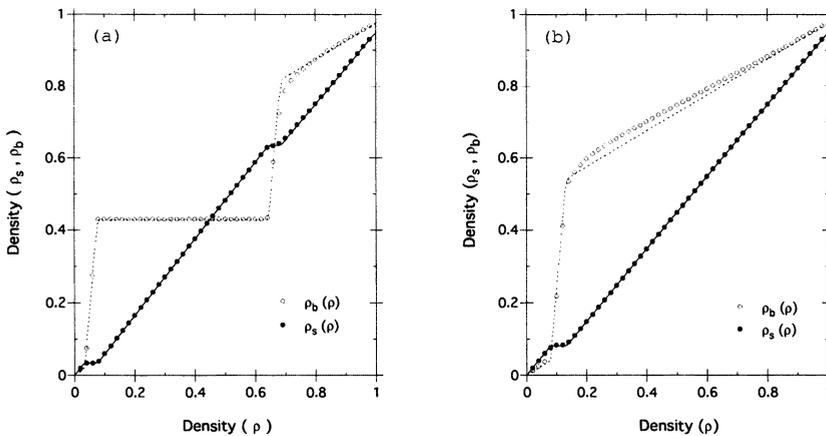
The presence of a tollbooth sets an upper bound to the traffic flow. The saturated flow is much less than the maximum capacity of the highway, especially in case (a). To restore the full capacity of a highway, conventionally two tollbooths per lane are set up to enhance the traffic flow. The system configuration is shown in Fig. 2(a). Near the tollbooths, the single lane highway is divided into two branches. On each branch, a tollbooth is set up. After the tollbooths, the two branches merge into a single lane. Parameter  $M$  denotes the length of the lane expansion. The length of the single-lane part of the highway becomes  $(L - M)$ . The global density is still denoted by  $\rho = N/L$ , where  $N$  is the number of vehicles. The resultant traffic flow is shown in Fig. 2(b). With naive expectation and comparing to Fig. 1(a), the saturated flow should be doubled, as in case (a). In cases (b) and (c), the maximum flow is restored since the saturated flow shown in Fig. 1(a) is larger than a half of the maximum.



**Fig. 2.** (a) System configuration of the lane expansion near the tollbooths. (b) The fundamental diagram with lane expansion for various types of toll collection. Parameters  $M = 100$  and  $L = 1900$ .

In case (a), a plateau is still observed in the fundamental diagram. As the height of the plateau is doubled, the width of the plateau is shortened. We denote

the average density on the single-lane part as  $\rho_s$ . The average densities on the separated two branches are equal and denoted as  $\rho_b$ . The numerical results are shown in Fig. 3(a), which can be well described by mean-field approximation. In the extreme densities, very low or very high, both  $\rho_s$  and  $\rho_b$  increase linearly with the increase of  $\rho$ . However, in the low density phase, we have  $\rho_s > \rho_b$  while in the high density phase, we have  $\rho_s < \rho_b$ . In the middle range of density, their behavior can be further divided into two portions. In the first portion, the vehicle queueing is confined within the separated branches; as  $\rho$  increases,  $\rho_s$  remains constant while  $\rho_b$  increases quickly. In the second portion, the vehicle queue is extended into the single lane highway; as  $\rho$  increases,  $\rho_s$  begins to increase again while  $\rho_b$  remains constant.



**Fig. 3.**  $\rho_s$  and  $\rho_b$  as functions of  $\rho$ , parameters  $M = 100$  and  $L = 1900$ . The analytical results for  $\rho_s$  and  $\rho_b$  are shown by the solid line and dotted line, respectively. (a) The case of  $T_w = 10$ . (b) The case of  $T_w = 1$ , or  $v_r = 1$ .

In cases (b) and (c), the maximum flow is restored. The effects of the tollbooths can be neglected. There is no phase separation. The free flow in the low density region transforms into the congestion phase in the high density region directly. The numerical results are shown in Fig. 3(b), which can also be well described by the analytical results. Similarly, we have  $\rho_s > \rho_b$  in the low density region and  $\rho_s < \rho_b$  in the high density region. In the transition region,  $\rho_b$  increases much faster than  $\rho_s$  does, which can be well approximated by a constant  $\rho_s$  and an linearly increasing  $\rho_b$ . The underestimate of  $\rho_b$  in the middle range of density is attributed to the overestimate of traffic flow in the mean-field approximation.

## 4 Conclusions

The setup of a tollbooth influences the traffic flow significantly. As no vehicles are allowed to bypass the tollbooth, the effect is not short ranged but can extend to the whole lattice especially when periodic boundary conditions are enforced. The tollbooth is characterized either by a waiting time  $T_w$  or a reduced speed  $v_r$ , determined by whether a vehicle were required to stop on the tollbooth. For  $v_r \geq 2$ , the traffic flow is unaffected. On the contrary, even  $T_w = 0$  will reduce the traffic flow significantly. The tollbooth serves as a bottleneck on the highway and the traffic flow is limited to a saturated value. For extreme densities, the effect of the tollbooth can be neglected. In the middle range of density, a plateau is observed in the fundamental diagram. Similar effects had also be reported for the highway with defects [6]. The synchronized flow near the on-ramp is another example, where the traffic flow remains constant with the increase of density [7]. Even the pedestrian flow reveals the same effect [8]. As a single tollbooth has already saturated the traffic flow, increasing the number of tollbooth has only negligible effects.

As the tollbooth represents a bottleneck, the reduction of traffic flow can be effectively restored by lane expansion near the tollbooth. With the conventional two tollbooths per lane, the maximum flow can be fully restored for  $v_r \geq 1$  or  $T_w \geq 3$ . Lane expansion also serves as a buffer to maintain the maximum flow over a wider range of density. The traffic flow in the high density region is enhanced accordingly.

## References

1. D. Chowdhury, L. Santen, and A. Schadschneider, *Phys. Rep.* **329**, 199 (2000).
2. D. Helbing, H.J. Herrmann, and M. Schreckenberg (Eds.), *Traffic and Granular Flow '99*, New York: Springer Verlag, 2000;  
M. Schreckenberg and D.E. Wolf (Eds.), *Traffic and Granular Flow '97*, New York: Springer Verlag, 1998;  
D.E. Wolf, M. Schreckenberg, and A. Bachem (Eds.), *Traffic and Granular Flow*, Singapore: World Scientific, 1996.
3. V. Privman (Ed.), *Nonequilibrium Statistical Mechanics in One Dimension*, Cambridge University Press, Cambridge, 1997.
4. K.A. Small, *Transportation* **19**, 287 (1992);  
P. Ferrari, *Transportation Research*, **B29** 357 (1995);  
H. Yan and W.H.K. Lam, *Transportation Research* **A30**, 319 (1996).
5. K. Nagel and M. Schreckenberg, *J. Phys. I (France)* **2**, 2221 (1992);  
M. Schreckenberg, A. Schadschneider, K. Nagel, and N. Ito, *Phys. Rev.* **E51**, 2939 (1995).
6. D.W. Huang, *Chin. J. Phys.* **36**, 553 (1998);  
M. Bengrine, A. Benyoussef, H. Ez-Zahraouy, and F. Mhirech, *Phys. Lett.* **A253**, 135 (1999).
7. B.S. Kerner and H. Rehborn, *Phys. Rev. Lett.* **79**, 4030 (1997);  
B.S. Kerner, *Phys. Rev. Lett.* **81**, 3797 (1998);  
B.S. Kerner, *J. Phys. A* **33**, L221 (2000).
8. Y. Tajima, K. Takimoto, and T. Nagatani, *Physica* **A294**, 257 (2001).

# **Networks / Internet**

# Packet Transport and Load Distribution in Scale-Free Networks

K.-I. Goh, B. Kahng, and D. Kim

School of Physics and Center for Theoretical Physics, Seoul National University,  
Seoul 151-747, Korea

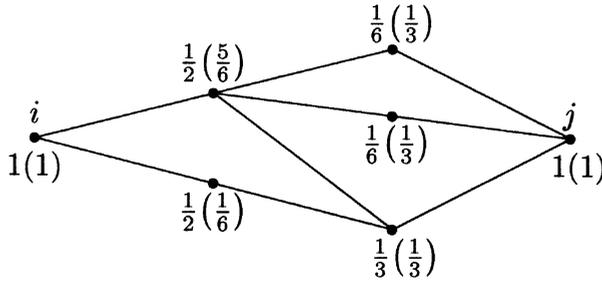
**Abstract.** We study a problem of data packet transport in scale-free networks whose degree distribution follows a power-law with the exponent  $\gamma$ . We define load at each vertex as the accumulated total number of data packets passing through that vertex when every pair of vertices sends and receives a data packet along the shortest path connecting the pair. It is found that the load distribution follows a power-law with the exponent  $\delta \approx 2.2(1)$ , insensitive to different values of  $\gamma$  in the range,  $2 < \gamma \leq 3$ , and different mean degrees, which is valid for both undirected and directed cases. Thus, we conjecture that the load exponent is a universal quantity to characterize scale-free networks.

Complex systems consist of many constituents such as individuals, substrates, and companies in social, biological, and economic systems, respectively, showing cooperative phenomena between constituents through diverse interactions and adaptations to the pattern they create [1,2]. Interactions may be described in terms of graphs, consisting of vertices and edges, where vertices (edges) represent the constituents (their interactions). This network approach was initiated by Erdős and Rényi (ER) [3]. In the ER model, the number of vertices is fixed, while edges connecting one vertex to another occur randomly with certain probability. However, the ER model is too random to describe real complex systems. Recently, Watts and Strogatz (WS) [4] introduced a small-world network, where a fraction of edges on a regular lattice is rewired with probability  $p_{WS}$  to other vertices, and compared its properties with those of real-world networks such as actor networks, power grid networks, and the neural network of *C. Elegans*. More recently, Barabási and Albert (BA) [5–7] introduced an evolving network where the number of vertices  $N$  increases linearly with time rather than fixed, and a newly introduced vertex is connected to  $m$  already existing vertices, following the so-called preferential attachment rule. The number of edges  $k$  incident upon a vertex is called the degree of the vertex. In the BA model, the probability for the new vertex to connect to an already existing vertex is proportional to the degree  $k$  of the selected vertex. Then the degree distribution  $P_D(k)$  follows a power-law  $P_D(k) \sim k^{-\gamma}$  with  $\gamma = 3$  for the BA model, while for the ER and WS models, it follows a Poisson distribution. Networks whose degree distribution follows the power-law are ubiquitous in real-world networks such as the world-wide web [8–10], the Internet [11–13], the citation network [14] and the author collaboration network [15] of scientific papers, and the metabolic networks in biological organisms [16]. On the other hand, there also exist random networks

such as the actor network whose degree distribution follows a power-law but has a sharp cut-off in its tail [17]. Thus, it has been proposed that the degree distribution can be used to classify a variety of diverse real-world networks [17]. The power-law behavior implies absence of a characteristic scale, so that networks following a power-law in the degree distribution are called scale-free (SF) networks [18]. In SF networks, one may wonder if the exponent  $\gamma$  is universal in analogy with the theory of critical phenomena; however, the exponent  $\gamma$  turns out to be sensitive to the detail of network structure. Thus, a universal quantity for SF networks is yet to be found.

A common feature between the WS and SF networks would be the small-world property that the mean separation between two vertices, averaged over all pairs of vertices (called the diameter hereafter), is shorter than that of a regular lattice. Even for small values of  $p_{WS}$  in the WS model, the diameter  $d$  scales with the number of vertices  $N$  logarithmically as in the ER random graph;  $d \sim \log N$  [19]. For SF networks, the diameter also shows the small-world behavior. For example, the diameter of the world-wide web is proportional to logarithm of the total number of web-pages [8]. Even more, for the metabolic networks in biological organisms, the diameter is independent of system size, *i.e.*, the number of vertices, making the biological organisms react to changing external conditions as quickly as possible [16]. As a whole, diameters of many complex networks are small, allowing objects transmitted through the network such as neural spikes on neural network, or data packets on Internet, to travel from one vertex to another quickly along the shortest path. The small-world property in SF networks results from the presence of a few vertices with high degree. In particular, the hub, the vertex whose degree is the largest, plays a dominant role in reducing the diameter of the system. Since the role of the hub in transport is much more dominant, compared with the contribution by the normalized degree, we study topological feature of shortest paths between two vertices. The transport from one position to another is mainly carried along the shortest path(s) between them. When a data packet is sent from one vertex to another through SF networks such as Internet, it is efficient to take a road along the shortest path between the two. Then vertices with higher degrees should be heavily loaded and jammed by lots of data packets passing along the shortest paths. To prevent such Internet traffic congestions, and allow data packets to travel in a free-flow state, one has to enhance the capacity, the rate of data transmission, of each vertex to the extent that the capacity of each vertex is large enough to handle appropriately defined "load."

In this manuscript, we define and study such a quantity, which we simply call load, to characterize the transport dynamics in SF networks. To be specific, we suppose that a data packet is sent from a vertex  $i$  to  $j$ , for every ordered pair of vertices  $(i, j)$ . For a given pair  $(i, j)$ , it is transmitted along the shortest path between them. If there exist more than one shortest paths, the data packet would encounter one or more branching points. In this case, we assume that the data packet is divided evenly by the number of branches at each branching point as it travels. A simple example of defining the load due to a unit packet transfer

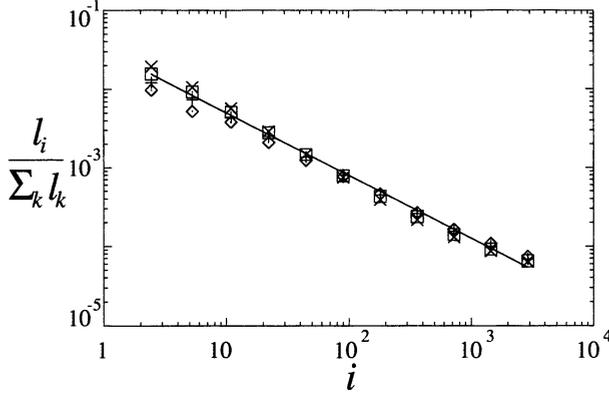


**Fig. 1.** The load at each vertex due to a unit packet transfer from the vertex  $i$  to the vertex  $j$  (from  $j$  to  $i$ ). In this diagram, only the vertices along the shortest paths between  $(i, j)$  are shown.

between a given pair  $(i, j)$  is depicted in Fig. 1. Note that the contribution to the load from the path  $i \rightarrow j$  may be different from that of  $j \rightarrow i$  even for the undirected network. Then we define the load  $\ell_k$  at a vertex  $k$  as the total amount of data packets passing through that vertex  $k$  when all pairs of vertices send and receive one unit of data packet between them. Here, we do not take into account the time delay of data transfer at each vertex or edge, so that all data are delivered in a unit time, regardless of the distance between any two vertices. However, the result we obtained is valid even for the case when data travel with constant speed, so that the time delay of data transfer is proportional to the distance between two vertices. We pay attention to the probability distribution of the load for various SF networks. We find numerically that the load distribution  $P_L(\ell)$  follows a power-law  $P_L(\ell) \sim \ell^{-\delta}$ . Moreover, the exponent  $\delta \approx 2.2$  we obtained is insensitive to the detail of the SF network structure as long as the degree exponent is in the range,  $2 < \gamma \leq 3$ . When  $\gamma > 3$ ,  $\delta$  increases as  $\gamma$  increases, however. The universal behavior is also valid for directed networks, when  $2 < \{\gamma_{in}, \gamma_{out}\} \leq 3$ . Since the degree exponents in most of real-world SF networks satisfy  $2 < \gamma \leq 3$ , the universal behavior is interesting.

We construct a couple of classes of SF networks both in the static and evolving ways. Each class of networks includes a control parameter, according to the degree exponent. First, we deal with the static case. There are  $N$  vertices in the system from the beginning, which are indexed by an integer  $i$  ( $i = 1, \dots, N$ ). We assign the weight  $p_i = i^{-\alpha}$  to each vertex, where  $\alpha$  is a control parameter in  $[0, 1)$ . Next, we select two different vertices  $(i, j)$  with probabilities equal to the normalized weights,  $p_i / \sum_k p_k$  and  $p_j / \sum_k p_k$ , respectively, and add an edge between them unless one exists already. This process is repeated until  $mN$  edges are made in the system. Then the mean degree is  $2m$ . Since edges are connected to a vertex with frequency proportional to the weight of that vertex, the degree at that vertex is given as

$$\frac{k_i}{\sum_j k_j} \approx \frac{(1 - \alpha)}{N^{1-\alpha} i^\alpha}, \tag{1}$$



**Fig. 2.** Plot of the normalized load  $l_i / \sum_k l_k$  versus vertex index  $i$  in double logarithmic scales for the scale-free networks with different degree exponents  $\gamma = 2.25$  ( $\times$ ),  $2.5$  ( $\square$ ),  $2.75$  ( $+$ ), and  $3.0$  ( $\diamond$ ). The solid line is the linear fit and has a slope  $-0.80$ . Simulations are performed for  $N = 10,000$  and  $m = 2$  and all data points are averaged over 10 configurations.

where  $\sum_j k_j = 2mN$ . Then it follows that the degree distribution follows the power-law,  $P_D(k) \sim k^{-\gamma}$ , where  $\gamma$  is given by

$$\gamma = (1 + \alpha) / \alpha. \tag{2}$$

This is a generic scaling noticed by Dorogovtsev *et al.* [20]. Thus, adjusting the parameter  $\alpha$  in  $[0,1)$ , we can obtain various values of the exponent  $\gamma$  in the range,  $2 < \gamma < \infty$ .

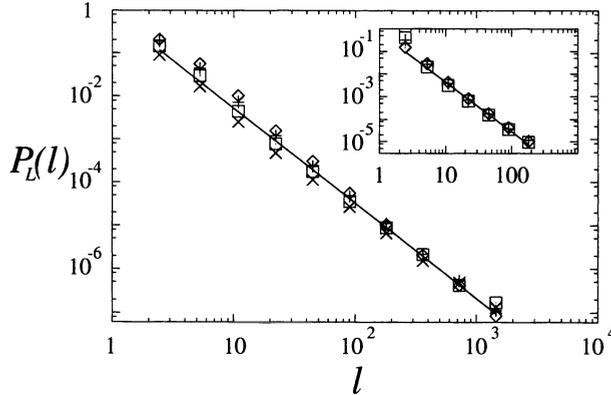
Once a SF network is constructed, we select an ordered pair of vertices  $(i, j)$  on the network, and identify the shortest path(s) between them and measure the load on each vertex along the shortest path using the modified version of the breath-first search algorithm introduced by Newman [21]. We have measured the load  $l_i$  for the networks with various  $\gamma$ . It is found numerically that the load  $l_i$  follows the formula,

$$\frac{l_i}{\sum_j l_j} \sim \frac{1}{N^{1-\beta} i^\beta}, \tag{3}$$

with  $\beta = 0.80(5)$ . This value of  $\beta$  is insensitive to different values of the exponent  $\gamma$  in the range,  $2 < \gamma \leq 3$  as shown in Fig. 2. The total load,  $\sum_j l_j$  scales as  $\sim N^2 \log N$ . This is because there are  $N^2$  pairs of vertices in the system and the sum of the load contributed by each pair of vertices is equal to the distance between the two vertices, which is proportional to  $\log N$ . Therefore, the load  $l_i$  at a vertex  $i$  is given as

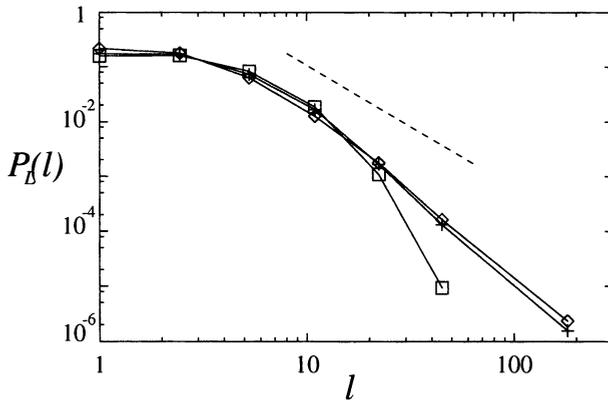
$$l_i \sim (N \log N)(N/i)^\beta. \tag{4}$$

From (4), it follows that the load distribution scales as  $P_L(\ell) \sim \ell^{-\delta}$ , with  $\delta = 1 + 1/\beta \approx 2.2(1)$ , independent of  $\gamma$  in the range,  $2 < \gamma \leq 3$ . Direct measure of

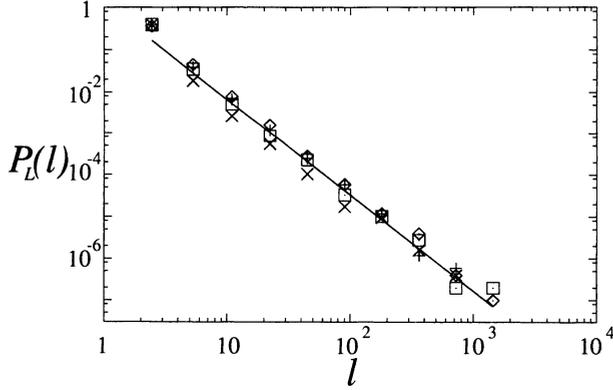


**Fig. 3.** Plot of the load distribution  $P_L(\ell)$  versus  $\ell$  for various  $\gamma = 2.25$  ( $\times$ ),  $2.5$  ( $\square$ ),  $2.75$  ( $+$ ) and  $3.0$  ( $\diamond$ ) in double logarithmic scales. The linear fit (solid line) has a slope  $-2.2$ . Simulations are performed for  $N = 10,000$  and  $m = 2$  and all data points are averaged over 10 configurations. Inset: Plot of  $P_L(\ell)$  versus  $\ell$  for different  $m = 2, 4$  and  $6$ , but for the same  $\gamma = 2.5$ .

$P_L(\ell)$  also gives  $\delta \approx 2.2(1)$  as shown in Fig. 3. We also checked  $\delta$  for different mean degrees  $m = 2, 4$  and  $6$ , but obtained the same value,  $\delta \approx 2.2(1)$  as shown in the inset of Fig. 3. Thus, we conclude that the exponent  $\delta$  is a generic quantity for this network. However, for  $\gamma > 3$ , the exponent  $\delta$  depends on the exponent  $\gamma$  in a way that it increases as  $\gamma$  increases. Eventually, the load distribution decays exponentially for  $\gamma = \infty$  as shown in Fig. 4. Thus, the transport properties of the SF networks with  $\gamma > 3$  are fundamentally different from those with  $2 < \gamma \leq 3$ .



**Fig. 4.** Plot of the load distribution  $P_L(\ell)$  versus  $\ell$  in double logarithmic scales for  $\gamma = 4$  ( $\diamond$ ),  $5$  ( $+$ ), and  $\infty$  ( $\square$ ). The dotted line having a slope  $-2.2$  is drawn to compare the data with the case for  $2 < \gamma \leq 3$ . Simulations are performed for  $N = 10,000$  and  $m = 2$  and all data points are averaged over 10 configurations.



**Fig. 5.** Plot of the load distribution  $P_L(\ell)$  versus  $\ell$  for the evolving model. The data are obtained for  $\gamma = 2.25$  ( $\times$ ),  $2.5$  ( $\square$ ),  $2.75$  ( $+$ ) and  $3.0$  ( $\diamond$ ). The fitted line has a slope  $-2.2$ .

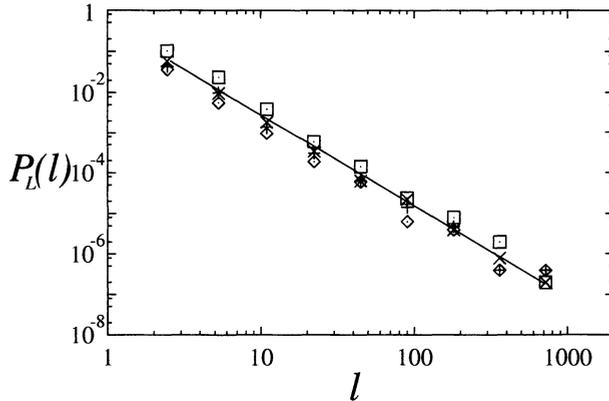
This is probably due to the fact that for  $\gamma > 3$ , the second moment of  $P_D(\ell)$  exists, while for  $\gamma \leq 3$ , it does not. Note that (1) and (4) combined give a scaling relation between the load and the degree for this network as

$$\ell \sim k^{(\gamma-1)/(\delta-1)}. \quad (5)$$

Thus, when and only when  $\gamma = \delta$ , the load at each vertex is directly proportional to its degree. Otherwise, it scales nonlinearly.

We generate SF networks in an evolving way according to the method proposed by Kumar *et al.* [22], which is similar to the method proposed by Simon in 1955 in their idea [23,24]. The stochastic rule includes two ingredients, the duplication and the mutation. At each time step, a new vertex is introduced and it creates  $m$  edges connecting to existing vertices by the following rule: Select an existing vertex randomly. Associated with it are  $m$  vertices to which edges were added previously at its creation. Add an edge to the selected vertex or to any one of the associated vertices with probability  $p_K$  and  $1 - p_K$ , respectively. Repeat this  $m$  times. The network generated in this way exhibits a power-law in its degree distribution, where the degree exponent is given by  $\gamma = (2 - p_K)/(1 - p_K)$ . The BA model is the case when  $p_K = 0.5$ . Through this model, we also obtain the load exponent  $\delta \approx 2.2$  for different values of the degree exponent in  $2 < \gamma \leq 3$  as shown in Fig. 5, which confirms the previous result. The load-degree scaling, (5), is also satisfied.

Next, we consider the case of directed SF network. The directed SF networks are generated following the static rule. In this case, we assign two weights  $p_i = i^{-\alpha_{\text{out}}}$  and  $q_i = i^{-\alpha_{\text{in}}}$  ( $i = 1, \dots, N$ ) to each vertex for outgoing and incoming edges, respectively. Both control parameters  $\alpha_{\text{out}}$  and  $\alpha_{\text{in}}$  are in the interval  $[0, 1)$ . Then two different vertices  $(i, j)$  are selected with probabilities,  $p_i / \sum_k p_k$

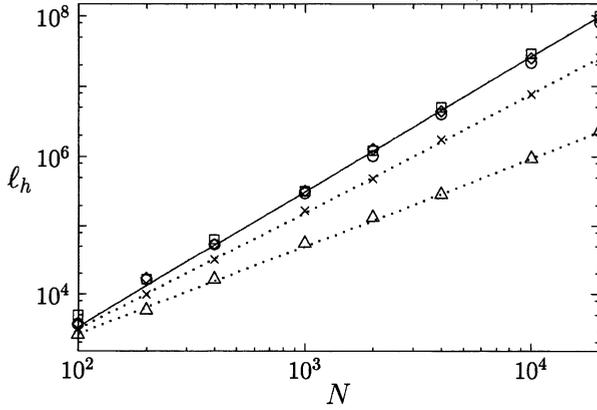


**Fig. 6.** Plot of the load distribution  $P_L(\ell)$  versus  $\ell$  for the directed case. The data are obtained for  $(\gamma_{\text{in}}, \gamma_{\text{out}}) = (2.1, 2.3)$  ( $\diamond$ ),  $(2.1, 2.7)$  ( $+$ ),  $(2.5, 2.7)$  ( $\square$ ) and  $(2.5, 2.2)$  ( $\times$ ). The fitted line has a slope  $-2.3$ .

and  $q_j / \sum_k q_k$ , respectively, and an edge from the vertex  $i$  to  $j$  is created with an arrow,  $i \rightarrow j$ . The SF networks generated in this way show the power-law in both outgoing and incoming degree distributions with the exponents  $\gamma_{\text{out}}$  and  $\gamma_{\text{in}}$ , respectively. They are given as  $\gamma_{\text{out}} = (1 + \alpha_{\text{out}}) / \alpha_{\text{out}}$  and  $\gamma_{\text{in}} = (1 + \alpha_{\text{in}}) / \alpha_{\text{in}}$ . Thus, choosing various values of  $\alpha_{\text{out}}$  and  $\alpha_{\text{in}}$ , we can determine different exponents  $\gamma_{\text{out}}$  and  $\gamma_{\text{in}}$ . Following the same steps as for the undirected case, we obtain the load distribution on the directed SF networks. The load exponent  $\delta$  obtained is  $\approx 2.3(1)$ , as shown in Fig. 6, consistent with the one for the undirected case, also being independent of  $\gamma_{\text{out}}$  and  $\gamma_{\text{in}}$  in  $2 < \{\gamma_{\text{out}}, \gamma_{\text{in}}\} \leq 3$ . Therefore, we conjecture that the load exponent is a universal value for both the undirected and directed cases.

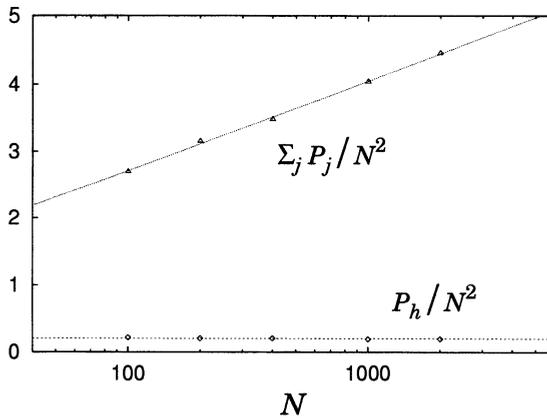
Now we examine the system-size dependent behavior of the load at the hub,  $\ell_h$  for the static model. According to (4),  $\ell_h$  behaves as  $\ell_h \sim N^{1.8} \log N$  in the range,  $2 < \gamma \leq 3$ , while for  $\gamma > 3$ ,  $\ell_h$  increases with  $N$  but at a much slower rate than that for  $2 < \gamma \leq 3$  as shown in Fig. 7. That implies that the shortest pathways between two vertices become diversified, and they do not necessarily pass through the hub for  $\gamma > 3$ . That may be related to the result that epidemic threshold is null in the range  $2 < \gamma \leq 3$ , while it is finite for  $\gamma > 3$  in SF networks, because there exist many other shortest paths not passing through the hub for  $\gamma > 3$ , so that the infection of the hub does not always lead to the infection of the entire system. Thus, epidemic threshold is finite for  $\gamma > 3$  [25].

We also study the topological feature of shortest paths between two vertices by introducing the ‘‘involving number’’ defined as follows [26]. First we define a set, composed of the vertices on the shortest path(s) from one position to another. Then, since there are  $N(N - 1)/2$  pairs of vertices, the same number

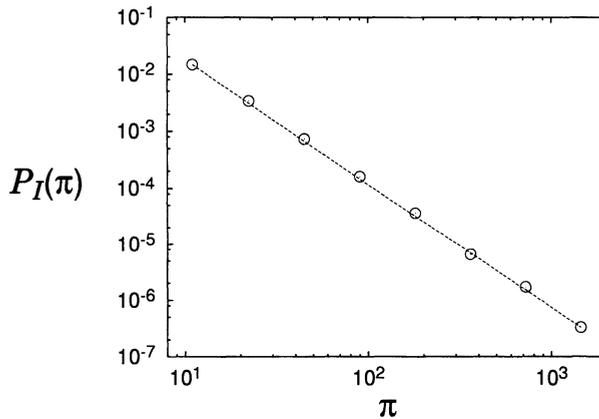


**Fig. 7.** Plot of the system-size dependence of the load at the hub versus system size  $N$  for various  $\gamma = 2.2$  ( $\circ$ ),  $2.5$  ( $\square$ ),  $3.0$  ( $\diamond$ ),  $4.0$  ( $\times$ ), and  $\infty$  ( $\triangle$ ). The solid line is  $N^{1.8} \log N$  and dotted lines have slope 1.70 and 1.25, respectively, from top to bottom. Simulations are performed for  $m = 2$  and all data points are averaged over ten configurations.

of sets exist in the system. Among them, we are interested in the amount of different sets a certain vertex  $j$  belongs to. This number is called the involving number  $P_j$ , and the normalized involving number  $\pi_j$  is defined as  $\pi_j \equiv P_j / \sum_l P_l$  for each vertex  $j$ . In particular, the involving number at the hub  $P_h$  is found numerically to scale as  $P_h \sim N^2$ , while the total involving number summed over all vertices is found numerically to scale as  $\sum_j P_j \sim N^2 \log N$  (see Fig. 8). So the normalized involving number at the hub  $\pi_h = P_h / \sum_j P_j$  behaves as  $\sim 1 / \log N$ , decaying much more slowly, compared with the normalized degree decreasing

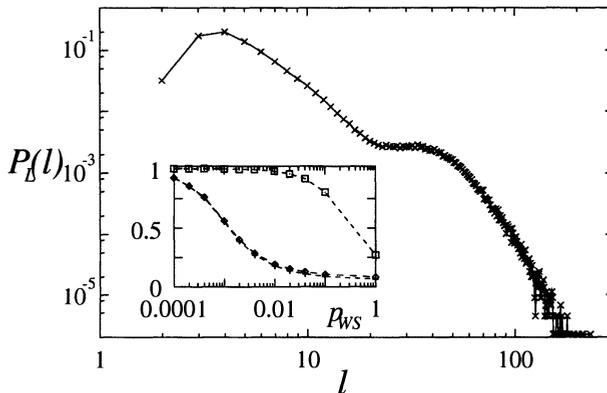


**Fig. 8.** Semi-logarithmic plot of  $P_h/N^2$  ( $\circ$ ) and  $\sum_j P_j/N^2$  ( $\triangle$ ) versus the system size  $N$  for the BA model.



**Fig. 9.** Involving number distribution  $P_I(\pi)$  for the BA model. The dashed line has a slope  $-2.2$ , being a guide to the eye.

as  $\sim N^{-0.5}$  in the case of the BA model [18]. Therefore, the contribution of the hub to shortest paths is much larger than the naive estimate based on the normalized degree at the hub,  $\mathcal{O}(1/\sqrt{N})$  as noticed before for the BA model. Fig. 9 shows the involving number distribution  $P_I(\pi)$ , showing a power-law behavior,  $P_I(\pi) \sim \pi^{-\eta}$  with  $\eta \approx 2.2$ .



**Fig. 10.** Plot of the load distribution  $P_L(\ell)$  versus load  $\ell$  for the small-world network. Simulations are performed for system size  $N = 1,000$ , and average degree  $\langle k \rangle = 10$ , and the rewiring probability  $p_{WS} = 0.01$ , averaged over 500 configurations. Inset: Plot of the average load ( $\diamond$ ), diameter ( $+$ ), clustering coefficient ( $\square$ ) versus the rewiring probability  $p_{WS}$ . All the data are normalized by the corresponding values at  $p_{WS} = 0$ . Dotted lines are guides to the eye.

Finally, we mention the load distribution of the small-world network of WS which is not scale-free. It is found that its load distribution does not obey a power-law, but shows a combined behavior of two Poisson-type decays resulting from short-ranged and long-ranged connections, respectively, as shown in Fig. 10. We also find the average load,  $\bar{\ell}(p_{WS}) \equiv (1/N) \sum_i \ell_i(p_{WS})$ , as a function of the rewiring probability  $p_{WS}$  decays rapidly with increasing  $p_{WS}$ , behaving similar to the diameter in the WS model, as shown in the inset of Fig. 10.

In conclusion, we have considered a problem of data packet transport on scale-free networks generated according to preferential attachment rules and introduced a physical quantity, load  $\{\ell_i\}$  at each vertex. We found that the load distribution follows a power-law,  $P_L(\ell) \sim \ell^{-\delta}$ , with the exponent  $\delta \approx 2.2(1)$ , which turns out to be insensitive to the detail of the network structure when the degree exponent is in the range  $(2, 3]$ . Moreover, it is also the same for both directed and undirected cases within our numerical uncertainties. Therefore, we conjecture that the load exponent is a generic quantity to characterize scale-free networks. Since the degree exponents are in the range  $2 < \gamma \leq 3$  for most of real-world scale-free networks, the universal behavior we found in this range may have interesting implications to the interplay of their structure and dynamics. We also find that the load scales linearly with the degree only when  $\gamma = \delta$ . For  $\gamma > 3$ , however, the load exponent  $\delta$  increases as the degree exponent  $\gamma$  increases, and eventually the load distribution decays exponentially as  $\gamma \rightarrow \infty$ . However, since the scale-free networks for  $\gamma > 3$  are fundamentally different from those for  $\gamma \leq 3$ , the universal behavior observed in  $2 < \gamma \leq 3$  could be taken as a generic nature of the scale-free networks.

This work is supported by grants No.2000-2-11200-002-3 from the BRP program of the KOSEF.

## References

1. S.H. Strogatz, *Nature* **410**, 268 (2001).
2. N. Goldenfeld and L.P. Kadanoff, *Science* **284**, 87 (1999).
3. P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **5**, 17 (1960).
4. D.J. Watts and S.H. Strogatz, *Nature* **393**, 440 (1998).
5. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
6. P.L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000); P.L. Krapivsky and S. Redner, *Phys. Rev. E* **63**, 066123 (2001).
7. S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, *Phys. Rev. Lett.* **85**, 4633 (2000).
8. R. Albert, H. Jeong, and A.-L. Barabási, *Nature* **401**, 130 (1999).
9. D. Butler, *Nature* **405**, 112 (2000).
10. A. Broder *et al.*, *Computer Networks* **33**, 309 (2000).
11. E.W. Zegura, K.L. Calvert, and M.J. Donahoo, *IEEE/ ACM Trans. Network* **5**, 770 (1997).
12. M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Comput. Commun. Rev.* **29**, 251 (1999).

13. R. Pastor-Satorras, A. Vázquez, and A. Vespignani, (cond-mat/0105161).
14. S. Redner, *Eur. Phys. J. B* **4**, 131 (1998).
15. M.E.J. Newman, *Proc. Natl Acad. Sci. USA* **98**, 404 (2001).
16. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvani, and A.-L. Barabási, *Nature* **407**, 651 (2000).
17. L.A. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, *Proc. Natl Acad. Sci. USA* **97**, 11149 (2000).
18. A.-L. Barabási, R. Albert, and H. Jeong, *Physica A* **272**, 173 (1999).
19. B. Bollobás, *Random Graphs* (Academic Press, New York, 1985).
20. S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, (cond-mat/0011115).
21. M.E.J. Newman, *Phys. Rev. E* **64**, 016131 (2001); **64**, 016132 (2001).
22. R. Kumar *et al.*, in *Proceedings of the 41th IEEE Symp. on Foundations of Computer Science* (2000).
23. H. Simon, *Biometrika* **42**, 425 (1955).
24. S. Bornholdt and H. Ebel, (cond-mat/0008465).
25. R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **63**, 066117 (2001).
26. K.-I. Goh, B. Kahng, and D. Kim, *Phys. Rev. E* **64**, 051903 (2001).

# Phase Transition of Three-Directional Traffic-Flow in 2D Network

Y. Honda

Dept. Computer Sciences and Systems Engineering, Muroran Institute of Technology, Muroran, 050-8585, Japan

**Abstract.** A traffic-flow model with objects moving towards three directions is investigated by numerical simulations and the mean-field approach. This traffic-flow model includes four kinds of objects in two-dimensional network, each chooses one of three directions to move. Average velocity and flow are calculated and phase transitions are observed increasing the density of objects. It is found that there are three kinds of traffic jams in a range of density of objects just above the critical value.

## 1 Introduction

Traffic-flow problems are important subjects not only in the field of transportation on freeways or city roads but also in the field of packet flow in computer networks. In the traffic-flow problems, the nature of a phase transition between a free-moving phase and a traffic-jam phase has to be clarified from the viewpoint of the statistical physics for a non-equilibrium system.

A cellular automaton (CA) is one of the useful tools to describe traffic-flow models [1]. Biham *et al.* [2] have proposed a CA model in order to describe the traffic flow of two kinds of cars in a two-dimensional network of city roads; it is called the BML model here after. The BML model has a discontinuous phase transition between the free-moving phase with lower density of cars and the traffic-jam phase with higher density of cars. A typical configuration of cars in the traffic-jam phase has a shape like a herringbone (see Fig. 1(b) in Ref. [3]). Since in the BML model, the cars do not change their moving direction, Cuesta *et al.* [4] introduced a probability of changing the direction. The discontinuous phase transition is observed in this model, too.

These models for two-dimensional traffic flow have two kinds of cars. We have recently proposed a traffic-flow model with four kinds of cars [3]. In contrast with models with two kinds of cars, we have found a sort of continuous transition between the free-moving phase and the traffic jammed phase. We have noticed that core configurations of cars play an important role to bring a traffic-jam; those core configurations are different from the herringbone type configuration in the BML model. In our previous study [3], we assume a two-way traffic on each pair of two lanes; the cars are able to move towards only a fixed direction on each lane. We do not permit a two-way traffic on one lane. Then we have found that the core configurations are never dissolved.

In the present paper, The word “object” instead of “car” is used, since an object can move backwards according to rules given in section 2. In contrast to

the previous model [3], the two-way traffic on one lane is permitted in the present model. Any core configuration for a traffic jam does not appear in the present model. It is found that there is a continuous phase transition whose behavior is different from that in the previous model given in Ref.[3].

In section 2, a three-directional traffic-flow model is described. In section 3, numerical results are shown for average velocities and the fundamental diagram. In section 4, a critical density of objects is estimated by considering a stability of an interface of a cluster. Section 5 is devoted to a summary of the present study.

## 2 A Traffic-Flow Model with Objects Moving Towards Three Directions in a Two-Dimensional Network

The square lattice is considered as a network in which objects move. A site of the square lattice is denoted by  $\mathbf{r} = (x, y)$ , where  $x \in \{1, 2, \dots, L\}$  and  $y \in \{1, 2, \dots, L\}$ ;  $L$  is a linear size of the square lattice. Each object can occupy

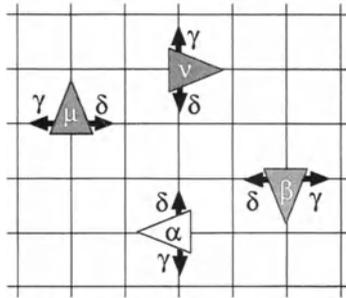


Fig. 1. Square lattice with four kinds of moving objects.

a site of the lattice as shown in Fig. 1, but can not be located on a bond between sites. In the present study, the two-way traffic on a bond is considered instead of the one-way traffic; for example an object can move from the west to the east or from the east to the west on the same bond.

Four types of objects are considered, namely  $\alpha$ -type,  $\beta$ -type,  $\mu$ -type and  $\nu$ -type, on the square lattice. Objects of  $\alpha$ -type try to move towards the west with a probability  $1 - \delta - \gamma$ , towards the south with a probability  $\gamma$ , and towards the north with a probability  $\delta$ . Objects of  $\beta$ -type try to move towards the south with a probability  $1 - \delta - \gamma$ , towards the east with a probability  $\gamma$ , and towards the west with a probability  $\delta$ . Objects of  $\mu$ -type try to move towards the north with a probability  $1 - \delta - \gamma$ , towards the west with a probability  $\gamma$ , and towards the east with a probability  $\delta$ . Objects of  $\nu$ -type try to move towards the east with a probability  $1 - \delta - \gamma$ , towards the north with a probability  $\gamma$ , and towards

the south with a probability  $\delta$ . Arrows attached to triangles mean directions to which the object is possible to move.

If a site to which an object tries to move is empty and a traffic signal permits that movement, the object can move to the site. If a site to which an object tries to move is occupied by another object or a traffic signal does not permit that movement, the object can not move to the site. There is a possibility that two objects simultaneously try to move into the same site from different sites. In this case, we select one of these two objects to move to the site with the same probability for each.

In order to describe microscopic equations of motion, we introduce four Boolean variables  $\alpha_r^t, \beta_r^t, \mu_r^t$  and  $\nu_r^t$  according to the type of objects. For example, an object of  $\mu$ -type at a site  $r$  and at time step  $t$  is denoted by  $\mu_r^t = 1$ . We introduce an auxiliary variable  $\kappa_r^t \in \{0, 1, 2, 3, 4\}$  whose value means a state of occupancy at site  $r$  and at time  $t$  as follows:

$$\kappa_r^t = \begin{cases} 0 & (\alpha_r^t = \beta_r^t = \mu_r^t = \nu_r^t = 0) \\ 1 & (\alpha_r^t = 1, \beta_r^t = \mu_r^t = \nu_r^t = 0) \\ 2 & (\beta_r^t = 1, \alpha_r^t = \mu_r^t = \nu_r^t = 0) \\ 3 & (\mu_r^t = 1, \alpha_r^t = \beta_r^t = \nu_r^t = 0) \\ 4 & (\nu_r^t = 1, \alpha_r^t = \beta_r^t = \mu_r^t = 0) \end{cases} \quad (1)$$

Using these variables, a microscopic equation of motion for an object of  $\mu$  type is expressed as follows:

$$\mu_r^{t+1} = \begin{cases} 1 & (S_{\mu,r}^{t+1} = 1, M_r^{t+1} = 0) \\ \delta_{\kappa_r^{t+1}, 3} M_{\mu,r}^{t+1} & (S_{\mu,r}^{t+1} = 0, M_r^{t+1} > 0) \\ 0 & (S_{\mu,r}^{t+1} = M_r^{t+1} = 0) \end{cases}, \quad (2)$$

$$\begin{aligned} S_{\mu,r}^{t+1} &\equiv \mu_r^t (\bar{\xi}_r^t \bar{\theta}_r^t \bar{\sigma}^t + \xi_r^t \bar{\theta}_r^t \sigma^t + \bar{\xi}_r^t \theta_r^t \sigma^t) \\ &+ \mu_r^t \{ \bar{\xi}_r^t \bar{\theta}_r^t \sigma^t (\mu_{r+\hat{y}}^t + \nu_{r+\hat{y}}^t + \alpha_{r+\hat{y}}^t + \beta_{r+\hat{y}}^t) \\ &+ \xi_r^t \bar{\theta}_r^t \bar{\sigma}^t (\mu_{r-\hat{x}}^t + \nu_{r-\hat{x}}^t + \alpha_{r-\hat{x}}^t + \beta_{r-\hat{x}}^t) \\ &+ \bar{\xi}_r^t \theta_r^t \bar{\sigma}^t (\mu_{r+\hat{x}}^t + \nu_{r+\hat{x}}^t + \alpha_{r+\hat{x}}^t + \beta_{r+\hat{x}}^t) \}, \end{aligned} \quad (3)$$

$$\begin{aligned} M_{\mu,r}^{t+1} &\equiv \bar{\mu}_r^t \bar{\nu}_r^t \bar{\alpha}_r^t \bar{\beta}_r^t (\mu_{r-\hat{y}}^t \bar{\xi}_r^t \bar{\theta}_r^t \bar{\sigma}^t \\ &+ \mu_{r-\hat{x}}^t \bar{\xi}_r^t \theta_r^t \bar{\sigma}^t + \mu_{r+\hat{x}}^t \xi_r^t \bar{\theta}_r^t \bar{\sigma}^t), \end{aligned} \quad (4)$$

$$M_r^{t+1} \equiv M_{\mu,r}^{t+1} + M_{\nu,r}^{t+1} + M_{\alpha,r}^{t+1} + M_{\beta,r}^{t+1}, \quad (5)$$

where  $\delta_{\kappa,3}$  is Kronecker's delta and  $\bar{a} = 1 - a$  for a Boolean variable  $a$ .  $\hat{x}$  and  $\hat{y}$  mean unit vectors in  $x$  and  $y$  direction, respectively.  $\xi_r^t \in \{0, 1\}$  and  $\theta_r^t \in \{0, 1\}$  are random variables that determine the direction in which an object tries to move. Frequencies for  $\xi_r^t$  and  $\theta_r^t$  are given in Table 1. The value of the Boolean variable  $\sigma^t$  means the state of traffic signals at time step  $t$ . When  $\sigma^t = 1$ , movement of objects is permitted from the north to the south or in inverse direction. On the other hand, when  $\sigma^t = 0$ , movement of objects is permitted from the west to the east or in inverse direction.  $S_{\mu,r}^{t+1} = 1$  means that the object of  $\mu$ -type has to stay at site  $r$  at time step  $t + 1$ .

Table 1.

$\xi_r^t$	$\theta_r^t$	Frequency
0	0	$(1 - \gamma - \delta)$
0	1	$\delta$
1	0	$\gamma$
1	1	0

Microscopic equations for another type of objects are constructed by giving attention to the type and directions to which they try to move. For example for  $S_{\nu,r}^{t+1}$ , in rhs of (3) we replace  $\mu_r^t$  by  $\nu_r^t$ ,  $\sigma^t$  by  $\bar{\sigma}^t$ ,  $\bar{\sigma}^t$  by  $\sigma^t$ ,  $r + \hat{y}$  by  $r + \hat{x}$ ,  $r - \hat{x}$  by  $r + \hat{y}$  and  $r + \hat{x}$  by  $r - \hat{y}$ , respectively. For  $M_{\nu,r}^{t+1}$ , in rhs of (4) we replace  $\mu_{r-\hat{y}}^t$  by  $\nu_{r-\hat{x}}$ ,  $\mu_{r-\hat{x}}^t$  by  $\nu_{r+\hat{y}}$ ,  $\mu_{r+\hat{x}}^t$  by  $\nu_{r-\hat{y}}$ ,  $\sigma^t$  by  $\bar{\sigma}^t$ ,  $\bar{\sigma}^t$  by  $\sigma^t$ ,  $r - \hat{y}$  by  $r - \hat{x}$ ,  $r - \hat{x}$  by  $r + \hat{y}$  and  $r + \hat{x}$  by  $r - \hat{y}$ , respectively. Likewise, we can obtain microscopic equations by such replacements for  $S_{\alpha,r}^{t+1}$ ,  $S_{\beta,r}^{t+1}$ ,  $M_{\alpha,r}^{t+1}$  and  $M_{\beta,r}^{t+1}$ .

When  $M_r^{t+1} > 0$ ,  $\kappa' \in \{0, 1, 2, 3, 4\}$  is a random variable whose frequency distribution  $P(\kappa')$  is defined as follows:

$$P(\kappa') = \frac{\delta_{\kappa',1} M_{\alpha,r}^{t+1} + \delta_{\kappa',2} M_{\beta,r}^{t+1} + \delta_{\kappa',3} M_{\mu,r}^{t+1} + \delta_{\kappa',4} M_{\nu,r}^{t+1}}{M_r^{t+1}}. \tag{6}$$

The frequency distribution means that one of two objects moves to the site with probability 1/2. Note that a multi-occupation of one site is not permitted.

### 3 Results of Numerical Simulations

Numerical simulations are carried out by using the microscopic equation of motion described in previous section. A periodic boundary condition is assumed in all of the present simulations. Average velocity and flow are calculated from the numerical simulations. The total number of objects is expressed by  $N$ . We assume that the number of objects of each type is  $N/4$ . The density of objects  $n$  is defined as follows:

$$n = \frac{N}{L^2}. \tag{7}$$

In the present study, the velocity of objects  $v(t)$  is defined as follows:

$$v(t) = \frac{1}{2N} \sum_r \{ (\alpha_r^{t+1} - \alpha_r^t)^2 + (\beta_r^{t+1} - \beta_r^t)^2 + (\mu_r^{t+1} - \mu_r^t)^2 + (\nu_r^{t+1} - \nu_r^t)^2 \}. \tag{8}$$

This velocity corresponds to the total number of objects which move in a time step from  $t$  to  $t + 1$ . Time averages are calculated as follows:

$$\langle v \rangle_t = \frac{1}{T_M - T_1} \sum_{t=T_1+1}^{T_M} v(t), \quad (9)$$

where  $T_M$  and  $T_1$  mean maximum time step and initial time step of observations, respectively.  $T_1 = 10000$  is used in order to obtain equilibrium states and  $T_M = 50000$  in order to calculate the time averages. Moreover, an average over initial configurations is taken as follows:

$$\langle v \rangle = \sum_{i=1}^R \langle v \rangle_t. \quad (10)$$

In the following, we set  $R = 50$  for the calculation of the velocity and the flow.

In the present study the flow of objects  $f(t)$  is defined as follows:

$$f(t) = nv(t). \quad (11)$$

Average values of the flow are calculated in the same way as in (9) and (10). The meaning of the flow is the number of moved objects per site.

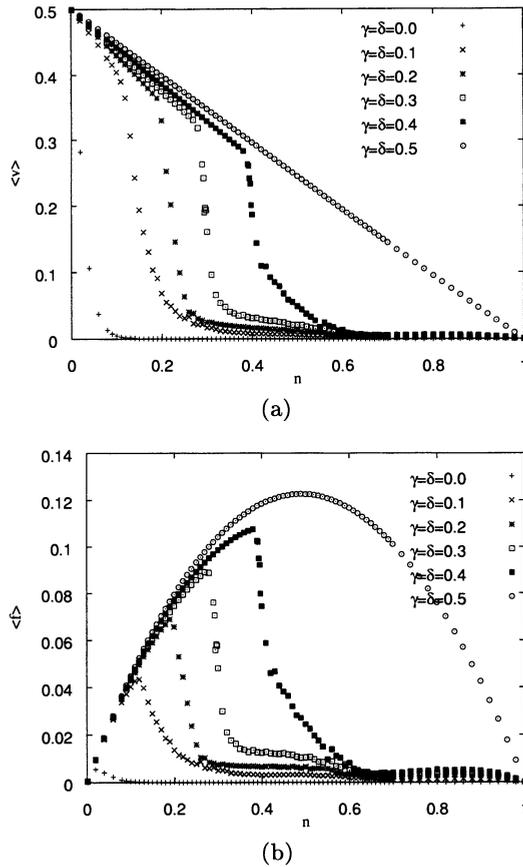
Figure 2(a) shows results for the velocity as a function of the density of objects in the case of  $L = 64$  and  $\gamma = \delta$ . We find a phase transition between the free-moving phase and the traffic-jam phase for  $0 < \gamma < 0.5$ . In this parameter range, increasing the value of  $\gamma$ , the density of objects at the transition point increases. In contrast to this parameter range, for  $\gamma = 0.5$  there is no phase transition.  $\gamma = 0.5$  means that objects can move only in two directions.

In Fig. 2(b), average values of the flow are shown as functions of the density of objects. For every value of  $\gamma$ , the flow becomes maximum at transition points of the density of objects. Note that the velocity and the flow have finite values even in the traffic-jam phase. This property is different from that for the model with one-way traffic [3].

In Fig. 3, results for  $\langle f \rangle_t$  are shown in detail. There are three branches of the flow in a range  $0.4 \lesssim n \lesssim 0.6$  which are pointed out in (b), (c) and (d), respectively in Fig. 3. By observing some snap shots of the configuration of objects it is found that the branch denoted by (b) corresponds to a zero-dimensional cluster, (c) to a one-dimensional cluster and (d) to a two-dimensional cluster.

## 4 Discussion for a Phase Diagram

Let us estimate the critical density of objects  $n_c$  as a function of  $\delta$  and  $\gamma$ , by considering a stability of an interface of the zero-dimensional cluster as shown in Fig. 4. Assuming that the zero-dimensional cluster forms a circle with radius  $R$  and the shaded area is occupied by objects of type  $\nu$ , (see Fig. 5) the length

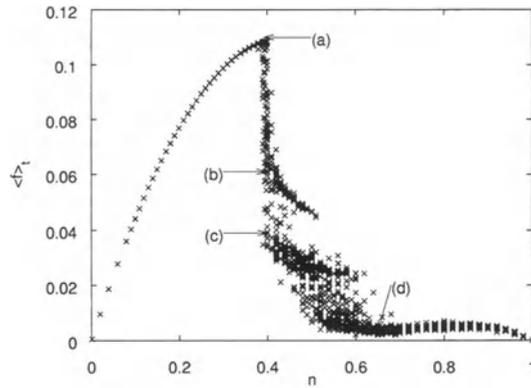


**Fig. 2.** (a) Average value of the velocity  $\langle v \rangle$  as a function of the density of cars  $n$ . (b) Average value of the flow  $\langle f \rangle$  as a function of  $n$ . These results are obtained from simulations for  $L = 64$ .

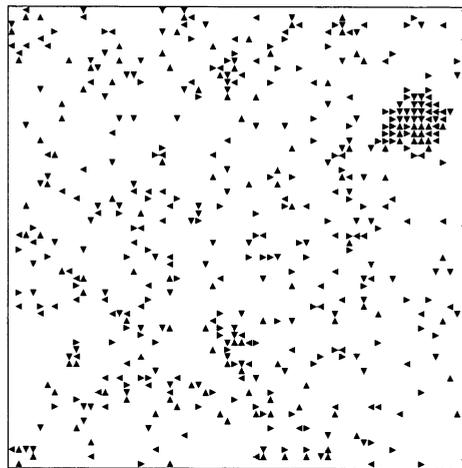
of the interface between inside and outside of the cluster becomes  $\frac{\pi R}{2}$  for the shaded area. We concentrate our attention only on the shaded part. Objects in the interface are depicted by small open circles attached to short arrows.

At first let us estimate the number of objects which break away from the interface of the cluster. Objects on an upper part of the interface (shown by (a) in Fig. 5) can go outside of the cluster by moving towards a direction shown by the arrow with a probability  $\gamma$ . Then the number of objects which try to break away from the interface (a) is given by  $\frac{\pi R}{4}\gamma$ . On the other hand, the number of objects which try to break away from the interface (b) is given by  $\frac{\pi R}{4}\delta$ . Therefore their total number  $N_b$  is denoted as follows:

$$N_b = \frac{\pi R(\gamma + \delta)}{4}. \tag{12}$$



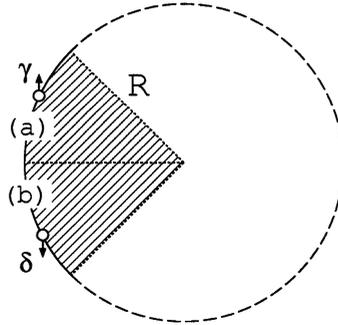
**Fig. 3.** Calculated values of the flow for  $\delta = \gamma = 0.4$ . Each symbol corresponds to result with a different initial configuration of objects.



**Fig. 4.** A snap shot of the configuration of objects for  $L = 64, \delta = \gamma = 0.1$  and  $n = 0.12$ .

Because just outside of the cluster, the density of objects is close to  $n$ , only a fraction  $(1 - n)$  of  $N_b$  is permitted to move to the outside of the cluster. The true number of objects which can break away from the interface is given by  $N_b(1 - n)$ .

Now let us estimate the number of objects which try to attach to the interface of the cluster. We denote the number of objects of type  $\alpha, \beta, \mu$  and  $\nu$  which try to attach to the interface by  $N_{a,\alpha}, N_{a,\beta}, N_{a,\mu}$  and  $N_{a,\nu}$ , respectively. Concerning the objects of type  $\alpha$ , the fraction which tries to attach to the interface is given



**Fig. 5.** A schematic representation of a zero-dimensional cluster of objects. Shaded area occupied by objects of type  $\nu$ . (a) denotes an upper part of an interface for type  $\nu$  and (b) a lower part.

by  $\frac{n}{4}\frac{1}{2}\gamma + \frac{n}{4}\frac{1}{2}\delta$ . Hence,  $N_{a,\alpha}$  is expressed as follows:

$$N_{a,\alpha} = (\delta + \gamma) \frac{\pi R n}{16}. \tag{13}$$

Concerning the objects of type  $\beta$ , the fraction which tries to attach to the interface is given by  $\frac{n}{4}\gamma + \frac{n}{4}\frac{1}{2}(1 - \delta - \gamma)$ . Hence,  $N_{a,\beta}$  is expressed as follows:

$$N_{a,\beta} = (1 - \delta + \gamma) \frac{\pi R n}{16}. \tag{14}$$

With respect to the objects of type  $\mu$ , the fraction which tries to attach to the interface is given by  $\frac{n}{4}\delta + \frac{n}{4}\frac{1}{2}(1 - \delta - \gamma)$ . Hence,  $N_{a,\mu}$  is expressed as follows:

$$N_{a,\mu} = (1 + \delta - \gamma) \frac{\pi R n}{16}. \tag{15}$$

With respect to the objects of type  $\nu$ , the fraction which tries to attach to the interface is given by  $\frac{n}{4}(1 - \delta - \gamma) + \frac{n}{4}\frac{1}{2}(\delta + \gamma)$ . Hence,  $N_{a,\nu}$  is expressed as follows:

$$N_{a,\nu} = (2 - \delta - \gamma) \frac{\pi R n}{16}. \tag{16}$$

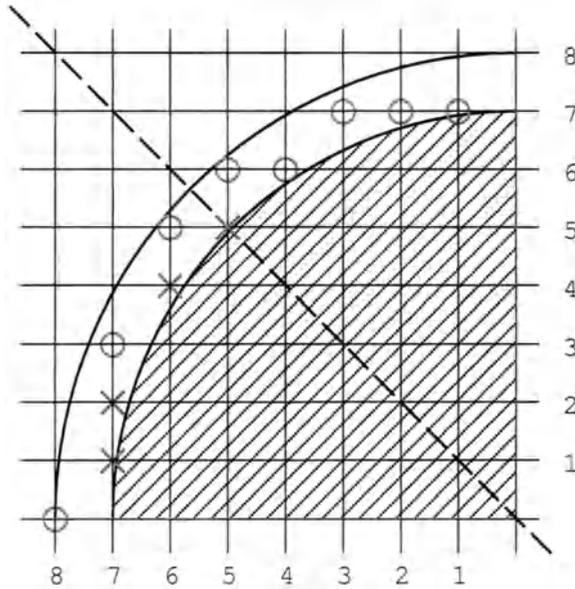
A situation where the value of  $N_b(1 - n)$  equals  $N_{a,\alpha} + N_{a,\beta} + N_{a,\mu} + N_{a,\nu}$  corresponds to  $n_c$ . Therefore the value of  $n_c$  is approximated as follows:

$$n_c \simeq \frac{\delta + \gamma}{1 + \delta + \gamma}. \tag{17}$$

There are two factors which can cause difference the between approximated values and simulated values. One of them is an effect of the zero-dimensional cluster on the density of objects  $n$ . By forming a cluster, the true density of the objects  $n'$  outside of the cluster should be smaller than  $n$ . This factor affects

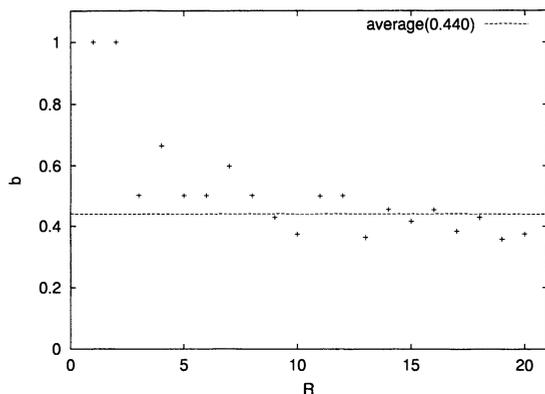
$N_b(1 - n)$ , so that the approximated value of  $n_c$  should increase by this factor. For system size  $L \rightarrow \infty$ , keeping  $R$  finite, this factor should vanish.

The second factor is an effect of the shape of the interface. In the present argument, the shape of the interface is assumed to be a circle and every object at the interface has the possibility to break away from the cluster. However, some of the objects can not break away from the interface, because the true shape of the interface is different from a circle. In Fig. 6, a real configuration



**Fig. 6.** A true configuration of objects at an interface of a cluster whose radius is eight. A shaded area means an inside of the cluster. Circles denote objects which can break away from the cluster towards the north. Crosses denote objects which can not break away from the cluster towards the north.

of objects at an interface with  $R = 8$  is shown as an example. A shaded area means an inside of the cluster. Circles denote objects which can break away from the cluster towards the north. Crosses denote objects which can not break away from the cluster towards the north. There are six objects at the boundary which corresponds to the interface (a) in Fig. 5. Among them three objects can not break away from the cluster towards the north. The parameter  $b$  denotes the fraction of objects which can not break away from the cluster by this factor. In the case of a cluster with radius eight we have  $b = 0.5$ . For a cluster of  $R = 20$  the critical density of objects is reached. In fact, we check up the value of  $b$  for up to  $R = 20$ . In Fig. 7, values of  $b$  are shown as a function of  $R$ . The smallest value of  $b$  is 0.357 for  $R = 19$ . In real situations of the present traffic-flow model,  $b$  takes a value in a range  $0.357 \lesssim b \leq 1$ .

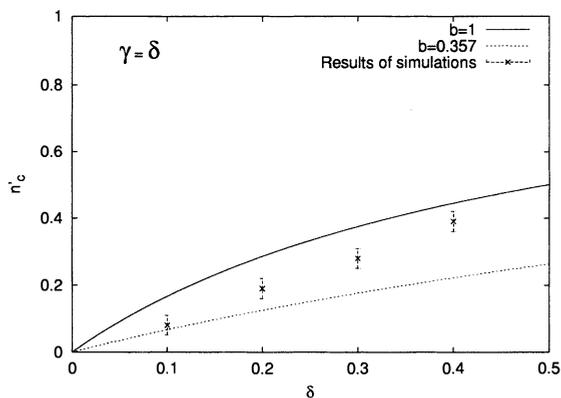


**Fig. 7.** Values of  $b$  as a function of radius of cluster. A broken line denotes an average  $b = 0.440$ .

Because of this factor, the value of  $N_b(1 - n)$  is modified to be  $N_b(1 - n)b$ . Then the approximated value of  $n_c$  should be modified to  $n'_c$  as follows:

$$n'_c \simeq \frac{(\delta + \gamma)b}{1 + (\delta + \gamma)b}. \tag{18}$$

In Fig. 8, we show the value of  $n'_c$  calculated by (18) as a function of  $\delta$  in the case of  $\gamma = \delta$  and obtained results from simulations. The solid line denotes the value of  $n'_c$  with  $b = 1$  which is the largest value of  $b$ . The dotted line corresponds



**Fig. 8.** The values of  $n'_c$  as a function of  $\delta$  in the case of  $\gamma = \delta$ . The solid line shows the calculated value by (18) with  $b = 1$ . The dotted line corresponds to  $b = 0.357$ .

to  $b = 0.357$  which is the smallest value in a range  $1 \leq R \leq 20$ . It is found that critical densities  $n'_c$  obtained by simulations are within a region bounded by  $b = 1$  and  $b = 0.357$ .

## 5 Summary

In the present study, a traffic-flow model with objects moving towards three directions on a two-dimensional network is proposed. Two-way traffic on the square lattice is assumed in this traffic-flow model. From numerical simulations, it is found that there are three kinds of structures of traffic-jams. This property is different from that the traffic-flow model with a one-way traffic [3] exhibits. The phase diagram is roughly understood by considering a stability of an interface of clusters.

The determination of an accurate boundary of the phase diagram and a properties of the phase transition are left as interesting problems.

## References

1. S. Wolfram, *Rev. Mod. Phys.* **55**, 601 (1983).
2. O. Biham, A.A. Middleton, and D. Levine, *Phys. Rev. A* **46**, R6124 (1992).
3. Y. Honda and T. Horiguchi, *J. Phys. Soc. Jpn.* **69**, 3744 (2000).
4. J.A. Cuesta, F.C. Martínez, J.M. Morera, and Sánchez, *Phys. Rev. E* **48**, R4175 (1993).

# Analysis of Minimal Model of Internet Traffic

K. Fukuda<sup>1</sup>, M. Takayasu<sup>2</sup>, and H. Takayasu<sup>3</sup>

<sup>1</sup> NTT Network Innovation Laboratories, 3-9-11, Midori-cho, Musashino, 180-8585, Japan

<sup>2</sup> Department of Complex Systems, Future University-Hakodate, 116-2, Kameda-Nakano, Hakodate, 041-0803, Japan

<sup>3</sup> Sony Computer Science Laboratories, 3-14-13, Higashi-Gotanda, Shinagawa, 141-0022, Japan

**Abstract.** We focus on the physical mechanism of phase transition between non-congested and congested phases in Internet traffic. We analyze the simplest model of packet transport protocol called stop-and-wait, which is an element of widely-used TCP/IP protocol. Our simulation clarifies that the stop-and-wait protocol can reproduce the same phase transition as appears in actual Internet traffic. The most important result is that the exponent of power-laws at the critical point has the same value as one in real traffic trace. We also show that a simpler model, which omits feedback control, shows a phase transition behavior, but, the power law exponent is different from the observed one. Our simulation clarifies the minimal network topology for generating  $1/f$  fluctuation when using the stop-and-wait algorithm.

## 1 Introduction

The Internet is becoming an essential tool in our daily lives, and more than  $10^8$  hosts are currently connected to this computer network. It is obviously the most complicated artificial object in the history of mankind. It is known as a totally open system with autonomous properties because it is not based on any central control.

The Internet consists of two types of nodes and links connecting them. One type of nodes is called hosts which are located at the peripheral leaves of the Internet corresponding to the computers that people use usually. From the viewpoint of information flow, hosts are both sources and sinks of information. The other type of nodes is called routers which are responsible for delivering information from a sender host to a receiver host.

When a sender host transmits some information to a receiver host, the information is sent as a number of packets and they pass through routers along the path from the sender to the receiver. Each router has a data table called the routing table containing information about the receiver addresses and neighboring routers. Because each packet has a header that includes both the sender's and receiver's addresses, routers can find the next hop router (or host) of the received packet by using its header information and the routing table. The routing table is periodically autonomously reconstructed by the communication with the neighboring routers. Unlike telephone networks in which a connection takes sole possession of a line, routers and links are always shared equally by all related

hosts in the Internet, namely there are no special packets – they are treated in the same manner.

When an enormous number of packets arrive at a router, a jam occurs like in highway traffic or granular flow. The special property of the Internet is that some packets are dropped at jammed routers in the worst case. Because of such properties, the macroscopic view is better for explaining the dynamics of Internet traffic than a view focusing on individual user behaviors. There have been many studies of the statistical characterization of Internet traffic. It has been shown that the fluctuation of traffic density in the Internet is well modeled by the self-similar property (i.e., long-range dependency), deviating clearly from the Poisson property (short-range dependency) [5,6,2,1]. Recently, Takayasu et al. showed that network traffic behavior is characterized by the phase transition between non-congested and congested phases [9,8,3]. They found that the self-similar property appears when the mean traffic flow density is close to the critical density, while the network behavior is well modeled by the usual Poisson property when the flow density is low.

Although observational studies have proved the occurrence of the phase transition, its mechanism has not been clarified. Takayasu et al. pointed out that even in the simplest situation of a single buffer, self-similar fluctuation can occur at the critical point. However, they reported that the exponent of the power law of the single buffer model, -0.5, was different from the observed one, -1.0. Fukuda et al. focused on the effect in Ethernet, which is widely used in local area networks [4], and clarified by numerical simulation that Ethernet can reproduce the  $1/f$  fluctuation with exponent -1.0 at the critical point. They also pointed out that the essence of the  $1/f$  fluctuation of Ethernet is the exponential backoff effect for avoiding packet collisions. The Ethernet model is useful for explaining the network's  $1/f$  behavior in a local area network, but the phase transition has also been observed in wide area networks. Therefore, we need a more general model to explain the  $1/f$  behaviors of a wide area network.

In order to elucidate the minimal model of phase transition in a wide area network, this paper focuses on the effect of the widely used TCP/IP protocol. TCP is designed for reliable and efficient communication between sender and receiver hosts, and is based on a feedback control to adjust to the current network condition [7]. Since TCP itself is a mixture of several control algorithms, it is difficult to clarify the effect of TCP as a whole. Thus, we concentrate on the effect of feedback control called the stop-and-wait algorithm as the essence of TCP/IP by using simple computer simulation.

First we show that the stop-and-wait protocol can reproduce the same statistical behavior including the exponent as the observed one. Next, we show that the exponent of the power law changes from -1.0 to -0.5 when we modify the feedback control. Finally, we clarify that the essence of pure  $1/f$  fluctuation is in the mutual correlation among three routers with feedback control and does not depend on the network topology.

## 2 Simulation Model

### 2.1 Simplified TCP

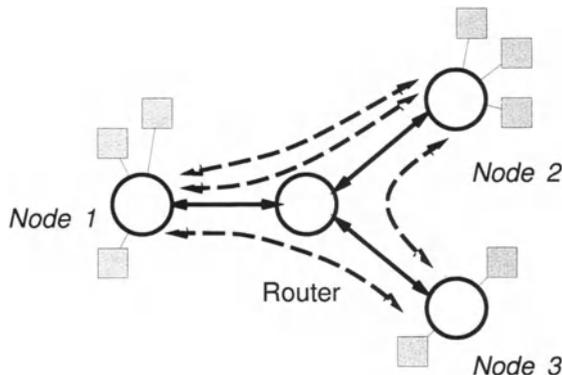
The TCP algorithm consists of several basic control algorithms to achieve the efficient communication. The main ones are:

- window-based flow control per connection
- retransmission algorithm
- transmission rate controls (such as slow start and congestion avoidance)

It is difficult to clarify the role of each component in the phase transition behaviors because the whole TCP is very complicated. In this paper, we focus on the most basic component of TCP; the window-based flow control contains a feedback effect, and we omit the effect of the last two algorithms because we have shown that they do not play essential roles in the phase transition.

In the TCP algorithm, basically the sender transmits packets to the receiver when the sender receives an acknowledgment packet from the receiver for previously sent packets. We pay attention to this acknowledgment algorithm and introduce a simple feedback control algorithm called the stop-and-wait algorithm. In this algorithm a sender transmits a packet only when it receives an acknowledgment packet responding to the sender's previous packet. Namely, only one packet or one acknowledgment packet can exist in the network at a time for each connection.

### 2.2 Simulation setup



**Fig. 1.** Network configuration. Circles indicate nodes (three hosts and a router). Gray boxes indicate senders and receivers.

We consider a network consisting of three hosts and one central router denoted by circles and transmission links connecting them (denoted by solid lines)

as shown in Figure 1. Gray boxes indicate senders and receivers at the host and dotted lines indicate connections among the senders and receivers. In this example, four connections are established. There are two connections between the Nodes 1 and 2, one between the Nodes 2 and 3, and one between the Nodes 1 and 3. The reason we treat this simple network topology is to minimize the effect of a complicated network topology.

We assume that buffer sizes of nodes and the router are infinite, so that no packet loss occurs. Each packet in the connection travels from the sender node to the receiver node passing through the router and the node's buffers. At each time step, every node tries to establish new connections with a certain probability. This probability (which we call the connection arrival rate  $r$ ) is the control parameter in phase transition studies. It is important that the connection creation event is temporally independent of others. When the trial succeeds, the new connection generates a data packet in the next time step. If there is a connection that received an acknowledgment packet or if it is newly established in the previous time step, the sender nodes put the next data packet into the buffer of the corresponding node. Also, when nodes and the router have packets in their buffers, they transfer the oldest packet in the buffer to the node or router. When a host receives a data packet, it generates an acknowledgment packet in the next time step. On the other hand, when the received packet is an acknowledgment packet, it generates the next data packet in the next time step, except when the acknowledgment is for the last data packet, in which case the connection is finished. We set the number of transferred packets in a connection to 20. It should be noted that because of an infinite buffer sizes in the nodes, the buffer length is expected to diverge when the connection arrival rate is too high.

Our simulation was done 20 times for each parameter value. Each run was executed for 1,310,720 simulated time steps. We focused on the density fluctuation of the number of packets transferred from the Router to Node 1 with granularity of 20 time steps. Thus, the result in each case is a time series of 65,536 data points.

### 3 Simulation Results

#### 3.1 Stop-and-Wait Algorithm

Figure 2 shows the relationship between connection arrival rate ( $r$ ) and the number of finished connections ( $c$ ).  $c$  linearly increased for  $r$  smaller than 3.7. This range corresponds to the non-congested phase. On the other hand, we can find a decrease of the number of finished connection for the region over  $r = 3.7$  corresponding to the congested phase. This figure clearly indicates that the network traffic produced by the simple stop-and-wait algorithm showed a phase transition between non-congested and congested phases with the critical point around  $r = 3.7$ .

To clarify the statistical property of traffic we analyzed the produced traffic density fluctuation by using congestion duration distribution analysis and power

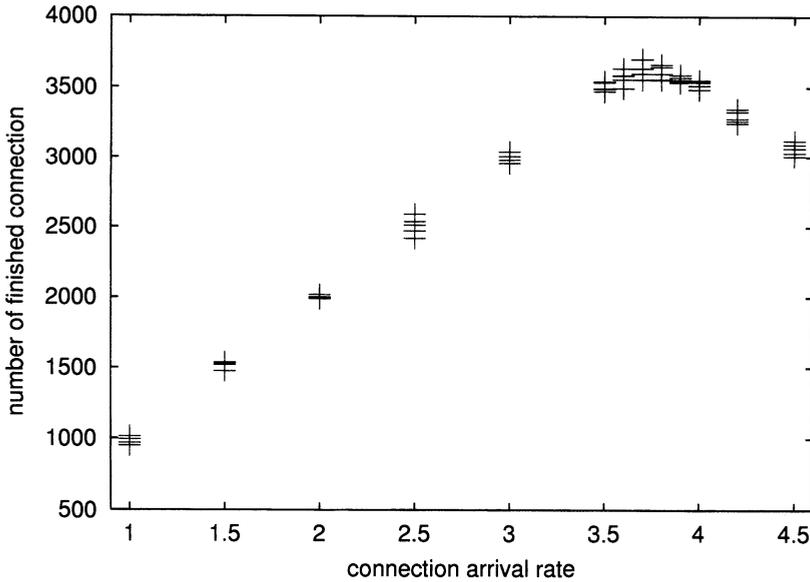


Fig. 2. Number of finished connection.

spectrum analysis. Figure 3 shows the congestion duration distribution of the three regions: non-congested phase, congested phase, and the critical point. The congestion duration length is defined as the consecutive number of congestion states (i.e. states having values larger than a certain threshold) multiplied by the bin size in traffic density fluctuation. It is well-known that the cumulative distribution of congestion length follows the power-law distribution  $P(> L) \propto L^{-\alpha}$  when the connection arrival rate is at the critical point. From this figure, we actually find that the distribution is approximated by a power law distribution at the critical point. It should be noted that the slope in the log-log scale is close to -1.0 which is consistent with the actual traffic analysis shown in Figure 4

As confirmed from Figures 3 and 4 the stop-and-wait simulation and real observation give consistent results not only at the critical point but also in both non-congested and congested phases. When the connection arrival rate was smaller than the critical value, the distribution decayed exponentially as a typical non-congested phase behavior. In the congested phase the distribution deviated from the power law of exponent -1.0, showing a plateau-like behavior.

Figure 5 shows the power spectra of the traffic density fluctuation produced by the stop-and-wait algorithm. The three plots correspond to the non-congested phase, the congested phase, and the critical point. Like the result of congestion duration analysis, the power spectrum followed the power law distribution  $S(f) \propto f^{-\alpha}$  at the critical point with an exponent approximately -1.0. In the non-congested phase the power spectrum deviated from the power law distribution but it was characterized nicely by a white-noise spectrum.

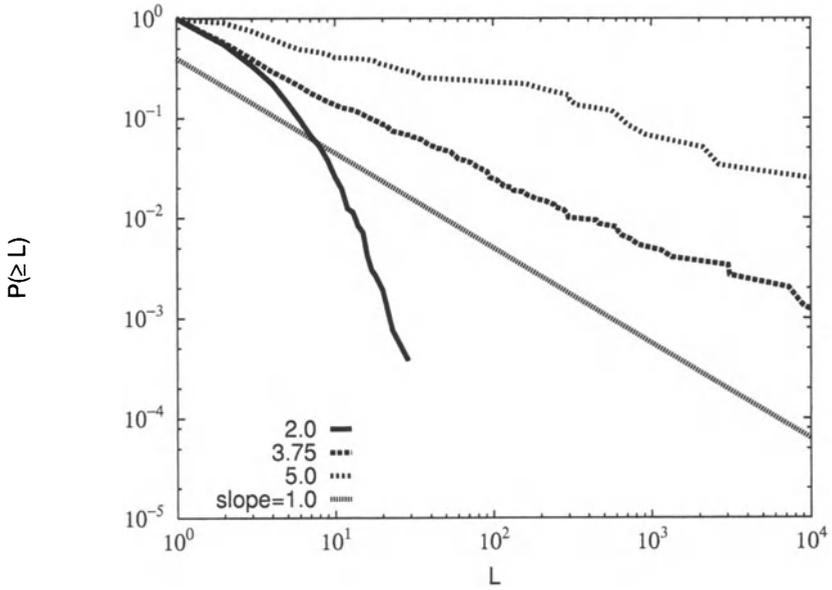


Fig. 3. Congestion duration distribution of the stop-and-wait algorithm.

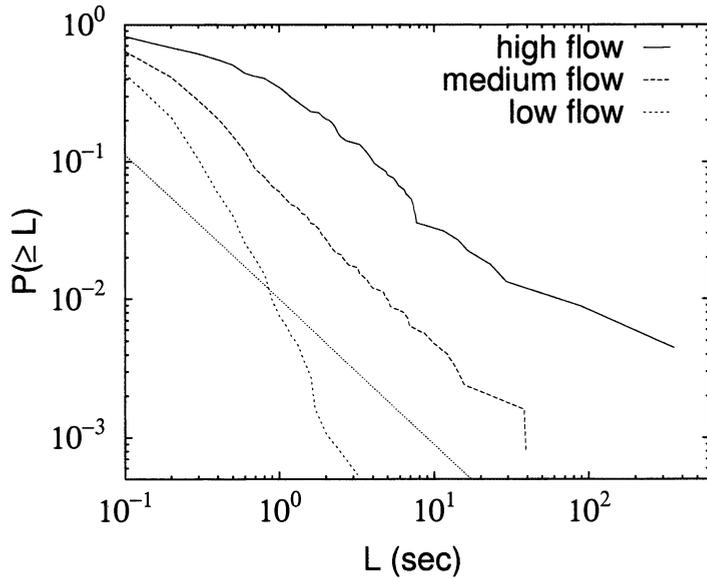


Fig. 4. Congestion duration distribution of real network traffic.

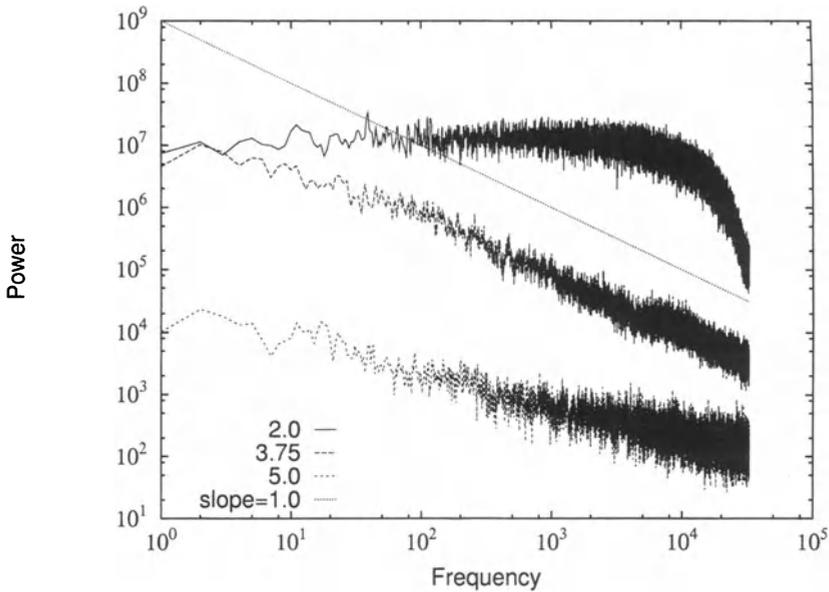


Fig. 5. Power spectrum of the stop-and-wait algorithm.

These results are completely consistent with the actual traffic's power spectrum, so we can conclude that the stop-and-wait algorithm extracts the essential mechanism of phase transition in the Internet traffic.

### 3.2 CBR Algorithm

In order to clarify the characteristics of the stop-and-wait algorithm, we performed a simulation based on the no-feedback algorithm called the CBR (Constant Bit Rate) algorithm. In the stop-and-wait algorithm, the sender node transmits a packet to the receiver node only when the acknowledgment packet for the previous packet is received. In CBR algorithm, the sender transmits packets to the receiver at a fixed time interval after a connection is established. Namely, there are no acknowledgment packets in the CBR algorithm. This algorithm cannot adapt to the current network status because it cannot change the packet transmission rate. We set the packet interval and the number of packets to transmit in a connection to 5 time steps and 20 packets, respectively.

The lines plotted in Figure 6 represent the power spectra of the traffic traces reproduced by CBR algorithm. We observe that the power spectrum changed from white noise to the power law as the connection arrival rate reached the critical value. Above the critical point, the flow density from the router was always saturated at the maximum level, because the buffer length of the router tended to diverge. Thus, we omit the plots above the critical point. Figure 7 indicates the cumulative distribution of congestion duration produced by the CBR algorithm. It is clear that the plotted line is close to a power law with

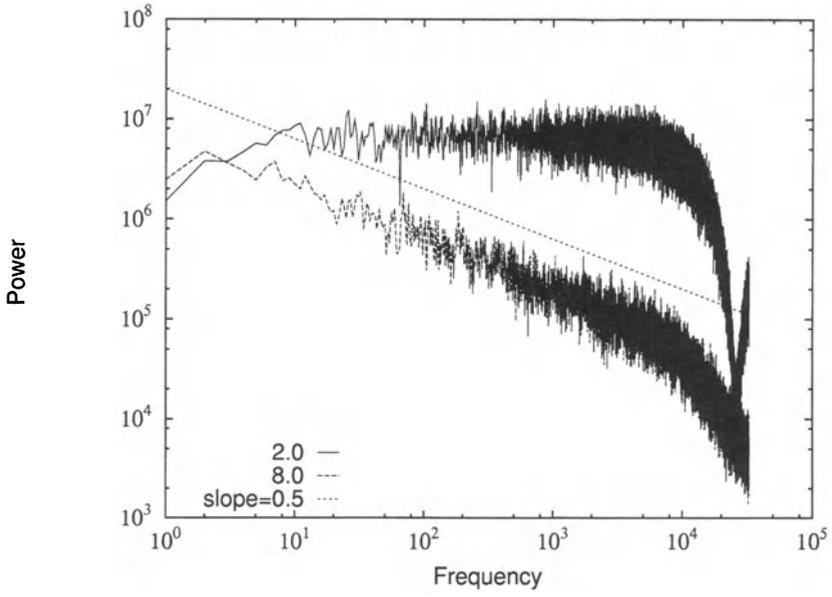


Fig. 6. Power spectrum of the CBR algorithm.

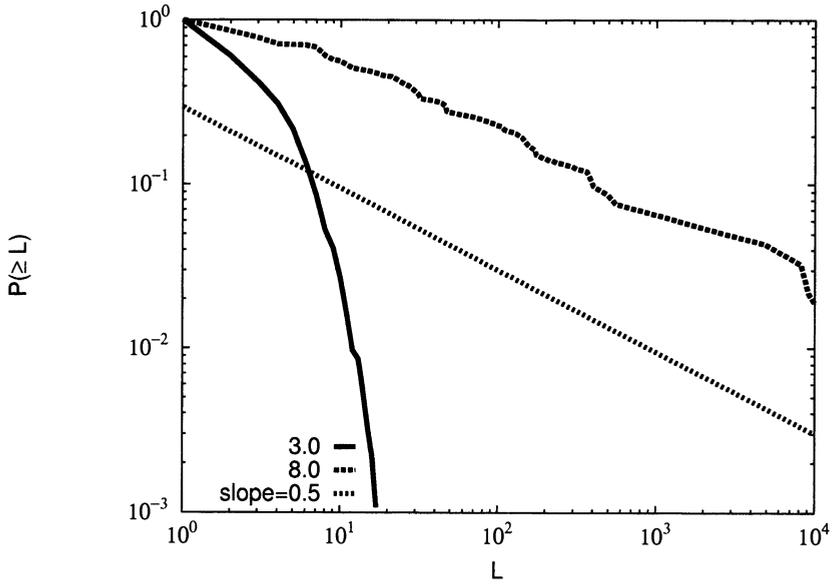


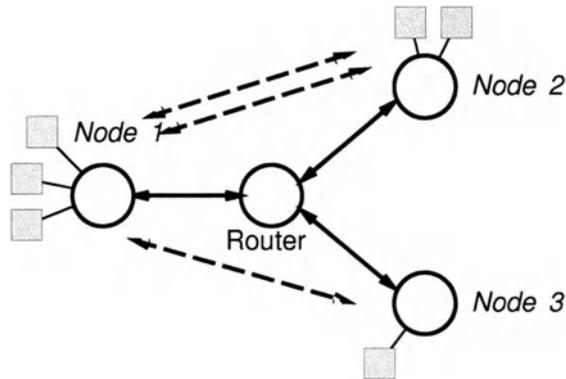
Fig. 7. Congestion duration of the CBR algorithm.

exponent of  $-0.5$  at the critical point though the plot decays exponentially below the critical point.

The important difference between the stop-and-wait and the CBR algorithms is the exponent of the power law. At the critical point, the slope of the plotted line for the CBR algorithm is approximately  $-0.5$ , not  $-1.0$ . Phase transitions characterized by the power-law with exponent of  $-0.5$  are known to occur for a single buffer with Poisson input traffic. Namely, the phase transition behavior of the traffic fluctuations reproduced by the CBR algorithm can be recognized as being produced when passing through the single buffer in the router.

### 3.3 Effect of Traffic Symmetry

Finally, we performed an extra simulation based on the stop-and-wait algorithm in order to determine the effect of traffic symmetry. Here, we restricted the direction of connection as shown in Figure 8, namely no connection was established between the Nodes 2 and 3 while connections appeared in the paths between the Nodes 1 and 2 and the Nodes 1 and 3 with equal connection arrival rate  $r$ . Figure 9 shows the power spectrum of this asymmetric case, and Figure 10



**Fig. 8.** Network configuration (2). Connections are not allowed between the Nodes 2 and 3 in this configuration.

shows the cumulative distribution of congestion duration. In both figures, the plotted lines are approximately power laws with exponent  $-0.5$  at  $r = 2.48$ , the estimated critical point. We also checked a simpler topology having two nodes and one router, and found that the power law exponent at the critical point was  $-0.5$ . These simulation results indicate that the phase transition mechanism of the asymmetric case is the same as that of the two-node case and the essence of the  $1/f$  fluctuation is due to the effect of other packets inserted into a packet stream because there are three nodes.

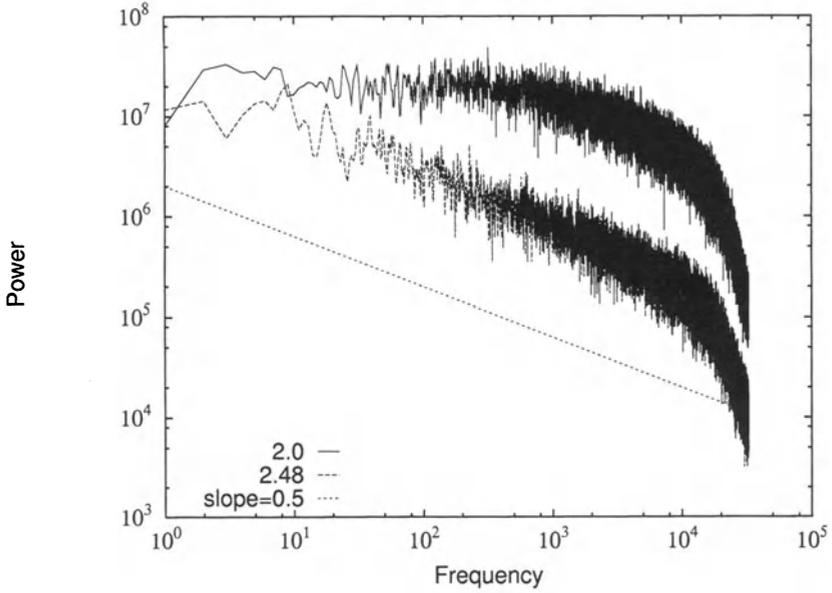


Fig. 9. Power spectrum of stop-and-wait algorithm (asymmetric case).

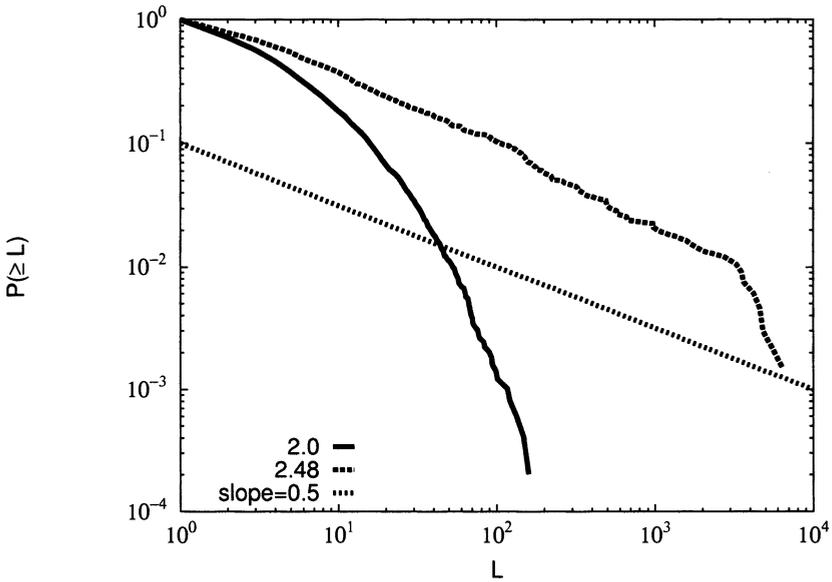


Fig. 10. Congestion duration of stop-and-wait algorithm (asymmetric case).

## 4 Discussion

Let us review our simulation results. First, we performed numerical simulations focusing on the effect of traffic control between the hosts. When the hosts deployed the stop-and-wait algorithm based on feedback control, the aggregated reproduced traffic behavior is completely consistent with the actual traffic behaviors. On the other hand, in the case of non-feedback control like the CBR algorithm, the exponent of power law at the critical point differed from the observed one. Because of the fact that TCP which is based on a feedback algorithm is more widely used in the current Internet than the CBR type non-feedback algorithm, we can conclude that this feedback effect plays an essential role in characterizing actual network traffic behaviors. In particular, it is important to note that  $1/f$  fluctuation appears without assuming any long-range dependence for input traffic.

We also found that the exponent at the critical point changed to  $-0.5$  when the connection arrival was asymmetric. Although the host used feedback control, we expected its phase transition behaviors to be identical to the single buffer case or the CBR algorithm case. Thus, we know that the most important point is the triangular correlation among three hosts by using the feedback control. The detailed analysis of the asymmetric case is an important subject for future work.

Finally, let us discuss the meaning of different exponent values for the Internet traffic. Figures 4 and 7 demonstrate that the occurrence probability of a large congestion is directly affected by the value of the exponent. We consider that a smaller probability of a large congestion is better for the network. Namely, an exponent of  $-1.0$  is better than  $-0.5$ . Thus, a well-controlled network using the stop-and-wait and Ethernet should be better than a network controlled by the CBR algorithm from the view of network performance.

## 5 Conclusion

We performed computer simulations in order to find the minimal model of phase transition in Internet traffic. Our results indicate that the feedback control among the nodes is the key factor for reproducing realistic network traffic behaviors. At the critical point, the exponent of the power law distribution was approximately  $-1.0$  without assuming any long-range dependence like power-law file size distributions. When nodes used non-feedback control like the CBR algorithm or when the number of active nodes was two, the exponent became close to  $-0.5$ , implying that a very large congestion occurred with a high probability. From the viewpoint of avoiding a large congestion, equivalent weights for the nodes and their feedback control are recommended.

In this paper, we neglected the effect of file sizes, but the file size distribution in real web traffic is known to follow a power-law distribution. In other words, input traffic in real systems is strongly correlated. Therefore, we should analyze such effects of correlated input traffic. Also, further analysis is needed to clarify

the effect of asymmetry in the network as this is now known to enhance the occurrence probability of large congestion.

## 6 Acknowledgment

This paper is partly supported by the Grants-in-Aid for Scientific Research, No.13831010, Japan Society for the Promotion of Science.

## References

1. M. Crovella and A. Bestavros, *IEEE/ACM Transactions on Networking* **5** (6), 835–846 (1997).
2. I. Csabai, *Journal of Physics A:Math.Gen.* **27**, 417–419 (1994).
3. K. Fukuda, H. Takayasu, and M. Takayasu, *Advances in Performance Analysis* **2** (1), 45–66 (1999).
4. K. Fukuda, M. Takayasu, and H. Takayasu, *Physica A* **287**, 289–301 (2000).
5. W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Willson, *IEEE/ACM Transactions on Networking* **2** (1), 1–15 (1994).
6. V. Paxson and S. Floyd, *IEEE/ACM Transactions on Networking* **3** (3), 226–244 (1995).
7. J. Postel, RFC 793 (Sep. 1981).
8. M. Takayasu, H. Takayasu, and K. Fukuda, *Physica A* **277**, 248–255 (2000).
9. M. Takayasu, H. Takayasu, and T. Sato, *Physica A* **233**, 924–934 (1996).

# Microscopic Modeling of Packet Transport in the Internet

T. Huisinga<sup>1</sup>, R. Barlovic<sup>1</sup>, W. Knospe<sup>1</sup>, A. Schadschneider<sup>2</sup>, and M. Schreckenberg<sup>1</sup>

<sup>1</sup> Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg, Lotharstr. 1, 47048 Duisburg, Germany

<sup>2</sup> Institute for Theoretical Physics, University of Cologne, Zùlpicher Str. 77, 50937 Cologne, Germany

**Abstract.** A microscopic description of packet transport in the Internet by using a simple cellular automaton model is presented. Therefore we introduced a generalized exclusion process which allows to study travel times of the particles ('data packets') along a path in the network. Computer simulations reveal the appearance of a free flow and a jammed phase separated by a (critical) transition regime. The power spectra are compared to empirical data for the RTT (Round Trip Time) obtained from measurements in the Internet. We find that the model is able to reproduce the characteristic statistical behavior in agreement with the empirical data for free flow and congested phase. Moreover the model shows, as observed in reality, critical behavior ( $1/f$ -noise) for paths with critical load.

## 1 Introduction

Nowadays computer networks play a decisive role in various fields of information transport. Especially the Internet has become a popular media for communication and the exchange of information. The enormous increase of data transfer caused by the growing demand leading to decreasing transfer rates makes it necessary to understand the basic properties of data transport for planing new connections and optimizing the usage of the existing resources. The evaluation of the statistical properties of real data measurements like the time-series of the round trip times (time a data packet needs to travel from a source to a destination host and vice versa) [1–4] or the load of a single router [5,6] is the basis for an effective analysis of the basic mechanism of data transport in the Internet. Moreover there are investigations on the social aspects of Internet and the “human factor” in the system [7]. A useful instrument in physics for analyzing time-series is the investigation of the power-spectra, which show in the case of the Internet for different paths characteristic dependencies, i.e.,  $1/f$ -noise, first presented by Csabai [1] and later by Takayasu *et al.* [2,8].

## 2 Model

In the Internet traffic data files are divided into small data packets of a definite size. These data packets move, for fixed source and destination hosts, due to the

structure of the Internet transportation protocols (TCP/IP), along a temporally fixed route. Therefore the transport between two specific hosts can be viewed as a one-dimensional process. In order to reproduce the statistical characteristics of Internet traffic we introduce a simple microscopic cellular automaton model with open boundary conditions based on the Asymmetric Simple Exclusion Process (ASEP) [10,12,15] by allowing a finite number  $B_n$  of particles (data packets) on each site (router)  $n$ . Hereby we take into account that each router has a buffer of finite capacity so that more than one data packet can be stored (multi-allocation of sites). The data packets move with a router specific probability  $p_n$  to the next router. This probability determines the amount of traffic at the network node (the current) as well as the statistical behavior of processing times. The dynamics of the system do not only depend on the probability a data packet moves to the next router, but also on the restriction of the buffers so that a data packet only moves to the next router as far as there is enough space left.

The model [13] is defined on a linear array of  $N$  sites (Fig. 1). Each site  $n = 1, \dots, N$  represents a router with a buffer which stores  $Z_n(t)$  particles. Each router has a finite capacity  $B_n$ , i.e.,  $Z_n(t) \leq B_n$ . A particle  $i$ , representing a data packet, moves with the probability  $p_n$  from site  $n$  to the next site  $n + 1$  as long as the buffer  $n + 1$  is not completely occupied. The update is performed in parallel for all packets and the travel times  $T_i$  of all packets are increased by the discrete time  $\Delta t$ . The data packets arriving at the last site  $N$  are removed with probability  $p_N$  and their travel times  $T_i$ , i.e., the times needed to travel through the system, are measured. At  $t = 0$  we start with empty buffers at all routers,

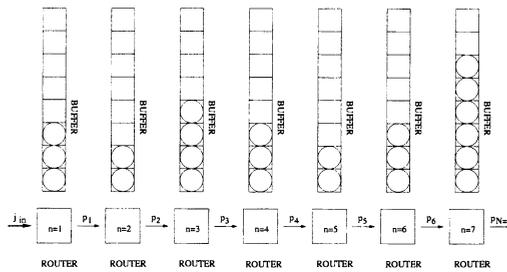


Fig. 1. System consisting of  $N = 7$  router with buffer size  $B = 8$ .

i.e.,  $Z_n(t = 0) = 0$ . In each time step the following update steps are applied in parallel:

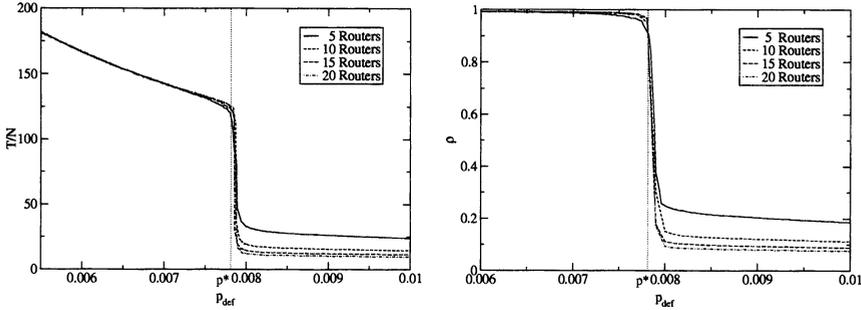
1. As long as the first router  $n = 1$  is not completely occupied  $j_{in}$  data packets are inserted:  $Z_1(t + 1) = \min(Z_1(t) + j_{in}, B_1)$ . The travel times of these packets are set to zero:  $T_i = 0$ .
2. The travel times  $T_i$  of those data packets  $i$  present in the system are increased by  $\Delta t = 1$ .
3. Data packets in the routers  $n = 1, \dots, N - 1$  move with probability  $p_n$  to the next router  $n + 1$  as long as this router is not completely occupied ( $Z_{n+1}(t) < B_{n+1}$ ).

4. Data packets in the last router which have not already been moved in the same time step are removed with probability  $p_N$  and their travel times  $T_i$  are measured.

Note that data packets in a buffer are stored in a waiting queue and therefore the packets with the highest waiting times try to move first. Moreover it is to mention that, due to the stochastic character of the movement and the multi allocation of sites particles can overtake each other which can not be found in the ASEP. Because of the parallel update each data packet can move only once during each time step. In contrast to [11] no data packets are lost. Note that for  $B_n = 1$  the model is identical to the ASEP with disorder in the hopping rates and boundary probabilities  $\alpha = 1$  and  $\beta = p_N$ .

### 3 Simulations

For the following investigations for all routers the buffer size is set to  $B = 128$ . Note that with regard to reality we restrict the number of routers to  $N = 15$  and the probabilities  $p_n$  are chosen in such a way to obtain a good agreement with empirical data. As in the ASEP [12] the state of the system is determined by the smallest of three currents, namely the maximal possible inflow, bulk flow and outflow. For  $j_n^{\text{in}} < j_n^{\text{max}}$  the dynamics of the system is governed by the dynamics of the collective behavior of the routers. In reality congestions occur, when the amount of traffic at a router exceeds its maximum capacity. In order to observe congestions in the simulation one single slow router with a lower moving probability  $p_{\text{def}}$  is inserted. This router then behaves like a bottleneck restricting the maximum flow to  $j_{\text{def}} = B p_{\text{def}}$ . Because there is no major influence of the unrestricted routers behind the bottleneck on the statistics of travel times in the system, we associate the bottleneck of a path with the boundary condition at the right end, i.e.,  $p_N = p_{\text{def}}$ . Since we are mainly interested in the impact of the slow router we restrict the inflow  $j_{\text{in}}$  to one data packet per update. Varying  $p_{\text{def}}$ , computer simulations reveal the existence of two phases which can be distinguished by the behavior of the travel times and the average density in the system (see Fig. 2). The travel times are obtained in simulations by summing up the waiting times  $\tau_{i,n}$  of every single data packet  $i$  in the routers along the path:  $T_i = \sum_{n=1}^N \tau_{i,n}$ . The behavior of the system is determined by the relation between  $p_{\text{def}}$  and  $p^*$  where  $p^*$  corresponds to the point of maximum bulk flow whereby  $p^*$  can be estimated with  $p^* = \frac{j_{\text{in}}}{B_{\text{def}}} = \frac{1}{B}$ . For  $p_{\text{def}} = p^*$  the inflow is equal to the maximum flow  $p_{\text{def}} B_{\text{def}}$  through the last router. For probabilities  $p_{\text{def}} > p^*$  the maximum flow through the bottleneck is higher than the inflow  $j_{\text{in}}$ . In this free flow system the mean flow only depends on the average capacity of each single router. Contrary, for probabilities  $p_{\text{def}} < p^*$  the mean flow through the bottleneck is lower than the inflow  $j_{\text{in}}$  and the system gets jammed. In the jammed state, the maximum system flow is determined by the maximum capacity of the bottleneck. Data packets can only move to the next router when a data packet left it a time step before. For  $p_{\text{def}} \approx p^*$  the mean flow through the

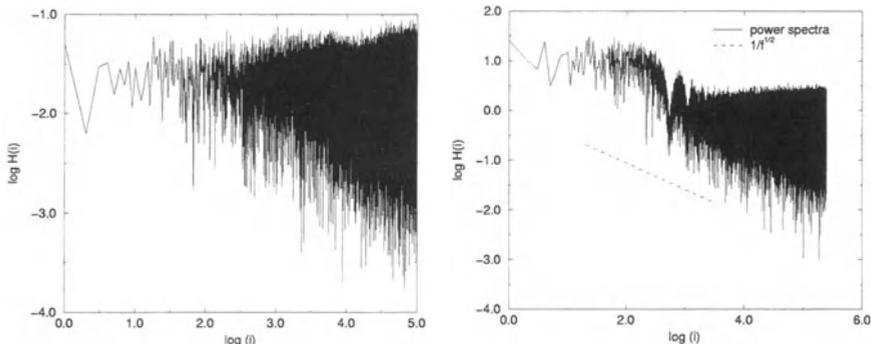


**Fig. 2.** **Left:** Diagram of the mean travel times  $T$  of data packets in dependence of the probability  $p_{\text{def}}$  of the last router for  $(N = 5, 10, 15, 20)$ ,  $B = 128$ , and  $p_n = 0.2$ . **Right:** Relation between the density  $\rho$  of data packets and the probability  $p_{\text{def}}$  of the last router for  $(N = 5, 10, 15, 20)$  routers. The other parameters correspond to the ones of the left diagram.

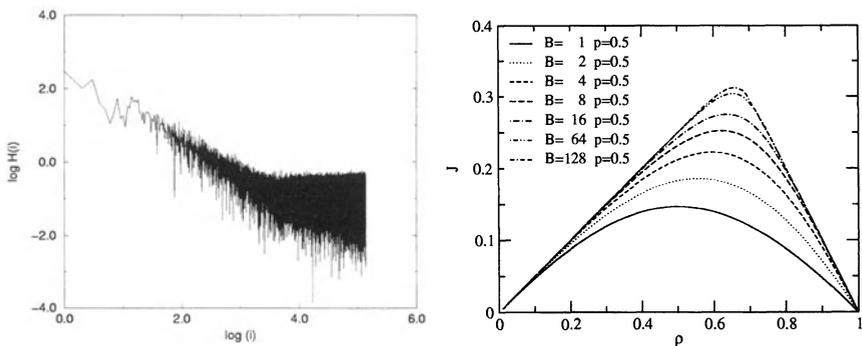
bottleneck is equal to the inflow  $j_{\text{in}}$  which means that the system operates at its maximum capacity. The existence of two well defined regimes in the presence of a defect router  $p_{\text{def}}$  is also confirmed by measurements of the mean density  $\rho$  of the system (see Fig. 2) which is defined by  $\rho = \sum_{n=1}^N \frac{Z_n}{B}$ . In Fig. 2 one can distinguish a free flow state with low density for  $p_{\text{def}} > p^*$  and a jammed state with high density for  $p_{\text{def}} < p^*$  in agreement with the results for the travel time. To compare the simulation results with empirical travel times from ping experiments, we investigated the statistics in the jammed and the free flow regime as well as in the transition region between these two regimes at  $p_{\text{def}} = p^*$ . Therefore we generated the power spectra of the travel times and analyzed the spectral density. The left part of Fig. 3 shows the power spectrum of a free flow system ( $p_{\text{def}} \gg p^* = 1/B$ ). White noise is found for the whole frequency range. This means that correlations in the travel times of the data packets are negligible small. The data packets move with probability  $p_n$  from one router to the next one without any limitation caused by the buffer restrictions. In contrast, jammed systems ( $p_{\text{def}} \ll p^*$ ) show an algebraic decay with an approximately  $1/f^{1/2}$  dependence at low frequencies (see right part of Fig. 3). Considering the occupancy of the buffer as a time dependent variable, the interval distribution of one jammed buffer corresponds to the first recurrence time in the random walk problem. Such a system then shows  $1/f^{1/2}$ -noise in the power spectrum and white noise at higher frequencies [2]. In the transition regime in the vicinity of  $p^*$  the power spectra of the travel times show characteristic  $1/f$ -noise (see Fig. 4) at low frequencies (long range correlations, critical behavior). All of the above findings of the statistical analysis of travel times generated by simulations of our simple model are in full agreement with the characteristic properties of measurements of ping time series in the Internet [2,4].

To show some other characteristic features of our model we further investigated the fundamental diagram for different buffer sizes. Therefore we initialized a system with periodic boundary conditions (data packets move according to the

update rules throughout a periodic system and are not deleted) with an arbitrary amount of particles and measured the system flow (Fig. 4). In contrast to other models where particles can overtake each other like cellular automaton models for multilane traffic (see [14]) the maximum of the fundamental diagram is shifted to densities higher than 1/2. This is caused by the fact that due to the noise only a part of packets try to move. However, the fraction of free cells in a router needed to permit unhindered movement is less than 1/2 of the buffer size.



**Fig. 3. Left:** Power spectrum of a free flow system showing white noise ( $N = 15$ ,  $B = 128$ ,  $p = 0.2$ , and  $p_{\text{def}} = 0.1$ ). **Right:** Power spectrum of a jammed system with  $1/f^{1/2}$ -noise at low and white noise at high frequencies ( $N = 15$ ,  $B = 128$ ,  $p = 0.2$ , and  $p_{\text{def}} = 0.003$ ).



**Fig. 4. Left:** Power spectrum of a system at critical load ( $N = 15$ ,  $B = 128$ ,  $p = 0.2$ , and  $p_{\text{def}} = 0.788$ ). One finds  $1/f$ -noise at low frequencies and white noise at high frequencies. **Right:** Fundamental diagram for a periodic system with  $B = 1, 2, 4, 8, 16, 64, 128$  and  $p = 0.5$  for all routers.

## 4 Discussion

A simple cellular automaton model for the Internet data packet transport along a fixed path in the Internet is introduced. This can be seen as a asymmetric exclusion process where occupation of sites (buffers) by more than one particle (data packet) is allowed. By means of computer simulations we found a jammed and free flow phase in the presence of a slow router. To compare our model with real Internet data we focused on the dynamic behavior of the travel times and their correlations. The analysis of travel times shows the typical power spectra of real Internet traffic in the two regimes, i.e., white noise for free flow and  $1/f^{1/2}$  for the jammed system. In the transition regime between these two phases the model shows a characteristic  $1/f$ -noise.

Future work should characterize the transition in more detail. However, our investigations revealed that many of the statistical properties of Internet traffic can already be understood by a simple one-dimensional model. Moreover there are investigations of a two dimensional extension of the model where the nodes can act as source and destination hosts showing temporal growth of perturbations.

## References

1. I. Csabai, J. Phys. **A27**, 417 (1994).
2. M. Takayasu and H. Takayasu, Physica **A233**, 924 (1996).
3. M. Takayasu, H. Takayasu, and K. Fukuda, Physica **A274**, 144 (1999); Physica **A277**, 248 (2000).
4. T. Huisinga, Diploma Thesis, Duisburg University (2000).
5. W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Willson, IEEE/ACM Trans. Networking **2**, 1 (1994).
6. W. Willinger, M.S. Taqqu, R. Sherman, and D.V. Wilson, IEEE/ACM Trans. Networking **5**, 71 (1997).
7. B.A. Huberman and R.M. Lukose, Science **277**, 535 (1997).
8. M. Takayasu, A.Yu. Tretyakov, K. Fukuda, and H. Takayasu, in: *Traffic and Granular Flow '97*, M. Schreckenberg and D.E. Wolf, (Eds.), p.57, (Springer, 1998).
9. K. Nagel and M. Schreckenberg, J. Phys. I France **2**, 2221 (1992).
10. D. Chowdhury, L. Santen, and A. Schadschneider, Phys. Rep. **329**, 199 (2000), and references therein.
11. N. Vandewalle, D. Strivay, H.P. Garnir, and M. Ausloos, in: *Traffic and Granular Flow '97*, M. Schreckenberg and D.E. Wolf, (Eds.), p. 75, (Springer, 1998).
12. G.M. Schütz, in: *Phase Transitions and Critical Phenomena*, Vol. 19, C. Domb and J.L. Lebowitz, (Eds.), (Academic Press, 2000), and references therein.
13. T. Huisinga, R. Barlovic, W. Knospe, A. Schadschneider, and M. Schreckenberg, Physica **A294**, 246 (2001).
14. W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, Physica **A265**, 614 (1999).
15. J. Krug, Phys. Rev. Lett. **67**, 1882 (1991).

**Granular**

# Avalanches and Flow Dynamics of a Collapsing Granular Pile

L.C. Jia<sup>1,2</sup>, P.-Y. Lai<sup>1</sup>, and C.K. Chan<sup>2</sup>

<sup>1</sup> Department of Physics and Center for Complex Systems, National Central University, Chung-Li, Taiwan 320, Republic of China

<sup>2</sup> Institute of Physics, Academia Sinica, Nankang, Taipei, Taiwan 115, Republic of China

**Abstract.** Results of an experimental investigation of the dynamics of flow in a collapsing granular pile are reported. The collapse of the pile is induced by reducing the length of the base support of the triangular pile. The dynamics are studied by using the statistics of the avalanches during the collapse. It is found that the collapse consists of two types of avalanches and two angles of repose are needed to describe the shape of the collapsing granular pile. Corresponding to these two angles of repose, the collapse of the granular pile can be characterized by local and global avalanches. A system-wide global avalanche will be triggered only when a large enough part of the base of the granular pile is removed. Otherwise, small avalanches localized near the tip of the pile will be produced. It is also found that the probability distributions of the avalanches depend on the sizes of the avalanches under consideration. The size of the localized critical tip corresponding to the local avalanche is independent of the size of the whole grain pile and can be estimated theoretically from a simple static friction model. Furthermore, the surface flow dynamics during the global avalanche is also measured. The mean velocity of the grains is found to decay exponentially into the depth of the surface layer. A slipping layer can be identified which has a non-trivial time dependence.

## 1 Introduction

It is well-known that a pile of grains can be packed to sustain a stable inclined free surface, and avalanches can be induced to occur on its free surface as a result of disturbances. There has been an increasing number of studies on the properties of sandpile because of its fundamental importance in self-organized criticality [1] and relation to practical problems of granular flow [2–4] and landslides. Usually, the creation of a granular pile is by adding the granular material on the top of the pile either randomly [5] or at a particular location [6]. The importance of these studies lies in the believe that the avalanches induced by the addition of a single grain of sand would obey a power law distribution [7]. This power law dependence has been found in rice piles [8]. There have been several theoretical attempts to investigate the properties of these surface granular flows during avalanches, most of them are phenomenological model that use the surface profile as the dynamic variable [9,10]. These models can account for some macroscopic properties of granular

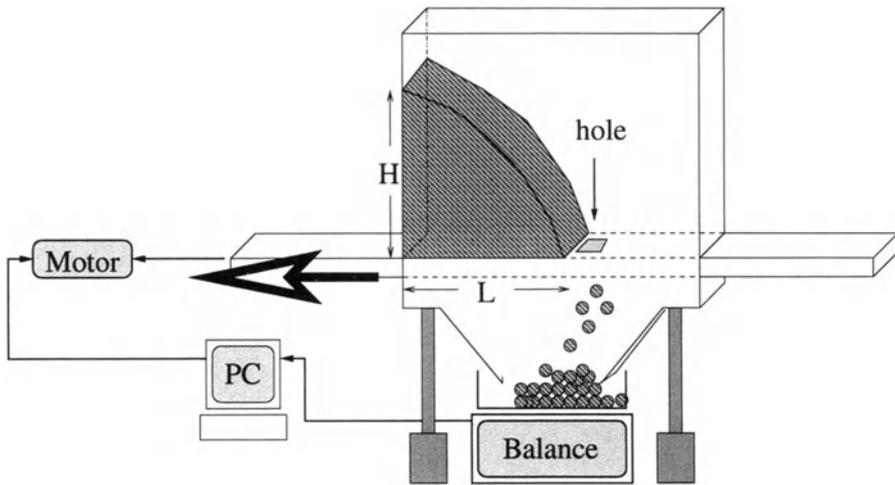
flow in some experiments such as rotating drum. However, there is little studies, especially experimentally, on the detailed dynamics of the grains during the avalanche.

On the other hand, in actual situations, another type of avalanches can also be produced in granular systems by the removal of grains or collapse of structure close to the bottom of the pile. For example, in the landslide [11] close to rivers, the collapses of nearby slopes are mainly due to the erosions of the river bed. Despite the practical importance of these avalanches, very little is known about the properties of these avalanches. Here we report results of an experimental investigation of the structures of these collapsing granular files as well as the flow properties of these avalanches induced by removing grain at the tip of the pile. Presumably, the transition from the static to the flowing states in the granular pile could reveal the nature of these avalanches. The detail dynamics, such as the velocity distribution in the fluidized flowing layer, is monitored by high speed imaging techniques. The transition from the static to flowing regimes can be studied in detail by observing the growth of the fluidized surface layer in the onset of the avalanches. The statistical properties of the avalanches are also investigated similar to the studies in Self-organized criticality. The mechanics of the grains in the flowing layer is related to the physics of granular materials under shear and in rapid flow.

## 2 Experimental Setup and Procedure

The experimental apparatus we used is similar to other experiments [8] in which the grains are confined in a Plexi-glass rectangular box of dimensions: 40cm (length)  $\times$  30cm (height)  $\times$  1cm (width). However, the bottom of the box is made of a movable Aluminium base plate with a hole (0.8cm  $\times$  2cm) as shown in Fig. 1. As in many other grain pile experiments, a triangular granular pile of height  $H$  and base length  $L$  is first built up to a steady state. Initially, a granular pile is built up by pouring rice into the Plexi-glass box slowly from the top at the left. At first, the grains will stay close to the position where they land. However, as more grains are added, they rest on top of one another, creating a pile. Avalanches will take place in the pile as grains are added. The base size of the pile will grow until a steady state is reached in which the amount of rice added is balanced, on average, by the amount of rice falling off the hole. Although  $L$  is fixed, the height  $H$  can still fluctuate in this steady state. The pile is said to be in the critical state when  $H$  assumes its maximum value. The avalanches are then induced by the removal of grains from the foot of the pile, not by addition; in contrast with other sand-pile or rice-pile experiments [2,3]. The granular pile built up in the box with a base distance of  $L$  determined by the position of the hole. With the base plate mounted on a computer controlled translation stage, the base size of the pile can be adjusted by the position of the hole. Collapses of the pile can then be triggered by moving the base plate from left to right in

Fig. 1 and hence reducing the base size  $L$  of the pile. Since grains will fall out from the box through the hole during the collapse, the size of the hole must be chosen to be large enough that there is no accumulation of grains around the hole. An analytic balance (Sartorius BP3100S), with a precision of 0.01 gram and a capacity of 3000 grams, is used to measure the total weight of grains dropped through the hole. The balance is interfaced to a PC which records the mass of the fallen grains.



**Fig. 1.** Schematic illustration of the experimental apparatus.

Most of the data presented here are from two types of rice grains, a roughly spherical round grain (sago) and a long grain (white rice), but we have performed experiments on six different rice grains (some are round and some are long), all of them show similar behaviors. The critical angles of repose,  $\theta$ , in the critical state is measured to be  $38.9^\circ$ . After reaching the critical state of the pile, no more grains are added to the pile. Experiments to study the behavior of the avalanches of the critical system can then be started by decreasing the base size of the pile with the computer controlled translation stage in the following manner. First, the position of the hole is shifted 1mm to the left. The reading of the balance is then checked to see if grains are falling from the pile. If grains are falling onto the balance, the reading of the balance will be increasing with time. After the reading of the balance has stabilized (no changes within 5 sec), the total mass of grains on the balance and the base size  $L$  are recorded. This process of shift-and-read can be repeated until  $L$  is smaller than 5cm. Note that the discharge mass  $M$  for every step can be computed from the recorded total mass. The critical state of the initial pile is usually prepared with  $L = 30\text{cm}$  (the order of 100 grains). The precision of the translation stage is 0.1mm which is much smaller

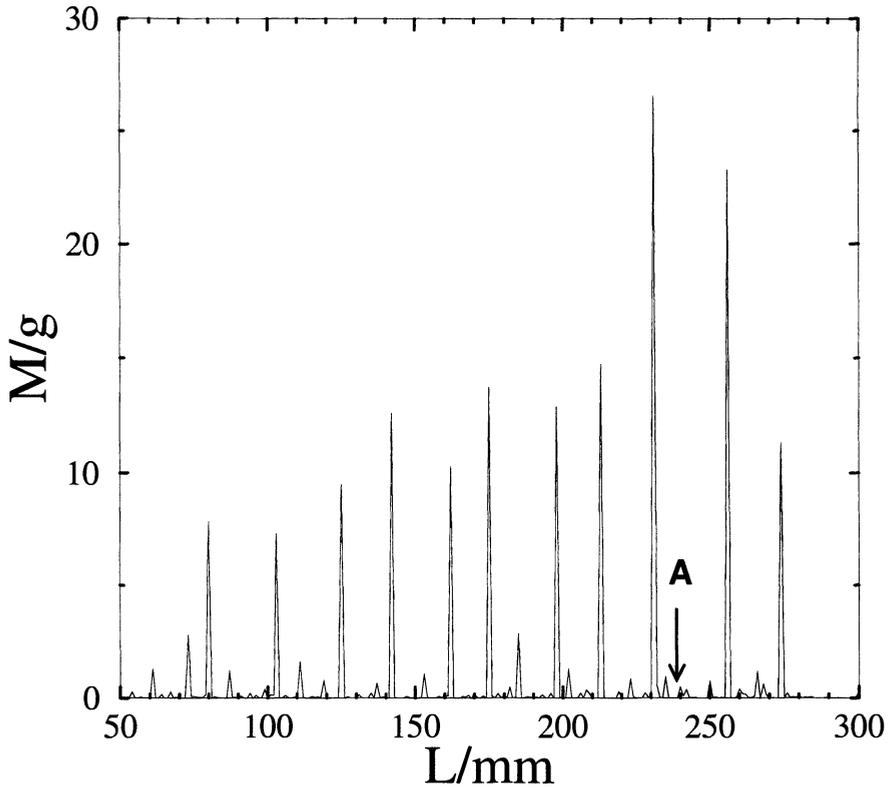
than the grain size. However, the sensitivity of the balance is not good enough to detect the presence of one single grain. The uncertainty of the discharge  $M$  is about 5 grains.

For visualization purposes, and for the ability to track the trajectory of grain during the avalanche process, a video measuring system was setup. The camera was a Kodak SR-500. Particle velocities are extracted from sequential images taken with a high speed camera at various time intervals  $\delta t$ . In order to measure particle velocities, the time intervals  $\delta t$  is chosen at 1/1000 sec. The resolution of the images is  $256 \times 240$  for  $\delta t = 1/1000$ .

### 3 Global and Local Avalanches

Fig. 2 shows a typical result of the above shift-and-read procedure from  $L = 30$  to 5cm. The diameter of the spherical grain is  $d \simeq 2$ mm. One remarkable feature is that there seems to be two types of avalanches, namely large and small. In fact in all the six different grains (some are round and some are elongated, with different aspect ratios) we used, all of them show the similar behaviors. It can be seen from the figure that the discharge  $M$  are usually small (less than 1 g). Especially in the beginning of the experiment ( $L \sim 30$ cm), most of the shifts produce zero or small  $M$ . That is: removals of grains from the bottom of the initial critical pile produce very little effect on the pile. However, a large  $M$  will occur when  $L$  has been substantially decreased. The pile seems to return to its initial state after the large avalanches because no large avalanches can be observed until  $L$  is again substantially decreased. From the figure, the small and then large avalanches of the system seem to be arranged in a more or less periodic cycle with the large avalanches separated by about 1 to 2 cm in  $L$ . The main characteristic in this repeating cycle of avalanches is that the size of the large discharges decrease with  $L$ .

To understand the observed behavior of Fig. 2, a CCD camera is used to monitor the discharge of the grains close to the foot of the pile. Video pictures from the bottom of the pile reveal that there are two angles of repose as shown in Fig. 3. This picture is taken after some small discharges have been taken place in the pile such as at point A in Fig. 2. These small discharges are induced by the local height differences when the grains on the foot of the pile fall off the box. After the small avalanches have stopped, the system reaches a meta-stable state. A local angle of repose,  $\phi$ , larger than the overall global one, will be formed. However, this local angle of repose can only sustain the mechanical balance of a small fraction of the pile close to the foot. Further small avalanches from the pile shown in Fig. 3 will trigger global avalanches in such a way that layers of grains from the whole length of the pile will be discharged and the initial global angle of repose will be restored. Therefore, the small and large avalanches are in fact local and global avalanches of the pile in response to the reduction of  $L$ . Once the global angle of repose is restored, this local and global avalanches cycle will repeat if  $L$  is further



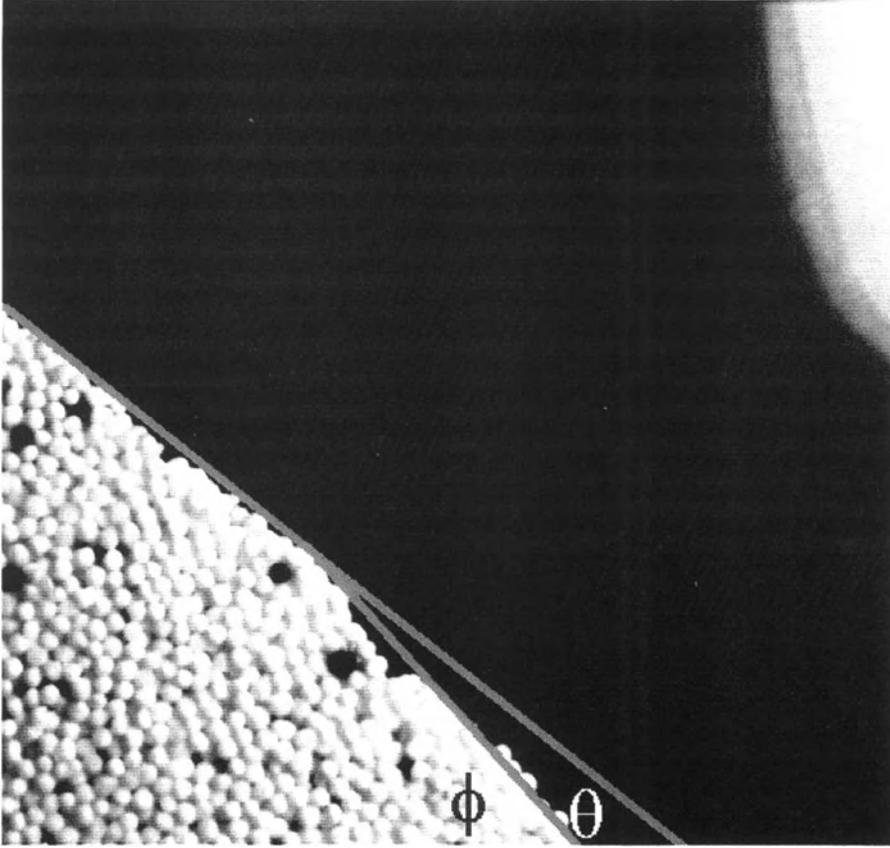
**Fig. 2.** Variation of the discharge mass  $M$  for different system size  $L$ .

decreased. In the experiment, it is found that the global angle of repose  $\theta$  is close to but less than  $\theta_c$  while the local angle of repose  $\phi$  is larger than  $\theta_c$ . Table 1 lists the measured values of  $\theta$  and  $\phi$  for six different types of grains.

**Table 1.** Critical angles of a collapsed pile.

Grain type	millet	sago	round rice	white rice (long grain)	dark rice (long grain)	oats
$\theta$ (deg.)	38.01	38.86	32.46	33.28	33.98	33.30
$\phi$ (deg.)	41.54	43.81	39.72	47.58	47.01	52.02

The observation of the local angle of repose suggests the existence of a slip plane close to the foot of the pile. The physical origin of this slip plane

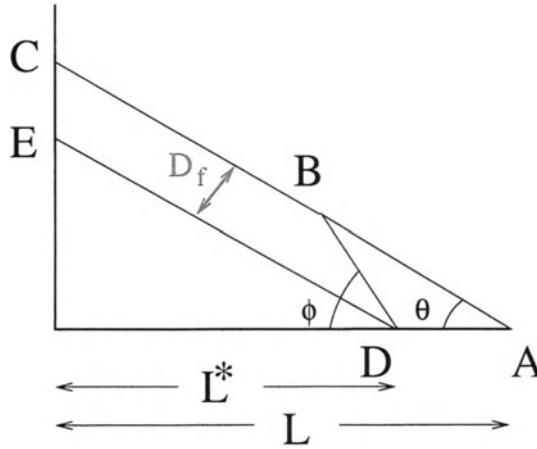


**Fig. 3.** A close up photograph of a granular pile of round grains showing the metastable slope at the foot of the pile. Grains of darker color are dyed for flow visualization.

is the nonuniform distribution of stresses inside the granular pile which is a common phenomenon for granular structures [12]. With this slip plane, the local and global avalanches can be easily understood in a model shown in Fig. 4. In the figure, the granular pile with base size  $L$  is originally at a global angle of repose,  $\theta$ . If there is a slip plane  $\overline{BD}$ , a decrease of the base size from  $L$  to  $L^*$  will trigger avalanches which involve grains only from the region BDA. Thus, a new local angle of repose,  $\phi$ , will be formed. However, further decrease of the base size from  $L^*$  will trigger the sliding of the whole block CEDB. The original global angle of repose,  $\theta$ , will be restored after the sliding. A consequence of this model is that the height of the pile,  $H$  will be a step function of  $L$ . The global avalanches seem to be governed by the location of the slip plane or the stress distribution in the pile. The statistical properties of these global and local avalanches have been reported in Ref.

[13]. The global avalanche size following an exponential distribution with a characteristic size that is proportional to the size of the whole grain pile. More interesting is the local avalanches which originates from the meta-stable state at the foot of the pile seem to have a non-trivial (power law) dynamics.

In Fig. 4, the granular pile with base size  $L$  is originally at a global angle of repose,  $\theta$ . An important prediction of the above model is that the size



**Fig. 4.** Schematic of the slip plane model for the global and local avalanches.

of the global avalanches will be proportional to  $L$ . The discharge mass is in proportion to the area  $A_l$  of BCDE. Define the depth of the fluidized layer is  $D_f$ , then

$$\overline{AD} = D_f / \sin \theta \quad (1)$$

and

$$\overline{AB} = D_f \cdot [\cot \theta + \cot(\phi - \theta)]. \quad (2)$$

From Fig. 4, we can obtain that the area ( $A_l$ ) of BCDE is

$$A_l = \frac{LD_f}{\cos \theta} - \frac{D_f^2}{2} \cdot [2 \csc 2\theta + \cot \theta + \cot(\phi - \theta)]. \quad (3)$$

So the discharge mass  $M$  of the global avalanches is given by

$$M \simeq \frac{m_0}{d^2} \cdot \left( \frac{LD_f}{\cos \theta} - O(D_f^2) \right) \quad (4)$$

Since  $L \gg D_f$ , it is useful to define the dimensionless quantity

$$\ell \equiv \frac{Md \cos \theta}{m_0 L} \simeq \frac{D_f}{d} \quad (5)$$

which is the effective number of granular layers discharged. Therefore, data from a single experiment can be used to describe the statistics of the global avalanches. From the global avalanche data in Fig. 2, one gets  $\ell \sim 5 - 6$  layers. From our data, the assumption that the global avalanche size is linear to the system size seems to be valid. Note that, for the small local avalanches, the triangle BDA in Fig. 4 is independent of system size,  $L$ . It has been shown in Ref. [13] that the global avalanches follow an exponential form:  $P(\ell) \sim \exp(\ell/\ell_0)$  for both round and long grains but with different values of the characteristic  $\ell_0$ . The existence of a characteristic  $\ell_0$  for both long and round grains suggests that the dynamics of the global avalanches is very different from that of self-organized criticality. The geometric shape of the grains which plays an important role in the self-organized critical behavior of rice piles seems not to be important here. Presumably, the position of the slip plane determines the value of this characteristic  $\ell_0$ . On the other hand, unlike the distribution of the global avalanches, the distributions for the local avalanches seem to follow a power law:  $P(M) \sim M^{-\alpha}$ .  $\alpha$  takes different values for long and round grains.

#### 4 Theoretical Estimate for the Sizes of Small Avalanches

Motivated by the fact that the size of the critical tip is independent of the size of the whole grain pile, the local frictional forces near the tip should be responsible for such a stable state. Here we attempt to estimate the size of this critical tip in terms of grain material properties. Suppose there are  $K + 1$  grains on the side  $\overline{BD}$  (labelled as  $0, 1, 2 \dots K$ ) in the critical tip of the pile just before the global avalanche occurs, as shown in Fig. 5. Expressing all forces in units of  $m_0g$  and balancing the forces acting on the  $j$ th-grain, one gets

$$\cos \phi = N_j + g_j - g_{j+1} \quad (6)$$

$$\sin \phi = P_j - P_{j+1} + f_j \quad j = 1, 2, \dots, K - 1 \quad (7)$$

$$f_j = g_j + g_{j+1} \quad (8)$$

$$f_j \leq \mu N_j \quad (9)$$

Similar equations holds for the  $j = K^{th}$  (last) grain with  $g_{K+1} = P_{K+1} = 0$ . Further impose the condition of  $f_j = \mu N_j$  for marginally frictional forces,  $g_j$ 's,  $N_j$ 's and  $P_j$ 's ( $j = 1, 2, \dots, K$ ) can be solved. In particular

$$P_1 \simeq K(\sin \phi - \mu) + \frac{\mu^2 \cos \phi}{2} \quad (10)$$

$$g_1 = \frac{\mu \cos \phi}{2} \left[ 1 - \left( \frac{\mu - 1}{\mu + 1} \right)^K \right] \simeq \frac{\mu \cos \phi}{2} \quad (11)$$

since in practice  $\left| \left( \frac{\mu-1}{\mu+1} \right)^K \right| \ll 1$ . For the  $0^{th}$  (bottom) grain, one has,

$$\cos \phi = N_o + P_o \cos \phi - f_w \sin \phi - g_1 \tag{12}$$

$$\sin \phi = P_o \sin \phi + f_o + f_w \cos \phi - P_1 \tag{13}$$

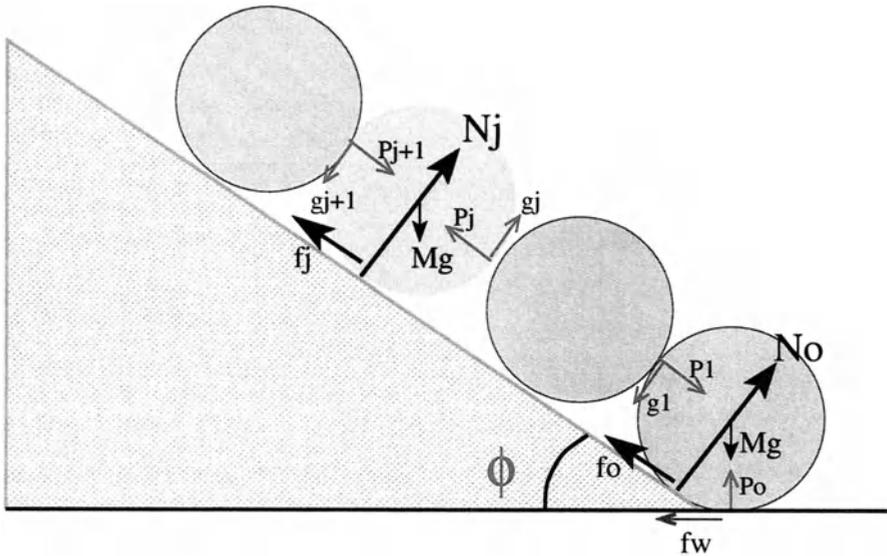
$$f_o \leq \mu N_o \tag{14}$$

$$f_w \leq \mu_w P_o \tag{15}$$

Finally, using the values of  $g_1$ ,  $P_1$ , solving for  $K$ , under the condition of marginal stability of  $N_o = 0$ , one gets

$$K \approx \frac{\mu(\tan \phi - \mu) + \mu_w [2 \sec^2 \phi + \mu(1 + \mu \tan \phi)]}{2(\tan \phi - \mu)(1 - \mu_w \tan \phi)}. \tag{16}$$

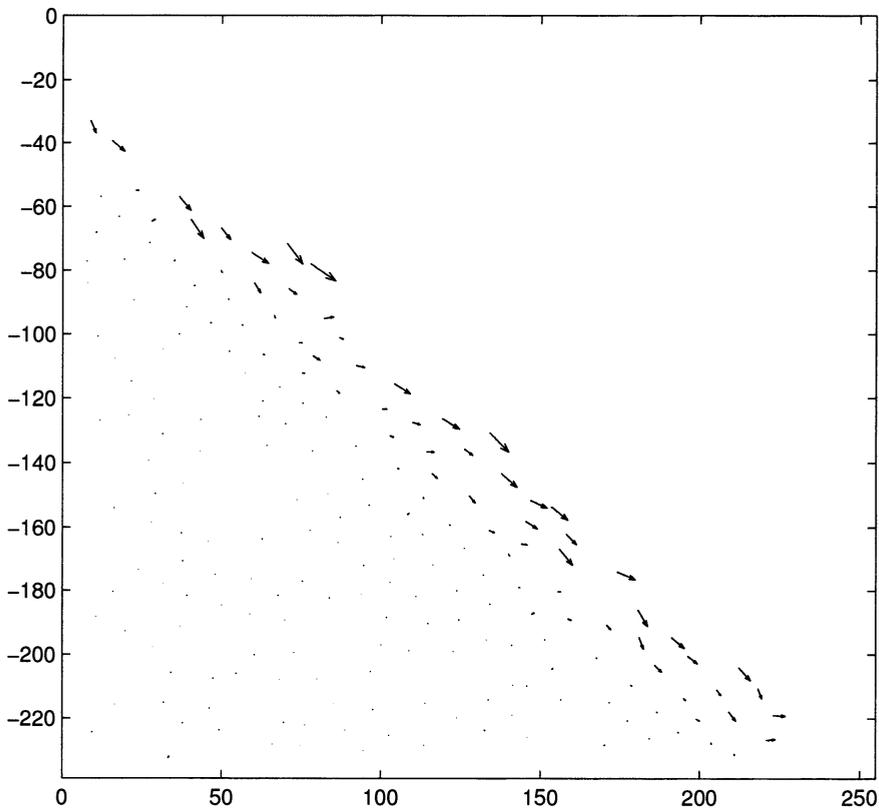
For the round grain we used in the experiment,  $\phi \simeq 43.8^\circ$ , so  $\tan \phi \simeq 0.959$  and the static sliding friction between the grains  $\mu$ , can be estimated to be  $\mu \approx \tan \theta \simeq \tan 38.9^\circ = 0.807$ .  $\mu_w$  is the friction coefficient between the grain and the base material which is aluminium. For an order of magnitude estimate, take  $\mu_w \sim 0.5$ , one gets  $K + 1 \sim 18$  which of the same order as the length  $\overline{BD}$  as estimated from the picture in Fig. 3. Once the length  $\overline{BD}$  is known, then the sum of all small avalanches can be obtained, which is proportional to the area of the  $\Delta ABD$ , and is given by  $\frac{1}{2}(K + 1)^2 \csc^2 \phi (\cot \theta - \cot \phi)$ .



**Fig. 5.** Schematic picture of  $K + 1$  grains lying on the critical tip of the pile just before the onset of a global avalanche.

## 5 Dynamics of the Global Avalanche

Landslide is induced by removing grains from the bottom of the heap as shown in Fig. 6. The mean velocity is determined from an ensemble average of particle displacement per unit time at different depth from the free surface. In Fig. 6, it seems that the mean velocity is nearly zero deep in the layer. The velocity of the surface of pile is faster and the speed of the grains decreases as one gets inside the layer. Fig. 6 also shows that the grains are almost frozen



**Fig. 6.** Velocity distribution of grain. System size  $L = 30$  cm. The axes are in units of mm.

deep in the layer. So we can obtain the thickness  $D$  from the free surface, at which the mean velocity is approximately zero. And it seems that the depth of the fluidized layer becomes thick with time and has a time dependence that is slower than an exponential approach to the long time limit  $D_\infty \simeq 1.2\text{cm} \simeq 6d$ . Clearly,  $D_\infty$  deduced from the dynamical measurement agrees well with

the layer thickness in the measurement of the global avalanches in previous section i.e.  $D_f \simeq D_\infty$ .

When the fluidized layer becomes steady, we monitored the velocities of grains to probe the properties of surface flow. The mean velocity appears to decay as the distance from the surface increases and vanishes at some finite depth. In the fast flowing regime near the surface, the velocity profile can be described by an exponential decay form, with a characteristic "skin depth". Such a behavior is not only specific to the system of spherical grains. For depth scale much larger than the skin depth, the particle motion changes from rapid surface flow to a slow stick-slip motion. In the experiment, we observed that the apparently frozen particles beneath the rapidly flowing surface are not stationary but slowly sliding. This suggests that there exists a boundary at which the particle motion changes from rapid surface flow to a slow creeping motion. Motivated by the experimental findings in previous section, we attempt to model the surface granular flow during the global avalanche in terms of a hydrodynamic model, with the viscosity of the flowing granular medium being dependent on the flow velocity. Such a model is able to predict an exponential decay velocity profile with the "skin depth" depending only on the material properties of the granular medium. More detailed of this model and experimental results will be published elsewhere [14].

## 6 Conclusion

It is clear from experiments and discussions that there are two types of avalanches in the collapse of the granular pile when the base length  $L$  is reduced. The global avalanches seem to be governed by the location of the slip plane and hence the avalanche size scales with the size of the whole pile. The scaling properties of these global avalanches are therefore different from those of Self-organized criticality. More interesting dynamics take place in the local avalanches. These small local avalanches which originates from the meta-stable state at the foot of the pile seem to have a non-trivial (power law) dynamics. Presumably, dynamics similar to those in Self-organized criticality might be present in the collapse of this meta-stable part. Geometric shapes of the rice do not seem to be important here in contrast to the case of Self-organized criticality. However, in view of the limited number of experiments performed, more thorough experimental investigation must be carried out to establish the form of the probability distributions for the local avalanches. We also found that there exists rapid surface flow and slow creep motion below this flow. It is possible to define the boundary between these two types of motion, across which the velocity profile changes continuously. The mean velocity of the grains decays exponentially into the layer depth in the rapid flowing. The fluidized surface layer of an inclined granular collapsing pile has a characteristic depth of about 5 to 6 layers. We believe that these behavior should be observable in other sheared granular systems [15].

## References

1. P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
2. G.A. Held *et al.*, Phys. Rev. Lett. **65**, 1120 (1990).
3. H.M. Jaeger, C.-h. Liu, and S.R. Nagel, Phys. Rev. Lett. **62**, 40 (1989).
4. A. Daerr and S. Douady, Nature **399**, 241 (1999).
5. V. Frette, Phys. Rev. Lett. **70**, 2762 (1993).
6. D.A. Head and G.J. Rodgers, Phys. Rev. E **56**, 1976 (1997).
7. M. Bretz, J.B. Chunningham, P.L. Kurczynski, and F. Nori, Phys. Rev. Lett. **69**, 2431 (1992).
8. V. Frette and K. Christensen *et al.*, Nature **379**, 49 (1996). Phys. Rev. Lett. **77**, 107 (1996).
9. J.-P. Bouchaud *et al.*, Phys. Rev. Lett. **74**, 1982 (1995).
10. T. Boutreux and P.G. de Gennes, J. Phys. I (France) **6**, 1295 (1996).
11. D.A. Noever, Phys. Rev. E **47**, 724 (1993).
12. R.M. Nedderman, *Statics and kinematics of granular materials*, (Cambridge, 1992).
13. L.C. Jia, P.Y. Lai, and C.K. Chan, Physica A **281**, 404 (2000).
14. L.C. Jia, P.Y. Lai, and C.K. Chan, to be published.
15. W. Losert *et al.*, Phys. Rev. Lett. **85**, 1428 (2000).

# Note on a Micropolar Gas-Kinetic Theory

H. Hayakawa

Graduate School of Human and Environmental Studies, Kyoto University, Kyoto  
606-8501, Japan

**Abstract.** The micropolar fluid mechanics and its transport coefficients are derived from the linearized Boltzmann equation of rotating particles. In the dilute limit, as expected, transport coefficients relating to microrotation are not important, but the results are useful for the description of collisional granular flow on an inclined slope.

## 1 Introduction

Micropolar fluids are fluids with micro-structures. They belong to a class of fluid with a non-symmetric stress tensor. Micropolar fluids consist of rigid, randomly oriented (or spherical) particles which have own spins and microrotations suspended in a viscous medium. The concept of microrotation, originally, has been proposed by Cosserat brothers in the theory of elasticity [1]. Condiff and Dahler [2], and Eringen [3] applied its concept to describe fluids with micro-structures in the middle of 60s. Recently, a comprehensive textbook on micropolar fluids has been published [4].

Physical examples of micropolar fluids can be seen in many fields where all of them contain intrinsic polarities. However, we believe that the most interesting application of micropolar fluid dynamics which would have potential is to characterize granular flows [5–7]. In fact, the granular flow is one of flows which have micro-structure and rotation of particles. Therefore, it is natural to introduce a fluid model which contains an equation for microrotation in addition to an equation of velocity. Along this idea, Kanatani [8] formulated micropolar fluid mechanics for granular flows, which contains the angle of repose within the frame work. Kano *et al.* [9] have confirmed the quantitative validity of a micropolar fluid model in a chute flow of granular particles from the comparison of their simulation of micropolar fluids with their experience. It is worthwhile to introduce that the velocity profile obtained from the micropolar fluid model [9] is far from the parabolic curve expected from the conventional Navier-Stokes flow.

Although we do not know how the micropolar fluid model is relevant in other situations of granular materials, it is worthwhile to investigate fundamental properties of micropolar fluids from the view of granular physics (see e.g. [10]). Thus, the micropolar fluid mechanics is not a nonsense generalization of the Navier-Stokes model, but is a physically relevant model which has many applications in physical systems.

Recently, Mitarai *et al.* [11] have analyzed the collisional flow of particles on an inclined slope. They suggested that the kinetic theory of granular particles is useful to derive micropolar fluid mechanics, and the fluid equation is quantitatively relevant to characterize the granular flow. Their success of the application of micropolar fluid mechanics to granular flows makes us understand microscopic origin of the micropolar fluid mechanics. Such attempts have been discussed within kinetic theory of polyatomic fluids [12,13] and there are some applications to granular fluid [14], but most of their works are not well accepted in these days.

The purpose of this paper is to summarize our current understanding of kinetic theory of micropolar gases. For the sake of simplicity we restrict our interest to two-dimensional perfectly rough particles in the dilute limit. To avoid complication which may be one of reasons for not appealing of the old kinetic theory of rough particles, we adopt the method of eigenvalue analysis of linearized Boltzmann equation [15].

The organization of this paper is as follows. In section 2 we explain the outline of micropolar fluid mechanics and its eigenvalue analysis. In section 3, we shortly summarize the classical mechanics of binary collisions of identical rough disks. In section 4, we explain the framework of kinetic theory of dilute gases of rough disks. We also summarize the relation between the eigenvalues of linearized Boltzmann equation and the transport coefficients. In section 5, we demonstrate how to obtain transport coefficients in micropolar fluids. In section 6 we will discuss our results.

## 2 Outline of Micropolar Fluid Mechanics and its Eigenvalue Analysis

### 2.1 Outline of Micropolar Fluid Mechanics

In this section, we explain the outline of micropolar fluid mechanics. The conventional fluid mechanics consists of equations of continuity of mass, linear momentum and energy. In the fluid without microstructure the conservation of angular momentum is automatically satisfied but it becomes a nontrivial conservation law in the fluid with microstructure. Therefore we need an extra equation of angular momentum in micropolar fluid mechanics. In collisional flows we can adopt the differential expansion of the strain field in the stress tensor, the couple stress and the heat flux. The constitutive equation is similar to that in the Newtonian fluid. In this paper we restrict our interest to two dimensional flows.

The equation of mass conservation is the same as that in usual fluid

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (1)$$

where  $\partial_t = \partial/\partial t$ , and  $\rho$  and  $\mathbf{u} = (u_x, u_y, 0)$  are the mass density and the macroscopic velocity field, respectively.

The continuity equation of linear momentum, on the other hand, becomes

$$\rho D_t \mathbf{u} = \nabla \cdot \sigma, \quad (2)$$

where  $D_t = \partial_t + (\mathbf{u} \cdot \nabla)$  is Lagrange's derivative. Here the stress tensor  $\sigma$  contains the asymmetric part. In collisional flows in two dimension, the stress tensor  $\sigma_{ij}$  can be expanded by strains

$$\begin{aligned} \sigma_{ij} = & (-p + \zeta \partial_k u_k) \delta_{ij} + \mu (\partial_i u_j + \partial_j u_i - \partial_k u_k \delta_{ij}) \\ & + \mu_r (\partial_i u_j - \partial_j u_i - 2\epsilon_{kij} w_k), \end{aligned} \quad (3)$$

where  $\epsilon_{kij}$  and  $w_k = w \delta_{k,3}$  are respectively Eddigton's epsilon and the micro-rotation (macroscopic spin field). Thus, the micropolar fluid mechanics has the spin viscosity  $\mu_r$  except for the usual viscosity  $\mu$  and the bulk viscosity  $\zeta$ .

The continuity equation of angular velocity is

$$nID_t w = \partial_j C_{j3} + \epsilon_{3ij} \sigma_{ij}, \quad (4)$$

where the component of the couple stress  $C_{ij}$  for two-dimensional micropolar flow is given by  $C_{13} = \mu_B \partial_x w$ ,  $C_{23} = \mu_B \partial_y w$  and  $C_{33} = 0$ .  $n$  and  $I$  are respectively the number density and the momentum of inertia of each disk. There is the relation between  $n$  and  $\rho$  as  $\rho = nm$  with the mass of a particle  $m$ .

The energy conservation law, in general, can be written as

$$\rho D_t e = -\partial_i q_i + \sigma_{ij} \partial_i u_j + C_{ij} \partial_i w_j - \epsilon_{kij} \sigma_{ij} w_k - \Gamma, \quad (5)$$

where  $e$ ,  $\Gamma$  and  $q_i$  are respectively the energy density, the dissipation rate by inelastic collisions and the heat flux. Now we may assume Fourier's law  $q_i = -\kappa \partial_i T$  with the heat conductivity  $\kappa$  and the granular temperature. In general, the energy density of micropolar fluid consists of two parts, the translational energy and the rotational energy. However, the mechanism of the energy transfer between two parts is not simple [16]. Here, we are only interested in the total energy density of both parts. We also assume  $\Gamma = 0$  to characterize the behavior of small dissipations. We will discuss the complete treatment including the energy transfer between the two parts and the energy loss of inelastic collisions elsewhere.

## 2.2 Linearized Hydrodynamics and its Eigenvalue Analysis

Here let us discuss the behavior of linearized hydrodynamics around the basic state  $\bar{\rho}$ ,  $\bar{e}$  and  $\bar{\mathbf{U}} = 0$ :

$$\rho = \bar{\rho} + \delta\rho, \quad \mathbf{U} = \delta\mathbf{U}, \quad e = \bar{e} + \delta e, \quad (6)$$

where  $\mathbf{U} = (\mathbf{u}, w) = (u_x, u_y, w)$  is the three dimensional velocity-microrotation vector. Assuming the relation between  $\delta e$  and  $\delta T$  as  $\delta e = C_V \delta T$  with the specific heat with constant volume  $C_V$ , the set of equations for linearized hydrodynamics in Fourier space thus becomes

$$\begin{aligned} \partial_t \rho_q &= -i\bar{\rho} \mathbf{q} \cdot \mathbf{u}_q \\ \partial_t \mathbf{u}_q &= -i\alpha \mathbf{q} \cdot \rho_q - i\beta \mathbf{q} \cdot T_q - \nu q^2 \mathbf{u}_q - \delta \mathbf{q} (\mathbf{q} \cdot \mathbf{u}_q) + 2i\nu_r \mathbf{q} \times \mathbf{w}_q \\ \partial_t w_q &= -\hat{\mu}_B q^2 w_q + 2\hat{\mu}_r (i\mathbf{q} \times \mathbf{u}_q - 2w_q) \\ \partial_t T_q &= -i\gamma \mathbf{q} \cdot \mathbf{u}_q - \eta q^2 T_q \end{aligned} \quad (7)$$

where  $\mathbf{w}_q = (0, 0, w_q)$  is the microrotation vector and the suffix  $q$  represents the Fourier component of the variable and

$$\begin{aligned} \alpha &= \frac{1}{\rho} \left( \frac{\partial p}{\partial \rho} \right)_T, & \beta &= \frac{1}{\rho} \left( \frac{\partial p}{\partial T} \right)_\rho, & \nu &= \frac{\mu + \mu_r}{\rho}, \\ \delta &= \frac{1}{\rho} (\zeta - \mu_r), & \gamma &= \frac{T}{\rho C_V} \left( \frac{\partial p}{\partial T} \right)_\rho, & \eta &= \frac{\kappa}{\rho C_V}, \\ \hat{\mu}_B &= \frac{\mu_B}{nI}, & \hat{\mu}_r &= \frac{\mu_r}{nI}, & \nu_r &= \frac{\mu_r}{\rho}. \end{aligned} \tag{8}$$

When we assume  $\mathbf{q} = q\hat{x} = (q, 0, 0)$ , the set of equations (7) can be summarized as

$$\partial_t \Psi_q = M_q \cdot \Psi_q, \tag{9}$$

where  $\Psi_q = {}^T(\rho_q, u_{x,q}, u_{y,q}, w_q, T_q)$  and

$$M_q = \begin{pmatrix} 0 & -iq\bar{\rho} & 0 & 0 & 0 \\ -i\alpha q & -(\nu + \delta)q^2 & 0 & 0 & -i\beta q \\ 0 & 0 & -\nu q^2 & -2i\nu_r q & 0 \\ 0 & 0 & -2i\hat{\mu}_r q & -\hat{\mu}_B q^2 - 4\hat{\mu}_r & 0 \\ 0 & -i\gamma q & 0 & 0 & -\eta q^2 \end{pmatrix}. \tag{10}$$

Therefore, the eigen equation can be written as

$$M_q \varphi_\alpha^q = \lambda_\alpha^q \varphi_\alpha^q. \tag{11}$$

With the aid of five eigenvectors  $\varphi_\alpha^q$ , the solution of linearized hydrodynamics (9) is represented by

$$\Psi_q(t) = \sum_{\alpha=1}^5 c_\alpha^q(t) \varphi_\alpha^q, \tag{12}$$

where  $c_\alpha^q(t)$  behaves as  $c_\alpha^q(t) = c_\alpha^q(0)e^{\lambda_\alpha^q t}$ .

The eigenvalues of (9) are connected with the transport coefficients, where the eigenvalues are the solution of

$$\det\{M_q - \lambda_\alpha^q I\} = 0. \tag{13}$$

Although the exact solution of (13) is difficult to obtain, the hydrodynamic solution near  $q = 0$  can be obtained easily. Following the method introduced in standard textbooks (see e.g. [15]), the five eigenvalues are obtained as follows: The first two eigenvalues are

$$\lambda_{1,2}^q = \pm ic_s q - \Gamma_s q^2, \tag{14}$$

where  $c_s = \sqrt{(C_p/C_V)(\partial p/\partial \rho)_T}$  with the specific heat with constant pressure  $C_p$ .  $\Gamma_s$  is the rate of sound absorption, given by  $\Gamma_s = \frac{1}{\rho} \left( \zeta + \mu + \kappa \left( \frac{1}{C_V} - \frac{1}{C_p} \right) \right)$ .

The third mode is

$$\lambda_3^q = -\frac{\mu}{\rho} q^2, \tag{15}$$

and the fifth mode is

$$\lambda_5^q = -\frac{\kappa}{\bar{\rho}C_p}q^2. \tag{16}$$

The fourth mode is the nontrivial mode which represents the relaxation of the microrotation as

$$\lambda_4^q = -\frac{4\mu_r}{nI} - \frac{1}{nI}\left(\mu_B + \frac{\mu_r I}{m}\right)q^2. \tag{17}$$

We should note that  $\lambda_4^q$  contains a constant which is independent of  $q$ . This means that the microrotation field cannot be connected with the zero eigenvectors of the linearized Boltzmann (or Enskog) equation.

### 3 Binary Collisions of Identical Rough Disks

In this section, we briefly summarize the result of binary collisions of two-dimensional, identical, circular disks obtained by Jenkins and Richman [17].

Here we assume that all variables are non-dimensionalized by the diameter of the disk  $d$ , the thermal velocity  $c_T \equiv \sqrt{2T_0/m}$ , where  $T_0$  and  $m$  are granular temperature and the mass of a particle, respectively. The angular velocity is assumed to be non-dimensionalized by  $\omega_T \equiv \sqrt{2T_0/I}$  with  $I = md^2/8$ . All dimensionless quantities are specified by hat. Thus, we introduce

$$\hat{x} = \frac{x}{d}, \quad \hat{\mathbf{c}} = \frac{\mathbf{c}}{c_T}, \quad \hat{\omega} = \frac{\omega}{\omega_T}, \tag{18}$$

which means that  $d\omega$  is reduced to  $2\sqrt{2}\hat{\omega}c_T$ .

Let us consider a binary collision in which the translational velocity  $\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2$  and the spin  $\hat{\omega}_1, \hat{\omega}_2$  prior to the collision become  $\hat{\mathbf{c}}'_1, \hat{\mathbf{c}}'_2$  and  $\hat{\omega}'_1, \hat{\omega}'_2$ . In order to characterize binary collisions of two identical disks we introduce two coefficients of restitution as

$$\hat{\mathbf{k}} \cdot \hat{\mathbf{v}}' = -e\hat{\mathbf{k}} \cdot \hat{\mathbf{v}} \quad \text{and} \quad \hat{\mathbf{k}} \times \hat{\mathbf{v}}' = -\beta_0(\hat{\mathbf{k}} \times \hat{\mathbf{v}}), \tag{19}$$

where  $\hat{\mathbf{k}}$  is the unit vector connecting the center of the particle 1 with that of 2, and  $\hat{\mathbf{v}}$  is the relative velocity of the points of contact defined by

$$\hat{\mathbf{v}} = \hat{\mathbf{g}} + \sqrt{2}\hat{\mathbf{z}} \times \hat{\mathbf{k}}(\omega_1 + \omega_2) \quad \text{with} \quad \hat{\mathbf{g}} = \hat{\mathbf{c}}_1 - \hat{\mathbf{c}}_2. \tag{20}$$

The coefficients of restitution are not constants in actual situation. In particular, the tangential restitution coefficient  $\beta_0$  strongly depends on the ratio of the tangential velocity to the normal velocity when the Coulomb slip takes place. However, they may be regarded as constants in a wide range approximately. The coefficient  $e$  may take values from 0 to 1, while  $\beta_0$  may vary between -1 and 1. The elastic hard core collision is characterized by  $e = 1$  and  $\beta_0 = -1$ , but the collisions of perfectly rough disks are characterized by  $e = \beta_0 = 1$ , where the energy is conserved in the binary collisions and there is still the time reversal symmetry.

In this paper, we are interested in the case  $e = \beta_0 = 1$ , because this situation allows the simplified description with time reversal symmetry but the spin can play nontrivial roles. In this situation, we have the changes in a binary collision of the relative velocity  $\hat{\mathbf{g}}$

$$\hat{\mathbf{g}}' - \hat{\mathbf{g}} = -\frac{2}{3}\hat{\mathbf{v}} - \frac{4}{3}\hat{k}(\hat{\mathbf{g}} \cdot \hat{k}), \tag{21}$$

the velocity

$$\hat{\mathbf{c}}'_1 - \hat{\mathbf{c}}_1 = -\frac{\hat{\mathbf{v}}}{3} - \frac{2}{3}\hat{k}(\hat{k} \cdot \hat{\mathbf{g}}), \quad \hat{\mathbf{c}}'_2 - \hat{\mathbf{c}}_2 = \frac{\hat{\mathbf{v}}}{3} + \frac{2}{3}\hat{k}(\hat{k} \cdot \hat{\mathbf{g}}), \tag{22}$$

and the spin

$$\hat{\omega}'_1 - \hat{\omega}_1 = -\frac{\sqrt{2}}{3}(\hat{k} \times \hat{\mathbf{v}})\hat{z}, \quad \hat{\omega}'_2 - \hat{\omega}_2 = -\frac{\sqrt{2}}{3}(\hat{k} \times \hat{\mathbf{v}})\hat{z}. \tag{23}$$

More general results for any  $e$  and  $\beta_0$  can be seen in the paper by Jenkins and Richman [17]. For example, let us present one result for any  $\beta_0$  as

$$\hat{\omega}_1^* + \hat{\omega}_2^* - \hat{\omega}_1 - \hat{\omega}_2 = -\frac{\sqrt{2}}{3\beta_0}(1 + \beta_0)\hat{z} \cdot (\hat{k} \times \hat{\mathbf{v}}), \tag{24}$$

where  $\hat{\omega}_i^*$  denotes the spin of  $i$  prior to a collision to become  $\hat{\omega}_i$  afterwards. Note that there is time reversal symmetry for any  $e$  and  $\beta_0$ .

## 4 Kinetic Theory of Dilute Gases

In this section, we discuss the kinetic theory of dilute rough disks. Since the Boltzmann equation should have modification if there is no time reversal symmetry, here we assume that the disks are perfectly rough, i.e.  $e = \beta_0 = 1$  to keep the time reversal symmetry. As will be shown, the result recovers usual Navier-Stokes equation in the dilute limit, but in some situation like in flows on an inclined rough plate the prediction of the dilute gas kinetics seems to be quantitatively useful [11]. The extension of this analysis for dense particles with dissipative collisions will be discussed elsewhere. Note that there are some treatments of three dimensional dissipation-less polyatomic fluids [12,13] based on Chapman-Enskog scheme. On the other hand, Jenkins and Richman [17] developed kinetic theory of rough inelastic disks in two dimensions based on the Grad expansion. It is difficult to check their argument because their calculation is long and complicated. In addition, their method cannot derive micropolar fluid mechanics as a closed form. Here we introduce a simpler method of calculation which can derive micropolar fluid mechanics and can determine the transport coefficients.

We begin with Boltzmann equation of particles with diameter  $d$ :

$$\partial_t f_1 + (\mathbf{c}_1 \cdot \nabla) f_1 = d \int d\mathbf{c}_2 \int_{-\infty}^{\infty} d\omega_2 \int d\hat{k} H(\hat{\mathbf{g}} \cdot \hat{k}) g \{f'_1 f'_2 - f_1 f_2\}. \tag{25}$$

Here we consider the collisions  $(\mathbf{c}^*, \omega^*) \rightarrow (\mathbf{c}, \omega) \rightarrow (\mathbf{c}', \omega')$ , but we note  $(\mathbf{c}^*, \omega^*) = (\mathbf{c}', \omega')$  because of time reversal symmetry for  $e = \beta_0 = 1$ . In (25), we use the following notations:  $\hat{\mathbf{k}}$  is the common unit normal vector at contact,  $\hat{\mathbf{g}} = \hat{\mathbf{c}}_1 - \hat{\mathbf{c}}_2$ ,  $\hat{g} = |\hat{\mathbf{g}}|$ ,  $H(x) = 1$  for  $x \geq 0$  and  $H(x) = 0$  for  $x < 0$ ,  $f'_1 \equiv f(\mathbf{r}, \mathbf{c}'_1, \omega'_1)$  and  $f_1 = f(\mathbf{r}, \mathbf{c}_1, \omega_1)$ .

Now let us analyze a linearly nonequilibrium situation. The distribution function can be expanded as

$$f(\mathbf{r}, \mathbf{c}, \omega, t) = n(\mathbf{r}, t) f_0(\mathbf{c}, \omega) (1 + \Psi(\mathbf{r}, \mathbf{c}, \omega)), \quad (26)$$

where  $n$  is the number density, and  $f_0$  is the Maxwell-Boltzmann distribution function which vanishes in the collisional integral in (25).

Now we adopt dimensionless quantities as in the previous section. Here we only introduce two variables as

$$\hat{n}(\hat{x}) = n(x) d^2, \quad f_0 = \frac{m\sqrt{I}}{(2T_0)^{3/2}} M(\xi), \quad (27)$$

where  $\xi = (\mathbf{c}, \hat{\omega})$  and  $M(\xi) = \exp(-\xi^2)/\pi^{3/2}$ .

It is easy to show that the zeroth order equation is given by  $d\hat{n}/d\hat{x} = 0$ . Thus, the distribution function  $f$  depends on the spatial coordinate only through  $\Psi$ .

The perturbative equation is now reduced to

$$\partial_t \Psi_1 + (\hat{\mathbf{c}}_1 \cdot \hat{\nabla}) \Psi_1 = \hat{n} \int d^3 \xi \int d\hat{\mathbf{k}} H(\hat{\mathbf{g}} \cdot \hat{\mathbf{k}}) \hat{g} M(\xi) \{ \Psi'_2 + \Psi'_1 - \Psi_2 - \Psi_1 \}. \quad (28)$$

This equation can be written as

$$\partial_t \Psi_1 + (\hat{\mathbf{c}}_1 \cdot \hat{\nabla}) \Psi_1 = \hat{n} L[\Psi_1], \quad (29)$$

where  $L[\Psi_1]$  is the collisional integral in (28). Introducing Fourier transform we may rewrite (29) as

$$\partial_t \Psi_q + iq\hat{c}_x \Psi_q = \hat{n} L[\Psi_q]. \quad (30)$$

The solution of (30) can be obtained from the non-Hermitian eigenvalue problem

$$(\hat{L} - iq\hat{c}_x) |\Psi_j^q\rangle = \lambda_j^q |\Psi_j^q\rangle. \quad (31)$$

Since we are interested in hydrodynamic behavior of (31), we adopt the expansion around  $q = 0$  as

$$\begin{aligned} |\Psi_j^q\rangle &= |\Psi_j^{(0)}\rangle + q |\Psi_j^{(1)}\rangle + q^2 |\Psi_j^{(2)}\rangle + \dots, \\ \lambda_j^q &= \lambda_j^{(0)} + q \lambda_j^{(1)} + q^2 \lambda_j^{(2)} + \dots \end{aligned} \quad (32)$$

Substituting (32) into (31) we obtain

$$\hat{n} L |\Psi_j^{(0)}\rangle = \lambda_j^{(0)} |\Psi_j^{(0)}\rangle. \quad (33)$$

Thus,  $|\Psi_j^{(0)}\rangle$  is represented by the linear combination of five fundamental eigenvectors as:

$$|\Psi_\alpha^{(0)}\rangle = \sum_{\alpha'=1}^5 c_{\alpha\alpha'} |\phi_{\alpha'}\rangle, \tag{34}$$

where  $|\phi_\alpha\rangle$  satisfies  $\hat{L}|\phi_\alpha\rangle = \lambda_\alpha^{(0)}|\phi_\alpha\rangle$ , and their explicit expressions are

$$\begin{aligned} |\phi_1\rangle &= 1, & |\phi_2\rangle &= \hat{c}_x, & |\phi_3\rangle &= \hat{c}_y, \\ |\phi_4\rangle &= \hat{\omega}, & |\phi_5\rangle &= \sqrt{\frac{2}{3}}(\xi^2 - \frac{2}{3}). \end{aligned} \tag{35}$$

Since  $|\phi_\alpha\rangle$  is the degenerated eigenvector, we need to use  $|\Psi_\alpha^{(0)}\rangle$ . The determination of  $|\Psi_j^{(0)}\rangle$  will be discussed later. We also assume that the eigenfunctions are orthonormal as

$$\langle \Psi_i^{(0)} | \Psi_j^{(0)} \rangle \equiv \int d\xi \Psi_j^{(0)} \Psi_i^{(0)} M(\xi) = \delta_{ij}. \tag{36}$$

Therefore we may introduce  $\bar{\lambda}_j^{(0)}$  as

$$\bar{\lambda}_j^{(0)} = \hat{n} \langle \Psi_j^{(0)} | L | \Psi_j^{(0)} \rangle. \tag{37}$$

This  $\bar{\lambda}_j^{(0)}$  is equivalent to  $\lambda_j^{(0)}$  if  $\lambda_j^{(0)}$  is independent of  $\xi$ . If  $\lambda_j^{(0)}$  is a function of  $\xi$ , two eigenvalues are different from each other. We believe that  $\bar{\lambda}_j^{(0)}$  plays fundamental roles in later discussion.

With the aid of (36) and (37) we obtain the relations at the first order:

$$|\Psi_j^{(1)}\rangle = \frac{iq\hat{c}_x + \lambda_j^{(1)}}{\hat{n}L - \lambda_j^{(0)}} |\Psi_j^{(0)}\rangle \tag{38}$$

and

$$\bar{\lambda}_j^{(1)} = i \langle \Psi_j^{(0)} | iq\hat{c}_x | \Psi_j^{(0)} \rangle, \tag{39}$$

where we use  $L = L^\dagger$  and  $\hat{n} \langle \Psi_4^{(0)} | L = \lambda_4^{(0)} \langle \Psi_4^{(0)} |$ .

From (31), (32) and (38) we obtain the second order correction of the eigenvalue as

$$\lambda_j^{(2)} = - \langle \Psi_j^{(0)} | (iq\hat{c}_x + \lambda_j^{(1)}) \frac{1}{\hat{n}L - \lambda_j^{(0)}} (iq\hat{c}_x + \lambda_j^{(1)}) | \Psi_j^{(0)} \rangle. \tag{40}$$

As the eigenvalues  $\lambda_\alpha^0 = 0$  are degenerate, we must be careful in starting from a proper basis that avoids the appearance of vanishing denominators. As in the case of quantum mechanics, we must solve exactly eigenvalue problem in the subspace spanned by  $|\phi_\alpha\rangle$ . From the comparison of (38) with (34) we obtain

$$\sum_{\alpha'=1}^5 c_{\alpha\alpha'} [\langle \phi_{\alpha'} | \hat{c}_x | \phi_\alpha \rangle - \bar{\lambda} \delta_{\alpha,\alpha'}] = 0, \tag{41}$$

where  $\tilde{\lambda} = i\lambda_\alpha^{(1)}/q$ . The required condition that nontrivial solutions of this equation exist is

$$\det[\langle \phi_\beta | c_x | \phi_\alpha \rangle - \tilde{\lambda} \delta_{\alpha\beta}] = 0. \tag{42}$$

Thus, we obtain

$$\tilde{\lambda}_1 = -\tilde{\lambda}_2 = \frac{1}{2}\sqrt{\frac{5}{3}}, \quad \tilde{\lambda}_3 = \tilde{\lambda}_4 = \tilde{\lambda}_5 = 0. \tag{43}$$

For later convenience, we introduce  $c_0 \equiv \tilde{\lambda}_1$ . Coming back to (41) we obtain

$$\begin{aligned} |\Psi_1^{(0)}\rangle &= \frac{1}{\sqrt{2}} \left[ \sqrt{\frac{3}{5}}|\phi_1\rangle + |\phi_2\rangle + \sqrt{\frac{2}{5}}|\phi_5\rangle \right], \\ |\Psi_2^{(0)}\rangle &= \frac{1}{\sqrt{2}} \left[ \sqrt{\frac{3}{5}}|\phi_1\rangle - |\phi_2\rangle + \sqrt{\frac{2}{5}}|\phi_5\rangle \right], \\ |\Psi_3^{(0)}\rangle &= |\phi_3\rangle, \\ |\Psi_4^{(0)}\rangle &= |\phi_4\rangle, \\ |\Psi_5^{(0)}\rangle &= \sqrt{\frac{2}{5}} \left[ -|\phi_1\rangle + \sqrt{\frac{3}{2}}|\phi_5\rangle \right]. \end{aligned} \tag{44}$$

These  $|\Psi_\alpha^{(0)}\rangle$  will be used as the basis of perturbative treatment.

Taking into account (40) we obtain the eigenvalues up to the second order as

$$\begin{aligned} \lambda_1 &= -ic_0q + q^2 \langle \Psi_1^{(0)} | (\hat{c}_x - c_0) \frac{1}{\hat{n}L} (\hat{c}_x - c_0) | \Psi_1^{(0)} \rangle, \\ \lambda_2 &= ic_0q + q^2 \langle \Psi_2^{(0)} | (\hat{c}_x + c_0) \frac{1}{\hat{n}L} (\hat{c}_x + c_0) | \Psi_2^{(0)} \rangle, \\ \lambda_3 &= q^2 \langle \Psi_3^{(0)} | \hat{c}_x \frac{1}{\hat{n}L} \hat{c}_x | \Psi_3^{(0)} \rangle, \\ \lambda_4 &= \lambda_4^{(0)} + q^2 \langle \Psi_4^{(0)} | \hat{c}_x \frac{1}{\hat{n}L} \hat{c}_x | \Psi_4^{(0)} \rangle, \\ \lambda_5 &= q^2 \langle \Psi_5^{(0)} | \hat{c}_x \frac{1}{\hat{n}L} \hat{c}_x | \Psi_5^{(0)} \rangle. \end{aligned} \tag{45}$$

Basically, the eigenvalues obtained in this section should be equivalent to those obtained in section 2 as the linear micropolar hydrodynamics. We have to note, however, that the results in this section is written in dimensionless forms but the results in section 2 include physical dimensions in terms of the thermal velocity  $c_T = \sqrt{2T_0/m}$  and the diameter of particles  $d$ . Thus we obtain the relations

$$\begin{aligned} c_s &= c_0 \sqrt{\frac{2T}{m}}, \\ \Gamma_s &= d \sqrt{\frac{2T_0}{m}} \langle \Psi_1^{(0)} | (\hat{c}_x - c_0) \frac{1}{\hat{n}L} (\hat{c}_x - c_0) | \Psi_1^{(0)} \rangle, \end{aligned}$$

$$\begin{aligned}
 \mu &= \rho d \sqrt{\frac{2T_0}{m}} \langle \Psi_3^{(0)} | \hat{c}_x \frac{1}{\hat{n}L} \hat{c}_x | \Psi_3^{(0)} \rangle, \\
 \mu_r &= -\frac{nI}{4d} \sqrt{\frac{2T_0}{m}} \lambda_4^{(0)}, \\
 \mu_B &= -\mu_r \frac{I}{m} - dnI \sqrt{\frac{2T_0}{m}} \langle \Psi_4^{(0)} | \hat{c}_x \frac{1}{\hat{n}L} \hat{c}_x | \Psi_4^{(0)} \rangle, \\
 \kappa &= -\rho d C_p \sqrt{\frac{2T}{m}} \langle \Psi_5^{(0)} | \hat{c}_x \frac{1}{\hat{n}L} \hat{c}_x | \Psi_5^{(0)} \rangle.
 \end{aligned} \tag{46}$$

## 5 Evaluation of Transport Coefficients

In this section, we evaluate the transport coefficients  $\mu$ ,  $\mu_r$  and  $\mu_B$  among many transport coefficients. Actually we can describe incompressible micropolar fluid mechanics in terms of these three coefficients.

### 5.1 Evaluation of $\mu_r$

As was shown in the previous section, to evaluate  $\mu_r$  we have to obtain  $\lambda_4^{(0)}$ . We here present the result for any  $\beta_0$ , since it is possible to obtain  $\lambda_4^{(0)}$  for any  $\beta_0$  in (19).

Substituting (24) into (28) we obtain

$$L[\hat{\omega}] = -\frac{\sqrt{2}(1 + \beta_0)}{3\beta_0} \{ \mathcal{L}_1[\hat{\omega}] + \mathcal{L}_2[\hat{\omega}] \} = \frac{\lambda_4^{(0)}}{\hat{n}} \hat{\omega}, \tag{47}$$

where

$$\mathcal{L}_1[\hat{\omega}] = -\int d^2 \hat{c}_2 \int d\hat{k} H(\hat{\mathbf{g}} \cdot \hat{\mathbf{k}}) (\hat{\mathbf{g}} \cdot \mathbf{k}) M_2(\hat{\mathbf{c}}) \hat{z} \cdot (\hat{\mathbf{k}} \times \hat{\mathbf{g}}) \tag{48}$$

and

$$\mathcal{L}_2[\hat{\omega}] = \sqrt{2} \hat{\omega} \int d^2 \hat{c}_2 \int d\hat{k} H(\hat{\mathbf{g}} \cdot \hat{\mathbf{k}}) (\hat{\mathbf{g}} \cdot \hat{\mathbf{k}}) M_2(\hat{\mathbf{c}}) \tag{49}$$

with  $M_2(\hat{\mathbf{c}}) = \exp[-\hat{c}_x^2 - \hat{c}_y^2]/\pi$ . With the aid of formulae for the modified Bessel function  $I_0(z)$  with zeroth order and the confluent Hypergeometric function  $F(a, b; z)$ :

$$I_0(z) = \frac{1}{\pi} \int_0^\pi dx e^{z \cos x}, \quad F(3/2, 1; c^2) = \frac{4}{\sqrt{\pi}} \int_0^\infty dr r^2 I_0(2rc) \tag{50}$$

it is possible to show

$$\mathcal{L}_2[\hat{\omega}] = \frac{\sqrt{2\pi}}{2} F(3/2, 1; \hat{c}^2) e^{-\hat{c}^2} \hat{\omega}, \quad \text{and} \quad \mathcal{L}_1[\hat{\omega}] = 0. \tag{51}$$

Therefore, we obtain the eigenvalue

$$\lambda_4^{(0)} = -\frac{\sqrt{\pi}(1 + \beta_0)\hat{n}}{3\beta_0} e^{-\hat{c}^2} F(3/2, 1; \hat{c}^2). \tag{52}$$

This result states that the eigenvalue depends on the velocity of particles.

As mentioned in section 4, if the eigenvalue depends on the velocity, the effective eigenvalue  $\bar{\lambda}_\alpha$  defined in (37) plays important roles. Thus,  $\bar{\lambda}_4^{(0)}$  is evaluated as

$$\bar{\lambda}_4^{(0)} = -\frac{\sqrt{\pi}(1 + \beta_0)\hat{n}}{3\beta_0} \int d^3\xi \frac{e^{-\xi^2}}{\pi^{3/2}} e^{-\hat{c}^2} F(3/2, 1; \hat{c}^2)\hat{\omega}^2. \tag{53}$$

The result of integration becomes

$$\bar{\lambda}_4^{(0)} = -\frac{\sqrt{2\pi}(1 + \beta_0)}{6\beta_0}\hat{n}, \tag{54}$$

where we use  $\int_0^\infty dcce^{-2c^2} F(3/2, 1; c^2) = 1/\sqrt{2}$ . We note that the result (54) is reduced to

$$\bar{\lambda}_4^{(0)} = -\frac{\sqrt{2\pi}}{3}\hat{n} \tag{55}$$

in the case of  $\beta_0 = 1$ .

From (45), (46) and (55) we obtain

$$\mu_r = \frac{\sqrt{2\pi}n^2}{96}md^3\sqrt{\frac{2T_0}{m}} \tag{56}$$

for  $e = \beta_0 = 1$ . Thus,  $\mu_r$  is proportional to  $n^2$  as predicted by Mitarai et al. [11]. Therefore, the effect of microrotation is negligible in usual situations of dilute gases. However, as demonstrated by Mitarai *et al.* [11] the shear near the boundary produces the relevant situation for microrotation.

### 5.2 Evaluation of $\mu$

$\mu$  can be evaluated from the third equation of (46). This equation may be rewritten as

$$\hat{\mu} = -\int d\xi \hat{c}_x \hat{c}_y \chi^{xy} M(\xi) = -\langle \chi^{xy} | Y^{xy} \rangle, \tag{57}$$

where  $|\chi^{xy}\rangle = \chi^{xy}$  and  $|Y^{xy}\rangle = \hat{c}_x \hat{c}_y$ .  $\hat{\mu}$  is defined by

$$\hat{\mu} = \mu / (\rho d \sqrt{\frac{2T_0}{m}}) \tag{58}$$

and  $\chi^{xy}$  is the solution of

$$\hat{n}L[\chi^{xy}] = \hat{c}_x \hat{c}_y \quad \text{or} \quad \hat{n}L|\chi^{xy}\rangle = |Y^{xy}\rangle. \tag{59}$$

Introducing the function  $|f^{xy}\rangle$  vertical to  $|\phi_\alpha\rangle$ , it is easy to show that  $\langle \bar{f} | \hat{L} | \bar{f} \rangle$  is minimum for  $|\bar{f}\rangle = |\chi^{xy}\rangle$  where  $|\bar{f}\rangle = \langle f^{xy} | Y^{xy} \rangle / \langle f^{xy} | \hat{n}L | f^{xy} \rangle$ . The minimum principle then becomes

$$\langle \chi | Y^{xy} \rangle \leq \langle \bar{f} | \hat{n}L | \bar{f} \rangle = \frac{|\langle f^{xy} | Y^{xy} \rangle|^2}{\langle f^{xy} | \hat{n}L | f^{xy} \rangle}. \tag{60}$$

Here we assume the following expansion:

$$|f^{xy}\rangle = \sum_{j=1}^m \beta_j^{(m)} |f_j\rangle, \tag{61}$$

where the coefficient  $\beta_j^{(m)}$  is determined by the minimum condition of the right hand side of (60). These conditions are summarized as

$$Y_j^{xy} = \sum_{l=1}^m \beta_l^{(m)} b_{jl}^{xy}, \quad \langle \chi^{xy} | Y^{xy} \rangle = -\frac{1}{\hat{n}} \sum_{j=1}^m \beta_j^{(m)} Y_j^{xy}, \tag{62}$$

where  $Y_k^{xy} = \langle f_k | Y^{xy} \rangle$  and  $b_{kl}^{xy} = -\langle f_k | L | f_l \rangle$ .

We assume that  $|f_j\rangle$  can be represented by

$$|f_j\rangle = S_2^{(j-1)}(\hat{c}^2) \hat{c}_x \hat{c}_y, \tag{63}$$

where  $S_2^{(j-1)}(x)$  is the Sonine polynomial which is defined by

$$S_l^{(r)}(x) = \sum_{j=0}^r \frac{(-1)^j \Gamma(l+r+1)}{\Gamma(l+j+1)(r-j)!} x^j \tag{64}$$

with the Gamma function  $\Gamma(x)$ . The Sonine polynomials satisfy the orthonormality condition  $\int_0^\infty dx x^l e^{-x} S_l^{(r)}(x) S_l^{(r')}(x) = \frac{\Gamma(r+l+1)}{r!} \delta_{r,r'}$ . In particular, the relations  $S_l^{(0)}(x) = 1$  is useful for later discussion. From the orthonormality the condition of  $\langle f^{xy} | \phi_\alpha \rangle = \sum_j^m \beta_j^{(m)} \langle f_j | \phi_\alpha \rangle = 0$  is automatically satisfied.

Therefore we obtain the expressions for  $Y_j^{xy}$  and  $b_{kl}^{xy}$  as

$$\begin{aligned} Y_j^{xy} &= \int d\xi \hat{c}_x \hat{c}_y S_2^{(j-1)}(\hat{c}^2) \hat{c}_x \hat{c}_y M(\xi), \\ b_{kl}^{xy} &= - \int d\xi M(\xi) S_2^{(k-1)}(\hat{c}^2) \hat{c}_x \hat{c}_y L \hat{c}_x \hat{c}_y S_2^{(l-1)}(\hat{c}^2). \end{aligned} \tag{65}$$

It is easy to calculate  $Y_j$  as

$$Y_j^{xy} = \frac{1}{4} \delta_{j,1}. \tag{66}$$

It is notable that  $Y_k^{xy}$  is zero except for  $Y_1^{xy}$ .

Since it is known that the expansion in terms of Sonine's polynomial is fast, the result with  $m = 1$  gives a good approximation. Adopting this approximation (62) becomes

$$\langle \chi^{xy} | Y^{xy} \rangle = -\frac{\beta_1^{(1)}}{\hat{n}} Y_1^{xy}; \quad Y_1^{xy} = \beta_1^{(1)} b_{11}^{xy}. \tag{67}$$

Eliminating  $\beta_1^{(1)}$  from (67) we obtain

$$\langle \chi^{xy} | Y^{xy} \rangle = -\frac{Y_1^{xy2}}{\hat{n} b_{11}^{xy}}. \tag{68}$$

From the definition of the operator  $L$   $b_{11}^{xy}$  can be written as

$$b_{11}^{xy} = \frac{1}{4\pi^3} \int d\xi_1 \int d\xi_2 \int_{-\pi/2}^{\pi/2} d\psi \cos \psi \hat{g} e^{-\xi_1^2 - \xi_2^2} \times [\hat{c}'_{1x} \hat{c}'_{1y} + \hat{c}'_{2x} \hat{c}'_{2y} - \hat{c}_{1x} \hat{c}_{1y} - \hat{c}_{2x} \hat{c}_{2y}]^2. \tag{69}$$

where  $\pi - \psi$  is the angle between  $\hat{g}$  and  $\hat{k}$ . To derive (69) we have used that the cross section ( $da/d\psi$  with the impact parameter  $a$ ) is  $\cos \psi$  and the time reversal symmetry of collisions.

From (21) and (69) we get the relation

$$b_{11}^{xy} = \frac{7}{18} \sqrt{\frac{\pi}{2}}. \tag{70}$$

From (57), (61) and (68) we obtain

$$\hat{\mu} = \frac{9}{56\hat{n}} \sqrt{\frac{2}{\pi}}. \tag{71}$$

From comparison of (46) with the definition of  $\hat{\mu}$ , the final expression of  $\mu$  is given by

$$\mu = \frac{9}{28d} \sqrt{\frac{mT_0}{\pi}}. \tag{72}$$

Note that this result is deviated from the result in ref. [17].

### 5.3 Evaluation of $\mu_B$

The method of calculation of  $\mu_B$  is similar to that of  $\mu$ . Let us introduce

$$\mu_c = - \int d\xi \hat{\omega} \hat{c}_x \chi^{\omega x} M(\xi) = - \langle \chi^{\omega x} | Y^{\omega x} \rangle, \tag{73}$$

where  $|Y^{\omega x}\rangle = \hat{\omega} \hat{c}_x$  and  $|\chi^{\omega x}\rangle = \chi^{\omega x}$  is the solution of

$$\hat{n}L|\chi^{\omega x}\rangle = |Y^{\omega x}\rangle. \tag{74}$$

Similar to the previous section  $\hat{\mu}_c$  is represented by

$$\mu_c = \frac{Y_1^{\omega x 2}}{\hat{n}b_{11}^{\omega x}} \tag{75}$$

in the lowest order approximation. Here  $Y_1^{\omega x}$  and  $b_{11}^{\omega x}$  are respectively given by

$$Y_1^{\omega x} = \int d\xi \hat{\omega} (\hat{c}_x)^2 M(\xi), \quad b_{11}^{\omega x} = - \int d\xi M(\xi) \hat{\omega} \hat{c}_x L \hat{\omega} \hat{c}_x. \tag{76}$$

It is easy to show

$$Y_1^{\omega x} = \frac{1}{4}. \tag{77}$$

On the other hand,  $b_{11}^{\omega x}$  is rewritten as

$$b_{11}^{\omega x} = \frac{1}{4\pi^3} \int d\xi_1 \int d\xi_2 \int_{-\pi/2}^{\pi/2} d\psi \cos \psi \hat{g} e^{-\xi_1^2 - \xi_2^2} \times [\hat{c}'_{1x} \hat{\omega}'_1 + \hat{c}'_{2x} \hat{\omega}'_2 - \hat{c}_{1x} \hat{\omega}_1 - \hat{c}_{2x} \hat{\omega}_2]^2. \quad (78)$$

From (22) and (23) we obtain

$$b_{11}^{\omega x} = \frac{11}{12} \sqrt{\frac{\pi}{2}}. \quad (79)$$

Thus  $\mu_c$  becomes

$$\mu_c = \frac{3}{44} \sqrt{\frac{2}{\pi}}. \quad (80)$$

From (46) and (73) we obtain

$$\mu_B = \frac{3}{176\sqrt{\pi}} d\sqrt{mT_0} - \frac{\sqrt{2\pi}(nd^2)^2}{768} d\sqrt{2mT_0}. \quad (81)$$

It is obvious that the second term is negligible in the dilute limit.

The method of derivation of  $\mu_B$  is relatively simple, when we compare the method based on Chapman-Enskog scheme in which  $\mu_B$  becomes the correction term of higher order [12].

## 6 Discussion

Here, we have demonstrated how micropolar fluid mechanics can be derived from the Boltzmann equation. Of course, the result is reduced to Navier-Stokes equation in the dilute limit. In this sense, at least, we will have to extend our work to the case of Enskog equation if we believe that the concept of micropolar fluid mechanics is useful. The effect of dissipation also plays important roles though we do not discuss it.

Roughly speaking the effect of microrotation is localized in the boundary layer. Therefore it will be important to analyze simple shear flows. Our preliminary result suggests that the micropolar fluid mechanics may not be enough to discuss the region close to a flat boundary. The success by Mitarai *et al.* [11] may come from their bumpy boundary condition in which there is no Knudsen's layer in the system due to random scattering of particles near the boundary.

There are many problems to be solved. Let us close the paper with our perspective whether micropolar fluid mechanics is useful. In the pessimistic view the analysis presented here is general nonsense, and micropolar fluid mechanics is useless. On the other hand, in the optimistic view, this work will be a milestone to discuss the fluid motion with microstructure. Although we are not sure which result we will see in the future, we hope that micropolar fluid mechanics is a useful concept to characterize the flow of particles.

The author thanks N. Mitarai and H. Nakanishi for fruitful discussion.

## References

1. E. and F. Cosserat, *Théorie des Corps Déformables* (A. Hermann, Paris, 1909).
2. D.W. Condiff and J.S. Dahler, *Physics of Fluids* **7**, 842 (1964).
3. A.C. Eringen, *J. Math. Mech.* **16**, 1 (1966).
4. G. Lukaszewicz, *Micropolar Fluids: Theory and Applications* (Birkhäuser, Boston, 1999).
5. H.M. Jaeger, S.R. Nagel, and R.B. Behringer, *Rev. Mod. Phys.* **68**, 1259 (1996).
6. H.J. Herrmann, J-P. Hapi, and S. Luding eds. *Physics of Dry Granular Media* (Kluwer Academic 1998).
7. Y. Kishino eds. *Powders and Grains 2001* (A.A. Balkema Pub., Rotterdam, 2001).
8. K. Kanatani, *Trans. Jpn. Soc. Mech. Eng. B* **45**, 507, 515 (1979).
9. J. Kano, A. Shimosaka, and J. Hidaka, *J. Soc. Powder Technol. Jpn.* **33**, 95 (1996).
10. H. Hayakawa, *Phys. Rev. E* **61**, 5477 (2000).
11. N. Mitarai, H. Hayakawa, and H. Nakanishi, to be published in *Phys. Rev. Lett.* (cond-mat/0108192). See also their paper in this proceedings.
12. B.J. McCoy, S.I. Sandler, and J.S. Dahler, *J. Chem. Phys.* **45**, 3485 (1966).
13. J.S. Dahler and M. Theodosopulu, *Adv. Chem. Phys.* **31**, 155 (1975).
14. C.K.K. Lun, *J. Fluid Mech.* **233**, 539 (1991).
15. P. Resibois and M. de Leener, *Classical Kinetic Theory of Fluids* (John Wiley & Sons, New York, 1977).
16. S. Luding, M. Huthmann, S. McNamara, and A. Zippelius, *Phys. Rev. E* **58**, 3416 (1998).
17. J.T. Jenkins and M.W. Richman, *Phys. Fluids*, **28**, 3485 (1985).

# Dynamics and Structure of Granular Flow Through a Vertical Pipe

O. Moriyama, N. Kuroiwa, T. Isoda, T. Arai, S. Tateda, Y. Yamazaki, and M. Matsushita

Department of Physics, Chuo University, Kasuga 1-13-27, Bunkyo-ku, Tokyo 112-8551, Japan

**Abstract.** We have experimentally shown how density waves of granular particles (ordinary sand) emerge, while they flow through a vertical glass pipe, by controlling air flow out of a flask attached to the bottom-end of the pipe. When a cock attached to the flask is fully open, air is dragged by falling granules and flows together with them. No density waves are observed for this situation. As the cock is gradually closed, however, the pressure gradient of air inside the pipe becomes gradually larger, inducing a velocity difference between granules and air. As a result, density waves emerge from the lower part of the pipe. The more the cock is closed, i.e., the smaller the rate of air flow, the higher the onset point (along the pipe) of the density waves. The onset of density waves is characterized by the growth of the lower frequency part of the power spectra of the time-series signals of the density waves. The power spectra of the density waves display a clear power-law form  $P(f) \sim f^{-\alpha}$  with the value of the exponent  $\alpha = 1.33 \pm 0.06$ , which is very close to  $4/3$ . The value of  $\alpha$  is robust even under the medium flow or variation of the pipe diameter, as far as density waves can be seen.

## 1 Introduction

Granular materials are ubiquitously seen in nature, industry and everyday life. To name but a few, sand, pebbles and stones on a sea or lake shore and a river bank, scree in a mountain, rubble and macadam at a quarry, snow flakes lying anywhere on a cold winter day, grain at a grain elevator, and salt, sugar and pepper on a dining table. Despite the importance of such materials to industry and everyday life, much about their peculiar behavior has only recently begun to be paid attention to by scientists. Static, dynamic and statistical properties of granular materials are in fact one of the most important topics in current science and its application to technology [1–4]. In some situations granular materials behave like ordinary solids or ordinary liquids. On the other hand, many unusual motions peculiar to granular materials have been found, such as size segregation [5], bubbling [6], standing waves and localized excitations under vertical vibrations [7,8], avalanches and other unusual motions in a rotating mill [9–11], chute flows down a slope [12], and a fluidized bed due to air injected inside a box containing granules [13–17].

Emergence of density waves of granules flowing through a vertical pipe is also a typical example of unusual features of granular motion [18–27]. The first

investigation on this type of granular flow was performed by numerical simulations [18–20]. They were followed by several experimental measurements, which were carried out by using sand as granules in air as medium [21,22,24–26], glass beads in air [23] and lead spheres in water and silicon oil [22,27].

Molecular dynamics (MD) [18] and lattice-gas automata (LGA) [19,20] methods were taken to reveal how density waves emerge.

One of their results was that the most essential factor for the onset of density waves was the dissipation due to both inelastic collisions between granules and the friction from a rough wall. It was also shown that the power spectrum of the density fluctuations have a clear power-law form  $P(f) \sim f^{-\alpha}$  with  $\alpha \simeq 4/3$  [19,20].

Horikawa *et al.* performed a first quantitative experiment [21,22]. They found an unambiguous situation where density waves emerge: Density waves of granular materials occurred when the hole at the bottom-end of a pipe was half closed, while they never occurred when the hole was completely open. Their results suggest that the most essential factor for the onset of density waves is the interaction between granules and medium (air). The importance of the interaction between granules and medium can be easily confirmed in the experiment using lead spheres as granules and water or silicon oil as the medium, where the back-flow of the medium can be easily realized [22,27].

In this article, by improving the experimental conditions, we show various kinds of flow patterns of granules (sand), the corresponding power spectra and the variation of flow behavior as we systematically vary both the measuring position along the pipe and the flow rate of air out of the pipe. We also estimate the precise value of the scaling exponent  $\alpha$  by taking a large number of independent trials.

## 2 Experimental Procedures

A hopper with open angle of  $60^\circ$  was attached to the top of a vertical glass pipe of 1500 mm length and 3 mm inner diameter. The bottom-end of the pipe was thrust into a flask whose outlet was followed by a flow-meter, as shown in Fig. 1, which can control the rate of air flow out of the pipe. The granular particles we used were rough sand whose individual diameter was about 0.5 mm. This means that the diameter ratio of the pipe to the granules is about 6 and the granular material is polydisperse. The reason why we did not use monodisperse particles such as glass beads is to avoid local and temporary crystallization which is clearly exhibited in two-dimensional flow experiments. Although we do not know how serious the crystallization is in three dimensions, we decided to keep this additional and irrelevant factor away from our present experiments and to use polydisperse rough sand for the sake of self-averaging. We performed experiments in a double-layered vinyl house with humidity fixed around  $60 \sim 70\%$  in order to avoid the static electricity which frequently makes granular particles stop in the pipe when humidity is low.

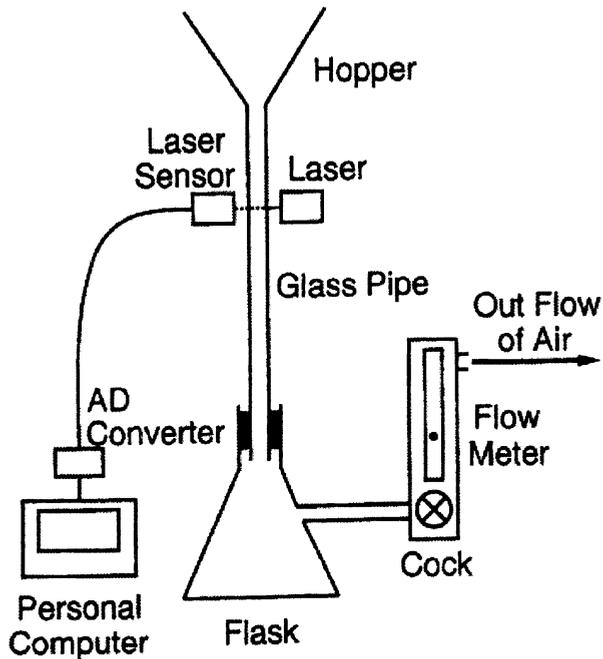


Fig. 1. Schematic illustration of the experimental setup.

We also treated the inner wall of the glass pipe to avoid the accumulation of static electricity.

We poured the rough sand into the hopper and let it drain from the hopper and flow through the pipe due to gravity. The rough sand finally falls into the flask, while air can exit through the flow-meter. It should be stressed that in our experiment the flowing patterns of granular particles were never affected by the characteristics of the hopper, since in any cases the flow once became uniform just below the hopper and then its behavior changed according to the flow rate of air. Corresponding data will be shown in the next section. Meanwhile we measured the density fluctuations as the transmission light intensity across the pipe by using laser light and detecting system (KEYENCE, LX-02). The laser light we used has a rectangular cross-section 10 mm wide and 1 mm high and is emitted with a pulse frequency of 4096 Hz. We varied both the measuring position along the pipe and the rate of air flow out of the pipe in order to investigate the qualitative variation of granular flow behavior. Output time-series signals were fed to an AD converter. The data from the AD converter were input to a personal computer in order to carry out fast Fourier transform (FFT) analysis and study the power spectrum of density waves of granules flowing through a pipe.

### 3 Experimental Results

#### 3.1 Uniform Flow and Density Waves

In order to investigate the formation of density waves, let us first define the measuring position  $x$  along the glass pipe as the distance from the hopper aperture down to the position at which we measured transmission light intensity across the pipe. Let us also define the rate  $V$  as the volume of air discharge out of the pipe per one minute, which can be controlled by the flow-meter. There are two extreme cases. In one case the flow-meter was removed from the apparatus: The bottom hole of the pipe was completely open. Hereafter, we refer this situation to *fully open*. The opposite extreme is *fully closed*, where the cock of the flow-meter was closed, i.e., did not allow any outflow of air. In addition to *fully open* and *fully closed* ( $V = 0$  ml/min), we investigated three intermediate flow rates;  $V \simeq 200, 400$  and  $600$  ml/min. In each case we traced the behavior of the sand flow at various heights from the top of the pipe ( $x = 1$  cm) to  $x = 120$  cm with intervals of 20 cm. (In *fully closed* in particular, we took a 10 cm interval.)

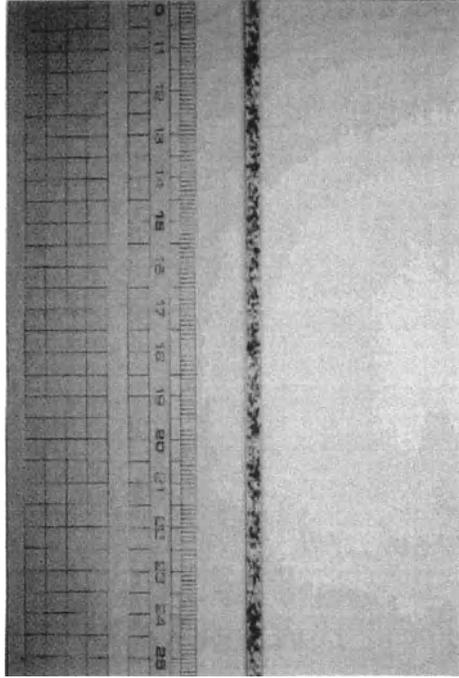
In *fully open* we did not see any density waves by the naked eye.

Granules flow rather freely and uniformly through the pipe like an ordinary fluid. The density of granules looks almost uniform, as seen in Fig. 2. On the other hand, we observed clear density waves in *fully closed*. A snapshot of the typical density waves is shown in Fig. 3. In contrast to the uniform flow in Fig. 2, one can find an intermittent structure of granular density in Fig. 3. In real flow patterns one can observe that dense clusters sometimes collide and merge into one, and sometimes one cluster splits into more than one, a behavior reminiscent of a chain of traffic jams on a crowded highway [28–30].

#### 3.2 Flowing Patterns, Organization of Density Waves and Power Spectra under Controlled Medium Flow

Before examining various kinds of flowing patterns, let us consider what happens by closing the cock of the flow-meter. Granular particles interact with the medium (air) due to viscosity. In *fully open*, however, granules can fall rather freely as if viscosity could be neglected, because both granules and air flow together through the pipe. As the cock is gradually closed, the air pressure in the flask rises and the effect of the viscous force becomes more severe. Let us consider the situation where the rate of air discharge from the flow-meter is zero. In such a case air goes a little upward in the pipe due to the conservation of the total volume (sand plus air) in the flask and strongly interacts with the granular particles. Thus the existence of air can never be neglected when the rate of air discharge is small.

*Fully open:* In the case of *fully open* time-series signals of flowing patterns at various measuring positions  $x$  are shown in Fig. 4. In this figure the higher (lower) voltage corresponds to the smaller (larger) granular density since we measured the laser light intensity transmitted across the pipe. The time-series signals have



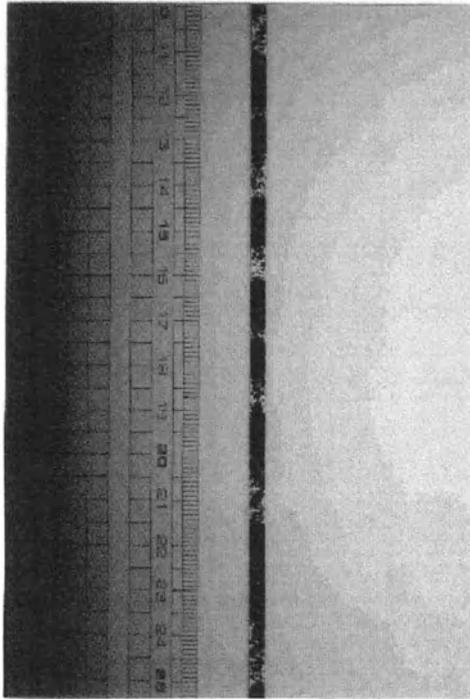
**Fig. 2.** Snapshot of a granular flow for *fully open*. Shutter speed is 1/4000 sec.

white-noise-like structure reflecting almost uniform density and random motion of granular particles. In fact, the power spectra exhibit no power-law form, as seen in Fig. 5.

Each spectrum was obtained from 8192 discrete points (two seconds in real time) and by averaging over 640 independent trials. The flow just below the hopper ( $x = 1$  cm) is slightly different from that of lower positions (below  $x = 20$  cm), although the flow looked everywhere uniform. The difference is the density of granular particles: Just below the hopper ( $x = 1$  cm), granular particles are dense and frequently collide with each other. As granular particles fall and are accelerated, the mean free path gets larger and the number of collisions decreases ( $x \geq 20$  cm). In spite of the existence of such difference, however, all the power spectra are more or less flat.

**Intermediate situations between *fully open* and *fully closed*:** Let us discuss the three intermediate situations ( $V \simeq 600, 400$  and  $200$  ml/min) between *fully open* and *fully closed*. The value of  $V$  is negligible for *fully closed* and is more than  $800$  ml/min for *fully open* since  $V$  is about  $800$  ml/min when the flow meter is attached and the cock is completely open.

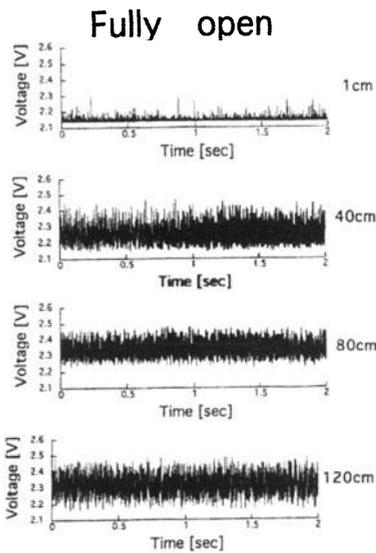
Figures 6(a), 6(b) and 6(c) show the power spectra in the cases of  $V \simeq 600, 400$  and  $200$  ml/min, respectively. Each spectrum was obtained by averaging



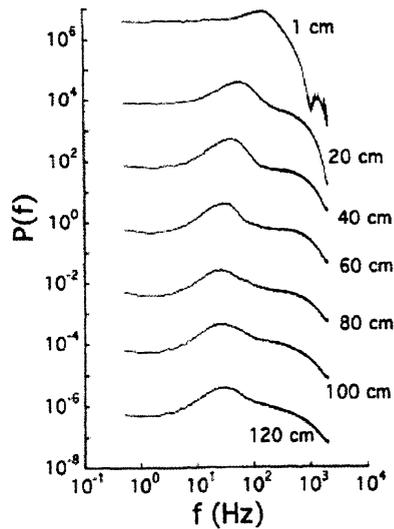
**Fig. 3.** Snapshot of density waves in *fully closed*. Shutter speed is 1/4000 sec. The measuring position of this snapshot was  $x \simeq 50$  cm, where  $x$  is the distance from the hopper. (The value of the scale in the photo is irrelevant to the actual measuring position.)

over 640 independent trials with 8192 discrete points each (two seconds in real time). All of the spectra with  $V \simeq 600$  ml/min (Fig. 6(a)) are almost the same as for *fully open* (Fig. 5) except for  $x = 120$  cm. The spectrum at  $x = 120$  cm in Fig. 6(a) is similar to the one at  $x = 80$  or 100 cm in Fig. 6(b) ( $V \simeq 400$  ml/min). Furthermore, the spectrum at  $x = 60$  cm in Fig. 6(b) is similar to that at  $x = 40$  cm in Fig. 6(c) ( $V \simeq 200$  ml/min), which is very similar to that at  $x = 20$  or 30 cm in Fig. 10 (*fully closed*;  $V \simeq 0$  ml/min). Thus the variation of the power spectra in Fig. 6 suggests that the point where the uniform flow becomes unstable rises upwards as the discharge of air out of the pipe decreases.

Figures 7 and 8 show, respectively, the time-series signals and corresponding power spectra of typical granular flows at fixed  $x$  ( $= 120$  cm) for these three cases of the flow rate, in addition to *fully open* and *fully closed*. As the value of  $V$  decreases, the number of density waves seen in a given time interval increases. The air discharge reduces the air pressure in the flask, since the pressure in the flask must increase when the volume of injected granules is larger than that of



**Fig. 4.** Time-series signals measured at various positions  $x$  for fully open. Each signal is shifted by one volt in order to avoid the data overlap. The higher (lower) voltage corresponds to the lower (higher) density of granules.



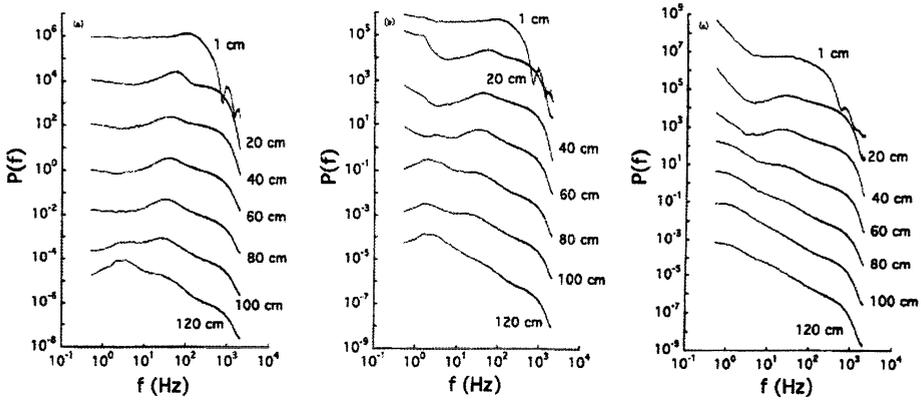
**Fig. 5.** Log-log plot of power spectra of *fully open*. Each line is appropriately shifted in order to avoid the data overlap. The spectrum at each  $x$  was obtained by averaging over 640 independent trials with the length of 8192 discrete points each (2 seconds in real time.)

air discharged from the bottom-end of the pipe. The smaller the value of  $V$ , the larger the pressure gradient in the pipe.

As a result, the emergence point of density waves rises upwards, which was exactly what we observed (Fig. 6). Thus we conclude that the most important factor for the emergence of density waves is the interaction between granules and the medium (air), as mentioned at the beginning of subsection 3.2, and there exists some threshold of pressure gradient for the emergence of density waves.

Density waves were clearly observed at  $x = 120$  cm with  $V \simeq 400$  ml/min and  $x = 80 \sim 120$  cm with  $V \simeq 200$  ml/min. They were sometimes observed at  $x = 120$  cm even with  $V \simeq 600$  ml/min. The slope of the power spectra of these density waves was found to be about  $-1.3$ , consistent with the scaling exponent  $\alpha = 4/3$  of the power-law  $P(f) \sim f^{-\alpha}$ .

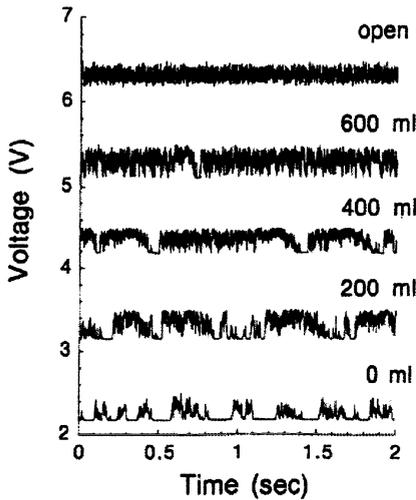
*Fully closed:* In *fully closed*, one can observe the formation process of density waves. Figure 9 shows typical time-series signals measured at various positions  $x$  for *fully closed*. Each signal has the length of 32 seconds and is shifted by one volt in order to avoid the data overlap. The signal at  $x = 10$  cm looks the



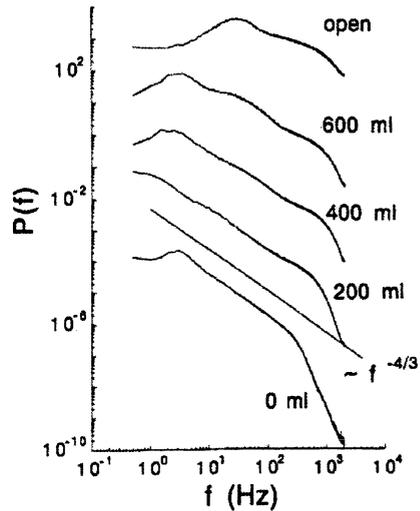
**Fig. 6.** Log-log plot of power spectra of the three intermediate situations between *fully open* and *fully closed*. Each line is appropriately shifted in order to avoid the data overlap. The spectrum at each  $x$  was obtained by averaging over 640 independent trials with the length of 8192 discrete points each (2 seconds in real time).  $V \simeq$  (a) 600 ml/min, (b) 400 ml/min, and (c) 200 ml/min.

random noise and is very similar to that of *fully open*. In fact, the signal at  $x = 1$  cm is also very similar to that of *fully open* shown in Fig. 4. As the values of  $x$  increases, density waves emerge and the number of dense clusters (regions with low voltage) also increases. For  $x \geq 50$  cm the time-series signals do not differ so much from the one for *fully closed* shown in Fig. 7 ( $x = 120$  cm) where an intermittent structure is observed. It should also be noted that the time-series signal at  $x = 40$  cm still consists of both dense clusters (10 ~ 20 sec in Fig. 9) and relatively long (several seconds, for example, 28 ~ 32 sec in Fig. 9) periods of uniform flow. Thus it is considered that Fig. 9 clearly shows the self-organized critical (SOC) behavior for the occurrence of density waves, and the transition point from the uniform flow to density waves is located near  $x \simeq 40$  cm.

Figure 10 shows the power spectra  $P(f)$  of time-series signals on different measuring positions  $x$  for *fully closed*. The spectrum at each  $x$  was obtained by averaging over 640 independent trials with 8192 discrete points each (two seconds in real time). At  $x = 10$  cm, the spectrum is very close to that at the same position for *fully open*. This means that only white-noise-like signals can be detected at  $x = 10$  cm even for *fully closed*. As the value of  $x$  increases, the power spectrum begins to show a power-law form. Especially, the difference between the power spectrum at  $x = 30$  cm and the one at  $x = 40$  cm is drastic: The power spectrum at  $x = 30$  cm is not so different from that at  $x = 10$  cm where only uniform flows were observed. On the other hand, the scaling region of the power-law form is extended to a wide frequency range at  $x = 40$  cm. The difference is also observed in Fig. 9: Uniform flow is dominant at  $x = 30$  cm, while it is exceptional at  $x = 40$  cm. Again, it is expected from the power spectra that the present granular flow clearly exhibits SOC-like behavior, and the onset



**Fig. 7.** Typical time series signals of granular flows with different rate  $V$ . Measuring position was at  $x = 120$  cm. Each signal is shifted by one volt in order to avoid the data overlap.



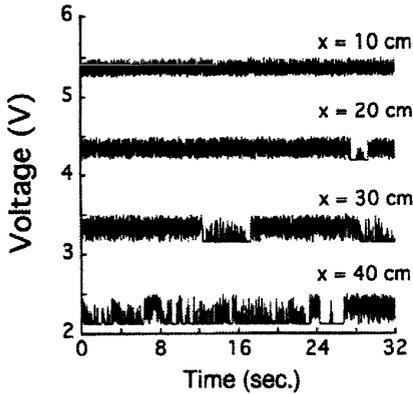
**Fig. 8.** Power spectra  $P(f)$  for the granular flows in Fig. 7. Each spectrum is appropriately shifted in order to avoid the data overlap. The straight line with a slope of  $-4/3$  is an eye-guide.

point of density waves is located between  $x = 30$  and  $40$  cm. The emergence of density waves is characterized by the growth of lower frequency components in the power spectrum.

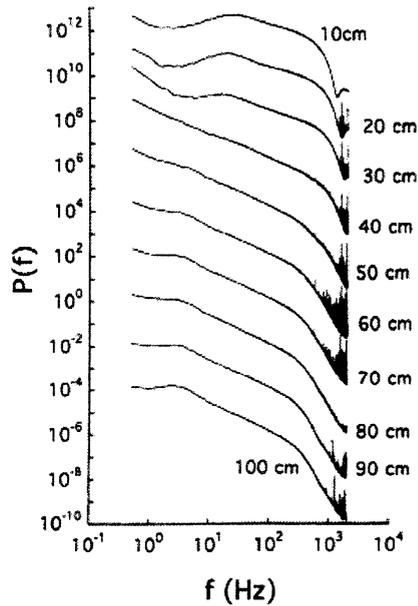
Let us next discuss the scaling exponent  $\alpha$  of the power-law  $P(f) \sim f^{-\alpha}$ . Figure 11 shows the values of the exponent  $\alpha$  as a function of the measuring position  $x$ , which seems to rapidly converge to a constant value as  $x$  increases. Averaged value of the exponent  $\alpha$  over  $x = 30 \sim 130$  cm is  $1.33 \pm 0.06$ , which is again very close to  $4/3$ . We conclude here that the value is consistent with  $\alpha = 4/3$ , which is in accord with the numerical simulations [19,20] and the theoretical prediction [24].

#### 4 Summary and Discussions

In this article we have shown how density waves of granular particles emerge by controlling air flow out of the bottom-end of a pipe: When the cock is not attached (*fully open*), i.e., the rate of the air flow is large, both granules and air flow together through the pipe and no density waves are observed. As the cock is gradually closed, the air pressure in the flask and the velocity difference between granules and air become gradually larger. As a result, density waves emerge. The smaller the rate of air flow, the higher the onset point (along the



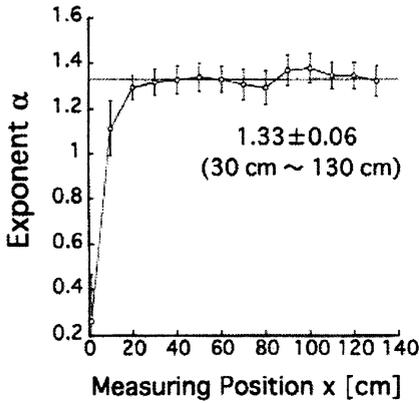
**Fig. 9.** Time-series signals and formation process of density waves in the case of *fully closed*. Length of each signal is 32 seconds. The signals are shifted by one volt from each other in order to avoid the data overlap.



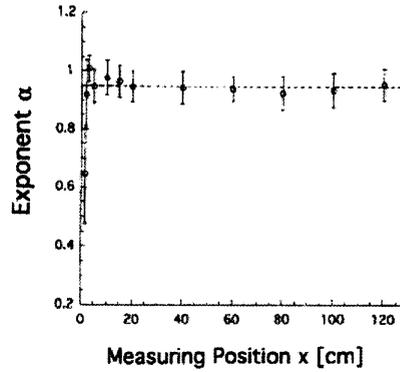
**Fig. 10.** Log-log plot of power spectra for *fully closed*. Each spectrum is appropriately shifted in order to avoid the data overlap. The spectrum at each  $x$  was obtained by averaging over 640 independent trials with the length of 8192 discrete points each (2 seconds in real time).

pipe) of density waves. The onset of density waves is characterized by the growth of the lower frequency part of the power spectra. We have also shown that power spectra of density waves display a clear power-law form  $P(f) \sim f^{-\alpha}$  and precisely estimated the scaling exponent  $\alpha$  as  $1.33 \pm 0.06$ , consistent with  $\alpha = 4/3$  obtained from the numerical simulations [19,20] and the theoretical prediction [24].

We have also confirmed that the value of  $\alpha$  is robust even under medium flow or variation of the pipe diameter, as far as density waves can be seen. However, when water (or silicone oil with various values of viscosity) is used as the medium, the situation is very different. In the fully closed case with lead spheres as granules and water as medium, density waves already emerge at  $x \simeq 1.5$  cm, very close to the hopper aperture. The power spectra of the density waves exhibit power-law behavior  $P(f) \sim f^{-\alpha}$ . However, the value of the exponent  $\alpha$  is close to 1, very different from  $4/3$ . Figure 12 shows the value of  $\alpha$  as a function of the measuring position  $x$ , which immediately converges to a constant value as  $x$  increases.



**Fig. 11.** Values of the exponent  $\alpha$  as a function of measuring positions  $x$ . This is the case where the interaction between granules and air is the strongest in this article (i.e., *fully closed*). Each value is estimated by the least mean square method in the frequency range 2.5 ~ 120 Hz. Averaged value of  $\alpha$  over  $x = 30 \sim 130$  cm is  $1.33 \pm 0.06$ .



**Fig. 12.** Values of the exponent  $\alpha$  as a function of measuring positions  $x$  for *fully closed* in the case of lead spheres as granules and water as medium. Averaged value of  $\alpha$  over  $x = 2 \sim 130$  cm is  $0.95 \pm 0.05$ .

When granules start falling from a hopper, they convert their potential energy to kinetic one and gain velocity. While they fall, they relax their kinetic energy among themselves and dissipate it to themselves and eventually to the medium air or pipe wall as thermal energy. This is a plausible scenario of single-path, unidirectional energy flow of falling granules when the cock is closed and the medium air does not flow with the granules. Density waves of granular flow emerging when the air is at rest is thought to be the consequence of this single-path, unidirectional energy flow. It does not matter whether the final sink of the energy flow is the medium air or the pipe wall. This may be the reason that our experimental result of the exponent  $\alpha \simeq 4/3$  is consistent with that of the computer simulations obtained by Peng and Herrmann [19,20]. However, if water is used as the medium, the viscosity is so strong that a considerable amount of kinetic energy of the granules may be dissipated directly to water, instead of dissipating first among themselves. In other words, if water (or silicone oil) is used as the medium, the energy flow is not single-path but multi-path. And the degree of multi-path energy flow may increase as the medium viscosity increases. This may be the origin of viscosity dependence of the exponent  $\alpha$  [22,27]. We believe that one should use air as the medium in order to obtain a universal value of the exponent  $\alpha$  for the single-path, unidirectional energy flow of granules.

So far we have mainly treated experimental data with Fourier analysis. In spite of the numerical efficiency of FFT, the information we can obtain from

Fourier analysis is limited since it is based on expansions using sinusoidal functions which extend over the whole space; one cannot relate local events in the time series signal to particular Fourier coefficients. Well-defined filtering technique (wavelet transform) is suitable to compensate for such a weakness of Fourier analysis. We have, therefore, performed wavelet analysis of granular flows [25].

## References

1. A. Mehta, (Ed.), *Granular Matter*, (Springer, New York, 1994).
2. H. Hayakawa, H. Nishimori, S. Sasa, and Y-h. Taguchi, *Jpn. J. Appl. Phys.* **34**, 397 (1995).
3. H.M. Jaeger, S.R. Nagel, and R.P. Behringer, *Rev. Mod. Phys.* **68**, 1259 (1996).
4. L.P. Kadanoff, *Rev. Mod. Phys.* **71**, 435 (1999).
5. A. Rosato, K.J. Strandburg, F. Prinz, and R.H. Swendsen, *Phys. Rev. Lett.* **58**, 1038 (1987).
6. H.K. Pak and R.P. Behringer, *Nature* **371**, 231 (1994).
7. F. Melo, P.B. Umbanhowar, and H.L. Swinney, *Phys. Rev. Lett.* **72**, 172 (1994).
8. P.B. Umbanhowar, F. Melo, and H.L. Swinney, *Nature* **382**, 793 (1996).
9. H.M. Jaeger, C-h. Liu, and S.R. Nagel, *Phys. Rev. Lett.* **62**, 40 (1989).
10. M. Nakagawa, *Chem. Eng. Sci.* **49**, 2540 (1994).
11. E.E. Ehrichs, H.M. Jaeger, G.S. Karczmar, J.B. Knight, V. Yu. Kuperman, and S.R. Nagel, *Science* **267**, 1632 (1995).
12. P.A. Thompson and G.S. Grest, *Phys. Rev. Lett.* **67**, 1751 (1991).
13. J.F. Davidson, R. Glift, and D. Harrison, (Eds.), *Fluidization*, (Academic, London, 1985).
14. G.K. Batchelor, *J. Fluid Mech.* **193**, 75 (1988).
15. T. Tanaka, T. Kawaguchi, and Y. Tsuji, *Int. J. Mod. Phys. B* **7**, 1889 (1993).
16. S. Sasa and H. Hayakawa, *Europhys. Lett.* **17**, 685 (1992).
17. H. Hayakawa and S. Sasa, in: *Complex Fluids*, L. Garrido, (Ed.), p. 319, (Academic, London, 1995).
18. J. Lee, *Phys. Rev. E* **49**, 281 (1994).
19. G. Peng and H.J. Herrmann, *Phys. Rev. E* **49**, R1796 (1994).
20. G. Peng and H.J. Herrmann, *Phys. Rev. E* **51**, 1745 (1995).
21. S. Horikawa, A. Nakahara, T. Nakayama, and M. Matsushita, *J. Phys. Soc. Jpn.* **64**, 1870 (1995).
22. S. Horikawa, T. Isoda, T. Nakayama, A. Nakahara, and M. Matsushita, *Physica A* **233**, 699 (1996).
23. T. Raafat, J.P. Hulin, and H.J. Herrmann, *Phys. Rev. E* **53**, 4345 (1996).
24. O. Moriyama, N. Kuroiwa, M. Matsushita, and H. Hayakawa, *Phys. Rev. Lett.* **80**, 2833 (1998).
25. O. Moriyama, N. Kuroiwa, M. Kanda, and M. Matsushita, *J. Phys. Soc. Jpn.* **67**, 1603 (1998).
26. O. Moriyama, N. Kuroiwa, M. Kanda, and M. Matsushita, *J. Phys. Soc. Jpn.* **67**, 1616 (1998).
27. A. Nakahara and T. Isoda, *Phys. Rev. E* **55**, 4264 (1997).
28. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Phys. Rev. E* **51**, 1035 (1995); T.S. Komatsu and S. Sasa, *Phys. Rev. E* **52**, 5574 (1995).
29. B.S. Kerner and P. Konh user, *Phys. Rev. E* **48**, 2335 (1993).
30. D.A. Kurtze and D.C. Hong, *Phys. Rev. E* **52**, 218 (1995).

# Asymmetric Random Average Process: Aggregation and Fragmentation on Continuous State Space

F. Zielen and A. Schadschneider

Institute for Theoretical Physics, University of Cologne, Germany

**Abstract.** A simple analytically treatable stochastic process on continuous state space, the asymmetric random average process, with a broad range of applications in traffic flow theory, internet modeling and granular media is presented. We concentrate mainly on the analysis of the basic properties of the model. The set of all exact mean field solutions is determined and we point out that it represents a class of very good approximants even for non-product-measure processes. Furthermore we study a truncated process that shows the occurrence of a nonsymmetric ergodicity breaking in the thermodynamic limit. This interesting phenomenon shows similar properties as spontaneous symmetry breaking, but without any explicit symmetry.

## 1 Model

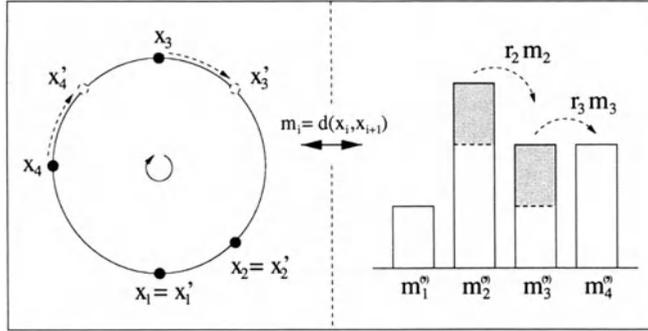
The asymmetric random average process (ARAP) [1,2] is defined on a one-dimensional periodic lattice with  $L$  sites. Each site  $i$  carries a non-negative continuous mass variable  $m_i$ . In every discrete time step  $t \rightarrow t + 1$  for each site a random number  $r_i \in [0, 1]$  is generated from a time-independent probability density function (pdf)  $\phi$ , sometimes called fraction density, that may depend on the actual configuration  $m = (m_1, \dots, m_L)$ . So the universal form can be written as  $\phi = \phi(r_1, \dots, m_1, \dots)$ . The fraction  $r_i$  determines the amount of mass  $r_i m_i$  transported from site  $i$  to site  $i + 1$ . The transport is completely asymmetric, i.e. no mass is transported in the opposite direction  $i + 1 \rightarrow i$  and we obtain

$$m_i \rightarrow (1 - r_i)m_i + r_{i-1}m_{i-1} . \quad (1)$$

These update rules correspond to a parallel time dynamics. Due to the conservation of the total mass  $M = \sum_i m_i$  the density  $\rho = \frac{M}{L}$  is fixed.

We would like to add that this so called stick representation can easily be mapped onto a particle picture: by considering  $L$  particles on a ring of length  $L\rho$  and denoting their positions by  $x_i$ , both representations become equivalent by identifying the masses  $m_i$  as the headway of particles, i.e.  $m_i = d(x_i, x_{i+1})$ . Due to the fact that the pdf  $\phi$  depends only on the particle distances  $m_i$ , the ARAP is closely related to the class of car-following models.

For both pictures the update procedure is visualized in fig. 1, but in the following we will work in the stick representation only.



**Fig. 1.** Asymmetric Random Average Process: the random numbers  $r_i$  are generated according to a probability density  $\phi = \phi(r_1, \dots, m_1, \dots)$ . The state variables  $x_i, m_i$  and  $x'_i, m'_i$  correspond to the times  $t$  and  $t + 1$ , respectively.

## 2 Fundamental Results

The basic quantity we are interested in is the mass pdf  $P(m, t)$ . The knowledge of this function corresponds to the complete solution of the problem. In the following we restrict our considerations to steady state dynamics only, i.e. the master equation

$$\begin{aligned}
 P(m'_2, \dots, m'_k) &= \int_0^\infty d^k m P(m) \int_0^1 d^k r \phi(r, m) \\
 &\times \prod_{i=2}^k \delta(m'_i - [r_{i-1} m_{i-1} + (1 - r_i) m_i]) \quad (2)
 \end{aligned}$$

with  $P(m) = \lim_{t \rightarrow \infty} P(m, t)$ . Determination of a  $(k-1)$ -site mass distribution presumes the calculation of the  $k$ -site distribution. Thus we are confronted with a hierarchy of an infinite number of equations in the limit  $L \rightarrow \infty$ .

The simplest version of the ARAP is obtained if we use the state independent uniform distribution defined by the fraction density  $\phi = 1$ . We will refer to this system as the free ARAP. Some results have been derived for this model so far whereby the exactness of product measure for  $L \rightarrow \infty$  and its form [3]

$$P(m) = \prod_i P(m_i) \quad \text{with} \quad P(m_i) = \frac{4m_i}{\varrho^2} e^{-2m_i/\varrho} \quad (3)$$

is an important one within the scope of this work. Equation (3) has been derived in the context of the q model [3] that has been developed for the description of force fluctuations in bead packs and shares many properties with the ARAP.

Another significant result is the factorization of the stationary two-point mass correlations for  $\phi = \prod_i \phi(r_i)$ , i.e. for ARAPs given by state and site independent pdfs [1]. This indicates that mean field calculations rather generally yield

good approximations and that the class of exact mean field mass distributions represents an important subset of the solution space.

Further findings related to the ARAP, e.g. the consideration of other kind of updates, a discrete state space or symmetric mass movements, can be found in [2,4-6].

### 3 Exact Mean Field Solutions

In this section we determine all exact stationary mean field (MF) solutions for  $\phi = \prod_i \phi(r_i)$  in the limit  $L \rightarrow \infty$ . This has been done explicitly in [6] and we would like to sketch the basic strategy here. Furthermore we show that our findings also represent very good approximants for arbitrary  $\phi$ -functions.

Although the MF ansatz  $P(m) = \prod_i P(m_i)$  decouples the infinite set of equations (2) we have to check consistency of (2) by inserting a MF solution into all equations. However, by  $k$ -dimensional Laplace transforming the  $k$ -site mass densities and introducing the functional map

$$F_Q(s, \bar{s}) \equiv \int_0^1 dr \phi(r) Q((1-r)s + r\bar{s}) \tag{4}$$

we reduce the infinite set of equations to the MF criterion (MFC)

$$F_Q(s_1, s_2) = F_Q(s_1, 0) \cdot F_Q(0, s_2) . \tag{5}$$

For  $s_1=s_2=s$ , (5) acts as a conditional equation for the determination of a MF solution and yields

$$Q(s) = \int_0^1 dr \phi(r) Q(rs) \int_0^1 dr \phi(1-r) Q(rs) . \tag{6}$$

Putting (6) in (5) we can easily check the exactness of the MF solution. By expanding the single site Laplace transformed  $Q(s)$  in a series around zero – this is always possible [6] – and representing the fraction density  $\phi$  by its moments  $\mu_n \equiv \int dr r^n \phi(r)$ , the set  $\mathcal{M}$  of all density functions  $\phi$  resp.  $\{\mu_n\}$  yielding exact MF solutions is calculated:

$$\mathcal{M} = \left\{ \{\mu_n\}_{n \in \mathbb{N}} \mid \mu_n = \frac{\Gamma(n + \lambda_1)}{\Gamma(\lambda_1)} \frac{\Gamma(\lambda_2)}{\Gamma(n + \lambda_2)} \right\} \tag{7}$$

with

$$\lambda_1 = \mu_1 \frac{\mu_1 - \mu_2}{\mu_2 - \mu_1^2} \quad \text{and} \quad \lambda_2 = \frac{\mu_1 - \mu_2}{\mu_2 - \mu_1^2} . \tag{8}$$

So  $\mathcal{M}$  is only parameterized by  $\mu_1$  and  $\mu_2$  that have to be chosen with respect to the general moment properties  $1 > \mu_1 > \mu_2 \geq \mu_1^2$ . We would like to note that the derivation of (7) is exact up to a conjecture which has been verified only partially until now [6].

The case  $\mu_2 = \mu_1^2$ , corresponding to  $\phi(r) = \delta(r - \mu_1)$ , is not covered by (8) and leads directly to  $P(m) = \delta(m - \varrho)$ . For  $\mu_2 > \mu_1^2$  we are able to calculate the general form of the MF mass distributions and obtain

$$P(m) = \frac{\lambda_2^{\lambda_2}}{\Gamma(\lambda_2)} \frac{1}{\varrho} \left(\frac{m}{\varrho}\right)^{\lambda_2-1} e^{-\lambda_2 \frac{m}{\varrho}} . \tag{9}$$

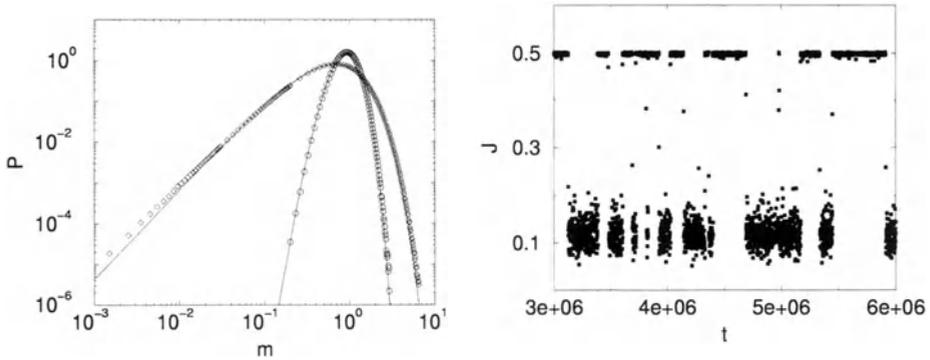
In contrast to (7) depending on both  $\lambda_1$  and  $\lambda_2$ ,  $P(m)$  is a function of  $\lambda_2$  only. So MF ARAPs with fixed  $\lambda_2$  and arbitrary  $\lambda_1$  have identical mass distributions.

For arbitrary ARAPs given by  $\phi = \prod_i \phi(r_i)$  with  $\phi \notin \mathcal{M}$  we are able to construct appropriate approximants with the help of (8) and (9). This is done by calculating the parameter  $\lambda_2$  that determines the shape of  $P(m)$  from the first and second moment  $\mu_1$  and  $\mu_2$  of  $\phi$  and inserting this value in (9).

As an example we present the simplest version of an ARAP with truncation. It is based on the free ARAP but enhanced by an additional parameter  $R \in (0, 1]$ . The cutoff  $R$  controls the movement in the following way: A stick fragment  $r_i m_i$  is only transferred if  $r_i \leq R$ . The corresponding density function takes the form

$$\phi_R(r) = (1 - R)\delta(r) + \theta(R - r) \tag{10}$$

where  $\theta$  is the Heaviside step-function. One can easily prove that  $\phi_R \notin \mathcal{M}$  for  $R < 1$ , so mean field is not exact. Nevertheless we obtain by the construction presented above approximants that match the Monte Carlo simulation data very well (see fig. 2a). Only for small masses  $m$  systematical differences occur.



**Fig. 2.** (a) Analytical [–] and numerical [ $R = 0.1(\circ), 0.5(\diamond)$ ] single site mass distributions of the truncated ARAP  $\phi_R(r) = (1 - R)\delta(r) + \theta(R - r)$ . (b) Current-time diagram of the TARAP in the mixed phase ( $L = 100, \hat{\Delta} = 0.846$ ).

### 4 Nonsymmetric Ergodicity Breaking

In this section a truncated ARAP with site dependent fraction function of the form  $\phi = \prod_i \phi(r_i, m_i)$  is presented. A phase characterized by the occurrence of

nonsymmetric ergodicity breaking is observed. In the following we summarize primarily the main results of [7].

The free ARAP is equipped with a simple additional update rule: transfers of masses  $r_i m_i$  larger than a cutoff  $\Delta > 0$  are rejected. This represents a truncated ARAP (TARAP) with fraction density

$$\phi(r, m) = [1 - R(m)] \delta(r) + \Theta(R(m) - r) , \tag{11a}$$

$$R(m) = \min(1, \Delta/m) . \tag{11b}$$

By substituting  $m \rightarrow \varrho$  in (11b) we obtain the site independent fraction density (10) with  $R = R(\varrho)$ . We identify two phases in this toy model: For  $\Delta \geq \varrho$  we obtain  $\phi = 1$  and the system is exactly solved by (3). For  $\Delta < \varrho$  we are in a phase described by the existence of truncation.

By restricting to the case  $\varrho = 1$  (without loss of generality) and introducing the rescaled cutoff

$$\tilde{\Delta} \equiv 2L^{-\frac{1}{2}} \Delta \tag{12}$$

we also identify two phases in the TARAP. Our results base on Monte Carlo simulations and analytical approximations.

For  $\tilde{\Delta} > \tilde{\Delta}_c \equiv 1$  the system remains in a high flow phase, i.e. the mass is distributed homogeneously and approximatively given by (3). The current, i.e. the average mass shift per site, is independent of the rescaled cutoff and equal to the flow of the free system ( $J_{\text{high}} = J_{\text{high}}^{\text{max}} \equiv \frac{1}{2}$ ).

For  $\tilde{\Delta} < \tilde{\Delta}_c$  the TARAP is in a mixed phase. Here the system can either exist in a high flow state (with properties as described above) or a low flow state. The latter is given by an macroscopic condensate on one site which corresponds to a jam in the particle picture. Simulations have shown that configurations with two or more aggregates are not stable and merge into one column after a while. The flux of the low flow state is given by  $J_{\text{low}} = J_{\text{low}}^{\text{max}} (1 - (1 - \tilde{\Delta}^2)^{\frac{1}{2}})$  with  $J_{\text{low}}^{\text{max}} \equiv \frac{1}{4}$ .

In finite systems the average lifetimes  $\tau_{\text{low}}, \tau_{\text{high}}$  of the low and high flow states are finite. So a single system switches between these states while evolving in time and an alternating current-time relation is obtained (fig. 2b).

In the thermodynamic limit simulations and analytical estimations indicate a diverging behaviour of the lifetimes, i.e.  $\tau_{\text{low}}, \tau_{\text{high}} \rightarrow \infty$  for  $L \rightarrow \infty$ . So ergodicity is broken because in dependence of the initial condition the system evolves into the high flow or the low flow state and will stay there forever. It is remarkable that low and high flow state are obviously not related by symmetry, so we call this effect a nonsymmetric ergodicity breaking.

We would like to remark that the scaling of the average lifetimes with the system size is different for high and low flow states. Although it is very difficult to fit the numerical curves appropriately  $\tau_{\text{low}}$  seems to increase algebraically with  $L$  while the  $L$ -dependence of  $\tau_{\text{high}}$  seems to be exponential. This behaviour reflects different switching mechanisms. The transition from the homogeneous high to the singular low flow state is mainly driven by fluctuations while the opposite transition is based on slow dynamics – it takes a long time to reduce an infinite aggregate.

## 5 Discussion

In this work the ARAP has been studied on a fundamental level. In section 3 exact solutions and approximations for state and site independent fraction densities  $\phi = \prod_i \phi(r_i)$  have been derived. In section 4 we have provided evidence for an interesting new feature, nonsymmetric ergodicity breaking, which can be observed in the truncated variant of the ARAP.

However, if we try to model problems of traffic or granular flow in terms of the ARAP we will in general be confronted with more complicated fraction densities. But there are some questions that can be answered by the use of simple ARAPs, i.e. simple  $\phi$ -functions. The q model [3] is a good example here. Furthermore the TARAP can be treated as a stochastic toy model of internet transport: a ring of routers  $\{i\}$  with actual storage  $m_i$  shuffles data packages in a preferred direction. Multiple and uncorrelated requests are realized by transporting only random fractions  $r_i m_i$  to the next neighbour and a maximum band width is represented by the cutoff  $\Delta$ . Based on this model we are able to compute a quantity like  $\tilde{\Delta}_c$  that reflects a critical band width considered as a lower bound for uncongested traffic. However, more realistic internet models must incorporate more features like bounded router capacities, open boundaries or re-sending processes.

So the ARAPs studied here fulfill mainly the role of toy models that give qualitative estimations, e.g. by the  $m \rightarrow \varrho$  substitution for site dependent  $\phi$ 's, or explain physical effects of more complicated processes. As an example we would like to mention the Krauss model [8] for traffic flow that represents a continuous version of the Nagel-Schreckenberg model. This system shows a similiar phase coexistence like the TARAP in the mixed phase, but the underlying dynamics are complex which complicates an analytical treatment. On the other hand, the essentials supporting the occurrence of the mixed phase are identical in both models: truncation on a continuous state space. So the TARAP might catch some of the fundamental physics behind the Krauss model.

## References

1. J. Krug and J. García, J. Stat. Phys. **99**, 31 (2000).
2. R. Rajesh and Satya N. Majumdar, J. Stat. Phys. **99**, 943 (2000).
3. S.N. Coppersmith *et al.*, Phys. Rev. E **53**, 4673 (1996).
4. G.M. Schütz, J. Stat. Phys. **99**, 1045 (2000).
5. R. Rajesh and Satya N. Majumdar, Phys. Rev. E **64**, 036103 (2001).
6. F. Zielen and A. Schadschneider, J. Stat. Phys. **106**, 173, (2002).
7. F. Zielen and A. Schadschneider, in preparation.
8. S. Krauss, P. Wagner, and C. Gawron, Phys. Rev. E, 3707 (1996).

# Shape Segregation for Bidisperse Mixtures of Ellipses in Two Dimensions

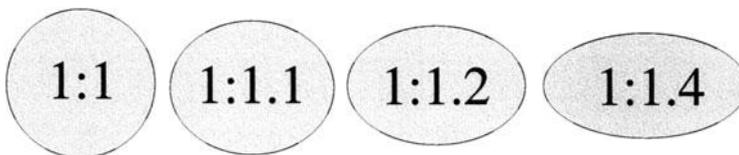
H.-G. Matuttis, N. Ito, and H. Watanabe

Department of Applied Physics, School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan

**Abstract.** Whereas size segregation is a fairly common phenomenon in granular material research, up to now shape segregation has not been observed. We present simulations of bidisperse shaped (elliptic and round) granular particle systems which exhibit a demixing of round and elongated particles. The effect is not only controlled by the particle elongation, but also by the Coulomb friction coefficient in the system.

## 1 Introduction

Spherical particle simulations are very common in granular materials research. Nevertheless, there are phenomena which seem not to be accessible by round particles. Our aim in studying systems of elliptic particles [1] was, that this shape seems to be the next simplest shape after spherical particles, allowing a continuous transition from round to elliptic shapes. Size segregation has been a common phenomenon in granular materials [2], sometimes under names like "Brazilnut-effect" or "physics of muesli" [3]. Our research interest was to investigate, whether segregation could not only be caused by different size particles, but also by particles with different shape.



**Fig. 1.** Example of ellipses with the same area but different elongation.

## 2 Setup and Simulation Parameters

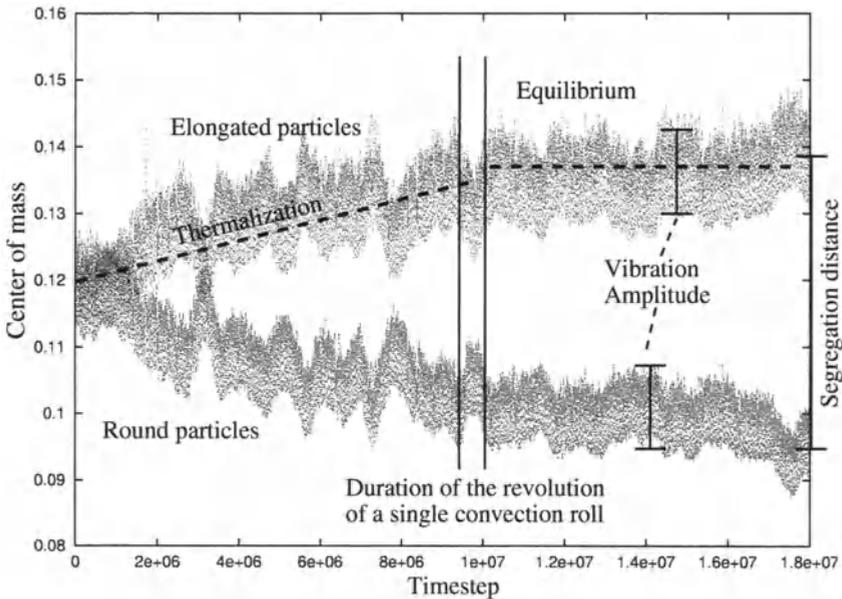
We simulated a vibrated box of bidisperse particles (round and elliptic) with **identical** volume/ area in two dimensions. The mixture consists of 50 % round, 50 % elongated particles. The box was driven with a frequency of 90 Hz and an amplitude of 25 mm. The particle diameter was about 8 mm. In the simulation, we monitored the center of mass of the spherical particles and the elongated

particles. We investigated the dependence of the shape segregation from different elongations (see Fig. 1) and from the Coulomb friction coefficient.

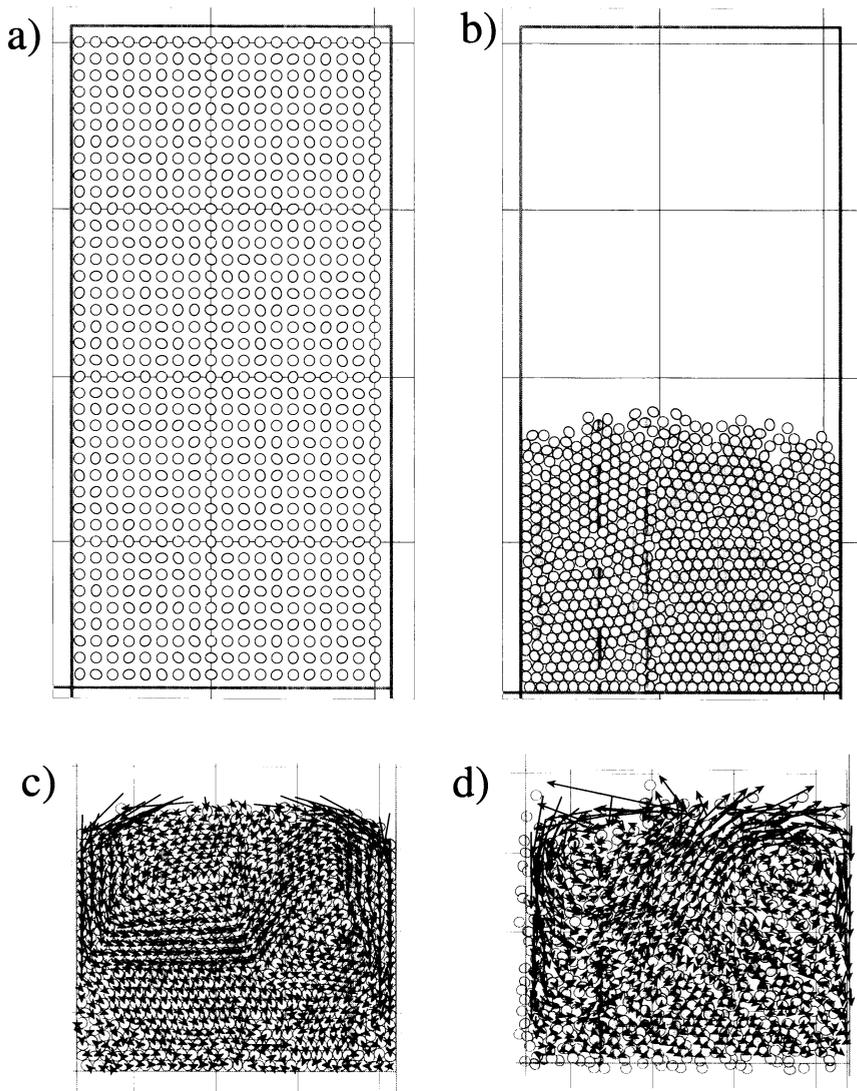
The interaction law has a "hardcore repulsion" (Young's modulus =  $10^6 \text{N/m}$ ), the normal damping/dissipation (coefficient of restitution) is so large that particle-particle collisions are practically inelastic (we varied the numerical value, but the effect was negligible), tangential Coulomb friction after Cundall and Strack [4] have been implemented using Coulomb coefficients between  $\mu = 0.0$  and  $\mu = 0.6$ . The force law was derived for polygons [5] but could be adapted for ellipses without modifications [1].

### 3 Time Evolution

The system was chosen so that the volume occupied by the particles was of approximately square shape. If the system is too shallow, the segregation is suppressed. An example of the quantitative time evolution, i.e. the motion of the center of mass of the elongated and round particles, is shown in Fig. 2. In Fig. 3, some snapshots from during the simulation are shown.



**Fig. 2.** Typical time evolution of the center of mass of the spherical particles and of the elongated particles.

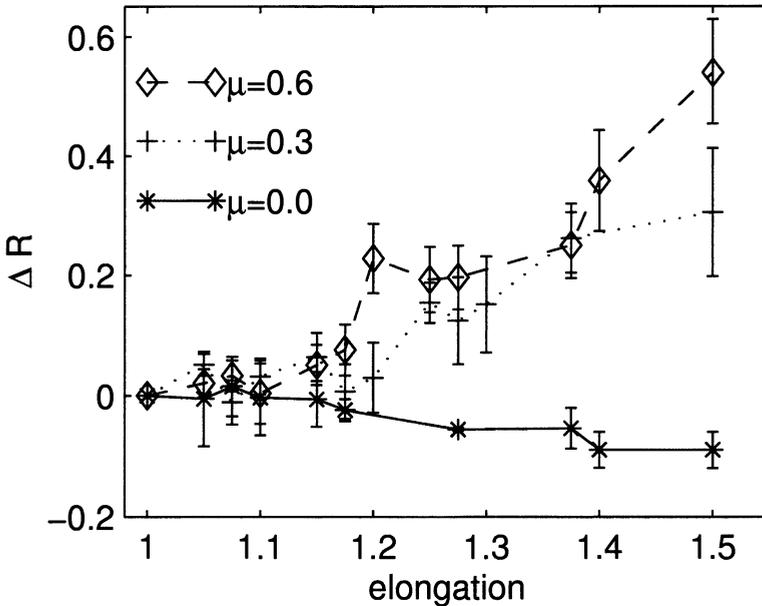


**Fig. 3.** Time evolution of the shape segregation in elliptic particles: a) Setting up the system with a random distribution of round (blue) and elongated (yellow) particles. b) Initial relaxation leads to a random packing. c) Onset of the convection after the particles are sufficiently compacted. d) Segregation result: Elongated particles (yellow) rise to the top. Black arrows show the velocities in arbitrary units.

#### 4 Influence of the Friction Coefficient

It turned out that the determining parameter of the segregation was not the elongation of the particles alone, but the Coulomb friction coefficient (we use the same friction coefficient for static and dynamic friction) was also crucially influencing the outcome.

The direction of the convection rolls is the same for all friction coefficients  $\mu = 0.0 - 0.6$ . Nevertheless, for friction coefficients  $\mu = 0.3 - 0.6$ , the elongated particles rise to the top, whereas for friction coefficient  $\mu = 0.0$ , the spherical particles rise to the top, see Fig. 4. The error bars indicated in Fig. 4 and Fig. 5 are the fluctuations around the equilibrium position as indicated in Fig. 2.

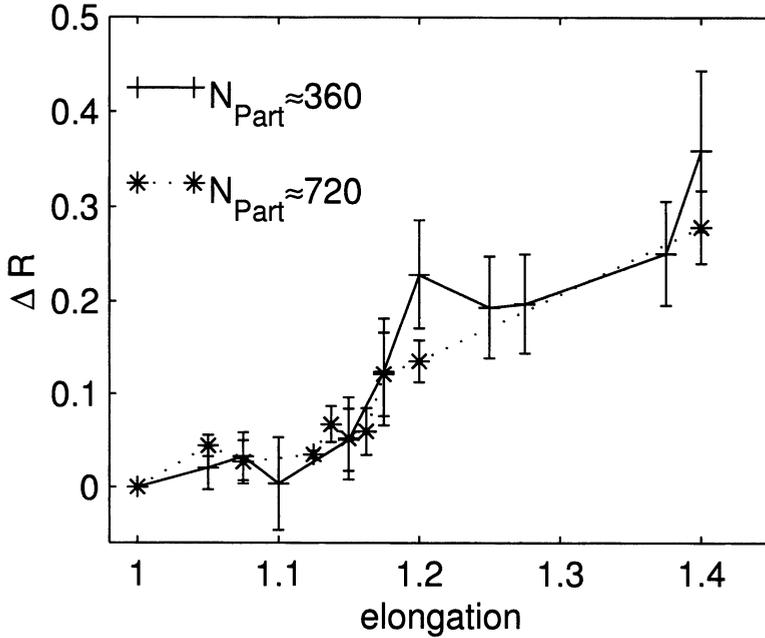


**Fig. 4.** Coulomb friction dependence of the shape segregation effect for friction coefficients  $\mu = 0.0, 0.3, 0.6$ . The  $\Delta R$  for the  $y$ -Axis is the height of the distance between the center of mass of the elliptic and the round particles, rescaled by the center of mass.

It turns out that both friction and elongation can be used to control the transition from a non-segregating system to a segregating system. For systems with 350 and 700 particles, the transition occurred at practically the same elongation of about 1.175. Bidisperse systems for a given elongation  $l$  of the longer particles show clearer segregation than polydisperse mixtures with an average elongation of  $\langle l \rangle = l$ .

No shape segregation could be found for bidisperse mixtures of spherical particles and triangles and for spherical particles and square particles. In the simulation

using elliptic particles Shape turns out to be stronger than size: As long as the longer axis of the elongated particles is longer than about 1.15 of the diameter of the round particles, the elongated particles (at least for our test at  $\mu \geq 0.3$ ) will rise to the top, even if the area of the round particles is larger than the area of the elliptic particles.

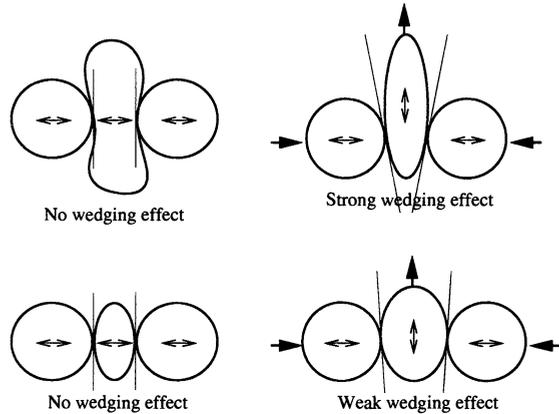


**Fig. 5.** Negligible size effect of the shape segregation for a friction coefficient of  $\mu = 0.6$ . The  $\Delta R$  for the  $y$ -Axis is the height of the distance between the center of mass of the elliptic and the round particles, rescaled by the center of mass.

## 5 Summary

We have found that a shape segregation effect in granular materials in two dimension exists. It is not only dependent on the shape alone, but depends also on the friction in the system. Fig. 6 shows a tentative explanation on geometrical grounds (wedging effect) why elongated particles can rise over round particles. This explanation should work independent of the friction coefficient, and is therefore insufficient.

We have up to now not been able to derive a microscopical explanation, neither interface energy (between clusters of spherical and elongated particles) nor entropical arguments are tractable to include the effects of friction.



**Fig. 6.** Possible geometry-dependent segregation mechanism for the shape segregation: The elongated particles are wedged upwards by the round particles. If the long axis of the elongated particles is too short, no wedging can take place. If the particle is of convex shape or has parallel sides, also no wedging can take place.

Friction has been found to be able to reverse the dynamic behavior of the frictionless system, i.e. the round particles are rising to the top for sufficiently large systems. This poses the interesting question, up to which point the investigation of a statistical system (with normal dissipation, but without tangential friction) is really relevant for the investigation of a granular system (with tangential friction).

A macroscopic explanation of the shape segregation may give insights into the existence of a "configurational entropy". We further have to narrow down the regime for numerical value of the elongation for which the onset of the segregation can be observed. Currently, we find this "critical elongation" to be of  $\approx 1.175 \pm 0.025$ . We further will try to find out whether there is a "critical friction coefficient" for which no segregation occurs, and whether it is universal for all elongations. We are also looking for suitable materials which will allow an experimental realization for 3 dimensions.

## References

1. H.-G. Matuttis, N. Ito, H. Watanabe, and K. M. Aoki, 'Vectorizable overlap computation for ellipse-based discrete element method', in: *Powders & Grains 2001*, Y.Kishino, (Ed.), pp. 173–176, (Balkema, Rotterdam 2001).
2. R.L. Brown, *The Institute of Fuel* **13**, 15 (1939).
3. G. Barker and M. Grimson, *New Scientist* **26**, 37 (1990).
4. P.A. Cundall and O.D.L. Strack, *Géotechnique* **29** (1), 47 (1979).
5. H.-G. Matuttis, *Granular Matter* **1** (2), 83 (1998).

# Bifurcations of a Driven Granular System under Gravity

M. Isobe

Department of Computational Science, Kanazawa University, Kakuma, Kanazawa 920-1192, Japan.

**Abstract.** The molecular dynamics study on the granular bifurcation in a simple model is presented. The model consists of hard disks, which undergo inelastic collisions; the system is under the uniform external gravity and is driven by the heat bath. The competition between the two effects, namely, the gravitational force and the heat bath, is carefully studied. We found that the system shows three phases, the condensed phase, the locally fluidized phase, and the granular turbulent phase, upon increasing the external control parameter. We conclude that the transition from the condensed phase to the locally fluidized phase is distinguished by the existence of fluidized holes, and the transition from the locally fluidized phase to the granular turbulent phase is understood by the destabilization transition of the fluidized holes due to mutual interference.

## 1 Introduction

The dynamics of fluidized granular systems has attracted much attention in physics communities as a non-equilibrium statistical system. In most experimental situations, the systems are excited by a vibrating plate (granular vibrated bed), in which the effect of gravity is very large. In the conventional setting of vibrated bed, the frequency of external driving plays an important role and gives variety of pattern phases. However, the vibration complicates the situation. Since energy input from the vibration depends on both the amplitude  $a$  and the angular frequency  $\omega$ , the accurate quantities of input energy cannot be estimated from the external control parameter  $\Gamma = a\omega^2/g$ , which is usually used as the control parameter in the granular vibrated bed. From the numerical point of view, if we want to obtain the data for the statistical properties during the DEM simulation, the sampling data must be taken at the same phase of external vibration cycle. The granular vibrated bed has several disadvantages to study the statistical characters of the inelastic particle system in the non-equilibrium steady state from the aspects of both theory and simulation.

The main purpose of our study is to examine the statistical and dynamical characters of the inelastic hard disk system in the non-equilibrium steady state under gravity, and to provide extensive numerical results for theoretical studies to construct a base of the universal framework. In this work, we have performed event-driven molecular dynamics [1] on the granular bifurcation in a simple model [2,3]. The system consists of inelastic hard disks driven by the heat bath under the uniform gravity, which can be completely characterized by

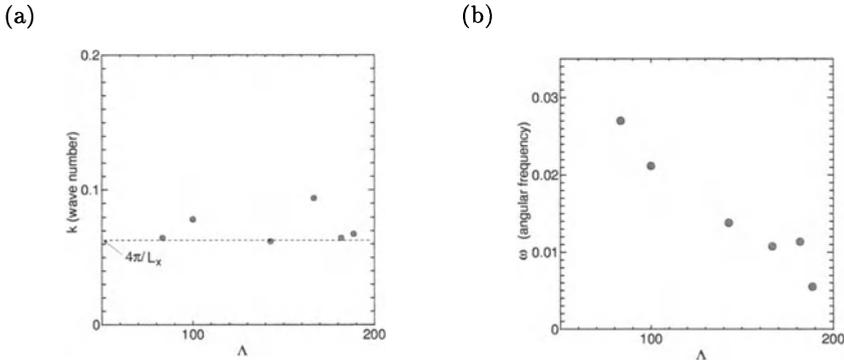
the following four dimensionless parameters; the restitution coefficient  $r$ , the driving intensity  $\Lambda \equiv k_B T_w / mgd$ , the system width in the unit of disk diameter  $N_w \equiv L_x / d$ , and the number of layer  $N_h \equiv N / N_w$ , where  $L_x$  is the system width and  $N$  is the total number of disks. Note that  $k_B$  is the Boltzmann constant and the wall temperature  $T_w$  here is just a parameter to characterize the external driving and is not related to the thermal fluctuation. The present system is analogous to the ordinary granular vibrated bed, but simpler because it does not have an external time scale, or a period of external driving vibration. We particularly focus on the competition between the effect of external driving and that of the gravitational field, systematically. We found that the system shows three phases, namely, the condensed phase (CP), locally fluidized phase (LFP), and granular turbulent phase (GTP), upon increasing the external control parameter.

## 2 Surface Wave in CP

For  $\Lambda$  in CP, there is a closed packing layer and the state is not very dynamic at the particle level, because excitation is low and the potential energy is dominant. However, it is apparent that collective motion appears on the surface of the layer. What is the mechanism of the surface wave-like motion in CP? There must be no surface tension in the macroscopic dissipative system, therefore the restoring force should originate from the balance between the gravity and the excitation, but it is not clear if it could be described as an ordinary gravitational wave in fluid.

To study this collective motion quantitatively, we calculate the dynamical structure factor of the surface wave in CP. The dynamical structure factor  $S(k, \omega)$  is described as

$$S(k, \omega) = \frac{1}{2\pi N_{gx}} \int_{-T}^T \sum_{i,j} \langle h_y(x_i, t) h_y(x_j, 0) \rangle_T e^{-ik(x_i - x_j)} e^{i\omega t} dt, \quad (1)$$



**Fig. 1.** The  $\Lambda$  dependence of the most strong peak ( $k_{max}, \omega_{max}$ ) is plotted. (a) The wave number  $k$  vs. external driving  $\Lambda$ . (b) The angular frequency  $\omega$  vs. external driving  $\Lambda$ .

where  $h_y(x, t)$  is the surface height of condensed layer;  $N_{gx}(= 200)$  is the total number of grids in the horizontal direction and the discrete positions are taken by  $x_i = (L_x/N_{gx})/2 + (i - 1)(L_x/N_{gx})$ , ( $i = 1 \dots N_{gx}$ ).

In Fig. 14 of Ref. [3], the spatio-temporal pattern of surface wave in CP near the CP-LFP transition point is shown. There are many domains, which means the collective motion of the surface appears. When  $\Lambda$  becomes low, we found the domain size in the vertical direction became small. While, the domain size in the horizontal direction was not changed very much. This means that the periodical cycle of the surface becomes shorter for low  $\Lambda$ . These results are also confirmed by the computation of the dynamical structure factor. When the strongest peak ( $k_{max}, \omega_{max}$ ) was plotted in  $k - \omega$  space for each  $\Lambda$ , we found that the  $k_{max}$  stays around  $k = 2\pi/(L_x/2)$  and increases only slightly upon decreasing  $\Lambda$  (Fig. 1(a)), but  $\omega_{max}$  becomes large as  $\Lambda$  approaches 0 (Fig. 1(b)).

We conclude the periodical behavior of the surface wave in CP shows the critical slowing down when  $\Lambda \rightarrow \Lambda_{CL}$ . We also found the amplitude of the surface fluctuation becomes large near the transition point  $\Lambda_{CL}$ . This is an interesting point that the system shows large fluctuation around the transition like the critical fluctuation in the second order phase transition in equilibrium.

### 3 Dynamics of Fluidized Holes

The localized excitation in LFP should resemble a circle in 3D system and reminds us of an oscillon, but they are different; the external vibration frequency is essential for the oscillon dynamics, but the localized excitation here does not have such a characteristic frequency. In the large  $N_w$  simulation, we can observe the merging process of two or more excitations, but their mode of interaction is not clear. An interesting question here is if the transition from LFP to CP and/or the transition to GTP can be understood in terms of the local excitation; if LFP transform to CP when the distance between the excitations diverges? Does LFP transform to GTP at the point where they condense?

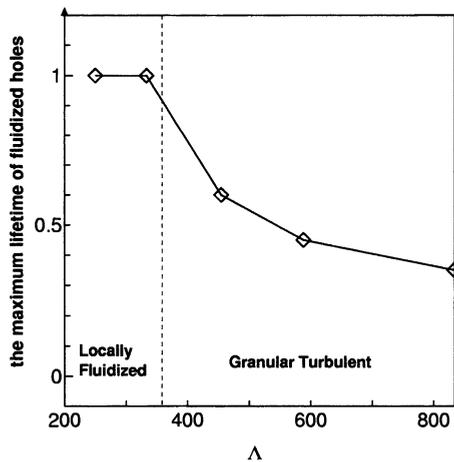
To investigate the excitation dynamics, we calculate the density fluctuation for each position with the width of particle diameter  $d$  in the system. The driving intensity  $\Lambda$  was varied from LFP to GTP. Near the critical point between CP and LFP ( $\Lambda \sim \Lambda_{CL}$ ), the life time of a fluidized hole does not continue through the simulation. When  $\Lambda$  is increased, the region of fluidized holes also increases. Above the LFP-GTP transition point, the local excitation domain fluctuates largely and sometimes expands to the length of the system. We found that the spatio-temporal patterns of fluidized holes changes significantly through the transition from LFP to GTP. The interference destroys the stable particle flow within an isolated fluidized hole, and the holes are destabilized dynamically. In Figs. 15(c) and 17(d),(e),(f) of Ref. [3], which are the spatio-temporal patterns in GTP, we found that many high-speed moving small clusters appeared. These clusters divide the region of the fluidized holes (black region) into several parts. In the patterns of LFP, there are no such moving clusters. Therefore, these high-speed moving small clusters are thought as the peculiar characters in GTP. When

these clusters move across the fluidized holes in the spatio-temporal patterns, the lifetime of the fluidized holes becomes shorter. Although there is still room for the definition of quantitative values to distinguish the phases between LFP and GTP, we show the maximum lifetime of fluidized holes in the spatio-temporal patterns as a function of  $\Lambda$ . In Fig. 2, the maximum lifetimes of the fluidized holes in  $N_w = 400$  around  $\Lambda \sim \Lambda_{LG}$  are shown. When the maximum lifetime of fluidized holes reaches the simulation length, we normalize them at unity. We found the maximum lifetime of fluidized holes changes drastically between LFP and GTP. Thus the LFP-GTP transition is interpreted as the dynamical destabilization transition of the fluidized holes.

## 4 Discussion

### 4.1 CP-LFP Transition

We found that the value of the maximum flatness shows a sharp rise around the transition point between CP and LFP [2,3]. This unexpected large value comes from the fact that the velocity probability distribution in LFP consists of two components; the narrow Gaussian distribution that comes from the dense layer and the broad stretched exponential distribution in the fluidized holes. Therefore, CP and LFP are distinguished by the existence of the fluidized holes. In LFP, the majority of the particles are in the condensed layer, and it is punctured by the fluidized holes, from which the high-speed particles are blown. The stable flow stabilizes the isolated fluidized hole in condensed layer. The size of the holes gets larger when  $\Lambda$  is increased. This result can be understood by the fact that the size is determined by the distance that the blown up particles reach, which is determined by  $\Lambda$ . The time scale of the surface fluctuation in CP becomes larger



**Fig. 2.** The maximum lifetime of fluidized holes vs. the driving intensity  $\Lambda$ .

as  $\Lambda$  approaches the critical point from below. On the other hand, the length stay constant. From the analysis of the dynamical structure factor of the surface wave, the typical frequency does not vanish at the critical point, which suggests the transition between CP and LFP is subcritical.

## 4.2 LFP-GTP Transition

At the LFP-GTP transition point, neighboring holes start interfering with each other. At the transition point, the blown-up particles at one hole barely reach the neighboring holes. In the larger systems, we observed more than one hole in the system, and found two different holes interfere with each other in GTP. The interference between holes destroys the stable particle flow within an isolated fluidized hole, and the holes are destabilized dynamically. The LFP-GTP transition is interpreted as the dynamical destabilization transition of the fluidized holes. The particle dynamics in GTP are very different from that of LFP. The average density is quite low, but it is very different from the ordinary molecular gas phase. The density fluctuation is very large and this fluctuation causes turbulent motion due to the gravity. Even for very large  $\Lambda$ , the high-density clusters that look like the condensed layer are formed temporally.

## 5 Concluding Remarks

We conclude that the transition from the condensed phase to the locally fluidized phase is distinguished by the existence of fluidized holes, and the transition from the locally fluidized phase to the granular turbulent phase is understood by the destabilization transition of the fluidized holes due to mutual interference.

In the study of the dissipative structure system, such as Rayleigh-Bénard convection, it is well known that the several phase changes (bifurcations) occur when the external driving is increased. However, in dissipative discrete element (granular) systems, there are very few studies so far. We believe our study will offer a step for understanding the macroscopic universal characters in the studies of fluidization of dissipative (inelastic) discrete element system with a heat bath under gravity and constructing the non-equilibrium statistical mechanics.

## Acknowledgements

I acknowledge helpful discussion with Prof. H. Nakanishi, Prof. H. Hayakawa and Ms. N. Mitarai on granular material.

## References

1. M. Isobe, *Int. J. Mod. Phys. C* **10**, 1281 (1999).
2. M. Isobe and H. Nakanishi, *AIP Conf. Proc. (USA)* **469**, 182 (1999); *J. Phys. Soc. Jpn.* **68**, 2882 (1999); *Prog. Theor. Phys. Suppl.* **138**, 596 (2000); *AIP Conf. Proc. (USA)* **519**, 238 (2000).
3. M. Isobe, *Phys. Rev. E* **64**, 031304 (2001).

# Simulation of the Impact of an Elastic Disk

H. Kuninaka and H. Hayakawa

Graduate School of Human and Environmental Studies, Kyoto University, Sakyo-ku, Kyoto, 606-8501, Japan

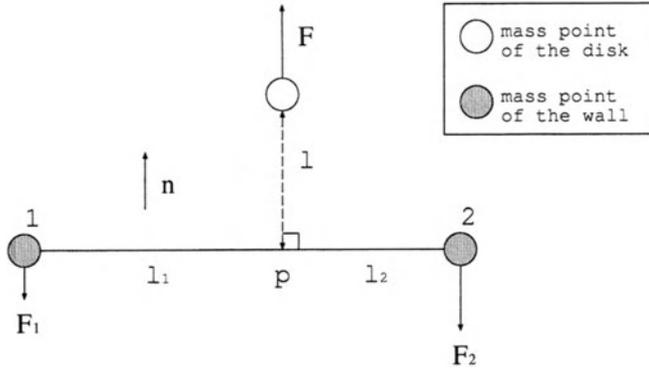
**Abstract.** The impact of an elastic disk with an elastic wall is numerically studied. We introduce a numerical model of the elastic disk and the elastic wall by mass points and linear springs. From our simulation, we find that the coefficient of normal restitution depends on the thickness of the elastic wall. We also calculate the coefficient of tangential restitution.

## 1 Introduction

The collision of particles with internal degrees of freedom is inelastic in general. Examples can be seen in collisions of atoms, molecules, elastic materials, balls in sports, and so on. The study of inelastic collisions will be able to be widely accepted as one of the fundamental subjects in physics, because they are almost always discussed in textbooks of elementary classical mechanics.

Physicists realize that inelastic collisions can be a fashionable subject in physics from recent extensive interest in granular materials [1,2]. In fact, granules consist of macroscopic dissipative particles. Therefore, the description of interaction among particles is obviously important. We believe that static interactions among granular particles can be described by the theory of elasticity [3,4]. For example, the normal compression may be described by the Hertzian contact force [5] and the shear force may be represented by the Mindlin force [6]. The dynamical part related to the dissipation, however, cannot be described by any reliable physical theory. Thus, the distinct element method [7] which is one of the most popular models to simulate collections of granular particles contains some dynamical undetermined parameters. In other words, to determine such parameters, is important for both granular physics and fundamental physics.

When elastic materials collide each other, a part of translational energy is transformed into various modes of vibrations. In our previous papers, we introduced two different numerical models to investigate the impact of an elastic disk with a structureless wall [8,9]. One of them is the lattice model consisted of mass particles and linear springs. Another is the continuum model which is identical to that by Gerl and Zippelius [10]. With these two models, we investigated the relation between the impact velocity and the coefficient of normal restitution, etc [8,9]. These two models have common features in that (i) they have no dissipative mechanism and (ii) the wall has no internal degrees of freedom. In actual situations of collision between a disk and a wall, however, a part of initial energy of the material is transferred into the wall. Thus we have to introduce the internal degrees of freedom of the wall to simulate impacts between the disk and the wall.



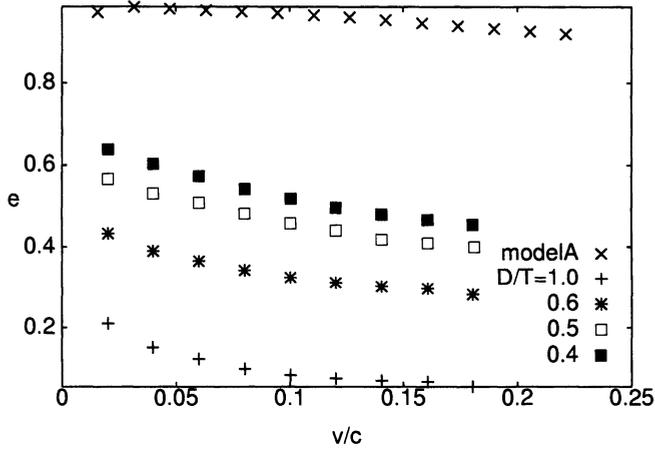
**Fig. 1.** The interaction between the mass points of the disk and the mass points of the wall.

To understand the macroscopic feature of impact phenomenon, two known parameters are introduced: the coefficient of normal restitution  $e$  and the coefficient of tangential restitution  $\beta$ . As for the normal impact, Sondergaard *et. al.* demonstrated that the coefficient of restitution depends on the thickness of the wall [11]. On the other hand, in the case of the oblique impact of the sphere, the coefficient of tangential restitution depends on the incident angle [12]. They also reported that there are two types of oblique collisions, slips and rotations.

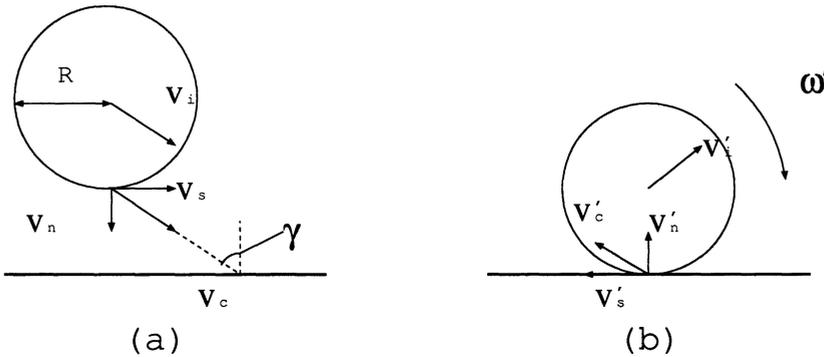
In this paper, we perform simulations of the normal and the oblique impacts of an elastic disk with an elastic wall. In the next section, we will introduce our two-dimensional model. In section 3, we will show the results of our simulation. In the last section, we will discuss and summarize our results.

## 2 Our Model

Let us introduce our numerical model. We construct an elastic disk of 1459 mass particles on a triangular lattice. In the continuum limit, Young's modulus  $E$  and Poisson's ratio  $\sigma$  correspond to  $2\kappa/\sqrt{3}$  and  $1/3$  respectively, where  $\kappa$  is the spring constant. We also introduce the elastic wall of 6400 mass particles on a square lattice. We introduce two spring constants:  $k_1$  for nearest neighbor interaction and  $k_2$  for next-nearest neighbor interaction for the wall. Controlling  $k_1$  and  $k_2$ , we can change the material constant of the elastic wall [13]. We set  $k_1$  and  $k_2$  as the Young's modulus of the elastic wall is equal to that of the elastic disk when the force is applied to the wall in the normal direction. The width of the wall is 4 times as long as the diameter of the disk. Two sides of the wall are fixed. We can choose any thickness of the wall. The interaction between the disk and the wall during a collision is introduced as follows. Figure 1. is the schematic figure of the interaction of mass points of the disk and the wall. When the distance  $l$  between the edge of the disk and the surface of the wall is less than the cutoff length (we set it equal to the length of the linear spring),



**Fig. 2.** The relation between the normal impact velocity of the disk and the coefficient of restitution. Each curve has different value of  $D/T$ . The curve labeled “model A” is the result in the situation when the wall has no internal degrees of freedom [9].



**Fig. 3.** The geometry of a collision to the wall. (a) the image before collision and (b) the image after collision. All the parameters after collision are denoted by primes.

where  $v'_c$  and  $t$  are the velocity of the contact point after collision and the unit vector in the tangential direction respectively. It is known that  $\beta$  is a function of the angle of incidence  $\gamma$ , with possible values lying in the range between -1 and 1 [14,15]. The incident angle  $\gamma$  is defined as  $\gamma = \arctan(v_s/v_n)$ , where  $v_n$  and  $v_s$  are the normal component and the tangential component of the initial velocity of the center of mass respectively.

In our simulation, we control the incident angle from  $5.7^\circ$  to  $63.4^\circ$  and let the disk collide with the wall without any rotation. When controlling the incident angle, we set  $v_n = 0.03c$  as a constant value and give  $v_s$  various values. The thickness of the wall is equal to the diameter of the disk (consisted of 6400 mass particles). Two sides of the wall are fixed. We will calculate  $\beta$  for various  $\gamma$  with

the points of the disk feel the repulsive force,  $\mathbf{F}(l) = aV_0 \exp(-al)\hat{\mathbf{n}}$ , where  $a$  is  $100/d_0$ ,  $V_0$  is  $amc^2d_0/2$ ,  $m$  is the mass of the particle,  $d_0$  is the lattice constant,  $c = \sqrt{E/\rho}$ , and  $\rho$  is the density.  $\hat{\mathbf{n}}$  is the unit vector perpendicular to the line between the two surface points. The reaction forces applied to the two points of the surface of the wall (point 1 and 2) are decided by the balance of the torques as  $\mathbf{F}_1(l) = -F(l)\hat{\mathbf{n}}/(1+l_1/l_2)$  and  $\mathbf{F}_2(l) = -F(l)\hat{\mathbf{n}}/(1+l_2/l_1)$ , where  $l_i$  ( $i = 1, 2$ ) is the distance between the point  $p$  and the point  $i$  (see Fig. 1).

To perform the numerical simulation, we scale all parameters by the radius of the disk  $R$  and the speed of sound  $c = \sqrt{E/\rho}$ , where  $\rho$  is the density of the disk. We adopted the fourth order symplectic numerical method with the time step  $\Delta t = 10^{-2}R/c$  ( $R$  is the radius of the disk) as the numerical scheme of the integration.

### 3 Simulation

Sondergaard *et al.* investigated the influence of the ratio of the diameter of sphere to the plate thickness on the coefficient of restitution resulting in normal impacts [11]. They dropped ball bearings (steel or bronze) and glass spheres without any significant rotational velocity from various heights onto plates of lucite or aluminum of which two sides were solidly clamped to a relatively rigid support structure. From their experiment, the coefficient of restitution decreases as the ratio of the particle diameter to the plate thickness increase.

At first, we will carry out the simulation of the normal impact. We set the situation corresponding to the experiment by Sondergaard *et al.* [11]. Our situation is as follows. We control the ratio of the disk diameter  $D$  to the wall thickness  $T$ ,  $D/T$ , from 0.4 to 1.0. We make the disk hit the wall with various initial velocity ranging from  $0.02c$  to  $0.18c$  and calculate the coefficient of restitution (Fig. 2). At the same impact velocity, the coefficient of restitution of small  $D/T$  is larger than that of large  $D/T$ . The curve labeled "model A" is the result of the case that the lattice model of the disk and the wall have no internal degrees of freedom [9].

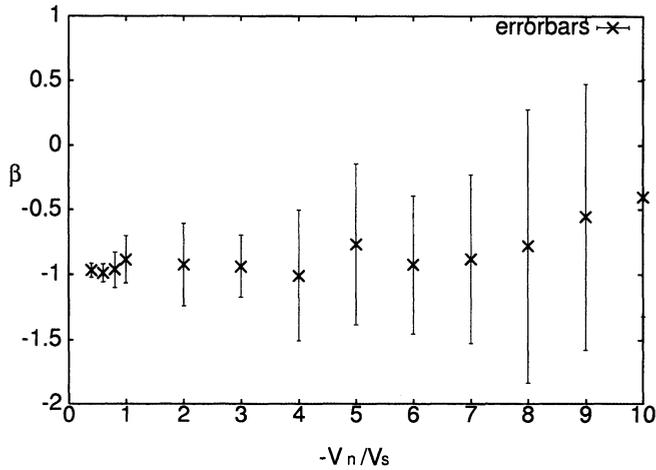
Second, we carry out the simulation of the oblique impact. Many experiments of the oblique impact focuses on the relation between the coefficient of tangential restitution and the incident angle.

Now let us introduce the coefficient of tangential restitution. We will consider the situation like Fig. 3. The coefficient of tangential restitution is defined by the ratio of the tangential component of the velocity of the contact point of the disk after collision to that before collision. The velocity of the contact point  $\mathbf{v}_c$  is defined as

$$\mathbf{v}_c = \mathbf{v}_i - \mathbf{R} \times \boldsymbol{\omega}, \quad (1)$$

where  $\mathbf{v}_i$  is the velocity of the center of mass of the disk,  $R$  is the radius of the disk, and  $\boldsymbol{\omega}$  is the angular velocity. Using the velocity of the contact point, the coefficient of tangential restitution  $\beta$  is defined as

$$\mathbf{v}_c' \cdot \mathbf{t} = -\beta \mathbf{v}_c \cdot \mathbf{t}, \quad (2)$$



**Fig. 4.** The relation between the incident angle and the coefficient of tangential restitution  $\beta$ .

this model. To eliminate the anisotropy of the disk, we set the initial angle of the disk  $\theta = \pi n/72$  ( $n = 0, 1, 2, \dots, 11$ ) as the initial state and average these 12 data.

The relation between  $\cot \gamma = -v_n/v_s$  and the coefficient of tangential restitution  $\beta$  is shown in Fig. 4. The cross points and the error bars represent the mean values and the standard deviations of the 12 data respectively. From this result,  $\beta$  is almost a constant negative value in the case of this model.

## 4 Discussion

We have performed both the normal and the oblique impact of the elastic disk with the wall. In the case of the normal impact, the coefficient of restitution  $e$  depends on the ratio of the diameter of the disk to the thickness of the wall  $D/T$ . If  $D/T$  is large,  $e$  is small. It is seen that the macroscopic vibration of the wall is so large that the initial energy of the disk decreases. Sondergaard *et. al.* introduced the impact parameter made by rescaling  $D/T$  and showed that the coefficient of restitution is a function only of the impact parameter [11]. In our case, however, we could not rescale our data in order to get an universal relation between  $D/T$  and  $e$ . To clarify the reason will be one of future subjects to be solved.

In the case of the oblique impact, we investigated the motion of the disk of each incident angle. According to Labous *et. al.* [12],  $\beta$  is believed to be represented by

$$\beta = \begin{cases} -1 - \mu_0 (1 + 1/K) (1 + e) v_n/v_s & (\gamma \geq \gamma_0) \\ \beta_0 & (\gamma \leq \gamma_0), \end{cases} \quad (3)$$

where  $K = 4I/md^2$  is a constant which is equal to  $1/2$  for disks ( $I$  is the moment of inertia,  $m$  is the mass,  $d$  is the diameter) and  $\mu_0$  is the friction coefficient and  $e$  is the coefficient of restitution. If  $\gamma$  is larger than the critical angle  $\gamma_0$ , the disk slips on the surface of the wall according to the Coulomb's law. In that case,  $\beta$  is a function of the incident angle while  $\mu_0$  is given by the ratio of the tangential to the normal component. If  $\gamma$  is smaller than the critical angle  $\gamma_0$ , the disk begins to roll after the collision with a constant value  $\beta_0$  as  $\beta$ . As for our model, the coefficient of tangential restitution  $\beta$  has a negative value in the range of incident angle  $5.7^\circ \leq \gamma \leq 63.4^\circ$ . In an actual case, Labous *et. al.* showed that  $\beta$  took a positive constant value  $\beta_0 = 0.5 \pm 0.1$  when the incident angle was greater than the critical angle (around  $37.5^\circ$ ) [12]. This means that the spheres begin to rotate with a rotational velocity of the same order as the velocity of the center of mass after the binary collision. In our case, however, the negative value of  $\beta_0$  means that a collision of the disk and the wall almost could not produce rotational motion of the disk, but the collision is slippery. Because the surfaces of the disk and the wall have no roughness, it can be seen that slippery collision is caused by the weak frictional force acting on the disk. We also need to consider initial rotation of the disk. How initial rotation affects the rotational motion of the disk after collision is an interesting problem. These are our near future tasks.

The results of this paper are as follows: (i) The coefficient of restitution depends on the ratio of the disk to the thickness of the wall. (ii) In the case of the oblique impact, the coefficient of tangential restitution  $\beta$  takes almost a negative constant value.

## References

1. L.P. Kadanoff: *Rev. Mod. Phys.* **71**, 435 (1999).
2. P.G. de Gennes: *ibid*, S367 (1999), and references therein.
3. A.E.H. Love: *A Treatise on the Mathematical Theory of Elasticity*, (Cambridge Univ. Press, Cambridge 1927).
4. L.D. Landau and E.M. Lifshitz: *Theory of Elasticity*, (2nd English ed.), (Pergamon, New York 1960).
5. H. Hertz: *J. Reine Angew. Math.* **92**, 156 (1882a).
6. R.D. Mindlin: *J. Appl. Mech.*, *Trans. ASME* **16**, 259 (1949).
7. P.A. Cundall and O.D.L. Strack: *Géotechnique* **29**, 47 (1979).
8. H. Kuninaka and H. Hayakawa: *J. of Phys. Soc. of Jpn.* **70**, 2220 (2001).
9. H. Hayakawa and H. Kuninaka: *Chem. Eng. Sci.* **57**, 239 (2002).
10. F. Gerl and A. Zippelius: *Phys. Rev. E* **59**, 2361 (1999).
11. R. Sondergaard, K. Chaney, and C.E. Brennen: *J. of Appl. Mech.*, *Trans. ASME* **57**, 694 (1990).
12. See for example: L. Labous, A.D. Rosato, and R.N. Dave: *Phys. Rev. E* **56**, 5717 (1997); N. Maw, J.R. Barber, and J.N. Fawcett: *J. of Lub. Tech.*, *Trans. ASME* **103**, 74 (1981).
13. Y. Hayakawa: *Phys. Rev. E* **49**, R1804 (1994).
14. O.R. Walton and R.L. Braun: *J. Rheol.* **30**, 949 (1986).
15. L. Labous, A.D. Rosato, and R.N. Dave: *Phys. Rev. E* **56**, 5717 (1997).

# Granular Flow in Vertical Pipes: Transition from Dilute to Dense

M. Hou, W. Chen, T. Zhang, and K. Lu

Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China.

**Abstract.** The transition from dilute flow to dense flow of a granular system in a vertical pipe, controlled by an electric field  $E$ , is studied. Depending upon its initial state and the applied field voltage the controlled flow rate remains in two phases, dilute flow or dense flow. For dilute flow, the electric field has no effect on the flow rate until  $V$  reaches a critical value  $V_1$ . At  $V = V_1$ , the flow rate drops abruptly and a transition of the particulate from dilute to dense flow occurs. For dense flow, the flow rate decreases monotonically with increasing  $V$ . A comparison of the flow rate controlled by directly changing the opening size is given. Similar behavior is observed. A 3-dimensional phase diagram with outflow rate, inflow rate and opening size as its three axes is also presented.

## Introduction

Granular matter is a subject of intense interest [1-3] in recent years. In this field, many important topics in nonlinear physics [4,5] - such as pattern formation [3], solitons [6], chaos [7], and cellular automata [8] - were studied. In particular, nonlinear waves in granular flow have been observed and computationally simulated [9-17].

This discrete, compressible system has distinct features when compared with classical fluids. Density fluctuation is an important character of granular flow and has been broadly noticed. Many interesting phenomena related to it were observed in different experiments using x-ray imaging [17], light detector, and spatiotemporal diagrams [9,11]. Intermittent and kinematic shock wave [9] was found in a small-angle two-dimensional funnel when the funnel angle was changed. Different kinds of wave regimes [11] in the vertical pipe were observed when the mass-flow rate was changed by adjusting the stopcock at the bottom end of the tube. The power spectra of density waves were shown to assume a stable power-law form,  $P(f) \sim f^{-\alpha}$ , when the air outflow rate was controlled [12,15]. Jamming phenomenon of granular flow in a two-dimensional hopper was studied experimentally [18].

In this paper, a new granular flow control mechanism - by applying a local, ac electric field - is introduced to study the nickel particle flow in a vertical pipe attached to a hopper. Due to the dipole-dipole interaction induced by the electric field, particle clusters were formed near the field and the granular flow rate was reduced nonlinearly with the applied voltage.

For a comparison we have examined the outflow rate of a dilute and a dense flow with adjustable opening at the bottom of the pipe. Similar behavior is observed: for a dilute flow there exists an abrupt transition from dilute to dense flow at a certain critical opening size; for a dense flow the rate changes with the opening size monotonically.

The effect of the electric field is like to apply a virtual bottleneck to the flow. Adjusting the voltage of the field is the same as adjusting the opening size of the flow bottleneck. A 3-dimensional phase diagram is given to demonstrate the dilute-dense-jammed phases and the transition from dilute to dense.

## Experimental Setup

The experimental setup is schematically shown in Fig. 1. Nickel spheres of an average diameter  $d = 0.25$  mm are filled in a glass hopper placed on top of a glass pipe with inner diameter of 3mm and 100mm in length. The metallic particle is chosen for less electrostatic effect. Two parallel flat copper electrodes 4.4mm apart, 15 mm in height and 12 mm in width, are attached to the outside walls of glass pipe. A 50 Hz ac electric voltage  $V$  is applied across the electrodes. The granular flow is initiated by pulling the stopper, which is inserted at the outlet of the hopper. A weighing sensor with sensitivity of 0.02g and recording rate of 0.02s is applied at the bottom of the pipe. By measuring the granular mass at each desired electric voltage, the dependence of flow rate vs. voltage can be obtained.

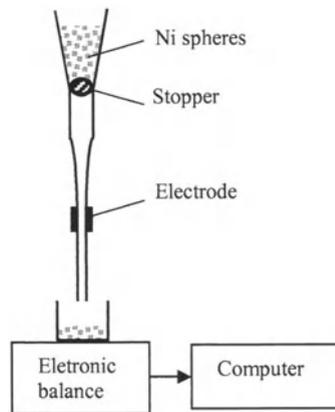
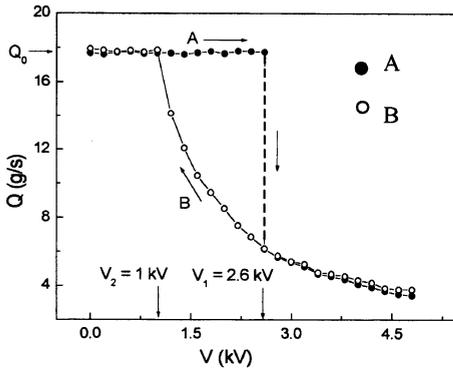


Fig. 1. Schematic diagram of E-field control setup.

## Experimental Results

In our experiment we change the applied voltage in steps to see the voltage influence on the granular flow. The following two processes are applied:

In the first process (denoted as A), we measure the flow rate at each desired voltage in a range from 0 V to 4.8 kV with an interval  $\Delta V = 0.2$  kV. In the second process (denoted as B), V is first fixed at a voltage 4.8 kV and then lowered to a desired voltage from 4.8 kV to 0 V with an interval of 0.2 kV. The flow rate Q vs. V is plotted in Fig. 2. It can be seen that in process A, the granular flow keeps a constant value  $Q_0$  with increasing V until  $V = V_1 = 2.6$  kV. The flow rate Q drops abruptly when we increase the voltage by a small amount (as small as 0.02 kV) higher than  $V_1$ . When we further increase the voltage, Q decreases gradually with increasing V. In process B when lowering the voltage, however, Q increases continuously without any sudden jump and reaches  $Q_0$  at a voltage  $V_2$  (1.0 kV). In the two processes the flow rates follow the same Q(V) curve in the region between  $V_1$  and 4.8 kV. The curve can be fitted approximately as  $Q = CV^{-1}$ , where C is a fitting parameter.

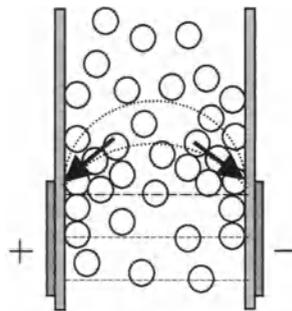


**Fig. 2.** Flow rate Q vs. the applied voltage V. A and B indicate different processes. The arrows represent the direction of applied voltage.

To understand the underlying mechanism, an analysis is proposed as follows: We first consider the case of process A. The force acting on the polarized particle in an inhomogeneous electric field can be expressed as

$$\vec{f} = -\nabla(-\vec{p} \cdot \vec{E}) = 4\pi\epsilon_0 R^3 \nabla E^2.$$

When the electric field is high enough to affect the motion of the particles, i.e. the interaction is comparable to the kinetic energy of the falling particles, the particles will be dragged by the gradient of the field and converge to the sidewall near the top of the electrodes as shown in Fig. 3. The arrows in Fig. 3 represent the direction of the dragging force. This force against the sidewall induces a friction between particle and wall, which will reduce the velocity of particles along the wall. The attraction of polarized particles cause the particles



**Fig. 3.** Sketch of the force acting on particles due to the field gradient in the region of electrodes. Dashed lines indicate electric lines in between electrodes.

moving closer and enlarges the local field between particles. The enhancement of the local field between the particles will further increase their mutual induction and cause particles to interact with each other and aggregate stronger. This positive-feedback effect, we believe, is the major contribution which causes the flow rate dropping dramatically at  $V_1$ . In fact there is a transition from dilute to dense flow occurring at  $V_1$ . When  $V < V_1$  the remaining opening at the central area of converged particles in the pipe is large enough for the particles passing through as a dilute state flow. The flow remains at a constant rate  $Q = Q_0 = \rho v S$ , where  $\rho$  and  $v$  are the average density and average velocity of particles, respectively, and  $S$  is the effective cross-sectional area. Although the  $v$  and  $S$  are reduced due to the electric field induced converging effect, the increase in  $\rho$  is able to assure  $Q (= \rho v S)$  to remain a constant  $Q_0$  until  $V$  reaches a critical voltage  $V_1$ . At this voltage the density  $\rho$  reaches its maximum value, and it becomes dense flow for  $V > V_1$ .

In process B, the flow rate vs. voltage shows a power law decay vs.  $V$  as described above,  $Q = CV^{-\beta}$  and  $\beta \approx 1$  in our experimental condition. Although the mechanism is still not yet clear [15], the rate of dense particle flow from a hopper can be expressed as  $Q \approx aD^{5/2}$ , where  $D$  is the diameter of the outlet. Therefore, the electric field plays a role here as an effective opening in the case of dense flow, i.e. increasing voltage corresponds to decreasing the size of the opening.

Therefore, by applying the electric field higher than a critical voltage the initial dilute flow can be transferred to a dense flow, corresponding to an abrupt drop of flow rate. This phenomenon may be similar to that of a first order phase transition. A quantitative explanation is still absent.

## Comparison with Opening-Size Controlling Method

In order to compare the electric-field controlling method with the adjustable opening-size controlling method, we have established an experiment of granular flow in a vertical pipe with controlled exit opening size. The experiment is done with glass particles of density =  $2.4 \text{ g/cm}^3$ . The experimental setup is similar to that of the E-field control, but with the pipe length = 62 cm and the inner diameter = 9.3 mm as shown in Fig. 4. A Copper plate with center opening of different sizes (from 3.0 mm to 9.4 mm) is placed at the exit of the pipe. The hopper is initially filled with glass particles of diameter  $d = 0.5 \text{ mm}$ . A dilute flow is initiated by pulling out vertically a stopper inserted at the hopper outlet and a dense flow is initiated by pulling a thin plate inserted at the end of the pipe.

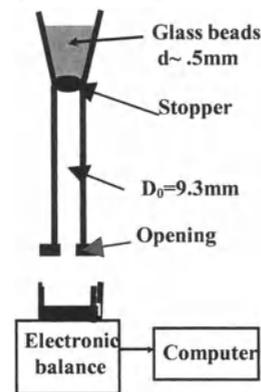


Fig. 4. scheme of opening-size control setup.

For a fixed inflow rate of a dilute flow, the measured outflow rate  $Q$  ( $= dM/dt$ ) at different exit opening sizes  $D$  is found, as shown in Fig. 5, to be practically independent of the width of the exit when the exit size is larger than some critical size  $D_c$ . At this critical opening size, the dilute flow turns to dense flow, and the flow rate drops abruptly. When the opening size is less than  $D_c$ , the flow becomes dense and  $Q$  decreases monotonically with the reducing size of the opening. The flow stops when the opening size is about three particle diameter wide. Here permanent arching occurs and causes jamming of the flow [18].

The critical opening size at which the transition from dilute to dense occurs is different for different inflow rate  $Q_0$ . As  $Q_0$  is lower, the critical size  $D_c$  is smaller. The  $Q(D)$  curves for different  $Q_0$  are shown with dashed lines in Fig. 5. The A-B curve represents a boundary line at which transition from dilute to dense flow begins for different  $Q_0$ .

For a fixed inflow rate of a dense flow, the outflow rate increases gradually with increasing  $D$ . The dense flow rate  $Q_d$  is along the curve from D to C in Fig. 5. There is no abrupt change in outflow rate as it appeared in the dilute flow.

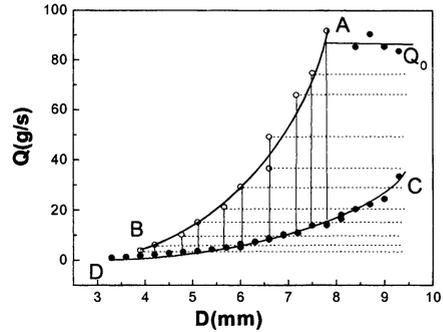


Fig. 5. The  $Q$  vs.  $D$  curves.

### The 3-D Phase Diagram

To obtain a clear picture of the dilute flow behavior, a 3-dimensional phase diagram is drawn in Fig.6. In the diagram it shows the two phases: dilute (or free) flow and dense (or congested) flow, and the transition from dilute flow to dense flow. At a given bottleneck the outflow rate is practically the same as  $Q_0$  until  $Q_0$  is sufficiently large and reaches the critical flow rate  $Q_c$  that dilute flow no longer holds and transition from dilute to dense occurs and the flow rate drops to the dense flow rate  $Q_d$ . The ratio of  $Q_0$  and  $Q_d$  depends on the inflow rate.

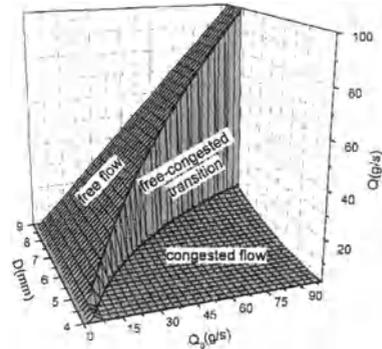


Fig. 6. The 3-D phase diagram.

When the inflow rate is higher, the ratio of  $Q_0$  and  $Q_d$  is larger. Therefore one way to increase the outflow rate at the bottleneck is to keep the flow dilute and decrease inflow rate a bit lower than  $Q_c$  at the exit.

## Conclusions

By applying an electric field to the granular flow in a vertical pipe we have observed some novel phenomena. Applying an electric field to a dilute flow, we find that the flow rate is unaffected by  $V$  and remains as  $Q_0$  till  $V$  reaches a critical value  $V_1$ . At  $V = V_1$  the flow rate drops abruptly, and decreases with increasing  $V$ . In a reverse process, when  $V$  decreases from a value  $V > V_1$ , the flow rate increases continuously with decreasing  $V$  and reaches  $Q_0$  at  $V_2$ . At  $V_1$  the transition from thin flow to dense flow occurs, which is similar to a first order phase transition. The observed flow behavior is comparable to opening-size controlling flow system. A 3-D phase diagram concludes the relation of the outflow rate with the inflow rate and the bottleneck size.

## References

1. L.P. Kadanoff, *Rev. Mod. Phys.* **71**, 435 (1999); P.G. de Gennes, *Rev. Mod. Phys.* **71**, S374 (1999); H.M. Jaeger, S.R. Nagel, and R.P. Behringer, *Rev. Mod. Phys.* **68**, 1259 (1996).
2. J. Duran, *Sands, Powders, and Grains*, (Springer, New York, 2000).
3. G.H. Ristow, *Pattern Formation in Granular Materials*, (Springer, New York, 2000).
4. *Introduction to Nonlinear Physics*, edited by L. Lam, (Springer, New York, 1997).
5. L. Lam, *Nonlinear Physics for Beginners*, (World Scientific, River Edge, NJ, 1998).
6. J. Hong, H. Kim, and J.-P. Hwang, *Phys. Rev. E* **61**, 964 (2000).
7. T. Shinbrot, A. Alexander, M. Moakher, and F.J. Muzzio, *Chaos* **9**, 611 (1999).
8. G. Peng and H.J. Herrmann, *Phys. Rev. E* **51**, 1745 (1995); T. Yanagita, *Phys. Rev. Lett.* **82**, 3488 (1999); L.C. Jia, P.-Y. Lai, and C.K. Chan, *Phys. Rev. Lett.* **83**, 3832 (1999).
9. S. Horluck and P. Dimon, *Phys. Rev. E* **60**, 671 (1999).
10. C. Denniston and H. Li, *Phys. Rev. E* **59**, 3289 (1999).
11. J.-L. Aider, N. Sommier, T. Raafat, and J.-P. Hulin, *Phys. Rev. E* **59**, 778 (1999).
12. Moriyama, N. Kuroiwa, M. Matsushita, and H. Hayakawa, *Phys. Rev. Lett.* **80**, 2833 (1998).
13. Moriyama, T. Isoda, N. Kuroiwa, M. Kanda, I. Rafols, and M. Matsushita, *J. Phys. Soc. Jpn.* **67**, 1616 (1998).
14. A. Nakahara and T. Isoda, *Phys. Rev. E* **55**, 4264 (1997).
15. T. Raafat, J.P. Hulin, and H.J. Herrmann, *Phys. Rev. E* **53**, 4345 (1996).
16. J. Lee, *Phys. Rev. E* **49**, 281 (1994).
17. G.W. Baxter, R.P. Behringer, T. Fagert, and G.A. Johnson, *Phys. Rev. Lett.* **62**, 2825 (1989).
18. K. To, P.-Y. Lai, and H.K. Pak, *Phys. Rev. Lett.* **86**, 71 (2001).

# Spatial Structure of $1/f$ Noise in Granular Flow Through a Pipe

A. Nakahara<sup>1</sup> and M. Nakahara<sup>2</sup>

<sup>1</sup> Laboratory of Physics, College of Science and Technology, Nihon University, Funabashi 274-8501, Japan

<sup>2</sup> Institute of Material Sciences, Tsukuba University, Tsukuba 305-8573, Japan

**Abstract.** We experimentally studied the spatio-temporal structure of non-uniform flow of granular materials which fall through a vertical pipe filled with liquid. As we increase the packing fraction of granular materials inside the pipe, the flow pattern changes from the low-packing free-fall flow to the high-packing slugging flow. At the intermediate packing fraction between the free-fall flow and the slugging flow, slugs come to appear, but since all these slugs are unstable, they emerge and disappear frequently. We find that the power spectrum  $P(f)$  of the density fluctuations with frequency  $f$  shows  $1/f$  noise at the intermediate packing fraction. As for the spatial structure, however, the structure factor  $S(k)$  with a wave number  $k$  does not obey a power law even at this intermediate packing fraction. These results indicate that  $1/f$  noise in pipe flow is not induced by the spatio-temporal criticality.

## 1 Introduction

The physics of  $1/f$  noise has attracted researchers in various fields, because it appears in most energy-dissipating systems, such as voltage or current fluctuations in electronic devices [2], density fluctuations in granular flow through a hopper and a pipe [10,4,3], and vehicle density fluctuation in traffic jams on highways [5]. The mechanism to produce  $1/f$  noise is still in hot debate, and the point is whether  $1/f$  noise is a kind of universal phenomena which does not depend on the precise value of physical parameters.

As an example to study the mechanism of  $1/f$  noise, we perform experiments on density fluctuations of granular flow through a pipe. When we pour granular materials through a vertical glass pipe, the falling materials form slugs of slow clusters. The dominant mechanism to produce density fluctuations is considered to be inelastic interactions due to both grain-grain collisions and grain-wall friction. Also the hydrodynamic interaction between grains and the surrounding fluid is important. To simplify the system, we use smooth metallic spheres with same size. We fill the pipe with liquid such as water and silicone oil, so that we can control the flow and the viscosity of the surrounding fluid easily [8].

As a control parameter in our experiments, we decide to use the mean packing fraction  $p$  of granular materials inside the pipe. The idea of packing fraction comes from the traffic flow problem, because, in traffic flows, the dominant parameter

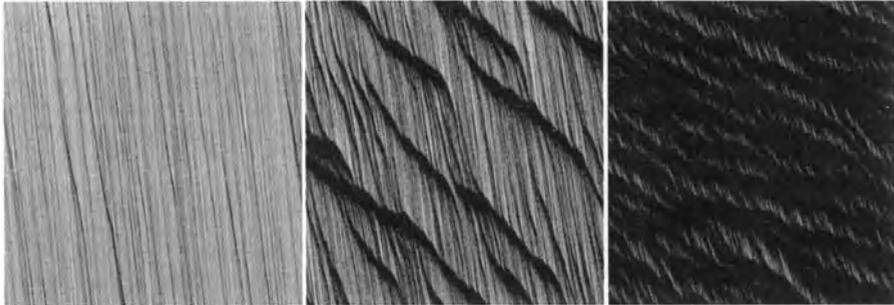
which governs the system is a number density of vehicles on the lane, and increasing vehicle density induces the jamming transition from free to jammed flows [6]. Considering that vehicles correspond to falling granular materials, we can expect that the jamming transition of traffic flow and the slugging transition in pipe flow are based on the same mechanism.

## 2 Flow Patterns and Fundamental Diagram

As metallic spheres, we use lead spheres with 1.7 mm in diameter, and a glass pipe with 9.0 mm in inner diameter. We systematically vary the mean packing fraction  $p$  by changing the inner diameter of the additional hopper which is set at the top of the vertical pipe.

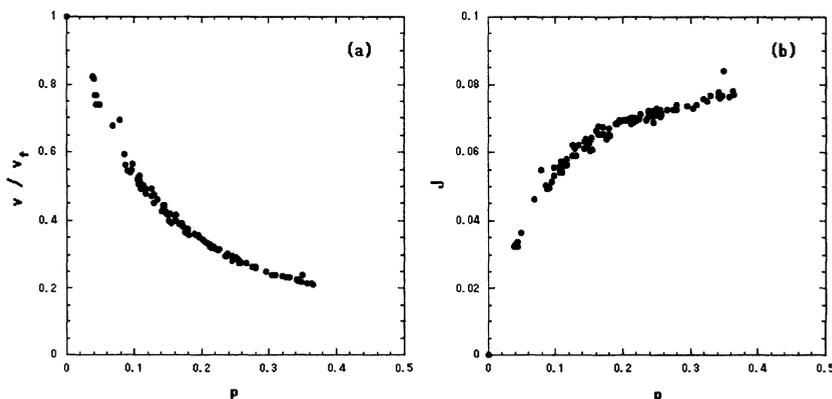
Figure 1 shows the spatio-temporal structure of flow patterns as we change the value of the mean packing fraction  $p$  from 0.05 in (a), 0.18 in (b), to 0.36 in (c). Here,  $t$  is the time, and  $x$  is the vertical distance down from the hopper. In these figures, the vertical direction is space, and metallic spheres fall downwards due to the gravity. The horizontal direction is time, and time is running from left to right.

In the low-packing-fraction region ( $p < 0.18$ ) as is shown in Fig. 1(a), the metallic spheres fall almost freely. On the other hand, in the high-packing-fraction region ( $p > 0.18$ ) as is shown in Fig. 1(c), the metallic spheres form slugs of large clusters and fall very slowly in groups. Note that at the intermediate packing fraction  $p = 0.18$  in Fig. 1(b), all these slugs are very unstable, they emerge and disappear frequently, and thus the flow of metallic spheres is rather intermittent.



**Fig. 1.** Spatio-temporal structure of flow patterns as functions of  $x$  and  $t$ .  $x$  is the distance down from the hopper and is shown as the vertical direction downwards.  $t$  is the time and is shown as the horizontal direction from left to right. So, each metallic sphere moves from upper-left to lower-right due to the gravity. The value of the mean packing fraction  $p$  in each figure is, from left to right, 0.05 in (a), 0.18 in (b), and 0.36 in (c), respectively. The size of each figure is  $\Delta x = 300$  mm in  $x$ -direction (downwards) and  $\Delta t = 8$  s in  $t$ -direction (from left to right).

Next, we measure the mean velocity  $v$  of the falling metallic spheres and the flux  $J$  as functions of the mean packing fraction  $p$ , and present these result in Fig. 2(a) and 2(b), respectively. In Fig. 2(a), the velocity  $v$  is normalized by the free-fall velocity  $v_f$  of a single metallic sphere. We find that the mean velocity  $v$  decreases monotonically as we increase the value of  $p$ . Fig. 2(b) represents a fundamental diagram. The flux  $J$  increases monotonically with a bend at around  $p=0.18$ , showing a crossover from the low-density free-flow to the high-density jammed flow.



**Fig. 2.** Normalized velocity  $v/v_f$  in (a) and the flux  $J$  in (b) as functions of the mean packing fraction  $p$ .

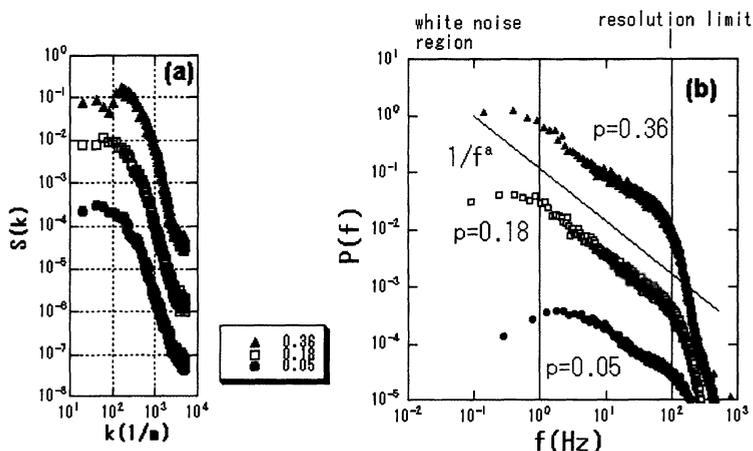
### 3 Structure Factor and Power Spectrum

First, we perform a fast Fourier transform (FFT) analysis of the spatial data of the pipe flow. Fig. 3(a) represents the structure factor  $S(k)$  of the density fluctuations as a function of the wave number  $k$ . In the high-packing slugging region ( $p > 0.18$ ) the structure factor  $S(k)$  has a peak at a characteristic wave number of about  $150 \text{ m}^{-1}$ , which corresponds to a mean distance between slugs and/or a mean size of slugs. On the other hand, since there is no macroscopic structure at the low-packing free-fall region ( $p < 0.18$ ), the functional form of the structure factor does not have any tendencies, such as a peak nor a power law.

Next, we perform a FFT analysis of the temporal data. Fig. 3(b) shows the power spectrum  $P(f)$  of the density fluctuations as a function of frequency  $f$ . The vertical guide line at  $f=100 \text{ Hz}$  is the resolution limit due to the finite size of the metallic spheres. The other guide line at  $0.5 \text{ Hz}$  denotes that in the lower-frequency region on the left side of this guide line all these power spectra have a white-noise form. Thus, we must study the functional form of the power spectra between these two vertical guide lines.

In the low-packing free-fall region ( $p < 0.18$ ) the functional form of the power spectrum shows a plateau of white noise at around  $f = 1$  Hz and does not have any tendencies. However, in the high-packing slugging region ( $p > 0.18$ ) the power spectrum is a monotonically decreasing function with a bend at 3 Hz. This bend at 3 Hz in the power spectrum  $P(f)$  is attributed to the formation of macroscopic structures such as slugs, because we can estimate the characteristic frequency of the slug motion by taking into account the size and the velocity of slugs.

We also note that, at the intermediate packing fraction  $p = 0.18$ , the power spectrum  $P(f)$  obeys a power law as  $P(f) = 1/f^\alpha$  with  $\alpha = 0.94$ . On the other hand, the structure factor  $S(k)$  of this  $1/f$  noise does not obey a power law.



**Fig. 3.** FFT analysis of the density fluctuations in pipe flow. The value of the mean packing fraction  $p$  is, from bottom to top,  $p = 0.05$ ,  $0.18$ , and  $0.36$ . (a) Structure factor  $S(k)$  as a function of wave number  $k$ . (b) Power spectrum as a function of frequency  $f$ .

## 4 Conclusions

We have performed experiments on pipe flow by changing the value of the mean packing fraction of granular materials inside the pipe. As we increase the packing fraction, the flow pattern changes from the low-packing free-flow to the high-packing slugging flow. We also obtained the fundamental diagram of pipe flow, and compared our result with that of traffic flow. Next, we performed signal and image analysis of the density fluctuations of pipe flow, and found that  $1/f$  noise appears only at the intermediate packing fraction. On the other hand, the spatial structure of  $1/f$  noise does not show any criticality, as was reported on  $1/f$  noise in voltage fluctuations of metal films [1]. These results indicate that  $1/f$  noise in pipe flow is not induced by the spatio-temporal criticality.

## References

1. R.D. Black, M.B Weissman, and F.M. Fliegel,  $1/f$  noise in metal films lacks spatial correlation, *Phys. Rev. B* **24**, 7454-7456 (1981).
2. C. Claeys and E. Simoen, *Noise in physical systems and  $1/f$  fluctuations*, (World Scientific, Singapore, 1997).
3. T. Kawaguchi, T. Tanaka, and Y. Tsuji, Numerical analysis of density wave in dense gas-solid flows in a vertical pipe, *Prog. Theor. Phys. Suppl.* **138**, 696-701 (2000).
4. O. Moriyama, N. Kuroiwa, M. Matsushita, and H. Hayakawa,  $4/3$  law of granular particles flowing through a vertical pipe, *Phys. Rev. Lett.* **80**, 2833-2836 (1998).
5. T. Musha and H. Higuchi, The  $1/f$  fluctuation of a traffic current on an expressway, *Jpn. J. Appl. Phys.* **15**, 1271-1275 (1976).
6. K. Nagel and M. Schreckenberg, A cellular automaton model for freeway traffic, *J. Phys. I France* **2**, 2221-2229 (1992).
7. A. Nakahara, Spatial structure of granular flow through a pipe, in: M. Tokuyama and H.E. Stanley, (Eds), *AIP Conference Proceedings 519: Statistical Physics*, American Institute of Physics, pp. 182-184 (2000).
8. A. Nakahara and T Isoda,  $1/f^\alpha$  density fluctuation at the slugging transition point of granular flow through a pipe, *Phys. Rev. E* **55**, 4264-4273 (1997).
9. G. Peng and H.J. Herrmann, Density waves of granular flow in a pipe using lattice-gas automata, *Phys. Rev. E* **49**, R1796-R1780 (1994).
10. K.L. Schick and A.A. Verveen,  $1/f$  noise with a low frequency white noise limit, *Nature* **251**, 599-601 (1974).

# Collisional Granular Flow on a Rough Slope and its Instability

N. Mitarai<sup>1</sup>, H. Nakanishi<sup>1</sup>, and H. Hayakawa<sup>2</sup>

<sup>1</sup> Department of Physics, Kyushu University 33, Fukuoka 812-8581, Japan

<sup>2</sup> Graduate School of Human and Environmental Studies, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

**Abstract.** Granular flow on a rough slope in the collisional flow region is investigated by numerical simulations. It is demonstrated that the uniform flow is only metastable; it is maintained for a while, but fluctuations trigger clustering of particles eventually. The micropolar fluid model is employed to analyze the properties of uniform flow. It is shown that the model can reproduce the deviation of the angular velocity field from the rotation of the velocity field.

## 1 Introduction

Granular materials flowing down a slope is one of the simplest and the most familiar situations to see the characteristics of their dynamics [1]. When the inclination angle of the slope is relatively large, the interaction between the grains is dominated by collisions among particles, and many researches have been performed in such a collisional or rapid flow regime [2]. Most of them have concerned the properties of the flow which is uniform along the slope. On the other hand, it is known that the uniform granular flow often becomes unstable. In the case of the granular flow in a vertical pipe, for example, the instability causes density wave formation, and the power spectrum of the density fluctuation obeys a power law [3]. The formation of waves has also been found in dense granular flow down flat frictional slope [4]. Recently, the instability of the collisional granular flow down a rough slope has been observed experimentally; the free surface of the flow shows a regular deformation in the direction perpendicular to the flow direction [5].

In this paper, we numerically investigate the collisional flow down a slope [6]. It is found that the flow has an instability in the flow direction that causes clustering, and this instability depends on the system size and the density. It is also found that the angular velocity field systematically deviates from the rotation of the linear velocity field. The micropolar fluid model [7] is used to analyze the uniform flow, and it is shown that the model with the coefficients of viscosity estimated by the elementary kinetic theory can reproduce that deviation [8].

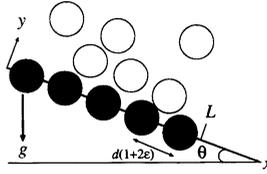


Fig. 1. The geometry of the simulation.

## 2 Results of Numerical Simulations

### 2.1 Simulation method

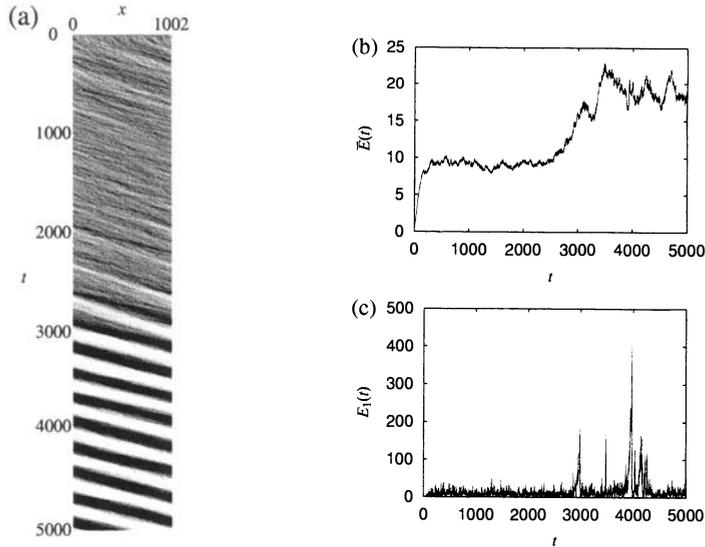
We employed the discrete element method (DEM) [9] with normal and tangential elastic forces and dissipation. Grains are modeled by two-dimensional disks with mass  $m$  and diameter  $\sigma$ . The system is two dimensional (Fig. 1), and each disk also has the rotational degree of freedom. Detailed description of the method is given in ref. [6]. Each particle is also subject to gravity, where the gravitational acceleration is given by  $\mathbf{g} = g(\sin \theta, -\cos \theta)$ . The surface of the slope is made rough by attaching particles identical to the moving ones with spacing  $2\epsilon\sigma$ , where  $\epsilon = 0.001$ . Periodic boundary conditions are imposed along the slope ( $x$ -direction). The system size  $L$  is determined by the number of particles attached to the slope  $n_s$  as  $L = (1 + 2\epsilon)\sigma n_s$ .

All quantities that appear in the following are given in dimensionless forms in terms of the length unit  $\sigma$ , the mass unit  $m$ , and the time unit  $\tau = \sqrt{\sigma/g}$ . The viscoelastic parameters in the DEM are chosen so that the coefficient of the normal restitution  $e_n$  becomes 0.7.

### 2.2 Results of Numerical Simulation: Instability of the Uniform Flow

Here we present simulation results with a fixed inclination angle,  $\sin \theta = 0.45$ . In Fig. 2, the data from the simulation with  $N = 1000$  and  $L = 1002$  are shown, where  $N$  is the number of flowing particles. Figure 2 (a) shows the spatiotemporal diagram of the local density along the slope; the vertical axis is the time  $t$ , the horizontal axis is the coordinate along the slope  $x$ , and the darker region represents the higher density region. The  $y$  coordinates of the particles are not taken into account in that diagram. The initial condition is chosen so that all the particles are distributed uniformly along the slope. We can see that the flow is uniform along the slope at the beginning, but then relatively large fluctuations of the density grow and finally the system separate into a high density region and a low density region, namely one large cluster is formed.

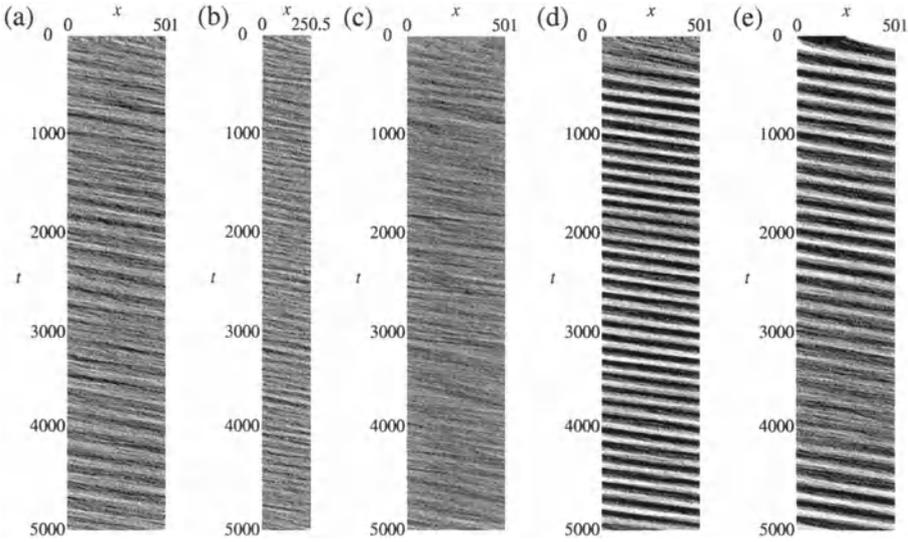
The time evolutions of the kinetic energy averaged over all particles,  $\bar{E}(t)$ , and that of one particular particle,  $E_1(t)$ , are also shown in Figs. 2 (b) and (c), respectively. We can see that, after the clustering,  $\bar{E}(t)$  grows and the fluctuation of  $E_1(t)$  becomes much larger. This is because the energy loss per unit time due



**Fig. 2.** (a) Spatiotemporal diagram of the local density. Darker regions indicate the higher density regions. (b) Time evolution of the averaged kinetic energy  $\bar{E}(t)$ . (c) Time evolution of the kinetic energy of one particular particle  $E_1(t)$ .

to inelastic collisions depends on the density. The particles near the front of the cluster escape to the low density region and get high kinetic energy, but they will be caught in the cluster again and lose the energy. This mechanism causes large fluctuation in  $E_1(t)$ , and because all the particles experiences the acceleration in the low density region,  $\bar{E}(t)$  becomes higher than the value in the uniform flow regime.

We also find that the instability depends on the system size  $L$  and the “density”  $\bar{\rho}$ . Here, we define the “density” as  $\bar{\rho} = N/L$ , namely the averaged number of particles per unit length along the slope. In Fig. 3, spatiotemporal diagrams of the local density with the different values of  $L$  and  $\bar{\rho}$  are shown. From the comparison of Fig. 2 (a), Figs. 3 (a) and (b), we can see that the instability becomes weaker for smaller system sizes. From Fig. 3 (a), (c), and (d), which are results with the same  $L$  but different  $\bar{\rho}$ , we find that the higher the density is, the more stable the uniform flow is. It can also be seen that, even though the clear cluster formation has not been found in Fig. 3 (a), the density fluctuation is quite large. Therefore we have also simulated the system with the same  $L$  and  $N$  as in Fig. 3 (a) but *with the initial configuration with large inhomogeneity along the flow direction* (Fig. 3 (e)). It is found that a cluster is formed after a while, but it almost disappears when the head and the tail of the cluster collide through the periodic boundary condition ( $3000 \leq t \leq 4000$ ).



**Fig. 3.** Spatiotemporal diagrams of local density with the different values of  $L$  and  $\bar{\rho}$ . The darkness is adjusted so that uniform distribution is shown by grey region in each figure. (a)  $N = 500$ ,  $L = 501$ ,  $\bar{\rho} = 1.0$ . (b)  $N = 250$ ,  $L = 250.5$ ,  $\bar{\rho} = 1.0$ . (c)  $N = 1000$ ,  $L = 501$ ,  $\bar{\rho} = 2.0$ . (d)  $N = 375$ ,  $L = 501$ ,  $\bar{\rho} = 0.75$ . (e)  $N = 500$ ,  $L = 501$ ,  $\bar{\rho} = 1.0$ , with the initial configuration with large inhomogeneity along the flow direction.

### 3 Properties of Uniform Flow

We expect that the angular velocity field deviates from  $\frac{1}{2}\nabla \times \mathbf{v}$  near the base because of the boundary effect. Since the flow is collisional, it should be described in terms of the kinetic theory, treating the angular rotation as an independent field. This leads to the micropolar fluid model, with which we examine the uniform flow. The detailed description of the model is given in ref. [8].

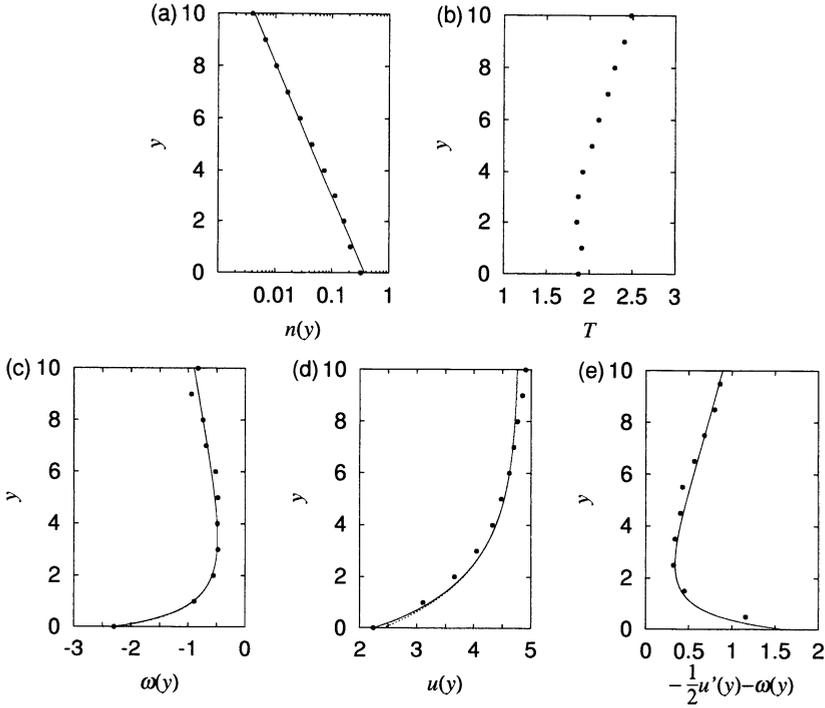
In the case of the two dimensional flow, the micropolar fluid model gives following equations under the assumption of a uniform steady flow, namely the mass density  $\rho = \rho(y)$ , the velocity  $\mathbf{v} = (u(y), 0, 0)$ , and the angular velocity  $\boldsymbol{\omega} = (0, 0, \omega(y))$ :

$$0 = \rho g \sin \theta - \frac{d}{dy} \left[ \mu \frac{du}{dy} + \mu_r \left( \frac{du}{dy} + 2\omega \right) \right], \quad (1)$$

$$0 = -\rho g \cos \theta - \frac{dp}{dy}, \quad (2)$$

$$0 = -2\mu_r \left( \frac{du}{dy} + 2\omega \right) + \frac{d}{dy} \left[ \mu_B \frac{d\omega}{dy} \right], \quad (3)$$

where  $p$  is the pressure. Here, eqs. (1) and (2) come from the equations of motion in the  $x$  and  $y$  direction, respectively. Equation (3) comes from the equation for the angular velocity. The second term in eq. (1) represents the force from the



**Fig. 4.** Properties of the uniform flow. The filled circles are obtained from the simulation data, and the solid line is the uniform steady solution of the micropolar fluid equations. (a) The number density profile  $n(y)$ . (b) The temperature profile  $T(y)$ . (c) The velocity profile  $u(y)$ . The dotted line represents the solution of Navier-Stokes equations. (d) The angular velocity profile  $\omega(y)$ . (e) The difference between the rotation of the bulk velocity field and the microrotation,  $\frac{1}{2}\nabla \times \mathbf{v} - \boldsymbol{\omega} = -\frac{1}{2}\partial_y u(y) - \omega(y)$ .

shear stress  $S_{xy}$ , and the first term in eq. (3) represents the torque due to the asymmetric part of  $S_{xy}$ . The second term in eq. (3) represents the torque from the couple stress  $C_{yz}$  which is the surface torque in the  $z$  direction acting on the plane perpendicular to the  $y$  axis. The coefficients of viscosity  $\mu$ ,  $\mu_r$ , and  $\mu_B$  can be estimated based on the elementary kinetic theory of the transport process [8], which gives the following estimation for two dimensional flow:

$$\mu \sim \frac{m}{\sigma} \sqrt{\frac{T}{m}}, \quad \mu_B \sim m\sigma \sqrt{\frac{T}{m}}, \quad \mu_r \sim mn^2\sigma^3 \sqrt{\frac{T}{m}},$$

where  $n$  is the number density, and  $T$  is the granular temperature. It should be noted that  $\mu_r$ , which represents the coupling between  $\nabla \times \mathbf{v}$  and  $\boldsymbol{\omega}$ , is proportional to  $n^2$ . More detailed arguments for the derivation can be found in ref. [10].

These equations can be solved analytically with the equation of state  $p = \rho T/m$  in the dilute limit under the assumption of constant temperature  $T = \bar{T}$ . Then, from eq. (2), we get  $\rho(y) = \rho_0 \exp(-y/h)$  with  $h = \bar{T}/mg \cos \theta$ . This ex-

ponential profile of density is consistent with the simulation data, as is shown in Fig. 4 (a). Here, the fitted value of the relaxation length  $h \sim 2.24$ , which gives  $\bar{T} \sim 2.0$ ; this value is also consistent with the simulation data (Fig. 4 (b)). Then we can solve eqs. (1) and (2) with the estimation of the viscosity. The obtained profiles of  $u(y)$  and  $\omega(y)$  are fitted to the profiles obtained from the simulation data of the uniform flow. They are shown in Figs. 4 (c) and (d). Because the particles near the boundary are forced to rotate by the collisions with the slope,  $|\omega(y)|$  becomes large near the base. This effect causes the systematic deviation of  $\omega$  from  $\frac{1}{2}\nabla \times v$  in the simulation, as is shown in Fig. 4 (e). The deviation is also large far from the slope, in which the localized spins of particles are much larger than the fluid rotation. The micropolar fluid model can describe these deviations successfully; due to coupling of the velocity field and the angular velocity field, such a deviation may cause the behavior different from the Newtonian fluid described by the Navier-Stokes equation. In the simulation here, however, the small deviation of the velocity field from the solution of the Navier-Stokes equation appears only near the base as is seen in Fig. 4 (d).

## 4 Summary and Discussion

We have shown that the uniform granular flow on a rough slope is unstable against clustering in the collisional flow region. The instability depends on the system size and the density. The angular rotation and velocity fields of the uniform flow can be successfully described by the micropolar fluid model. It is a future issue to perform the stability analysis of the uniform solution. It is also an interesting problem to clarify whether there are any common mathematical structures between the instability found here and the spontaneous cluster formation out of the uniform flow seen in other systems, such as the granular flow in a vertical pipe or the traffic flow on a freeway.

Part of the computation in this work has been done using the facilities of the Supercomputer Center, Institute for Solid State Physics, University of Tokyo.

## References

1. H.M. Jaeger, S.R. Nagel, and R.P. Behringer, *Rev. Mod. Phys.* **68**, 1259 (1996).
2. C.S. Campbell, *Ann. Rev. Fluid Mech.* **22**, 57 (1990).
3. O. Moriyama, N. Kuroiwa, M. Matsushita, and H. Hayakawa, *Phys. Rev. Lett.* **80**, 2833 (1998).
4. M.Y. Louge and S.C. Keast, *Phys. Fluids* **13**, 1213 (2001).
5. Y. Forterre and O. Pouliquen, *Phys. Rev. Lett.* **86**, 5886 (2001).
6. N. Mitarai and H. Nakanishi, *J. Phys. Soc. Jpn.* **70**, 2809 (2001).
7. G. Lukaszewicz, *Micropolar Fluids, Theory and Applications* (Birkhäuser, Boston, 1999) and references therein.
8. N. Mitarai, H. Hayakawa, and H. Nakanishi, to be published in *Phys. Rev. Lett.*
9. P.A. Cundall and O.D.L. Strack, *Geotechnique* **29**, 47 (1979).
10. B.J. McCoy, S.I. Sandler, and J.S. Dahler, *J. Chem. Phys.* **45**, 3485 (1966). See also H. Hayakawa in this proceedings.

# The Nature of Occurrence of Queued Flow at Capacity Bottleneck of Ordinary Section

T. Oguchi

Tokyo Metropolitan University, Department of Civil Engineering, Minamiosawa 1-1, Hachioji, 192-0397, Japan

**Abstract.** The nature of occurrence of *queued flow*, clearly defined here, is practically investigated. The headway and speed of each vehicle observed at one section of expressways can help to identify the occurrence of queued flow.

## 1 Congested Road Traffic Condition

The word *traffic congestion* includes many conditions of road traffic flow. Therefore the author dares to use the word *queued flow* as *real* congested or jammed traffic condition. *Queued flow* is defined as a traffic flow condition with queued vehicles upstream of a segment of a traffic capacity bottleneck, when traffic demand exceeds the bottleneck capacity. In other words, the queued flow condition is a traffic condition which only occurs in the upstream section of a capacity bottleneck, when traffic demand exceeds the traffic capacity of the bottleneck section.

There are many types of ‘bottlenecks’ of traffic capacity, even on expressways. The merging or diverging sections, weaving sections, and toll gates are some obvious bottlenecks. Some of road vertical alignment sag curve sections are known as bottlenecks as well as tunnel entrances, especially on Japanese expressways, even if those sections are mere ordinary, normal basic sections (as found in [1,3]). Of course, road sections with traffic accident, some vehicle brake down, and so on are also called bottlenecks, as well as very slowly moving vehicles (sometimes called as *moving bottleneck*).

## 2 The Nature of Bottleneck Phenomena

### 2.1 Bottleneck Caused by Excess Demand

Fig. 1 shows a typical example of bottleneck phenomena observed by vehicle detectors installed on upstream and downstream sections with five-minute traffic flow rate (represented by vehicles per hours for two-lane, one direction) and five-minute (harmonic) mean speed. The bottleneck is the so called *Torisawa sag* located around the 62.5 kilometer-post (kp), outbound of Chuo-Expressway, Japan.

Fig. 1 shows that the flow rate increases gradually at about, and speed suddenly decreases 7 a.m. at the upstream detector, without definite speed decrease

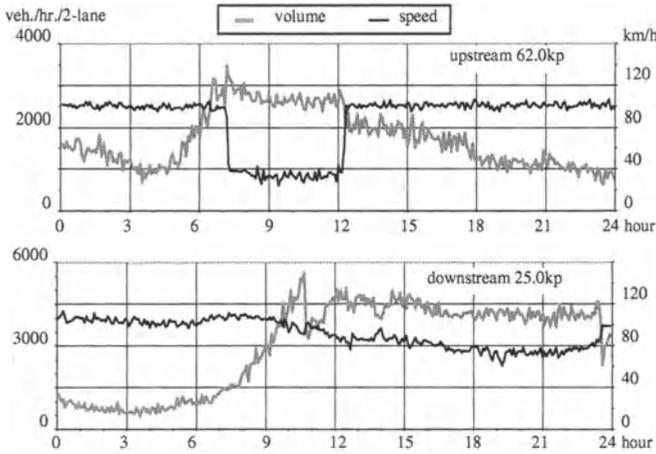


Fig. 1. Bottleneck caused by excess demand.

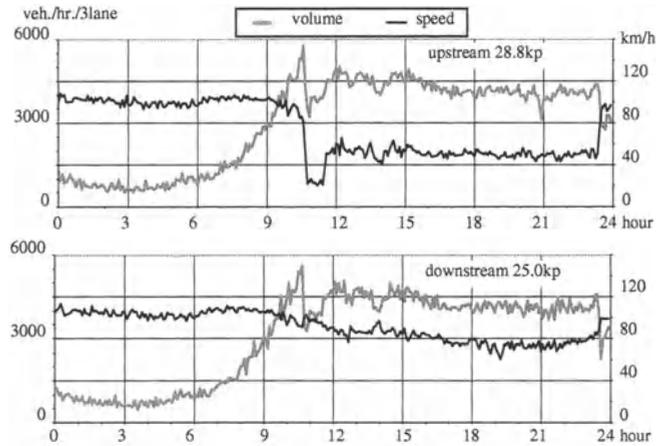


Fig. 2. Bottleneck caused by incident.

at downstream detector. This is one of the specific features of occurrence of *queued flow* or *bottleneck activation* at ordinary or basic section bottlenecks such as sags or tunnel entrances.

Maximum flow rate is observed at the start time of bottleneck activation; this is called ‘capacity before queued flow occurrence’ ( $Q_{levelA}$ ). After the activation of the bottleneck, *i.e.* a queue formed upstream of the bottleneck, flow rate becomes lower than  $Q_{levelA}$  and gradually decreasing; this is called ‘capacity after queue formed’ ( $Q_{levelB}$  and  $Q_{levelC}$ ) as found in [3].

Speed suddenly increases just after noon, showing disappearance of queued flow. The flow rate just after disappearance is definitely lower than  $Q_{levelB}$  or  $Q_{levelC}$ . This is also another feature of this kind of bottleneck.

## 2.2 Bottleneck Caused by an Accident

Fig. 2 shows the typical feature of bottleneck phenomena caused by accidents or other incidents, leading to a sudden capacity decrease. This figure shows data observed upstream and downstream of the accident section; one of most famous sag bottlenecks called 'Ayase sag' (around 29.5 kp) is located near this section, inbound of Tomei-Expressway, Japan.

Fig. 2 shows that flow rate increases gradually around 10 a.m., but speed suddenly decreases around 11 a.m. along with *decrease* of flow rate, and with *speed increase* (a little) at the upstream detector. This is one of the specific features of occurrence of bottleneck activation at accident bottlenecks.

Decreased flow rate shows the capacity of the incident section, which is definitely lower than  $Q_{levelA}$ ,  $Q_{levelB}$ , and  $Q_{levelC}$ . This flow condition lasts for about one hour (from 11 a.m. and 12 a.m.).

After the cause of the incident is removed (after 12 a.m.), flow condition changes to queued flow due to the sag bottleneck, continuing until around 24 o'clock. By the way, another small incident is seen at about 14 o'clock.

## 3 Detection of Bottleneck Activation

Fig. 3 shows an example of bottleneck activation observed just upstream of a sag bottleneck. It shows the speed of each vehicle (observed by a site-located video camera), flow rate and mean speed aggregated for one minute, and flow rate and mean speed aggregated for five minutes. The data were collected on the median lane.

Japan Highway Corporation adopted a rule of thumb for identifying the timing of bottleneck activation: if the five-minute mean speed observed upstream of the bottleneck becomes smaller than 40 km/h, the bottleneck is judged to be

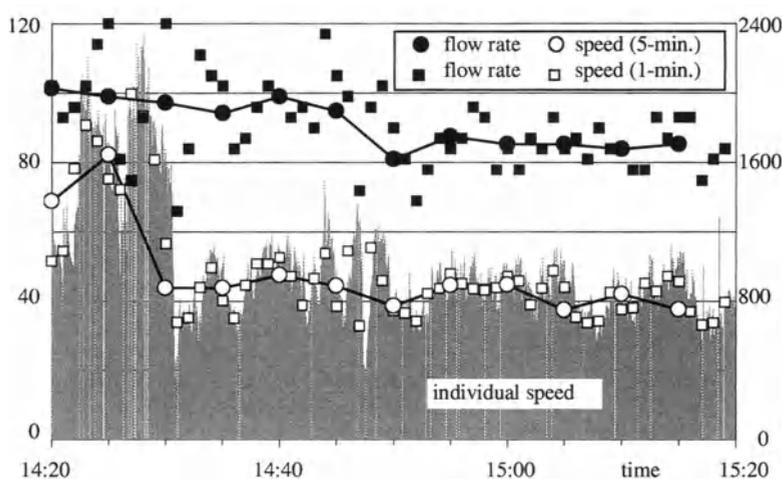


Fig. 3. Detection and judgment of bottleneck activation.

active. Therefore, Fig. 3 shows that the bottleneck activation has occurred at 14:50. Obviously, this time is too late to be judged as the bottleneck activation time, a miss-judgment.

When aggregation duration is shortened to one minute, the figure shows the bottleneck activation occurs at 14:31 using the same ‘rule of thumb’ of 40 km/h threshold for identifying activation time, and this time seems to be reasonable. But this is still ‘rule of thumb’.

There must be some definite differences between queued flow and non-queued flow condition on each vehicle behavior, therefore, there must be a more logical manner to identify the flow condition (already discussed in [2,4]). It is already known that the observed speed tends to become higher if the upstream vehicle detector is located nearer to a particular bottleneck section, especially located ‘in’ the bottleneck section (mentioned in [4]). Therefore, a judgment method of bottleneck activation should better not use a fixed threshold value of the speed.

## 4 State Variables of Each Vehicle

### 4.1 Nature of State Variables of Each Vehicle

Fig. 4 shows the speed-density relationship, the speed-flow rate relationship, and the flow rate-density relationship for each vehicle observed by the site installed

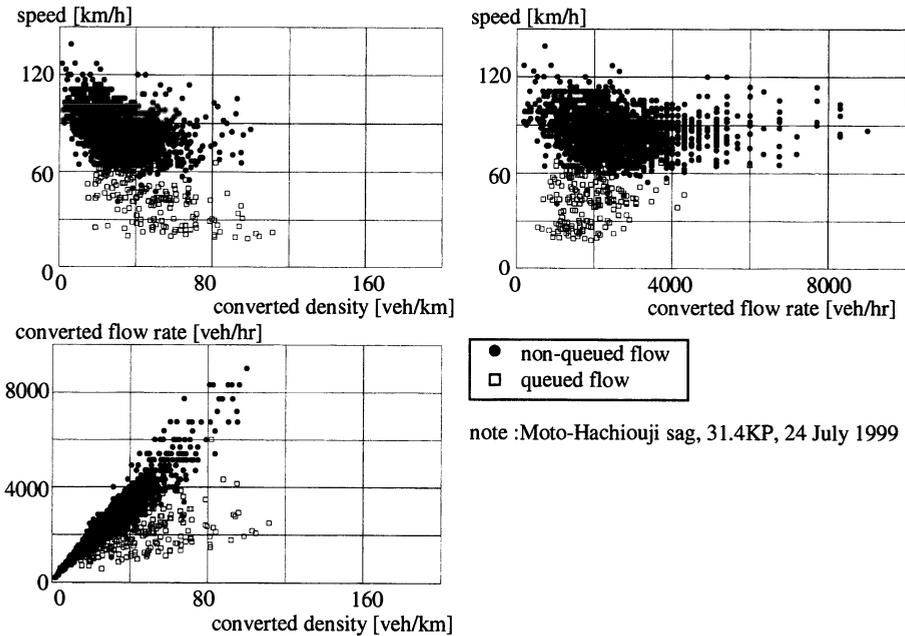


Fig. 4. Speed, flow and density relationship of each vehicle.

video camera (same as in Fig. 3). The speed of each vehicle ( $v_i$ ) is directly observed by the video camera, installing two hypothetical image sensors on the screen for each vehicle. The headway ( $h_i$ ) is also measured directly. On the other hand, the flow rate ( $q_i$ ) is calculated from the headway, and the density ( $k_i$ ) is calculated by the product of speed and headway.

The well-known  $q-k$ ,  $q-v$ , or  $k-v$  relationship for aggregated state variables (e.g. five-minute aggregation) shows some correlation between those pairs of state variables. On the other hand, those relationships of state variables for each vehicle do not show good correlations, except for the flow rate-density ( $q-k$ ) relationship in a non-queued flow condition (see Fig. 4).

### 4.2 Density Gap

The author introduces the idea of ‘density gap’  $D_i$  as shown below.

$$\begin{aligned}
 D_i &= k_i - k_{calc} && (q_i < q_{max}) \\
 &= 0 && (q_i \geq q_{max})
 \end{aligned}
 \tag{1}$$

where  $\exists k_{calc} \leq 91.7 \mid q_i = -0.6(k_{calc} - 91.7)^2 + 5041.7$ .

The function defining  $k_{calc}$  from  $q_i$  is determined from several observed data: observations were carried out for several days of bottleneck phenomena activated at two bottleneck sections of Moto-hachiouji sag, outbound of Chuo-Expressway and Ayase sag, inbound of Tomei-Expressway.

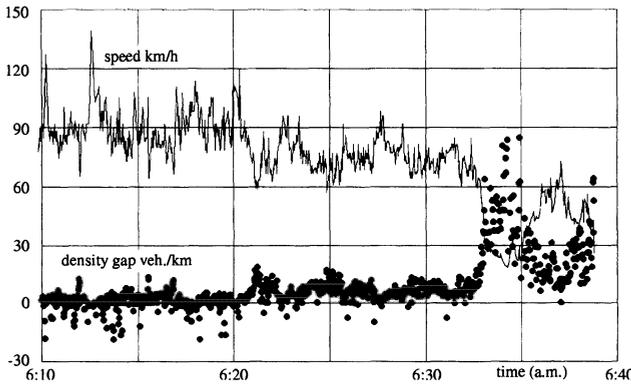
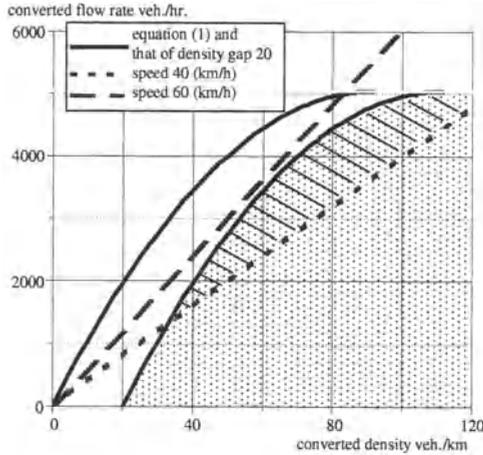


Fig. 5. Time series of speed and density gap.

Fig. 5 shows an example of speed and density gap for successive vehicles, around the time of bottleneck activation, from same data source as in Fig. 3. The density gap changes from still condition (almost zero) to big positive values with more fluctuations before and after the bottleneck activation.

Fig. 6 shows the comparison of the concept of density gap and definite value of speed threshold for identifying the difference of queued flow and non-queued flow condition (bottleneck activation and de-activation).



**Fig. 6.** Judgment in flow rate-density relationship.

The figure shows an example of density gap threshold of 20 veh./km. Compared with the idea of a speed threshold of 40 km/h, the idea of a density gap threshold of 20 veh/km can take the condition with high density but relatively high speed (more than 40 km/h) as queued flow condition. Therefore, this 'density gap' method might have possibility to be used as judgment index for bottleneck activation.

## 5 Conclusion

The nature of the queued flow condition and occurrence of this condition (bottleneck activation) is discussed, and the new idea of 'density gap' method is introduced to detect the bottleneck activation without any fixed threshold value of speed for identifying; queued or non-queued traffic flow condition.

## References

1. M. Koshi, *IATSS Research*, **9**, 50–56 (1985).
2. M. Koshi, M. Iwasaki, and I. Ohkura, 'Some findings and an overview on vehicular flow characteristics'. in: *Transportation and Traffic Theory*, pp. 403–426, (Univ. of Toronto Press, Toronto 1981).
3. M. Koshi, M. Kuwahara, and H. Akahane, *ITE Journal* **62–5**, 17–22 (1992).
4. M. Katakura, T. Oguchi, and S. Shikata, 'A detecting method of traffic congestion on expressways', in: *Proc. of 7th World Congress on Intelligent Transport Systems*, CD-ROM, (2000).

# **Pedestrian Dynamics**

# Bionics-Inspired Cellular Automaton Model for Pedestrian Dynamics

A. Schadschneider

Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany

**Abstract.** We present a 2-dimensional cellular automaton model for the simulation of pedestrian dynamics. Inspired by the principles of chemotaxis the interactions between the pedestrians are mediated by a so-called *floor field*. This field has a similar effect as the chemical trace created e.g. by ants to guide other individuals to food places. In our case the floor field modifies the transition rates to neighbouring cells. It has its own dynamics (diffusion and decay) and can be changed by the motion of the pedestrians. This means that in our model pedestrians follow a virtual rather than a chemical trace as in the case of chemotaxis. The approach is extremely efficient and makes faster-than-real-time simulations of large crowds possible. Already the inclusion of only nearest-neighbour interactions allows to reproduce many of the collective effects and self-organization phenomena (lane formation, flow oscillations at doors etc.) encountered in pedestrian dynamics.

## 1 Introduction

Methods from physics have been successfully used for the investigation of vehicular traffic for a long time [1,2]. On the other hand, pedestrian dynamics has not been studied as extensively. Due to its generically two-dimensional nature, pedestrian motion is more difficult to describe in terms of simple models. However, many interesting collective effects and self-organization phenomena have been observed (see [2,3] for an overview and a comprehensive list of references):

- **Jamming:** At large densities various kinds of jamming phenomena occur, e.g. when many people try to leave a large room at the same time. This clogging effect is typical for a bottleneck situation. Other types of jamming occur in the case of counterflow where two groups of pedestrians mutually block each other. This happens typically at high densities and when it is not possible to turn around and move back, e.g. when the flow of people is large.
- **Lane formation:** In counterflow, i.e. two groups of people moving in opposite directions, a kind of spontaneous symmetry breaking occurs (see Sec. 4). The motion of the pedestrians can self-organize in such a way that (dynamically varying) lanes are formed where people move in just one direction [4]. In this way, strong interactions with oncoming pedestrians are reduced and a higher walking speed is possible.
- **Oscillations:** In counterflow at bottlenecks, e.g. doors, one can observe oscillatory changes of the direction of motion. Once a pedestrian is able to pass the bottleneck it becomes easier for others to follow him in the same direction

until somebody is able to pass (e.g. through a fluctuation) the bottleneck in the opposite direction.

- **Patterns at intersections:** At intersections various collective patterns of motion can be formed. A typical example are short-lived roundabouts which make the motion more efficient. Even if these are connected with small detours the formation of these patterns can be favourable since they allow for a “smoother” motion.
- **Panics:** In panic situations many counter-intuitive phenomena (e.g. “faster-is-slower” and “freezing-by-heating” effects [5]) can occur. For a thorough discussion we refer to [3,6,7] and references therein. Elsewhere in these proceedings we will study the dynamics of evacuation processes in more detail [8] (see also [7,9–11]).

In recent years, several models for the description of pedestrian dynamics have been suggested:

- **Social force models:** So far continuum models have been most successful in modelling pedestrian dynamics. An important example are the social force models (see e.g. [2,4] and references therein). Here pedestrians are treated as particles subject to long-ranged forces induced by the social behaviour of the individuals. This leads to (coupled) equations of motion similar to Newtonian mechanics. There are, however, important differences since, e.g., in general the third law (“actio = reactio”) is not fulfilled.
- **Active walker models:** So-called active walker models [12,13] have been used to describe the formation of human or animal trails etc. Here the walker leaves a trace by modifying the underground on his path. This modification is real in the sense that it could be measured in principle. For trail formation, vegetation is destroyed by the walker and some animals (like ants) leave a chemical trace which can be detected by other animals.
- **Cellular automata:** Most cellular automata (CA) models for pedestrian dynamics proposed so far are rather simple [14–17] and can be considered as generalizations of the Biham-Middleton-Levine model for city traffic [18]. However, these models are not able to reproduce all the collective effects observed empirically. The same is true for more sophisticated discrete models, e.g. the one suggested by Gipps and Marksjös [19] or a discretized version of the social force model [20].

In [21,22] a new kind of CA model has been introduced which – despite its simplicity – is able to reproduce the observed collective effects. It takes its inspiration<sup>1</sup> from the process of chemotaxis (see [23] for a review) as used by some insects. They create a chemical trace to guide other individuals to food places. This is also the central idea of the active-walker models. In the approach of [21] the pedestrians also create a trace. In contrast to trail formation and chemotaxis, however, this trace is only virtual although one could assume that it corresponds to some abstract representation of the path in the mind of the

<sup>1</sup> Such “learning from nature” is the central idea of a field called *Bionics*.

pedestrians. Its main purpose is to transform effects of long-ranged interactions (e.g. following people walking some distance ahead) into a local interaction (with the “trace”). This allows for a much more efficient simulation on a computer.

## 2 Basic Principles of the Model

First we discuss some general principles applied in the development of the model [21,22]. To allow an efficient implementation for large-scale computer simulations a discrete model is preferable. Therefore a two-dimensional CA is used with a stochastic dynamics taking into account the interactions between the pedestrians. As mentioned above, similar to chemotaxis, we want to transform long-ranged interactions into local ones. This is achieved by the introduction of so-called *floor fields*. The transition probabilities for all pedestrians depend on the strength of the floor fields in their neighbourhood in such a way that transitions in the direction of larger fields are preferred.

Interactions between pedestrians are repulsive for short distances. One likes to keep a minimal distance to others in order to avoid bumping into them. This is taken into account through hard-core repulsion which prevents multiple occupation of the cells. For longer distances the interaction is often attractive. E.g. when walking in a crowded area it is usually advantageous to follow directly behind the predecessor. Large crowds may also be attractive due to curiosity and in panic situation often herding behaviour can be observed [6].

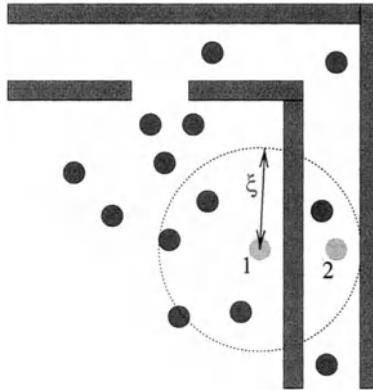
The long-ranged part of the interaction is implemented through the floor fields. We distinguish two kinds, a *static floor field* and a *dynamic floor field*. The latter models the dynamic interactions between the pedestrians, whereas the static field represents the constant properties of the surroundings.

The dynamic floor field corresponds to a virtual trace which is created by the motion of the pedestrians and in turn influences the motion of other individuals. Furthermore it has its own dynamics, namely through diffusion and decay, which leads to a dilution and finally the vanishing of the trace after some time.

The static floor field does not change with time since it only takes into account the effects of the surroundings. Therefore it exists even without any pedestrians present. It allows to model, e.g., preferred areas, walls and other obstacles. A typical example can be found in [8] where the evacuation from a room with a single door is examined. Here the strength of the static field decreases with increasing distance from the door.

The introduction of the floor fields allows for a very efficient implementation on a computer since now all interactions are local. We have translated the *long-ranged spatial interaction* into a *local interaction with “memory”*. The number of interaction terms in other long-ranged models, e.g. the social-force model, grows proportionally to the square of the number of particles whereas in our model it grows only linearly. Another advantage can be seen in the case of complex geometries, e.g. as shown in Fig. 1. Here it is assumed that the typical interaction range is  $\xi$ . However, due to the presence of walls not all particles within this range interact with each other. Therefore one needs an algorithm

to check whether two particles “see” each other and interact or whether the interaction is blocked by some obstacle.



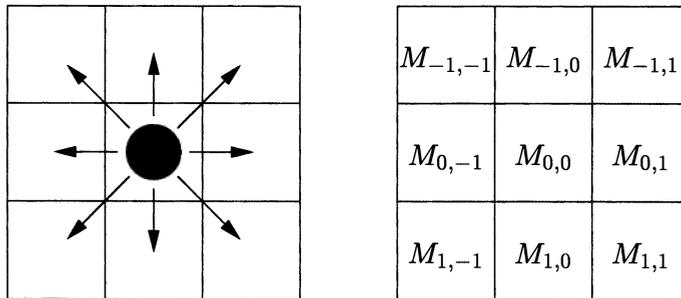
**Fig. 1.** Typical situation in a complex geometry: Particles 1 and 2 do not interact with each other due the wall between them.

Furthermore we do not need to provide the pedestrians with some sort of “intelligence”. The floor fields are sufficient to achieve the formation of complex structures and collective effects by means of self-organization. Even for the evacuation scenario [8,11,21] described above the knowledge about the floor fields is enough to find the exit. Therefore the pedestrians behave like simple ‘particles’ in a field without explicit intelligence. Since we also not make detailed assumptions about the human behaviour it allows us to keep the model simple. Nevertheless it is able to reproduce many of the basic phenomena.

In contrast to vehicular traffic the time needed for acceleration and braking is negligible in pedestrian motion. The velocity distribution of pedestrians is sharply peaked [24]. These facts naturally lead to a model where the pedestrians have a maximal velocity  $v_{\max} = 1$ , i.e. only transitions to neighbour cells are allowed. Furthermore, a larger  $v_{\max}$ , i.e. pedestrians can move more than just one cell per timestep, would be harder to implement in two dimensions and reduce the computational efficiency. The number of possible target cells increases quadratically with the maximal velocity. Furthermore it has to be checked whether the path is blocked by other pedestrians. This might even be ambiguous for diagonal motion and crossing trajectories. Also higher velocity models lead to timescales which are much too small (see Sec. 3).

For some applications it is useful to introduce another field called *matrix of preference* which encodes the preferred walking direction and speed of each pedestrian. From this direction, a  $3 \times 3$  matrix is constructed which contains the probabilities for a move of the particle. The central element describes the probability for the particle not to move at all, the remaining eight correspond to a

move to one of the neighbouring cells<sup>2</sup> (see Fig. 2). The probabilities can directly be related to observable quantities, namely the velocity and the longitudinal and transversal standard deviations (see [21,25] for details). In the simplest case the pedestrian is allowed to move in one direction only without fluctuations. In the corresponding matrix of preference only one element is one and all others are zero. In the following it is assumed that a matrix of preference is given at every timestep for each pedestrian. In principle, it can differ from cell to cell depending on the geometry and aim of the pedestrians. Therefore it can be combined with some model for route selection which assigns certain routes to each pedestrian.



**Fig. 2.** A particle, its possible transitions and the associated matrix of preference  $M = (M_{ij})$ .

### 3 Definition of the Model

CA are discrete in space, time and state variables. The area available for pedestrians is divided into cells of approximately  $40 \times 40 \text{ cm}^2$  which is the typical space occupied by a pedestrian in a dense crowd [26]. Each cell can either be empty or occupied by exactly one particle (pedestrian). For special situations it might be desirable to use a finer discretization, e.g. such that each pedestrian occupies four cells instead of one, but we restrict to the simplest case which is sufficient for the applications studied here.

The update is done in parallel for all particles. This introduces a timescale into the dynamics which can roughly be identified with the reaction time  $t_{\text{reac}}$ . The existence of a timescale in the dynamics of the model is essential if one wants to make *quantitative predictions* for real processes. In the deterministic limit, corresponding to the maximal possible walking velocity in our model, a single pedestrian (not interacting with others) moves with a velocity of one cell per timestep, i.e.  $40 \text{ cm}$  per timestep. Empirically the average velocity of a pedestrian is about  $1.3 \text{ m/s}$  [26]. This gives an estimate for the real time

<sup>2</sup> In most cases motion in diagonal directions is not considered so that the matrix of preference has at most five nonzero elements.

corresponding to one timestep in our model of approximately 0.3 *sec* which is of the order of the reaction time  $t_{\text{reac}}$  and thus consistent with our microscopic rules. It also agrees nicely with the time needed to reach the normal walking speed which is about 0.5 *sec*. This corresponds to at least  $v_{\text{max}}$  timesteps if the pedestrian can only accelerate by one unit per timestep. Therefore in models with large  $v_{\text{max}}$  a timestep would correspond to a real time shorter than the smallest relevant timescale. This makes the model more complicated than necessary and reduces the efficiency of simulations.

We now define the stochastic dynamics of the model by specifying the transition probabilities  $p_{ij}$  for a motion to a neighbouring cell in direction<sup>3</sup>  $(i, j)$ . As explained in Sec. 2 the long-ranged interactions with other pedestrians and the surrounding is incoded in two fields, the dynamic floor field  $D$  and the static floor field  $S$ , respectively. Furthermore a matrix of preference  $M$  can be used to specify the walking direction, speed and fluctuations for each individual. The *transition probability*  $p_{ij}$  in direction  $(i, j)$  is then determined by all three contributions. Explicitly it is given by

$$p_{ij} = N e^{k_D D_{ij}} e^{k_S S_{ij}} M_{ij} (1 - n_{ij}) \xi_{ij}. \quad (1)$$

$D_{ij}$  and  $S_{ij}$  are the strengths of the dynamic and static floor field at the target cell and  $M_{ij}$  is the matrix element of the matrix of preference for a motion in the direction  $(i, j)$ .  $N$  is a normalization factor to ensure  $\sum_{(i,j)} p_{ij} = 1$  where the sum is over the nine possible target cells. The factor  $1 - n_{ij}$ , where  $n_{ij}$  is the occupation number of the neighbour cell in direction  $(i, j)$ , takes into account that transitions to occupied cells are forbidden.  $\xi_{ij}$  is a geometry factor (obstacle number) which is 0 for forbidden cells (e.g. walls) and 1 else. Finally, we have introduced two coupling constants  $k_D$  and  $k_S$  so that we can vary the coupling strengths to each field individually.

The actual values of the parameters  $k_D$  and  $k_S$  depend on the situation. A large  $k_S$  implies that the pedestrians choose their path mainly due to the surrounding without being distracted too much by other people. A large coupling  $k_D$  to the dynamic field, on the other hand, corresponds to a strong herding behaviour. Here the pedestrian tries to follow the lead of others, e.g. in the case of panics or insufficient knowledge about the surroundings. The effects of these couplings will be discussed in more detail in [8].

As mentioned before, the dynamic floor field is created by the motion of the pedestrians and corresponds to a virtual trace. At  $t = 0$  it is zero everywhere. Whenever a particle moves from site  $(x, y)$  to one of its neighbours  $(x + i, y + j)$ , the field  $D_{xy}$  at the origin cell is increased by one ( $D_{xy} \rightarrow D_{xy} + 1$ ). Thus  $D_{xy}$  has only non-negative integer values which can be interpreted as the number of ‘bosons’ located at site  $(x, y)$ .

The dynamic floor field is not only changed by the motion of the pedestrians, but it is also subject to diffusion and decay which first leads a spreading and dilution of the trace and finally to its vanishing after some time. Diffusion and decay are controlled by two parameters  $\alpha \in [0, 1]$  and  $\delta \in [0, 1]$ . In each time step

<sup>3</sup>  $i, j \in \{-1, 0, 1\}$ , see Fig. 2.

of the simulation each boson of the dynamic field  $D$  decays with the probability  $\delta$  and diffuses with the probability  $\alpha$  to one of the neighbouring cells.

The update rules of the full model including the interaction with the floor fields then have the following structure:

1. The dynamic floor field  $D$  is modified according to its diffusion and decay rules.
2. Using equation (1), for each pedestrian the transition probabilities  $p_{ij}$  for a move to an unoccupied neighbour cell  $(i, j)$  are determined by the matrix of preference and the local dynamic and static floor fields.
3. Each pedestrian chooses a target cell based on the probabilities of the transition matrix  $P = (p_{ij})$ .
4. The conflicts arising by any two or more pedestrians attempting to move to the same target cell are resolved (see below).
5. The pedestrians which are allowed to move execute their step.
6. The pedestrians alter the dynamic floor field  $D_{xy}$  of the cell  $(x, y)$  they occupied before the move.

These rules have to be applied to all pedestrians at the same time (parallel dynamics). This introduces a timescale into the dynamics (see Sec. 2) which corresponds to approximately 0.3 *sec* of real time and allows to translate evacuation times measured in computer simulations into real times.

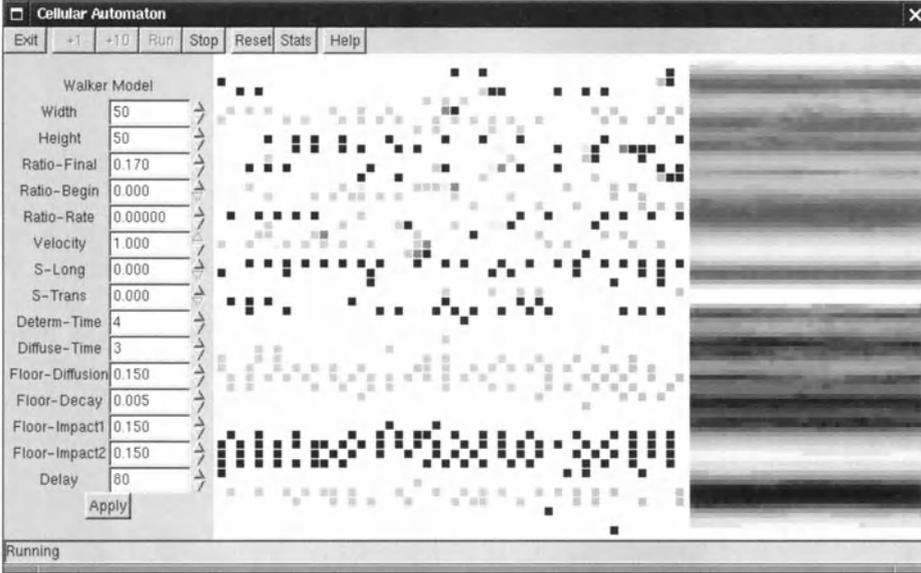
One detail is worth mentioning. If a particle has moved in the previous timestep the boson created then is not taken into account in the determination of the transition probability. This prevents that pedestrians get confused by their own trace.

Finally we like to comment on step 4, the resolution of conflicts which occur if  $m \geq 2$  particles choose the same destination cell in step 3. This is relevant for high density situations. In order to avoid multiple occupancies of cells only one particle is allowed to move while the others keep their position. There are different ways to choose this particle. E.g. the winning particle can be picked at random with probability  $1/m$  or according to the relative probabilities  $p_{ij}$  with which each particle chose their target [21]. For the problems studied here and in [8] the details of the conflict resolution turned out to play no important role.

## 4 Collective Phenomena

In this section we show that the model is indeed able to reproduce the collective effects mentioned in Sec. 1. As most prominent example we want to discuss lane formation out of a randomly distributed group of pedestrians. This corresponds to a spontaneous breaking of the symmetry of the particle number distribution in space. Simulations show that an even as well as an odd number of lanes may be formed. The latter corresponds to a spontaneous breaking of the left-right symmetry of the system.

Fig. 3 shows simulations of a rectangular corridor which is populated by two species of pedestrians moving in opposite directions. Parallel to the direction of



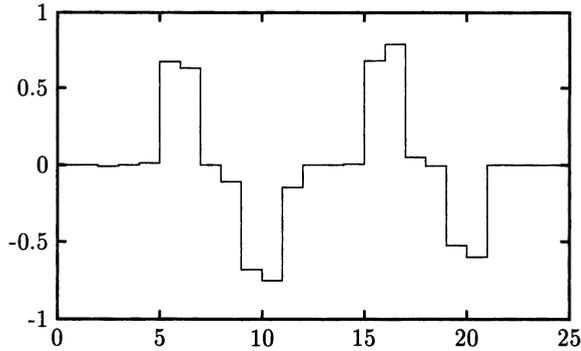
**Fig. 3.** Snapshot of a simulation of counterflow along a corridor. The left part shows the parameter control. The central window is the corridor and the light and dark squares are right- and left-moving pedestrians, respectively. The right part shows the floor fields for the two species.

motion the existence of walls is assumed. Orthogonal to the direction of motion both periodic and open boundary conditions have been investigated. With periodic boundary conditions, the density of pedestrians is fixed for each run. It is ensured that the overall number of pedestrians is evenly divided by the numbers for the different species, each interacting with its own dynamic floor field<sup>4</sup>. For open boundaries, we fix the rate at which pedestrians enter the system at the boundaries. The pedestrians leave the system as soon as they reach the opposite end of the corridor. Fig. 3 shows the graphical frontend running a simulation of a small periodic system. Lanes have already formed in the lower part of the corridor and can be spotted easily, both in the main window showing the cell contents and the small windows on the right showing the floor field intensity for the two species.

The formation of the lanes can also clearly be seen in the velocity profile (Fig. 4) which has been measured at a cross section perpendicular to the direction of flow. In a certain density regime, the lanes are metastable. Spontaneous fluctuations can disrupt the flow in one lane causing the pedestrians to spread and interfere with other lanes. Eventually the system can run into a jam by this mechanism.

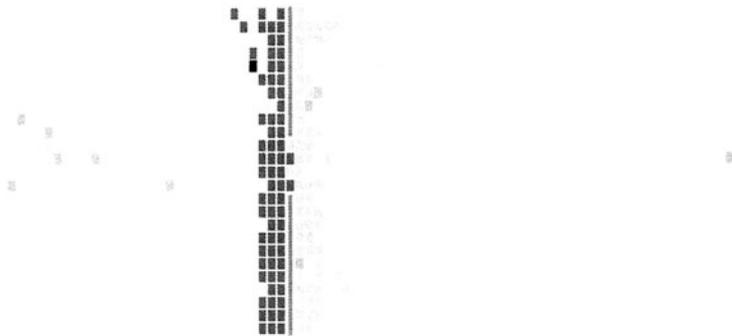
Another typical feature of pedestrian behaviour are oscillations of the direction of flow when two groups of opposite walking direction are facing each other at a tight spot like a narrow door [4]. In our simulations we can observe oscil-

<sup>4</sup> The static floor field is set to  $S_{xy} = 0$  in this example.



**Fig. 4.** Velocity profile of a periodic system with  $\rho = 0.12$ .

lations on two time scales (provided that the parameters are suitably chosen): An exchange between a blocked situation and a flow in both directions is the main result. Inside the blocked period small groups of only one species can break through. This is illustrated in Fig. 5. These breaks can alternate between the two species and can therefore be interpreted as oscillations.



**Fig. 5.** Oscillations of the direction of flow: A group of particles of the same species break through a blockade at a door (located in the center of the figure, with a width of 4 cells).

Apart from lane formation and oscillations of the direction of flow at doors also other phenomena, e.g. the formation of flow patterns at intersections, have been observed [9,21,25]. Since the model captures the main phenomena correctly and is extremely efficient in computer simulations it might be an ideal tool for evacuation simulations [11].

## 5 Conclusions

We have introduced a stochastic cellular automaton to simulate pedestrian behaviour, focussing on the general concept. The key mechanism is the introduction of floor fields which act as a substitute for pedestrian intelligence and lead

to collective phenomena. These floor fields make it possible to translate spatial long-ranged interactions into non-local interactions in time. The latter can be implemented much more efficiently on a computer. Another advantage is an easier treatment of complex geometries.

The general idea in our model is similar to chemotaxis. However, the pedestrians leave a virtual trace rather than a chemical one. This virtual trace has its own dynamics (diffusion and decay) which restricts the interaction range (in time). It is realized through a dynamic floor field which allows to give the pedestrians only minimal intelligence and to use local interactions. Together with the static floor field it offers the possibility to take different effects into account in a unified way, e.g. the social forces between the pedestrians or the geometry of the building.

We have shown that the approach is able to reproduce the fascinating collective phenomena observed in pedestrian dynamics. As an example the formation of lanes in counterflow and oscillations of the flow direction at bottlenecks have been discussed in Sec. 4.

In Part II [8] we investigate the influence of the coupling strengths on the dynamics. There it is also demonstrated that the approach can be used for simple evacuation simulations. Therefore, the model is a good starting point for realistic applications [11]. Further information can also be found on the webpage [27].

The description of pedestrians using a cellular automaton approach allows for very high simulation speeds. Pedestrian crowds of a few hundred people can be simulated 10 to 100 times faster than real time. Therefore, we have the possibility to extract the complete statistical properties of our model using Monte Carlo simulations.

Finally it should be emphasized that we have presented only the basic ideas of the approach. For realistic applications modifications might be appropriate, e.g. smaller cell sizes or inertia effects [21].

**Acknowledgement:** I like to thank my collaborators Carsten Burstedde, Ansgar Kirchner, and Kai Klauck for their input and many inspiring discussions.

## References

1. D. Chowdhury, L. Santen and A. Schadschneider, Phys. Rep. **329**, 199 (2000).
2. D. Helbing, cond-mat/0012229.
3. D. Helbing, I. Farkas, P. Molnar and T. Vicsek, in: M. Schreckenberg and S.D. Sharma (Eds.), "Pedestrian and Evacuation Dynamics" (Springer, 2001).
4. D. Helbing and P. Molnar, Phys. Rev. **E51**, 4282 (1995).
5. D. Helbing, I. Farkas and T. Vicsek, Phys. Rev. Lett. **84**, 1240 (2000).
6. D. Helbing, I. Farkas and T. Vicsek, Nature **407**, 487 (2000).
7. D. Helbing: these proceedings.
8. A. Kirchner and A. Schadschneider, *Cellular automaton simulations of pedestrian dynamics and evacuation processes*, these proceedings.

9. C. Burstedde, A. Kirchner, K. Klauck, A. Schadschneider and J. Zittartz, in: M. Schreckenberg and S.D. Sharma (Eds.), "Pedestrian and Evacuation Dynamics" (Springer, 2001).
10. N. Ohi, M. Ikai and K. Nishinari, these proceedings.
11. A. Kirchner and A. Schadschneider, in preparation.
12. D. Helbing, F. Schweitzer, J. Keltsch and P. Molnar, *Phys. Rev.* **E56**, 2527 (1997).
13. D. Helbing, J. Keltsch and P. Molnar, *Nature* **388**, 47 (1997).
14. M. Fukui and Y. Ishibashi, *J. Phys. Soc. Jpn.* **68**, 2861, 3738 (1999).
15. M. Muramatsu, T. Irie and T. Nagatani, *Physica* **A267**, 487 (1999).
16. M. Muramatsu and T. Nagatani, *Physica* **A275**, 281 (2000); **A286**, 377 (2000).
17. H. Klüpfel, T. Meyer-König, J. Wahle and M. Schreckenberg, in: *Theory and Practical Issues on Cellular Automata*, S. Bandini and T. Worsch (Eds.), Springer (2000); see also these proceedings.
18. O. Biham, A.A. Middleton and D. Levine, *Phys. Rev.* **A46**, R6124 (1992).
19. P.G. Gipps and B. Marksjö, *Math. and Comp. in Simulation* **27**, 95 (1985).
20. K. Bolay, Diploma Thesis, Stuttgart University (1998).
21. C. Burstedde, K. Klauck, A. Schadschneider and J. Zittartz, *Physica* **A295**, 507 (2001).
22. A. Schadschneider, in: M. Schreckenberg and S.D. Sharma (Eds.), "Pedestrian and Evacuation Dynamics" (Springer, 2001).
23. E. Ben-Jacob, *Contemp. Phys.* **38**, 205 (1997).
24. L.F. Henderson, *Nature* **229**, 381 (1971).
25. C. Burstedde, Diploma Thesis, Universität zu Köln (2001); available for download at <http://www.burstedde.de/carsten/diplom.html>.
26. U. Weidmann, *Transporttechnik der Fussgänger*, Schriftenreihe des IVT **80**, ETH Zürich (1992).
27. <http://www.thp.uni-koeln.de/~as/as.html>.

# Critical Discussion of “Synchronized Flow”, Simulation of Pedestrian Evacuation, and Optimization of Production Processes

D. Helbing<sup>1</sup>, I.J. Farkas<sup>2</sup>, D. Fasold<sup>1</sup>, M. Treiber<sup>1</sup>, and T. Vicsek<sup>2</sup>

<sup>1</sup> Institute for Economics and Traffic, Dresden University of Technology, Andreas-Schubert-Str. 23, D-01062 Dresden, Germany

<sup>2</sup> Department of Biological Physics, Eötvös University, Budapest, Pázmány Péter Sétány 1A, H-1117 Hungary

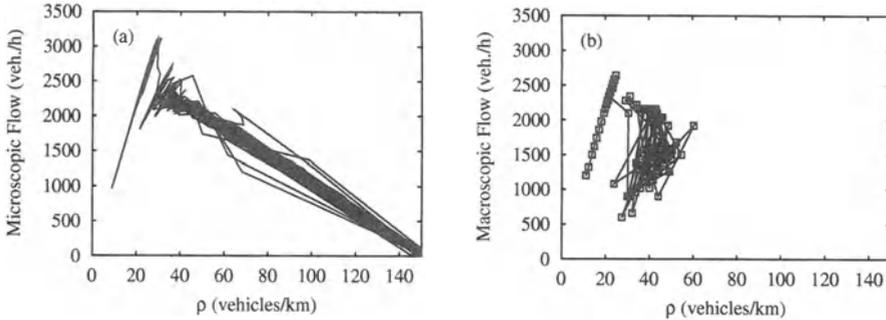
**Abstract.** We critically discuss the concept of “synchronized flow” from a historical, empirical, and theoretical perspective. Problems related to the measurement of vehicle data are highlighted, and questionable interpretations are identified. Moreover, we propose a quantitative and consistent theory of the empirical findings based on a phase diagram of congested traffic states, which is universal for all conventional traffic models having the same instability diagram and a fundamental diagram. New empirical and simulation data supporting this approach are presented as well. We also give a short overview of the various phenomena observed in panicking pedestrian crowds relevant from the point of evacuation of buildings, ships, and stadia. Some of these can be applied to the optimization of production processes, e.g. the “slower-is-faster effect”.

## 1 Freeway Traffic: “Synchronized Flow”, “Pinch Effect”, and Measurement Problems

### 1.1 What is New About “Synchronized Flow?”

Congested traffic has been investigated for many decades because of its complex phenomenology. Therefore, Kerner and Rehborn have removed the data belonging to wide moving jams (see MLC and TSG in Figs. 9 and 10) and found that the remaining data of congested traffic data still displayed a wide and two-dimensional scattering [1], see Fig. 4(c). By mistake (see Figs. 1, 4(c), and Sec. 1.2), they concluded that all models assuming a fundamental diagram were wrong and defined a new state called “*synchronized flow*” (“synchronized” because of the typical synchronization among lanes in congested traffic, see Fig. 2(a), and “flow” because of flowing in contrast to standing traffic in fully developed jams). Since then, Kerner suggests to classify *three phases*: (1) *free flow*, (2) “*synchronized flow*”, and (3) *wide moving jams* (i.e. *moving localized clusters* whose width in longitudinal direction is considerably higher than the width of the jam fronts). In some applied empirical studies, however, Kerner *et al.* additionally distinguish a fourth state of stop-and-go traffic [2,3].

*Free flow* is characterized by the average desired velocity  $V_0$  and, therefore, by a strong correlation and quasi-linear relation  $Q \approx \rho V_0$  between the local flow  $Q$  and the *local* density  $\rho$  [4]. It is also well-known that *wide moving jams*

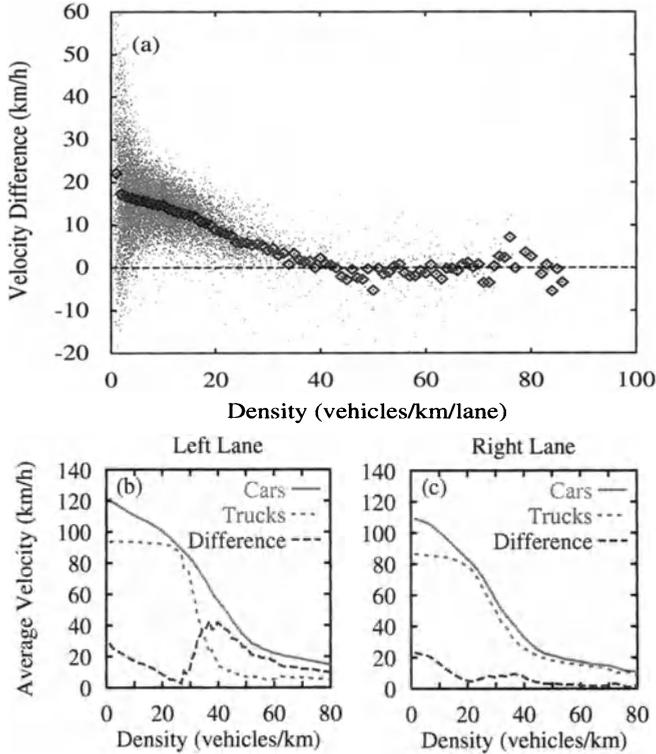


**Fig. 1.** (a) Flow-density relation for narrow moving jams simulated with a microscopic traffic model. (b) The aggregated (1-minute) data corresponding to the narrow moving jams displayed in (a) show a wide scattering and erratic movement in the flow-density plane. By mistake, this is used to characterize “synchronized flow”.

propagate with constant form and (phase) velocity  $C \approx -13\text{km/h}$  [5–7]. Kerner found that this propagation is not affected by bottlenecks or the presence of “synchronized flow”. Moreover, he showed that the outflow  $Q_{\text{out}}$  from wide jams is a self-organized traffic constant as well [6,8]. In contrast to wide moving jams, the flow inside of “synchronized flow” remains finite, and its downstream front is normally fixed at the location of some bottleneck, e.g. an on-ramp. Therefore, “synchronized flow” basically agrees with previous observations of queued or congested traffic (see, e.g., Refs. [9–11] and the references therein). In his patent [12], Kerner applies the queuing theory himself, which goes back to the fluid-dynamic traffic model by Lighthill and Whitham [13].

The synchronization of the average velocities among neighboring lanes has been already described by Koshi *et al.* [14] (but see also Refs. [5,15,16]). It is true on a *macroscopic* level for *all* forms of congested traffic including wide moving jams. Simulations have shown that this is a result of lane changes [17], while the assumed reduction in the lane changing rate [18] occurs only after the speeds in neighboring lanes have been successfully balanced [19]. On a *microscopic* scale, over-taking maneuvers continue to exist almost at all densities [20], see Figs. 2(b), (c). Nevertheless, the probability of lane changes drops considerably with increasing density, when most of the road is used up by the vehicles’ safety headways [20], but not in the postulated Z-shaped way [21]. Due to the reduced opportunities for overtaking and the related bunching of vehicles, the velocity variance goes down with increasing density as well [18,22,23].

The transition between free and congested traffic is of hysteretic nature, i.e. the inverse transition occurs at a lower density and a higher average velocity. This has been known for a long time [24–26]. Kerner specifies that the transition is usually caused by a localized and short perturbation of traffic flow that starts downstream of the on-ramp and propagates upstream with a velocity of about  $-13\text{km/h}$ . As soon as the perturbation passes the on-ramp, it triggers the



**Fig. 2.** (a) The difference of the average velocity in the left and the right lane vanishes at densities above 30 vehicles per kilometer, corresponding to a macroscopic synchronization of the speeds [22]. (b), (c) The difference in the empirically determined velocities of cars and trucks, however, show that overtaking maneuvers continue to exist even at higher densities [20].

breakdown which spreads upstream in the course of time. The congested state can then persist for several hours [18].

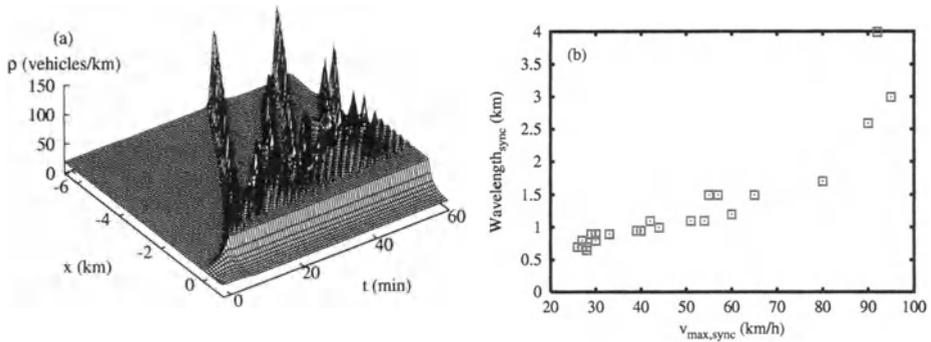
Moreover, Kerner and Rehborn distinguish three types of “synchronized flow” [1], which may be short-lived: (i) Stationary and homogeneous states where both the average speed and the flow rate are stationary (see, e.g., also Refs. [27–29]. Later on, we will call these “*homogeneous congested traffic*” (HCT) [30].

(ii) States where only the average vehicle speed is stationary, named “*homogeneous-in-speed states*” (see also Refs. [31,32]). We interpret this state as “*recovering traffic*” [33], as it bears several signatures of free traffic and mostly appears downstream of bottlenecks with congested traffic.

(iii) Non-stationary and non-homogeneous states (see also Refs. [31,34,35]). For these, we will use the term “*oscillating congested traffic*” (OCT) [30].

At least types (i) and (iii) are characterized by a considerably scattering and erratic change of the flow-density data, the various sources of which will be ad-

dressed in the following subsection. Continuous transitions between these types are probably the reason for the so-called “*pinch effect*” [36], see Fig. 3(a):

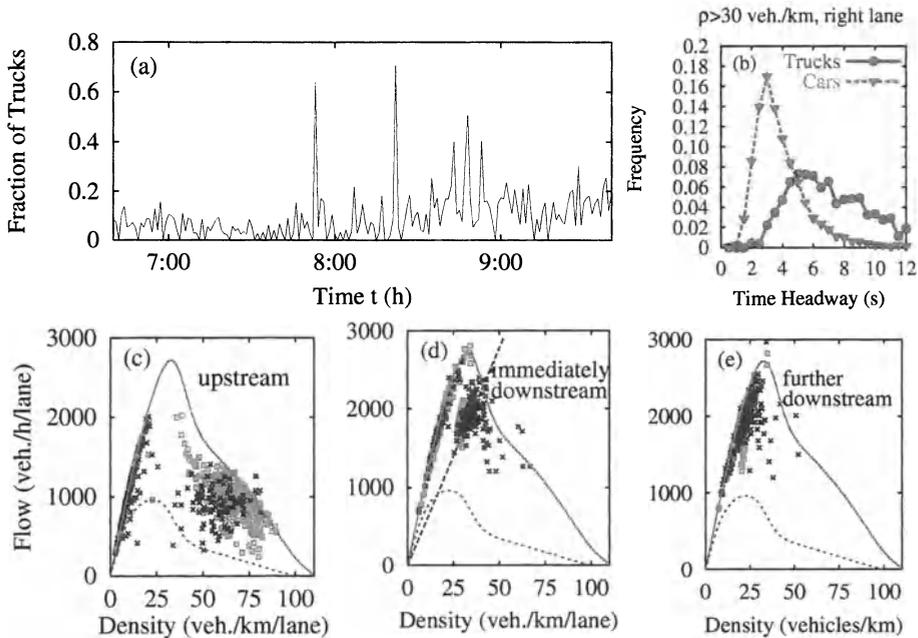


**Fig. 3.** (a) Simulation of the pinch effect with a deterministic microscopic model showing stable traffic at low and high densities, linearly unstable traffic at medium densities, and metastable traffic in between. The spatio-temporal density plot illustrates the breakdown to homogeneous congested traffic (HCT) upstream of a bottleneck, emerging oscillating congested traffic (OCT) further upstream, and a few stop-and-go waves (TSG) emanating from this region. The conditions for this spatial coexistence of congested traffic states are as follows [33,37]: The density in the congested region immediately upstream of the bottleneck should be in the linearly unstable, but convectively stable range, where perturbations are convected away in upstream direction [38,39]. In this case, traffic flow will appear stationary and homogeneous close to the bottleneck, but small perturbations will grow as they propagate upstream in the congested region starting at the bottleneck. If the perturbations propagate faster than the congested region expands, they will reach the area of free traffic upstream of the bottleneck. During rush hours, it is quite likely that this free flow is in the metastable range between the free and linearly unstable density region. Consequently, sufficiently large perturbations can trigger the formation of jams, which continue travelling upstream, while small perturbations are absorbed. (b) The wavelength (average distance between density maxima) determined from (a) qualitatively displays the empirical increase with the vehicle velocity observed by Kerner [36].

Upstream of a section with homogeneous congested traffic close to a bottleneck, there is a so-called “*pinch region*” characterized by the spontaneous birth of small and narrow density clusters, which are growing while they travel further upstream. Wide moving jams are eventually formed by the merging or disappearance of narrow jams, in which the cars move slower than in wide jams [36]. Once formed, the wide jams seem to suppress the occurrence of new narrow jams in between. Similar findings were already reported by Koshi *et al.* [14], who observed that “ripples of speed grow larger in terms of both height and length of the waves as they propagate upstream”.

### 1.2 Wide Scattering of Congested Flow-Density Data

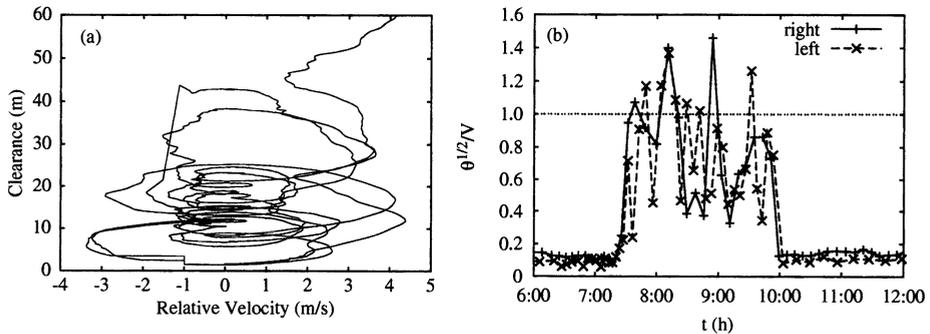
The collection and evaluation of empirical freeway data is a subject with often underestimated problems. To make reliable conclusions, in original investigations one should specify (i) the measurement site and conditions (including applied control measures), (ii) the sampling interval, (iii) the aggregation method, (iv) the statistical properties (variances, frequency distributions, correlations, survival times of traffic states, etc.), (v) data transformations, (vi) smoothing procedures, and the respective dependencies on them.



**Fig. 4.** (a) The empirical truck fraction varies considerably in the course of time. (b) The time headways of long vehicles (“trucks”) are on average much higher than those of short vehicles (“cars”). (c)-(e) Assuming a fundamental diagram for cars (solid line), a separate one for trucks (dashed line), weighting them according to the measured truck fraction, and using empirical boundary conditions allows to reproduce the observations in a (semi-)quantitative way [40]: Free traffic (at low densities) is characterized by a (quasi-)one-dimensional curve. (c) Data of congested traffic *upstream* of a bottleneck are widely scattered in a two-dimensional area. (d) *Immediately downstream* of the bottleneck, one observes homogeneous-in-speed states reflecting recovering traffic. (e) Further downstream the data points approach the curve describing free traffic. Dark symbols correspond to empirical one-minute data, light ones to corresponding simulation results.

The measurement conditions include ramps and road sections with their respective in- and outflows, speed limits, gradients, and curves with the respectively related capacities, furthermore weather conditions (like rain, ice, blinding

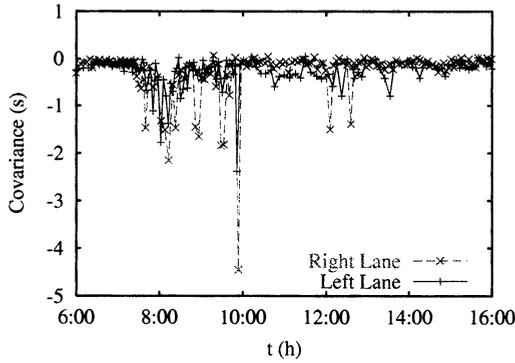
sun, etc.), presence of incidents (including the opposite driving direction), and other irregularities such as road works, which may trigger a breakdown of traffic flow. Moreover, one should evaluate the long vehicles (“trucks”) separately, as their proportion varies significantly, see Fig. 4(a). This can explain a considerable part of the wide scattering of congested traffic [40], see Figs. 4(c), (d). Presently, this explanation is still the only one for this observation that has been quantitatively checked with empirical data. Note that a considerable variation of the time headways is also observed among cars, see Fig. 4(b). This is partly due to different driver preferences and partly due to the instability of traffic flow, see Fig. 5(a). While vehicle platoons with reduced time headways imply an increase of the flow with growing density, a reduction in vehicle speeds is usually related with a decrease. According to Banks [41], this can account for the observed erratic changes of the flow-density data. We support this interpretation.



**Fig. 5.** (a) The measured oscillations of the clearance and the relative velocity [46] indicate an instability in the car-following behavior [47]. (b) The empirical standard deviation  $\sqrt{\theta(t)}$  of vehicle velocities divided by the average velocity  $V(t)$  is particularly large during the rush hour, where traffic flow is congested and unstable [33,48].

The strong variations of traffic flows imply that all measurements of macroscopic quantities should be complemented by error bars (see, e.g., Ref. [42]). Due to the relatively small “particle” numbers behind the determination of macroscopic quantities, the error bars are actually quite large. Hence, many temporal variations are within one error bar, when traffic flow is unstable, see Fig. 5(b). It is, therefore, very questionable whether it is possible to empirically prove the existence of small perturbations triggering a breakdown of traffic flow [18] or of the “birth” and merging of narrow density clusters in the “pinch region” [36]. At least, this would require a thorough statistical support. Consequently, we deny that such kind of data are presently suited as starting point for the development of new models [43] or traffic theories [21]. There is a considerable risk of overinterpreting particular (possibly statistical) features of the data recorded at special freeway sections and to construct new models that merely simulate what has been incorporated by means of the model assumptions. In fact, the only

reason why we believe in the correctness of these observations is the existence of plausible deterministic traffic models reproducing these hard-to-see effects without any special assumptions or extensions (see Fig. 3 and Refs. [17,37,44]).



**Fig. 6.** The covariance between headways  $d_\alpha$  and inverse velocities  $1/v_\alpha$  shows significant deviations from zero in congested traffic, while it approximately vanishes in free flow, compare Fig. 5(b). Even after traffic has recovered, there seem to remain weak correlations between headways and vehicle speeds for a considerable time. These are probably a reminiscence of congestion due to platoons which have not fully dissolved [33,48].

Because of the above mentioned problems, we would like to call for more refined measurement techniques, which are required for more reliable data. These must take into account *correlations* between different quantities, as is pointed out by Banks [45].

For example, approximating the vehicle headways by  $d_\alpha = v_\alpha \Delta t_\alpha$  (where  $v_\alpha$  is the velocity and  $\Delta t_\alpha$  the time headway of vehicle  $\alpha$ ) and determining arithmetic multi-vehicle averages  $\langle \dots \rangle$  at a fixed location, one obtains for the inverse vehicle flow

$$\frac{1}{Q} = \langle \Delta t_\alpha \rangle = \left\langle \frac{d_\alpha}{v_\alpha} \right\rangle = \langle d_\alpha \rangle \left\langle \frac{1}{v_\alpha} \right\rangle + \text{cov} \left( d_\alpha, \frac{1}{v_\alpha} \right). \tag{1}$$

Herein,  $\text{cov}(d_\alpha, 1/v_\alpha)$  denotes the covariance between the headways  $d_\alpha$  and the inverse velocities  $1/v_\alpha$ , which is negative and particularly relevant at large vehicle densities, as expected (see Fig. 6). Defining the local density  $\rho$  by

$$\rho = 1 / \langle d_\alpha \rangle \tag{2}$$

and the average velocity  $V$  via

$$\frac{1}{V} = \left\langle \frac{1}{v_\alpha} \right\rangle, \tag{3}$$

we obtain the fluid-dynamic flow relation

$$Q = \rho V \quad (4)$$

by the conventional assumption  $\text{cov}(d_\alpha, 1/v_\alpha) = 0$ . This, however, overestimates the density systematically, since the covariance tends to be negative due to the speed-dependent safety distance of vehicles. In contrast, the common method of determining the density via  $Q/\langle v_\alpha \rangle$  systematically underestimates the density [33,49]. *Consequently, errors in the measurement of the flow and the density due to a neglect of correlations partly account for the observed scattering of flow-density data in the congested regime.*

A similar problem occurs when the density is determined via the time occupancy of a certain cross section of the road. Considering that  $\Delta t_\alpha = T_\alpha + l_\alpha/v_\alpha$ , where  $T_\alpha$  is the (netto) time clearance and  $l_\alpha$  the length of vehicle  $\alpha$ , we have the relation

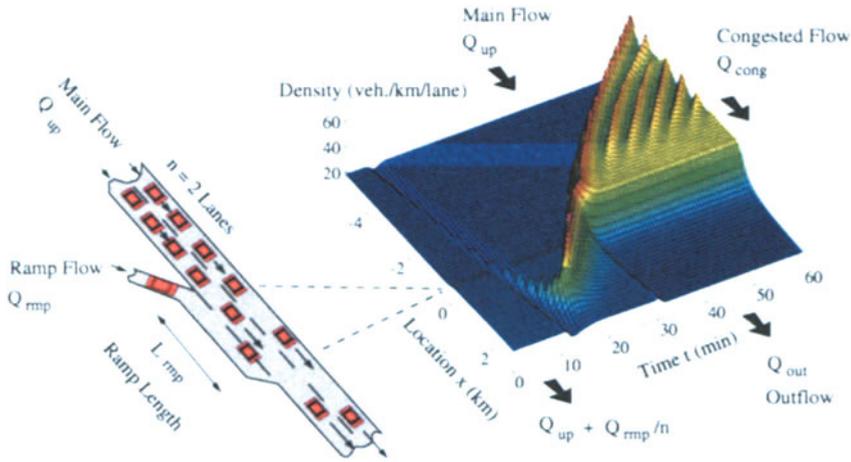
$$\rho = \rho_{\max} \frac{\langle l_\alpha/v_\alpha \rangle}{\langle \Delta t_\alpha \rangle} = \rho_{\max} Q \langle l_\alpha/v_\alpha \rangle = \frac{Q}{\langle l_\alpha \rangle} \langle l_\alpha/v_\alpha \rangle, \quad (5)$$

where  $\rho_{\max} = 1/\langle l_\alpha \rangle$  is the maximum density and  $\langle l_\alpha \rangle$  the average vehicle length. For a finite detector length  $L_D$ , we have to replace  $l_\alpha$  by  $l_\alpha + L_D$  [33,50]. The formula  $1/V = \langle l_\alpha/v_\alpha \rangle/\langle l_\alpha \rangle$  for the average velocity results in the correct expression  $1/V = \langle 1/v_\alpha \rangle$  only, if the individual vehicle lengths and velocities are not correlated, which is usually not the case.

### 1.3 A Quantitative Theory of Congested Traffic States

When Kerner started to question all traffic models with a fundamental diagram, physicists were used to simulate traffic in closed systems with periodic boundary conditions. With the Kerner-Konhäuser model, it was possible to produce free traffic, emergent stop-and-go waves, and triggered wide moving jams [51–53]. However, attempts to simulate “synchronized flow” failed even when small ramp flows were added to the system. They resulted in what we call a *pinned localized cluster* (PLC) located at the on-ramp [54] (see Figs. 9 and 10). Because of the sensitivity of the model and problems with the treatment of open systems, it was not possible to simulate open systems with considerable ramp flows. Other independent studies for periodic systems with localized bottlenecks produced either homogeneous vehicle queues (HCT) or *oscillating congested traffic* (OCT) [17,55–64], but at that time nobody could make sense of these apparently inconsistent findings. This situation changed, when Helbing *et al.* derived a phase diagram of congested traffic states. They managed to simulate a macroscopic traffic model with open boundary conditions even in extreme situations and investigated a freeway stretch with a single ramp [30]. Instead of the densities, they identified the main flow on the freeway and the on-ramp flow as the suitable control parameters for an open system and varied them systematically. In this way, they found that a perturbation could trigger different kinds of congested traffic states.

Moreover, the boundaries separating different states could be related to the *instability diagram* for homogeneous freeways and other characteristic quantities [30]. For this reason, they concluded that the phase diagram should be qualitatively the same, i.e. *universal*, for all microscopic, mesoscopic, or macroscopic traffic models having the same instability diagram. This has been supported in the meantime [30,35,65]. Apart from this, the phase diagram of traffic models with different instability diagrams can be directly derived [33]. Generalizations to other kinds of bottlenecks (e.g. gradients) have been developed as well [35].



**Fig. 7.** Negative perturbation triggering oscillating congested traffic. When the traffic density has sufficiently increased to reach the metastable regime, the “negative” perturbation will be amplified if it only exceeds the critical amplitude. While it is small, it will move downstream with the vehicles, so one could hope it would pass the bottleneck and leave the system. However, when the density wave grows larger, it will reduce its speed and even change its propagation direction. Once it is fully developed, it moves upstream with constant velocity, since vehicles leave the jam at the downstream front, while new ones join it at the upstream front. Hence, the perturbation returns to the bottleneck (see Fig. 8 for this “*boomerang effect*”), and it triggers a breakdown of traffic, when it passes the bottleneck in upstream direction, as it thereby reduces the effective capacity to  $Q_{out}$  [33,44,70].

In the following, we will sketch the basic ideas behind the phase diagram of congested traffic states (for a more detailed discussion see Ref. [33]). Let us assume our traffic model has a fundamental diagram

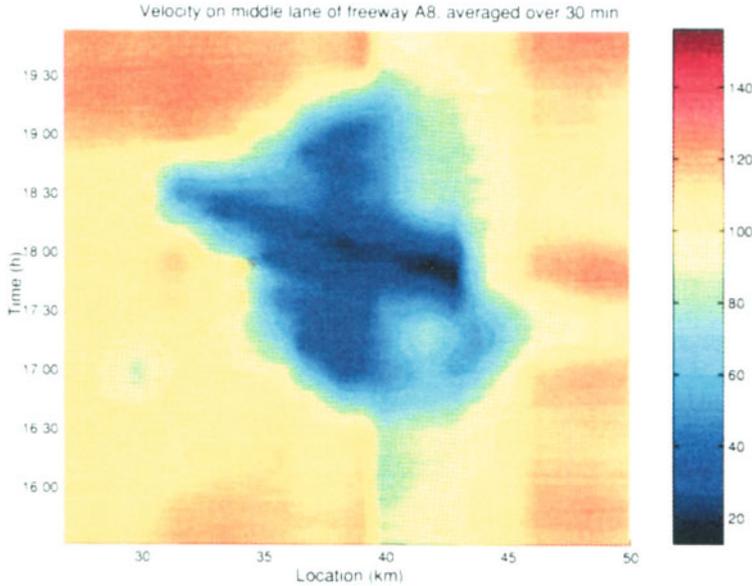
$$Q_f(\rho) = \rho V_f(\rho) \tag{6}$$

describing the relation between the vehicle flow  $Q$ , the *spatially averaged* density  $\rho$ , and the average velocity  $V$  in homogeneous and stationary traffic. (The flow-density relation of emergent stop-and-go waves is characterized by a linear relation, i.e. it looks different [53].) Moreover, let us assume the model has

ranges of stable traffic flow at small and high densities, a range of linearly unstable traffic flow at medium densities, and ranges of metastable traffic flow in between. This kind of instability diagram is, for example, found for the macroscopic model used by Kühne, Kerner and Konhäuser, or Lee *et al.* [17,51,53], for the microscopic optimal velocity model [66], for the non-local gas-kinetic-based traffic model [67], or the microscopic intelligent driver model [35] (among which the first two models are rather sensitive to parameter variations, but the latter two are quite robust).

In contrast to circular freeways, emergent “phantom traffic jams” are not typical for open homogeneous freeway stretches, as it is normally impossible to reach the linearly unstable density regime by feeding in vehicles at the upstream boundary. This is in agreement with empirical observations [68]. Most cases of traffic congestion on an  $n$ -lane freeway are observed upstream of on-ramps or other bottlenecks. They can be triggered by perturbations significantly below the theoretical capacity, as soon as the sum of the upstream freeway flow  $Q_{\text{up}}$  and the on-ramp flow  $\Delta Q = Q_{\text{rmp}}/n$  per lane exceeds the outflow  $Q_{\text{out}}$  from congested traffic: If a disturbance leads to temporary congestion, the drivers must accelerate again and suffer some time delay, which reduces the capacity to  $Q_{\text{out}}$ . Therefore, the following vehicles will queue up, and the temporary perturbation grows to form a persistent kind of congestion. The initial perturbation can even be a temporary *reduction* of the traffic flow and/or vehicle density, which can be caused by temporal variations of the traffic volume or even by vehicles leaving the freeway at some off-ramp [69,70], see Fig. 8.

If the total traffic volume  $Q_{\text{tot}} = (Q_{\text{up}} + \Delta Q)$  is greater than the dynamic capacity  $Q_{\text{out}}$ , we will automatically end up with a growing vehicle queue upstream of the on-ramp. The traffic flow  $Q_{\text{cong}}$  resulting in the congested area normally gives, together with the inflow  $\Delta Q$ , the outflow  $Q_{\text{out}}$ , i.e.  $Q_{\text{cong}} = (Q_{\text{out}} - \Delta Q)$  (but see the footnote on p. 1111 of Ref. [33] for exceptions). One can distinguish the following cases [30,35,33] (see Fig. 9): If the density  $\rho_{\text{cong}}$  associated with the flow  $Q_{\text{cong}} = Q_f(\rho_{\text{cong}})$  is stable, we find *homogeneous congested traffic* (HCT) such as typical traffic jams during holiday seasons. For a smaller on-ramp flow  $\Delta Q$ , the congested flow  $Q_{\text{cong}}$  is linearly unstable, and we either find *oscillating congested traffic* (OCT) or *triggered stop-and-go traffic* (TSG), which often emerges from a spatial sequence of homogeneous and oscillating congested traffic (so-called “*pinch effect*” [36]). In contrast to OCT, stop-and-go traffic is characterized by a sequence of moving jams, between which traffic flows freely. Each traffic jam triggers the next one by inducing a small perturbation at the ramp, which propagates downstream as long as it is small, but turns back when it has grown large enough (“*boomerang effect*”, cf. Figs. 7 to 10). This, however, requires the downstream traffic flow to be linearly unstable. If it is metastable instead (when the traffic volume is further reduced), a traffic jam will usually not trigger a growing perturbation. In that case, one finds either a single *moving localized cluster* (MLC), or a *pinned localized cluster* (PLC) at the location of the ramp. The latter requires the traffic flow in the upstream section to be sta-

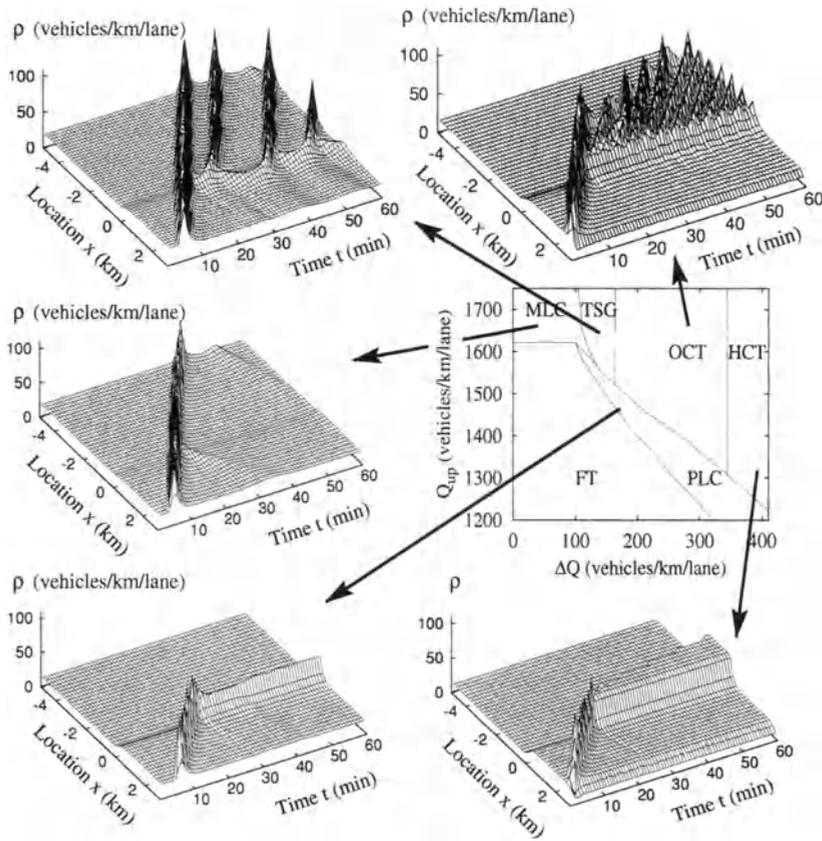


**Fig. 8.** The wide moving jam left of kilometer 43 starts with a “boomerang effect” and travels through the “synchronized” congested traffic flow left of kilometer 41 (dark area). (Reproduction with kind permission of Rudolf Sollacher, Siemens AG, Munich.)

ble, so that no traffic jam can survive there. Finally, for sufficiently small traffic volumes, we find *free traffic* (FT), as expected.

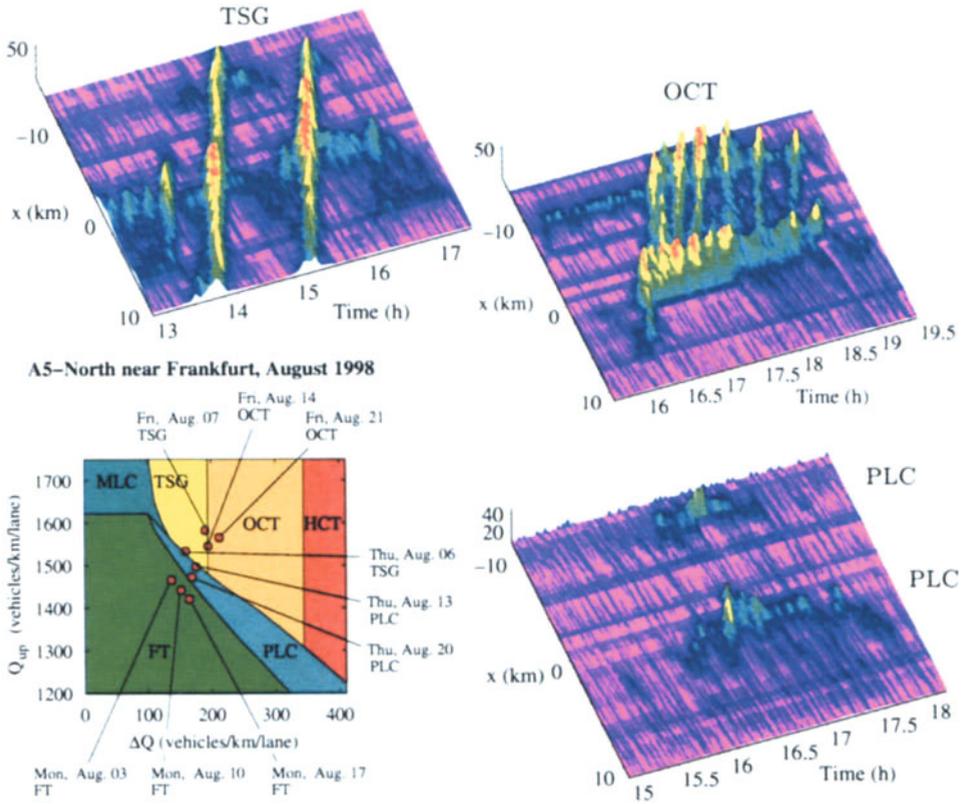
The different congested traffic states found in the microsimulations (as displayed in Fig. 9) could all be identified in real traffic data (see Fig. 10 for some examples). Moreover, according to our first investigation results, the traffic patterns observed on the German freeway A5 near Frankfurt have a typical dependence on the respective weekday and are even quantitatively consistent with the phase diagram (see Fig. 10). Of course, the empirically measured patterns look less regular, as the simulation results displayed in Fig. 9 are for a deterministic model with identical vehicle parameters.

We will now reply to some criticism and misunderstandings: (1) Although the phase diagram and the congested traffic states have been derived for identical driver-vehicle units and one bottleneck only, many observations already fit very well into this scheme. If several bottlenecks are present, the situation becomes more complicated, but can be addressed by similar methods. In such cases, we may find the spatial *coexistence* of states such as OCT and PLC, *transitions* between different states, extended congested traffic states (HCT, OCT, or TSG) with a *fixed upstream front*, and other phenomena. The phenomenon of multistability and coexisting states is, by the way, already found for the case of one single bottleneck (see Ref. [65] and Fig. 8 in Ref. [35]). (2) Because of a



**Fig. 9.** Simulated representatives (density-over-space-and-time plots) of the different congested traffic states, which were triggered by a big perturbation travelling upstream. Center right: Phase diagram of the traffic states as a function of the (upstream) traffic volume  $Q_{up}$  on the freeway and the strength  $\Delta Q$  of a bottleneck at location  $x = 0$  km, e.g. an on-ramp with inflow  $\Delta Q$  per freeway lane.

certain “penetration depth” [35], MLC states can propagate through small areas of stable traffic. (3) The variation and scattering of the flow-density data is well reproduced, if different driver-vehicle types are distinguished (and overtaking maneuvers are taken into account). (4) The “pinch effect” does not correspond to a transition between different phases (HCT, OCT, and TSG) in the course of time, but it corresponds to a certain area in the phase diagram, where the congested flow is convectively stable but linearly unstable (not shown). The empirically observed increase of the vehicle velocity in “synchronized flow” with the average oscillation wavelength [36] is qualitatively well reproduced, see Fig. 3(b). (5) The flow downstream of congested traffic can exceed the dynamical capacity



**Fig. 10.** Empirical examples of triggered stop-and-go traffic (TSG), oscillating congested traffic (OCT), and pinned localized clusters (PLC), and the location of the empirical data points in the phase diagram for the German freeway A5 near Frankfurt.

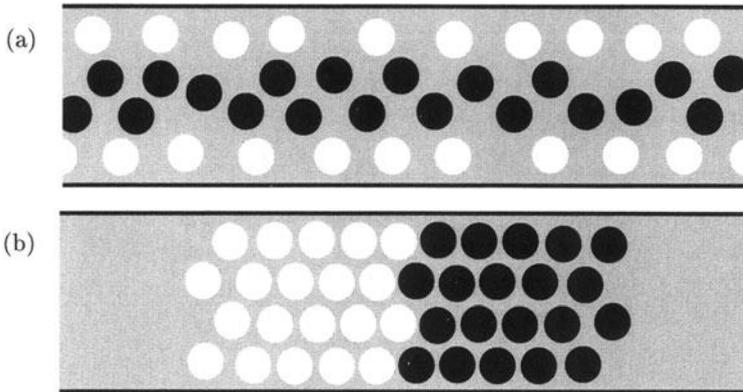
$Q_{out}$ , if vehicles can enter the freeway via the ramp downstream of the congestion front [21]. This is practically relevant for particularly long on-ramps like freeway intersections (see Footnote on p. 1111 of Ref. [33]).

## 2 Pedestrian Evacuation

The topics of pedestrian traffic and evacuation of buildings, stadia, and ships have recently attracted great interest among traffic scientists. Here, we will give a short overview only, as there are several detailed reviews available (see Refs. [22,33,71–73]). We will focus on the *social-force model* of pedestrian dynamics which describes the different competing motivations of pedestrians by separate force terms. It has the following advantages:

- (1) The social-force model takes into account the flexible usage of space (i.e. the compressibility of the crowd), but also the excluded volume and friction effects which play a role at extreme densities.
- (2) The model assumptions are simple

and plausible. Virtual fields [74] or other questionable model ingredients are not necessary to obtain realistic results. (3) There are only a few model parameters to calibrate. (4) The model is robust and naturally reproduces many different observations without modifications of the model. (5) Nevertheless, it is easy to consider individual differences in the behavior, and extensions for more complex problems such as trail formation [22,71,75] are possible.

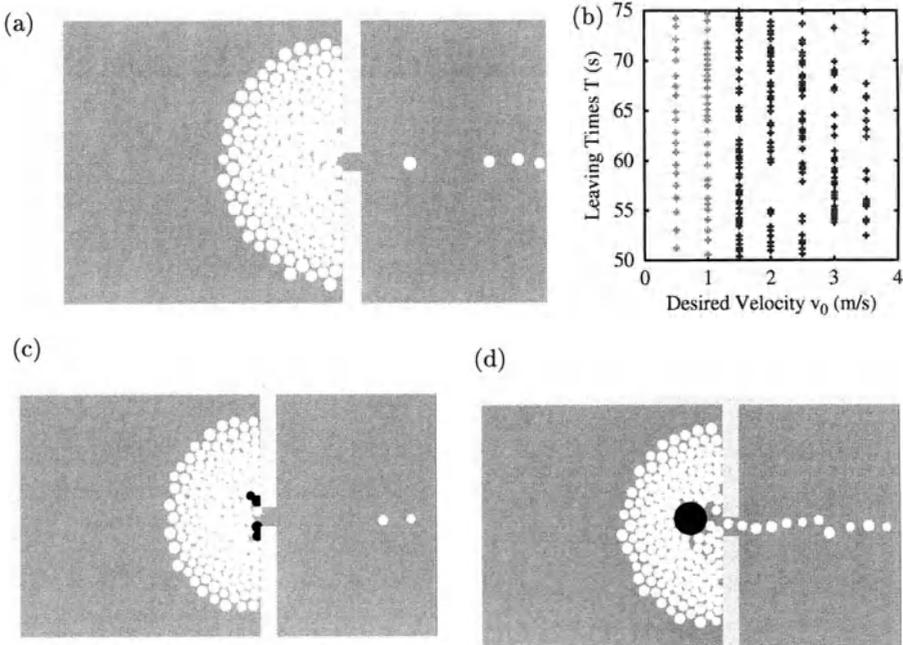


**Fig. 11.** (a) Formation of lanes in initially disordered pedestrian crowds with opposite walking directions and small noise amplitudes  $\eta_i$  (after [33,73,76]; cf. also [77,78]). White disks represent pedestrians moving from left to right, black ones move the other way round. (b) For sufficiently high densities and large fluctuations, we observe the noise-induced formation of a crystallized, “frozen” state (after [33,76]).

For normal, “relaxed” situations with small fluctuation amplitudes, our microsimulations of counterflows in corridors reproduce the empirically observed *formation of lanes* consisting of pedestrians with the same desired walking direction [33,76–78], see Fig. 11(a). If we do not assume periodic boundary conditions, these lanes are dynamically varying (see the Java applet at [www.helbing.org/Pedestrians/Corridor.html](http://www.helbing.org/Pedestrians/Corridor.html)). Their number depends on the width of the street [77], on pedestrian density, and on the noise level. Interestingly, one finds a *noise-induced ordering* [79]: Compared to small noise amplitudes, medium ones result in a more pronounced segregation (i.e., a smaller number of lanes). Large noise amplitudes lead to a “freezing by heating” effect characterized by a breakdown of “fluid” lanes and the emergence of “solid” blockages [76], see Fig. 11(b). Note that our model can explain lane formation even without assuming asymmetrical interactions or attraction effects [76–78]. It is an *optimal self-organization* phenomenon [78] resulting from the combination of driving and repulsive forces. The same model also reproduces the observed oscillations of the flow direction at bottlenecks [33,70,77] without the need of a virtual “floor field” [80,81]. *Cellular automaton* Java applets from 1998 are available in the internet to visualize these phenomena (see [www.helbing.org/Pedestrians/](http://www.helbing.org/Pedestrians/)

Corridor.html, Door.html). They are based on a discretization of the social force model that can be viewed as a discrete two-dimensional optimal velocity model [82]. For other cellular automata see Ref. [72,83–86].

“Freezing by heating” is one of the phenomena observed in crowd stampedes. Another one is the “faster-is-slower effect” (or “slower-is-faster effect”) [87]. It can trigger a “phantom panic” [87] and is caused by arching and clogging at bottlenecks like exits, see Fig. 12(a). The underlying reason is the friction effect

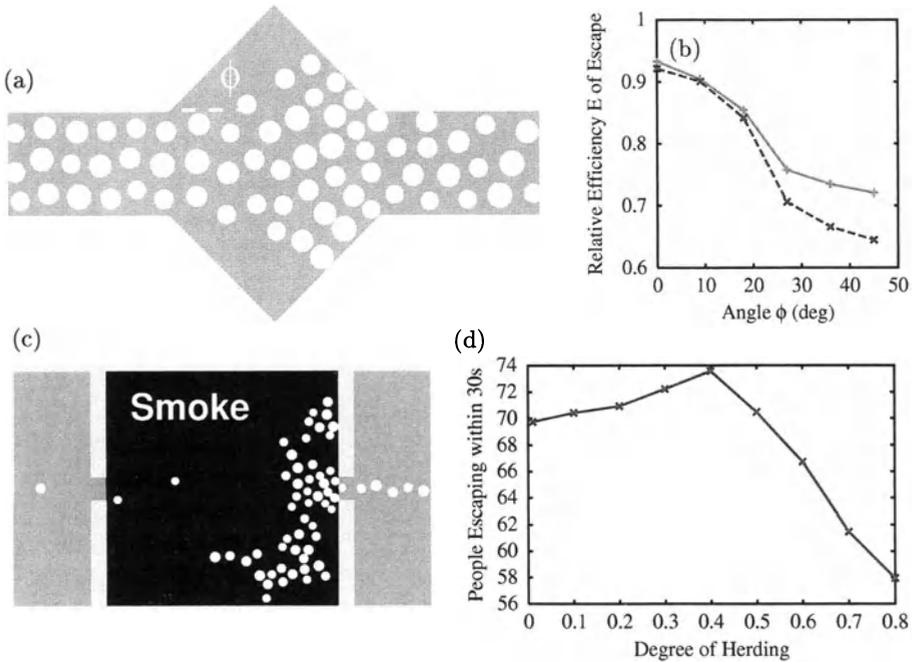


**Fig. 12.** (a) When the desired velocity  $v_0$  is too high (e.g. in panic situations), pedestrians come so close to each other, that their physical contacts cause the build up of pressure and obstructing friction effects, which results in temporary arching and clogging (for online Java simulations see <http://angel.elte.hu/~panic/>). (b) This is related with an inefficient and irregular outflow, while the outflow is regular for small enough desired velocities ( $v_0 \leq 1.5$  m/s) [33,73,87]. (c) In panicking crowds, high pressures build up due to physical interactions. This can injure people (black disks), who turn into obstacles for other pedestrians trying to leave [73]. (d) A column in front of the exit (large black disk) can avoid injuries by taking up pressure from behind. It can also significantly increase the outflow [73]. In large exit areas used by several hundred people, several randomly placed columns are needed to subdivide the crowd and the pressure. An asymmetric configuration of the columns is most efficient, as it avoids equilibria of forces which may temporarily stop the outflow. (See <http://angel.elte.hu/~panic/> for online Java simulations.)

occurring in dense crowds, if the desired velocity is so high that pedestrians have

physical interactions. In these situations, extreme pressure can build up in the crowd, and people may be crushed or trampled, in this way turning into obstacles for other fleeing people, see Fig. 12(c). These dangerous pressures can be reduced by columns, when suitably placed in front of the exits, see Fig. 12(d). Thereby, the number of injuries can be significantly reduced, and the outflows are considerably increased (see the Java applets at <http://angel.elte.hu/~panic/>).

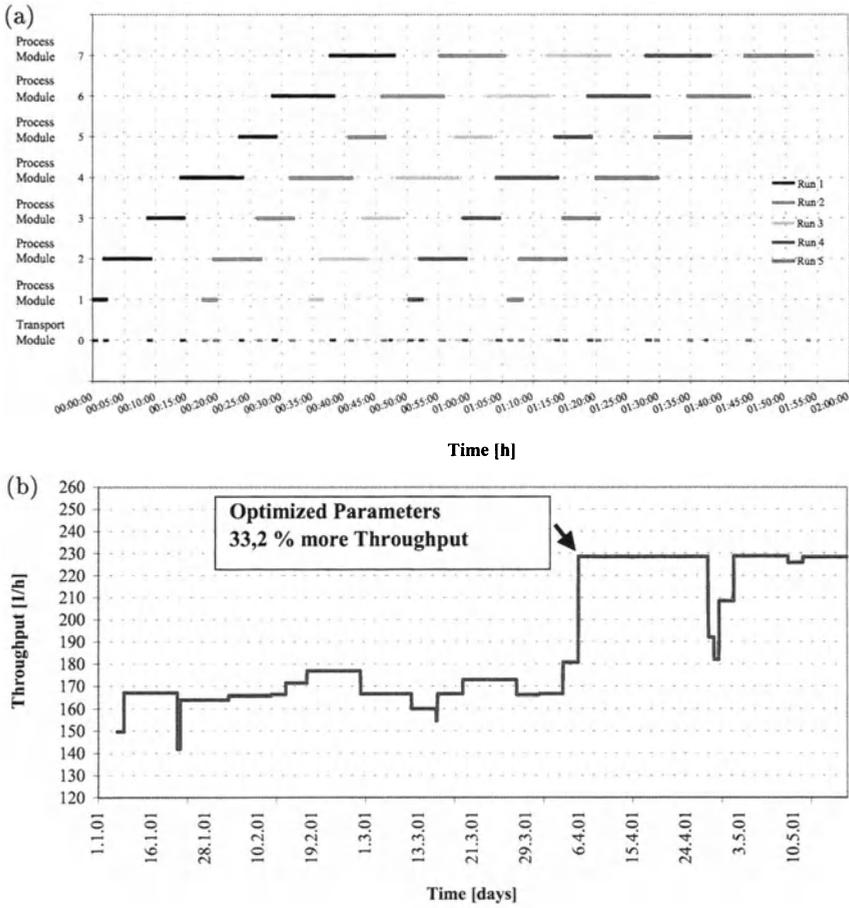
Not only bottlenecks are dangerous in panic situations, but also widenings [87]. These can reduce the efficiency of leaving, see Figs. 13(a), (b). Another problem is *herding behavior*, as it is responsible for an inefficient use of available exits [33,87], see Figs. 13(c), (d).



**Fig. 13.** (a), (b) Drop of the efficiency of leaving in corridors with widenings [73,87]. (c), (d) Herding behavior of panicking pedestrians in a smoky room (black), leading to an inefficient use of available escape routes [73,87]. (Online Java simulations of this effect are provided at <http://angel.elte.hu/~panic/>.)

### 3 Summary and Application to Optimal Production Processes

Nowadays, most aspects of traffic dynamics have been understood on the basis of self-driven many-particle models. The observed phenomena can be (semi-)quan-



**Fig. 14.** (a) Gantt diagram illustrating the treatment times in different modules, corresponding to different production steps. The limiting factor for increasing the throughput is the transport module. (b) Applying the knowledge of the “slower-is-faster effect” to the treatment times, we were able to increase the throughput of some production processes in a major Chip factory by 33% or even more [88].

tatively reproduced by simulations using measured boundary conditions [35]. Moreover, a universal phase diagram of congested traffic states for freeway sections with one bottleneck has been found, and the generalization to more complex situations is straightforward. The reproduction of fine details, however, will require a more detailed measurement of the interactive vehicle dynamics and the consideration of psychological aspects. Although these may also be described in a mathematical way [43], it will be hardly possible to prove or disprove the corresponding models, i.e. the criteria demanded in the natural sciences would have to be relaxed. A more promising research direction is the modelling and optimization of production processes. For example, applying the knowledge of the

“slower-is-faster effect” to the treatment times in a series of production steps, we were able to increase the throughput of production processes in a major Chip factory by up to 39% [88].

Many conclusions from traffic models are relevant for the organization of societies, administrations, companies, production processes, and so on, as the basic model ingredients agree: (1) The system consists of a large number of similar *units/entities* (individuals, pedestrians, cars, boxes, ...). (2) The units are externally or internally *driven*, i.e. there is some energy input, e.g., they can move. (3) Units are *non-linearly interacting*, i.e. under certain conditions, small variations can have large effects. The system behavior is dominated by interactions rather than by the boundary conditions (the external control). (4) There is a *competition for resources* such as time (slots), space, money. (5) Each unit has a certain *extension* in space or time. (6) When units come too close, they have *frictional effects* and obstruct each other.

## References

1. B. S. Kerner, H. Rehborn: *Phys. Rev. E* **53**, R4275 (1996).
2. B. S. Kerner, H. Rehborn, M. Aleksic, A. Haug, R. Lange: *Straßenverkehrstechn.* **10**, 521 (2000).
3. B. S. Kerner, H. Rehborn, M. Aleksic: in *Traffic and Granular Flow '99*, edited by D. Helbing, H. J. Herrmann, M. Schreckenberg, and D. E. Wolf (Springer, Berlin 2000), pp. 339–344.
4. L. Neubert, L. Santen, A. Schadschneider, M. Schreckenberg: *Phys. Rev. E* **60**, 6480 (1999).
5. H. S. Mika, J. B. Kreer, L. S. Yuan: *Highw. Res. Rec.* **279**, 1 (1969).
6. B. S. Kerner, H. Rehborn: *Phys. Rev. E* **53**, R1297 (1996).
7. M. J. Cassidy, M. Mauch: *Transpn. Res. A* **35**, 143 (2001).
8. B. S. Kerner, S. L. Klenov, P. Konhäuser: *Phys. Rev. E* **56**, 4200 (1997).
9. B. N. Persaud: ‘Study of a freeway bottleneck to explore some unresolved traffic flow issues’, Ph.D. thesis, University of Toronto (1986).
10. J. H. Banks: *Transpn. Res. Rec.* **1287**, 20 (1990).
11. J. H. Banks: *Transpn. Res. Rec.* **1320**, 83 (1991).
12. B. S. Kerner, H. Kirschfink, H. Rehborn: German Patent DE 196 47 127.3; US Patent US 5,861,820 (1999).
13. M. J. Lighthill, G. B. Whitham: *Proc. Roy. Soc. London, Ser. A* **229**, 317 (1955).
14. M. Koshi, M. Iwasaki, I. Ohkura: In: *Proceedings of the 8th International Symposium on Transportation and Traffic Flow Theory*, ed. by V. F. Hurdle, E. Hauer, G. N. Stewart (University of Toronto, Toronto, Ontario 1983) pp. 403.
15. L. C. Edie, R. S. Foote: *High. Res. Board Proc.* **37**, 334 (1958).
16. T. W. Forbes, J. J. Mullin, M. E. Simpson: In: *Proceedings of the 3rd International Symposium on the Theory of Traffic Flow*, ed. by L. C. Edie (Elsevier, New York, N. Y. 1967) pp. 97.
17. H. Y. Lee, H.-W. Lee, D. Kim: *Phys. Rev. Lett.* **81**, 1130 (1998).
18. B. S. Kerner, H. Rehborn: *Phys. Rev. Lett.* **79**, 4030 (1997).
19. V. Shvetsov, D. Helbing: *Phys. Rev. E* **59**, 6328 (1999).
20. D. Helbing, B. A. Huberman: *Nature* **396**, 738 (1998).
21. B. S. Kerner: *Networks and Spatial Economics* **1**, 35 (2001).

22. D. Helbing: *Verkehrsdynamik* (Springer, Berlin 1997).
23. D. Helbing: *Phys. Rev. E* **55**, 3735 (1997).
24. J. Treiterer, J. A. Myers: In: *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, ed. by D. Buckley (Reed, London 1974) pp. 13.
25. H. Payne: *Transpn. Res. Rec.* **971**, 140 (1984).
26. F. L. Hall: *Transpn. Res. A* **21**, 191 (1987).
27. F. L. Hall, K. Agyemang-Duah: *Transpn. Res. Rec.* **1320**, 91 (1991).
28. B. Persaud, S. Yagar, R. Brownlee: *Transpn. Res. Rec.* **1634**, 64 (1998).
29. D. Westland: In: *Proceedings of the 3rd International Symposium on Highway Capacity*, ed. by R. Rysgaard (Road Directorate, Denmark 1998) pp. 1095.
30. D. Helbing, A. Hennecke, M. Treiber: *Phys. Rev. Lett.* **82**, 4360 (1999).
31. B. S. Kerner: In: *Proceedings of the 3rd International Symposium on Highway Capacity*, ed. by R. Rysgaard (Road Directorate, Denmark 1998) Vol. 2, pp. 621.
32. H. Y. Lee, H.-W. Lee, D. Kim: *Phys. Rev. E* **62**, 4737 (2000).
33. D. Helbing: *Reviews of Modern Physics* **1067**, 1141 .
34. M. J. Cassidy, R. L. Bertini: *Transpn. Res. B* **33**, 25 (1999).
35. M. Treiber, A. Hennecke, D. Helbing: *Phys. Rev. E* **62**, 1805 (2000).
36. B. S. Kerner: *Phys. Rev. Lett.* **81**, 3797 (1998).
37. M. Treiber, D. Helbing: e-print cond-mat/9901239 (1999).
38. P. Manneville: *Dissipative Structures and Weak Turbulence* (Academic, New York 1990).
39. M. C. Cross, P. C. Hohenberg: *Rev. Mod. Phys.* **65**, 851 (1993).
40. M. Treiber, D. Helbing: *J. Phys. A: Math. Gen.* **32**, L17 (1999).
41. J. H. Banks: *Transpn. Res. Rec.* **1678**, 128 (1999).
42. F. L. Hall, B. L. Allen, M. A. Gunter: *Transpn. Res. A* **20**, 197 (1986).
43. B. S. Kerner, S. L. Klenov, *J. Phys. A: Math. Gen.* **35**, L31 (2002).
44. D. Helbing, M. Treiber: *Phys. Rev. Lett.* **81**, 3042 (1998).
45. J. H. Banks: *Transpn. Res. Rec.* **1510**, 1 (1995).
46. D. H. Hoefs: *Untersuchung des Fahrverhaltens in Fahrzeugkolonnen* (Bundesministerium für Verkehr, Abt. Straßenbau, Bonn-Bad Godesberg 1972).
47. D. Helbing, B. Tilch: *Phys. Rev. E* **58**, 133 (1998).
48. B. Tilch: ‘Modellierung und Simulation selbstgetriebener Vielteilchensysteme mit Anwendung auf den Straßenverkehr’, Ph.D. thesis, University of Stuttgart, in preparation.
49. B. Tilch, D. Helbing: In: *Traffic and Granular Flow '99*, ed. by D. Helbing, H. J. Herrmann, M. Schreckenberg, D. E. Wolf (Springer, Berlin 2000) pp. 333.
50. A. D. May: *Traffic Flow Fundamentals* (Prentice Hall, Englewood Cliffs, NJ 1990).
51. R. D. Kühne: In: *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, ed. by I. Volmuller, R. Hamerslag (VNU Science, Utrecht 1984) pp. 21.
52. R. D. Kühne: In: *Highway Capacity and Level of Service*, Proceedings of the International Symposium on Highway Capacity, ed. by U. Brannolte (Balkema, Rotterdam 1991) pp. 211.
53. B. S. Kerner, P. Konhäuser: *Phys. Rev. E* **50**, 54 (1994).
54. B. S. Kerner, P. Konhäuser, M. Schilke: *Phys. Rev. E* **51**, R6243 (1995).
55. P. K. Munjal, Y.-S. Hsu, R. L. Lawrence: *Transpn. Res.* **5**, 257 (1971).
56. W. F. Phillips: *Kinetic Model for Traffic Flow* (National Technical Information Service, Springfield, VA 22161), Technical Report DOT/RSPD/DPB/50-77/17 (1977).
57. M. Cremer: *Der Verkehrsfluß auf Schnellstraßen* (Springer, Berlin 1979).

58. Y. Makigami, T. Nakanishi, M. Toyama, R. Mizote: In: *Proceedings of the 8th International Symposium on Transportation and Traffic Flow Theory*, ed. by V. F. Hurdle, E. Hauer, G. N. Stewart (University of Toronto, Toronto, Ontario 1983) pp. 427.
59. K. H. Chung, P. M. Hui: *J. Phys. Soc. Jpn.* **63**, 4338 (1994).
60. H. Emmerich, E. Rank: *Physica A* **216**, 435 (1995).
61. Z. Csahók, Z., T. Vicsek: *J. Phys. A: Math. Gen.* **27**, L591 (1994).
62. M. Hilliges: *Ein phänomenologisches Modell des dynamischen Verkehrsflusses in Schnellstraßennetzen* (Shaker, Aachen 1995).
63. A. Klar, R. D. Kühne, R. Wegener: *Surv. Math. Ind.* **6**, 215 (1996).
64. T. Nagatani: *J. Phys. Soc. Jpn.* **66**, L1928 (1997).
65. H. Y. Lee, H.-W. Lee, D. Kim: *Phys. Rev. E* **59**, 5101 (1999).
66. M. Bando, K. Hasebe, K. Nakanishi, A. Nakayama, A. Shibata, Y. Sugiyama: *J. Phys. I France* **5**, 1389 (1995).
67. M. Treiber, A. Hennecke, D. Helbing: *Phys. Rev. E* **59**, 239 (1999).
68. B. S. Kerner: *J. Phys. A.: Meth. Gen.* **33**, L221 (2000).
69. C. F. Daganzo, M. J. Cassidy, R. L. Bertini: *Transpn. Res. A* **33**, 365 (1999).
70. D. Helbing: *Phys. Bl.* **57**, 27 (2001).
71. D. Helbing, P. Molnár, I. Farkas, K. Bolay: *Env. Plan. B* **28**, 361 (2001).
72. M. Schreckenberg, S. D. Sharma, Eds.: *Pedestrian and Evacuation Dynamics* (Springer, Berlin 2002).
73. D. Helbing, I. J. Farkas, P. Molnár, T. Vicsek: In: *Pedestrian and Evacuation Dynamics*, ed. by M. Schreckenberg, S. D. Sharma (Springer, Berlin 2002) pp. 21.
74. C. Burstedde, K. Klauck, A. Schadschneider, J. Zittartz: *Physica A* **295**, 507 (2001).
75. D. Helbing, J. Keltsch, P. Molnár: *Nature* **388**, 47 (1997).
76. D. Helbing, I. Farkas, T. Vicsek: *Phys. Rev. Lett.* **84**, 1240 (2000).
77. D. Helbing, P. Molnár: *Phys. Rev. E* **51**, 4282 (1995).
78. D. Helbing, T. Vicsek: *New J. Phys.* **1**, 13.1 (1999).
79. D. Helbing, T. Platkowski: *Int. J. Chaos Theor. Appl.* **5**, 47 (2000).
80. A. Schadschneider: In: *Pedestrian and Evacuation Dynamics*, ed. by M. Schreckenberg, S. D. Sharma (Springer, Berlin 2002) pp. 75.
81. C. Burstedde, A. Kirchner, K. Klauck, A. Schadschneider, J. Zittartz: In: *Pedestrian and Evacuation Dynamics*, ed. by M. Schreckenberg, S. D. Sharma (Springer, Berlin 2002) pp. 87.
82. K. Bolay: Master's thesis, University of Stuttgart (1998).
83. V. J. Blue, J. L. Adler: *Trans. Res. Rec.* **1644**, 29 (1998).
84. M. Fukui, Y. Ishibashi: *J. Phys. Soc. Jpn.* **68**, 2861 (1999).
85. M. Muramatsu, T. Irie, T. Nagatani: *Physica A* **267**, 487 (1999).
86. H. Klüpfel, M. Meyer-König, J. Wahle, M. Schreckenberg: In: *Theoretical and Practical Issues on Cellular Automata* ed. by S. Bandini, T. Worsch (Springer, London 2000) pp. 63.
87. D. Helbing, I. J. Farkas, T. Vicsek: *Nature* **407**, 487 (2000).
88. D. Fasold: Master's thesis, TU Dresden (2001).

# Cellular Automaton Simulations of Pedestrian Dynamics and Evacuation Processes

A. Kirchner and A. Schadschneider

Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany

**Abstract.** We present applications and numerical results for a bionics-inspired cellular automaton approach to pedestrian dynamics [1,2]. The model is able to reproduce collective effects and self-organization phenomena encountered in pedestrian traffic, e.g. lane formation in counterflow. Here we present an analysis of the quantitative impact of the so-called sensitivity parameters  $k_S$  and  $k_D$  during evacuation processes. Furthermore the model is applied to the problem of optimization of evacuation times.

## 1 Introduction

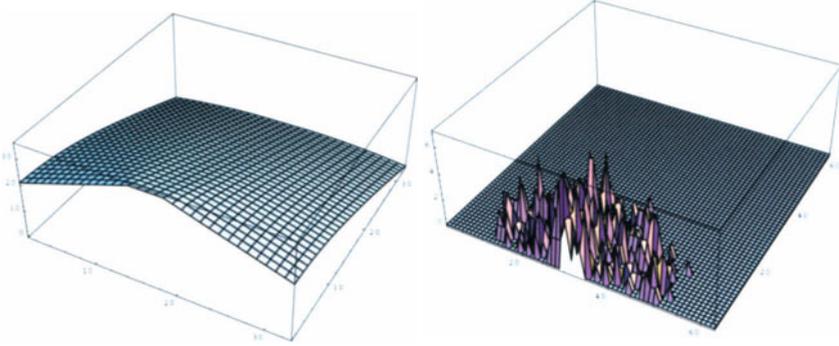
In [1] a new two-dimensional cellular automaton model for the description of pedestrian dynamics has been introduced. Further details, especially concerning the definition of the model, can be found elsewhere in these proceedings [2]. The basic idea is similar to the phenomenon of chemotaxis. The transition probabilities for the motion of the pedestrians are determined using so-called *floor fields* in such a way that a movement in direction of higher fields is more likely. The *static floor field*  $S$  does not evolve with time and is not changed by the presence of the pedestrians. Such a field can be used to specify regions of space which are more attractive. The *dynamic floor field*  $D$  is a virtual trace left by the pedestrians and has its own dynamics, i.e. diffusion (described by a diffusion constant  $\alpha \in [0, 1]$ ) and decay (with decay constant  $\delta \in [0, 1]$ ). It is used to model an attractive interaction between the particles.

Here we want to show the model's usefulness for deepening our understanding of the complex problems of evacuation processes. Realistic procedures of evacuations are in general not uniform, highly ordered behaviour of the pedestrians as well as panic situations are observed. As we will see in the following, the model is able to reproduce many of these collective social phenomena.

We will focus on evacuation processes from a large room with one or more doors. Fig. 1(a) shows  $S$  for a lattice of size  $33 \times 33$  with one exit of width five cells. For each lattice site  $S$  is calculated using a special distance metric so that the field values are increased in the direction to the door. Fig. 1(b) shows a 3-d plot of the dynamic field  $D$  in the final stages of an evacuation process.

## 2 Evacuation Simulations

In the following we describe results of simulations of a typical situation, i.e. the evacuation of a large room (e.g. in case of fire). At this we focus on the influence



**Fig. 1.** Evacuation from a large room with a single door: (a) static floor field  $S$ ; (b) dynamic floor field  $D$  in the final stages of the evacuation.

of the sensitivity parameters  $k_D$  and  $k_S$  on the evacuation times. As we will see the variation of these parameters leads to many interesting collective phenomena between the pedestrians. Concluding we give an example of the relevance of our model for practical applications at safety estimations in architectural planning.

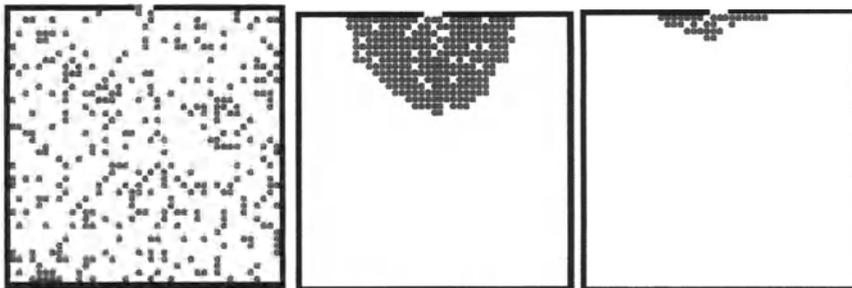
## 2.1 The Impact of Sensitivity Parameters

The value of  $k_S$ , the coupling to the static field, can be viewed as a measure of the knowledge of the pedestrians about the location of the exit. A large  $k_S$  implies a motion to the exit on the shortest possible path. For vanishing  $k_S$ , on the other hand, the people will perform a random walk and just find the exit by chance. So the case  $k_S \ll 1$  is relevant for processes in dark or smoke-filled rooms where people do not have full knowledge about the location of the exit.

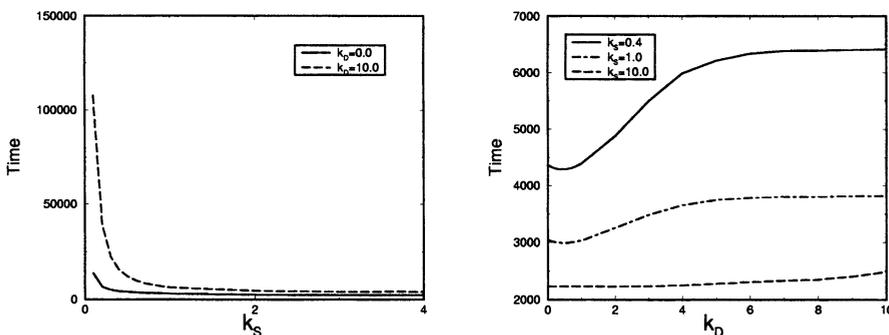
The parameter  $k_D$  for the coupling to the dynamic field controls the tendency to follow the lead of others. A large value of  $k_D$  implies a strong herding behaviour which has been observed in the case of panics [3].

We consider a grid of size  $63 \times 63$  sites with an exit of one cell in the middle of one wall. The particles are initially distributed randomly and try to leave the room. The only information they get is through the floor fields. Fig. 2 shows typical stages of the dynamics for an initial particle density of  $\rho = 0.3$ , which means 1116 particles. In the middle picture of Fig. 2 a half-circle jamming configuration in front of the door is easy to spot. A typical feature of the dynamics is a radial motion of ‘holes’ created by particles escaping through the door.

The evacuation times are averaged over 500 samples and strongly depend on the sensitivity parameters  $k_D$  and  $k_S$ . Fig. 3(a) shows the evacuation times for fixed sensitivity parameter  $k_D$  of the dynamic field and variable sensitivity parameter  $k_S$  of the static field. The averaged evacuation times are always measured in update time steps. With the generic timescale of the model [1,2] it is easy to translate that into a real time value. Let us first consider the case of  $k_D = 0$ , i.e. no coupling to the dynamic field. In Fig. 3(a) one can see the influence of  $k_S$ . For  $k_S \rightarrow 0$  the pedestrians do not sense the strength of the field.



**Fig. 2.** Typical stages of the dynamics: (a) beginning evacuation ( $t = 0$ ); (b) middle stages; (c) end stage of evacuation with only a few particles left



**Fig. 3.** Averaged evacuation times for a large room (Fig. 2) with an initial particle density of  $\rho = 0.3$  and  $\delta = 0.3$ ,  $\alpha = 0.3$  for (a) fixed  $k_D$ , and (b) fixed  $k_S$ .

Therefore they do not have any guidance through the surroundings and perform a pure random walk which leads to a maximal value of the evacuation time for  $k_S = 0$ . For  $k_S \rightarrow \infty$  they have full information about the shortest distance to the door and the evacuation time converges towards a minimal value. The movement of the particles becomes almost deterministic. Therefore  $k_S$  can be interpreted as some kind of inverse temperature for the degree of information about the inanimate surrounding of the pedestrians.

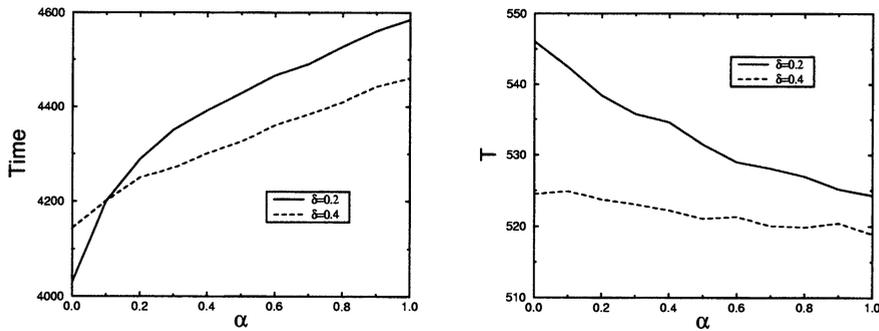
In the same way the sensitivity parameter  $k_D$  of the dynamic field works as an inverse temperature for the information about the virtual trace. If  $k_S$  is turned on from zero to infinity, a non-zero value of  $k_D$  only means additional noise to the pedestrians, the evacuation times are increased for higher coupling strength to  $k_D$  (see Fig. 3(a)).

Much more interesting is the behavior for fixed  $k_S$ . In this case the evacuation times for various sensitivity parameters  $k_D$  are shown in Fig. 3(b). They converge to maximal values for growing sensitivity parameter  $k_D$  of the dynamic field. The most interesting point is the occurrence of minimal evacuation times for

non-vanishing small values of the sensitivity parameter  $k_D$  of the dynamic field. This effect also depends on the value of the diffusion parameter  $\alpha$  and becomes most pronounced in the limit  $\alpha \rightarrow 0$ . Therefore, a small interaction with the dynamic field, which is proportional to the velocity-density of the particles, is of advantage. It represents some sort of minimal intelligence of the pedestrians. They are able to detect regions of higher local flow and minimize their waiting times.

If the coupling to the dynamic field is further increased, the evacuation times increase again from the small regime of the minimal times to maximal values. The interaction with other pedestrians becomes more and more unfavourable, one can compare this to the arising of a panic situation. The weaker the coupling to the static field  $S$  is, the higher are the evacuation times: the particles then have less information of the unanimate surrounding (for example they cannot find the way to the door because of smoke in a fire situation).

The influence of the diffusion parameter  $\alpha$  of the dynamic field in the regime of minimal evacuation times strongly depends on the particle density  $\rho$ . In



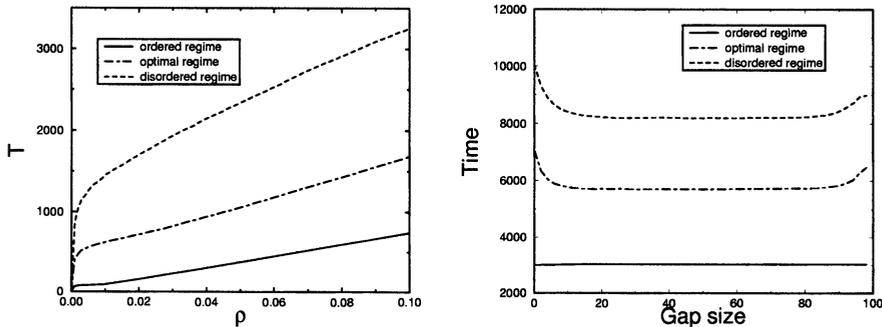
**Fig. 4.** Averaged evacuation times for a large room (Fig. 2) for two densities, variable  $\alpha$  and  $k_S = k_D = 0.4$ : (a)  $\rho = 0.3$ ; (b)  $\rho = 0.004$ .

Fig. 4(a) one can see that in the high density regime an increasing diffusion parameter  $\alpha$  increases the evacuation times for arbitrary decay parameter  $\delta$ . In that case diffusion of the field only means noise for the particles. In the low density regime one finds the opposite behaviour: an increasing value of  $\alpha$  leads to a minimization of the evacuation time. In the low density regime the diffusion of the dynamic field therefore leads to a higher degree of information for the particles and to favorably long-ranged interactions between them.

Thus three main regimes for the behaviour of the particles can be distinguished. For strong coupling to  $k_S$  and very small coupling to  $k_D$  we find an *ordered regime* where particles only react to the static floor field and the behaviour is in some sense deterministic. The *disordered regime* characterized by strong coupling to  $k_D$  and weak coupling to  $k_S$  leads to a maximal value of

the evacuation time. The behaviour here is typical for panic situations. Between these two regimes an *optimal regime* exists where the combination of interaction with the static and the dynamic floor fields minimizes the evacuation time.

Fig. 5(a) shows the influence of an increasing particle density for the evacuation times in the three main regimes. In all regimes the curves increase almost



**Fig. 5.** Average evacuation times for a large room (a) with one door as a function of the density and (b) with two doors as function of the gap between the doors.

linearly. Only for very small densities one finds a very strong increase of the evacuation times with increasing density. For a much more detailed investigation of all this, see [4].

## 2.2 Room with Two Doors

As a simple example for safety estimations in architectural planing we investigate how evacuation times change if a gap between two doors is increased from zero to a maximal value. For that we consider again an ordinary room with no internal structure of grid size  $102 \times 102$ . We start with one door of width two cells in the middle of one wall. Then the door is split into two doors of width one cell each which are separated by gaps ranging from 2 to 98 (Fig. 6(b) and(c)). Fig. 6 shows typical stages of the dynamics for different gap sizes. Averaged evacuation times are measured for all three main regimes introduced in Sec. 2.1. Depending on the regime, Fig. 5(b) shows a strong influence of the size of the gap for small and for large gaps. In the ordered regime, the evacuation time is almost uninfluenced by the gap size. The reason is the almost deterministic movement of the particles which do not interact with each other through the dynamic floor field  $D$ . In the other two regimes, however, the value of the gap can have a large influence due to the strong interaction effects through  $D$ . For both the optimal and the disordered regime one finds minimized evacuation times for a wide area of the gap length ranging from about 20 to 80 cells. For small gaps the evacuation times increase due to the interactions between the pedestrians. For large gaps



**Fig. 6.** (a) No gap between doors, (b) medium gap size between doors, (c) doors at the boundaries.

the presence of the side walls has a negative effect (Fig. 6(c)) and so again the evacuation times increase. For intermediate gaps the crowd of pedestrians will be subdivided into smaller parts (Fig. 6(b)), leading to more favorable interactions between them.

### 3 Conclusions

We have studied simple evacuation processes using a recently introduced stochastic cellular automaton for pedestrian dynamics which makes use of an idea similar to chemotaxis. Due to its simplicity the model allows very high simulation speeds and is very well suited for the optimization of evacuation procedures in complex situations.

We have focussed on the influence of the coupling strengths  $k_S$  and  $k_D$  on the evacuation time. Interesting nontrivial collective phenomena could be observed, e.g. the nonmonotonic dependence of the evacuation time on the coupling to the dynamic floor field. A more detailed account of our results can be found in [4].

### References

1. C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz, *Physica* **A295**, 507 (2001).
2. A. Schadschneider, *Bionics-inspired cellular automaton model for pedestrian dynamics*, these proceedings.
3. D. Helbing, I. Farkas, and T. Vicsek, *Nature* **407**, 487 (2000).
4. A. Kirchner and A. Schadschneider, in preparation.

# Modeling Pedestrians and Granular Flow in 2-Dimensional Optimal Velocity Models

Y. Sugiyama<sup>1</sup>, A. Nakayama<sup>2</sup>, and K. Hasebe<sup>3</sup>

<sup>1</sup> Division of Mathematical Science, City College of Mie, Tsu, Mie 514-0112, Japan

<sup>2</sup> Gifu Keizai University, Ohgaki, Gifu 503-8550, Japan

<sup>3</sup> Faculty of Business Administration, Aichi University, Miyoshi, Aichi 470-0296, Japan

**Abstract.** We extend the Optimal Velocity Model for 1-dimension car following to a model in 2-dimensional space. We present the possibilities of applying the model to the investigation of several kinds of phenomena of collective behaviors, such as granular flow and pedestrian dynamics.

## 1 General Formulation of 2-*d* OV Models

The Optimal Velocity Model is one of the successful models for describing the dynamics of traffic flow [1,2]. It can be extended into several directions. Here, we try to develop a 2-dimensional version. And we apply the model to the several kinds of phenomena of collective motion of interacting particles, such as granular flow in a pipe. Actually the pseudo 1-*d* model [3] is very similar to a 1-*d* OV model, or pedestrian dynamics in a passage and the bio-motion, etc. We express two characteristic formulae of the 2-*d* OV model as the extension of the formula of the 1-*d* OV model in 2-dimensional space.

• Model-1

$$\frac{d^2}{dt^2} \mathbf{x}_n(t) = a \left\{ \sum_m^{N(n)} V(r_{mn}(t)) P(\theta_{mn}(t)) \mathbf{e}_{mn}(t) - \frac{d}{dt} \mathbf{x}_n(t) \right\} \quad (1)$$

• Model-2

$$\frac{d^2}{dt^2} \mathbf{x}_n(t) = a \left\{ V_c \mathbf{e}_x - \sum_m^{N(n)} [v_{max} - V(r_{mn}(t))] P(\theta_{mn}(t)) \mathbf{e}_{mn}(t) - \frac{d}{dt} \mathbf{x}_n(t) \right\}. \quad (2)$$

where  $\mathbf{x}_m(t)$  is the 2-dimensional position of the  $m$ th particle.  $r_{mn}(t) = |\mathbf{x}_n(t) - \mathbf{x}_m(t)|$  is the distance between two particles, which determines the optimal velocity  $V(r_{mn})$ . The function  $P(\theta_{mn})$  gives the anisotropy of the interaction, where  $\theta_{mn}$  is the relative angle of two particles.

$\sum_m^{N(n)}$  denotes the summation with respect to  $m$  which satisfies  $m \in \{|\mathbf{x}_n - \mathbf{x}_m| < 2r, \text{ s.t. } r = \min_m |\mathbf{x}_n - \mathbf{x}_m|\}$ .  $N(n)$  means the number of particles with

the condition surrounding particle  $n$ . We note the summation is dependent on time according to the positions of all particles. Instead of this ‘simple summation’ we can define the ‘average summation’ as  $\{1/\sum_m^{N(n)} \rho(\theta_{mn})\} \sum_m^{N(n)} \rho(\theta_{mn})$ , where  $\rho(\theta_{mn})$  denotes the weight which depends on the direction to the  $m$ th particle. This means that the particle pays more attention to the particle just in front of it than to the particles in other directions. If we chose  $\rho(\theta_{mn}) = 1$ , the average summation turns to  $\{1/N(n)\} \sum_m^{N(n)}$ , which can be called ‘simple average’. But this summation is not realistic. Actually, the model with this summation results in the trivial motion. We have two possibilities for the choice of summations.

We can freely chose the functions  $V(r_{mn})$ ,  $P(\theta_{mn})$  and two types of summations with  $\rho(\theta_{mn})$ . We have much variates for construction of 2- $d$  OV models according to what kind of phenomena we want to apply.

In general, OV models have two characteristics, the following behavior and the exclusive behavior. Model-1 stresses the acceleration of the effect of following behavior of particles by the attractive force. This model may be suitable for the flow of bio-motion or evacuation dynamics. Model-2 stresses the deceleration of the exclusive effect by the repulsive force. This model may be suitable for granular flow in a liquid or pedestrians.  $V_0 \mathbf{e}_x$  is the desired velocity for free-moving with no interaction. For simplicity,  $V_0$  is a constant value.  $x$ -axis is for the the direction of pipe in granular flow or the passage in pedestrian flow.

In this paper we concentrate on the dynamics of granular flow in a liquid through a pipe and pedestrian flow in a passage. For this purpose model-2 is adequate and we perform the simulation using model-2.

## 2 A Model for Granular Flow in a Liquid

For a model of granular flow in a liquid through a pipe, we introduce the anisotropy of interaction as

$$P(\theta_{mn}(t)) = \frac{1}{2}(1 + \cos \theta_{mn}(t)), \quad (3)$$

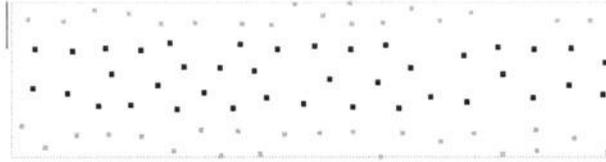
and choose the simple summation. The setting seems reasonable in the physical point of view, but this is one trial among various possibilities. The results of simulation are presented in the paper by A. Nakayama in this volume.

## 3 A Model for Pedestrian Flow

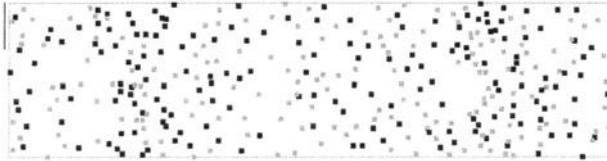
For a model of pedestrian flow in a passage, we introduce the anisotropy of interaction as

$$P(\theta_{mn}(t)) = \cos \theta_{mn}(t), \quad (4)$$

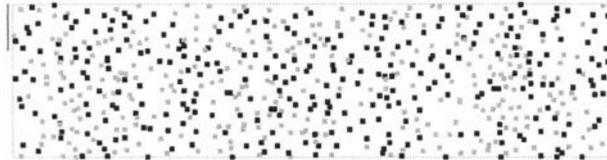
$N = 80$



$N = 350$



$N = 600$



**Fig. 1.** Snapshot of the simulation for pedestrians with uni-directional flow:  $a = 0.5$ , fast particles (light) and slow particles (dark).

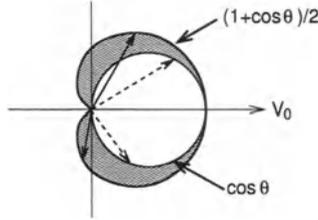
and chose the average summation with the weight function as

$$\rho(\theta_{mn}(t)) = \cos \theta_{mn}(t). \tag{5}$$

For the simulation of pedestrians we introduce two different kinds of particles, which have different OV functions. In this case we can observe the fast particles pass the slow particles. Figure 1 shows uni-directional flow of pedestrians.

In the low density regime the flows of slow particles and fast particles are sharply separated and each flow moves homogeneously. For intermediate density the bunching appears, which is similar to the phenomena in traffic flow. This results in the formation of an ‘island’ of jam cluster, in which the shape is qualitatively different from the case of granular matter. The jam of granular matter forms the ‘band’ perpendicular in the direction of the flow in fig. 6, which is caused by the difference of angular-dependence of sight,  $P(\theta_{mn})$  (cf. fig. 2) . For high density no structure appears for too many particles.

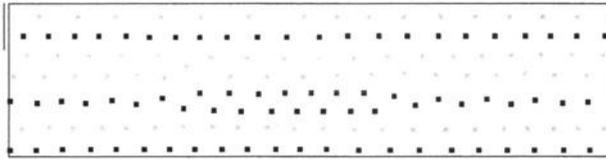
As for pedestrians, counter flow is possible. Figure 3 shows the results of simulations for a counter flow of pedestrians. For low density several thin



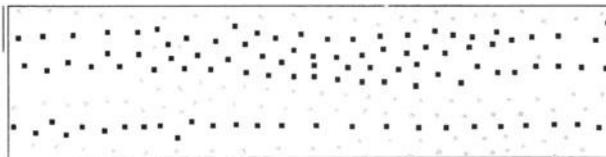
**Fig. 2.** The range of sight for granular matter,  $\frac{1}{2}(1 + \cos \theta_{mn})$  and for pedestrians,  $\cos \theta_{mn}$ . The axis of  $V_0$  is the desired direction for free-moving.

streams in the opposite direction are formed. This is the specific property for a counter flow of pedestrians. For intermediate density the streams become unstable and grow wide and move winding. But their stability is very sensitively dependent to the initial conditions. For high density the jam islands are formed clearly, but several particles find their way to escape from the jam island to free-moving region.

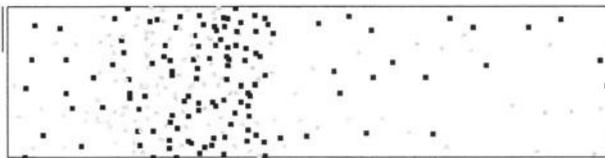
$N = 150$



$N = 200$



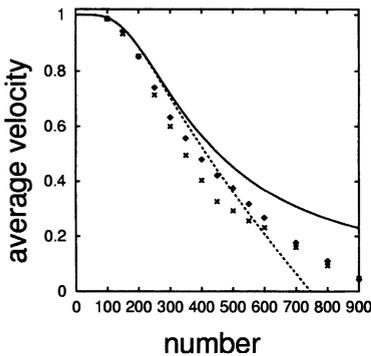
$N = 250$



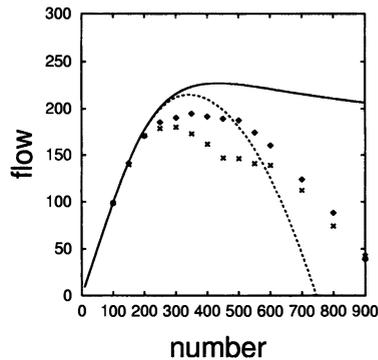
**Fig. 3.** Snapshot of simulations for pedestrians with counter flow:  $a = 0.5$ , fast particles (light) and slow particles (dark).

### 4 Comparison Between Models for Granular and Pedestrian Flow

We discuss the difference of the effect of the angular-dependence of sight,  $P(\theta_{mn})$  according to the relation between average velocity of flow and density in fig. 4 and the Q-k diagrams in fig. 5. The simulations have been performed for uni-directional flow of one kind of particles. In both figures, the solid curves represent the analytic result for homogeneous solution with nearest neighbor interaction. The dashed curves represent also the analytic results for the homogeneous solutions including next-nearest neighbor interaction. Analytic solution of homogeneous flow is the same for granular and pedestrian flows. If the number of particles exceeds 200, the homogeneous solution is not satisfied with the simulation. And the effect of long-range interaction, such as next-nearest neighbor interaction, becomes important. As we can see in both figures, the simulation for granular matter ( $\times$ ) differs larger than that for pedestrians ( $\diamond$ ) from the curve of homogeneous solution including long-range interactions in the region of  $N = 300 \sim 600$ . This indicates that the pedestrian flow does not show the clear congestion of flow in contrast of granular flow, which is caused by the difference of angular-dependence of sight,  $P(\theta_{mn})$ . This qualitative difference should be checked by the real data for pedestrian flow.

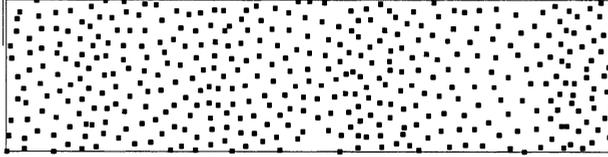


**Fig. 4.** The relation between average velocity of flow and the number of particles  $N$ .

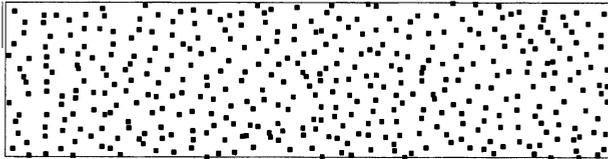


**Fig. 5.** The Q-k diagrams: the relation between the flow and the number of particles  $N$ .

The jam islands formed in the case of two kinds of particles; a uni-directional flow of fast and slow particles and a counter flow of left and right moving particles, are types of clusters different from traffic congestion, which is seen in the jam of granular flow. The jam islands of pedestrian flow are only caused by the bunching phenomena of two kinds of particles.



**Fig. 6.** Snapshot of a simulation for granular flow:  $a = 1$ ,  $N = 400$ .



**Fig. 7.** Snapshot of a simulation for pedestrian flow:  $a = 1$ ,  $N = 400$ .

## Acknowledgments

This work is partly supported by a Grant-in-Aid for Scientific Research (C) (No.12650065) and (C) (No.13640409) of the Japanese Ministry of Education, Science, Sports and Culture, and by the Center for Integrated Research in Science and Engineering of Nagoya University.

## References

1. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, *Japan J. of Ind. and Appl. Math.*, **11**, 203 (1994); *Phys. Rev. E* **51**, 1035 (1995). Y. Sugiyama, in: D.E. Wolf, M. Schreckenberg, and A. Bachem, (Eds.), *Traffic and Granular Flow*, (World Scientific, Singapur, 1996), p. 137; M. Schreckenberg, and D.E. Wolf (Eds.), *Traffic and Granular Flow '97*, (Springer, Berlin, 1998), p. 301.
2. D. Chowdhury, L. Santen, and A. Schadschneider, Statistical Physics of Vehicular Traffic and Some Related Systems, *Physics Reports* **329**, 4-6, 228 (2000).
3. O. Moriyama, N. Kuroiwa, M. Matsushita, and H. Hayakawa, *Phys. Rev. Lett.* **80**, 2833 (1998).
4. D. Helbing, in: M. Schreckenberg, D.E. Wolf, (Eds.), *Traffic and Granular Flow '97*, (Springer, Berlin, 1998), p. 21; D. Helbing, I.J. Frakas, and T. Vicsek, in: D. Helbing, H.J. Herrmann, M. Schreckenberg, and D.E. Wolf, (Eds.), *Traffic and Granular Flow '99*, (Springer, Berlin, 1998), p. 245; T. Vicsek, A. Czirók, and D. Helbing, p. 147: and references therein.

# Evacuation Analysis of Ship by Multi-Agent Simulation Using Model of Group Psychology

M. Katuhara, H. Matsukura, and S. Ota

National Maritime Research Institute, Japan, Shinkawa1-6-1, Mitaka, Tokyo 181-0004, Japan

**Abstract.** It is the most important problem of simulation how human selects the evacuation route. Multi-agent type simulation model is proposed. Man is an autonomous agent who chooses a course. Recognizing surroundings through senses and being mentally influenced, man judges the optimum course every moment. And evacuation action is considered to be group behavior. This model is verified by demonstrations on a real ship and is shown to give optimum evacuation route. Applications to large ships have succeeded.

## Introduction

It is one of the safety requirements for a passenger ship that passengers can evacuate smoothly to the refuge place. Japanese domestic ship regulation of evacuation safety is nearly equivalent to the international i.e. the resolution A.757 [1] of the international maritime organization (IMO). For example, the door width of the N planned persons passing through should be wider than  $N$  cm.

Ignited by the accident of the Estonia in September 1994, IMO improved the safety standard of RORO passenger ship of SOLAS treaty in November 1995. The duty that RORO passenger ship built on and after July 1, 1999 should carry out evacuation analysis in the early stage of a design is imposed. The IMO decided interim guidelines for a simplified evacuation analysis in response to this. This estimates human flow by manual calculation. This is provisional and is so rough, and then it is now under discussion till 2002 in FP subcommittee of IMO.

The purpose of this paper is to show the microscopic simulation, which uses model of group psychology, to verify it by the demonstration on the ship, and to apply it to a real ship design.

It is the most important problem of simulation how human selects the escape route. Even if there are two of the same arrangements of evacuation space, escape route of human are different because of their different psychology. For example those cases are likely that familiar friends are on the other deck, or not, that the party was held on the previous day, or not, and that there is a man of leadership at a room, or not. There should happen various psychological factors in the spot.

Most previous research [2] of evacuation took hypothesis of minimum waiting-time, and shortest course of selection etc, disregarding psychology. This paper can show various cases of evacuation.

Next, the model is verified by the demonstrations. And the optimum escape route is calculated.

## Simulation Model

### Network Model of Space

The interior of the ship is described by the network, which is constituted from nodes, paths and spaces as shown in Fig.1. Each has attributes such as width of passage, length and area of space. Walking speed is constant. Flow of human is limited by the opening width of passage and by the density of space (Fig.2).

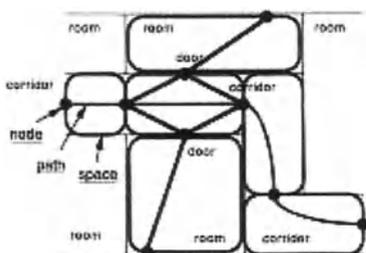


Fig. 1. Network of node-path-space.

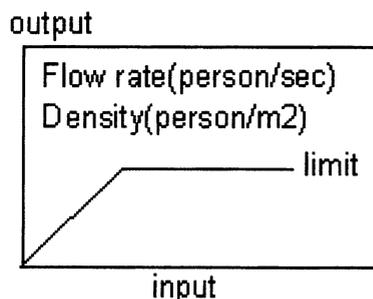


Fig. 2. Limit of flow rate and density.

Two kinds of simulations are carried out. The one is that the escape routes are appointed in advance by input files, and the other is that routes are selected indigenously in simulation. There is so little problem in the former case to be omitted in this paper.

### Simulation of Routes Selection

Man is an autonomous agent who chooses a course. As shown in Fig. 3, recognizing surroundings through senses and being mentally influenced, man judges the optimum course every moment. And evacuation action is considered to be group behavior [3].

The route selection model based on group psychology is as follows:

- 1) People desire to reach earliest to refuge place.
- 2) People get all information of the same deck as themselves, using every sense of sight, hearing and smell.
- 3) People only imagine the situation about other than their own deck.
- 4) Group psychology is dominant.

5) Walk speed and action start time depend on human type (it classifies into an adult, a child, an old person, a handicapped person, etc.).

As an amount of psychology imaginary distance shall be defined as the distance on the imagination from stairs to refuge, as shown in Fig. 4. Each will choose the route which is calculated as minimum of  $\{(walking\ time\ on\ self-deck) + (waiting\ time\ at\ congestion\ point\ on\ self-deck) + (walking\ time\ along\ the\ imaginary\ distance\ of\ other\ decks)\}$ . Group psychology means that imaginary distance of each person takes the same value. This hypothesis lessens number of mental variables and makes calculation remarkably simple.

As shown in Fig. 4 the network of route needs to be assembled for every deck, which is mentally same situation.

This model is based on the fact of real ship experiment mentioned later that there exists divergent point of human flow. One of them is shown in Fig. 5 This is the part of data of the track of every examinee for about 5 seconds from the beginning of the real ship experiment. It comes out from a room and is diverted into right and left. This is evidence that evacuation action is based on group psychology.

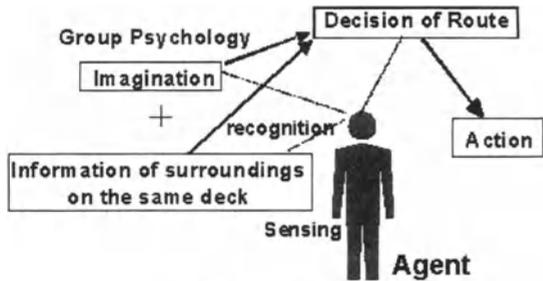


Fig. 3. Agent Model.

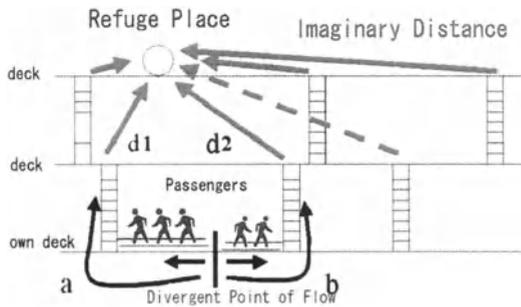


Fig. 4. Route Selection Model.

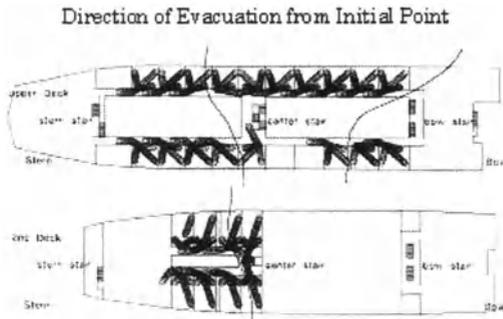


Fig. 5. Divergent Lines of Flow.

## Verification by Demonstration

### Actual Ship Demonstration

Eight times evacuation experiments were done alongside the pier of Tokyo Harbor by the training ship “Seiun-maru” of Institute for Sea Training, Ministry of Transport in December 1994, July 1995, and July 1996. “Seiun-maru” is 5,044 gross tonnage, 114.6m length, 256 persons (Officer 34, Crew 42, Cadets 180) capacity.

In Fig. 6 she has three decks as evacuation space, six stairs, refuge place at the most upper deck and rooms for cadets, dinning and lecture. Perpendicular evacuation routes are three, i.e. stern, central, and bow stairs, which are here named as route1, route2 and route3 respectively.

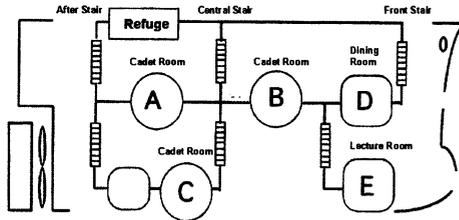


Fig. 6. Vertical Section of Model Ship.

Table 1 shows the conditions of eight experiments, i.e. the number of persons set initially.

Table 1. Number of Persons Setted Initially.

No.	A	B	C	D	E	Total	comment
H6-1	47	39	40	0	0	126	in room
H6-2	26	26	26	24	24	126	Dispersed
H6-3	18	18	18	17	17	88	Dispersed
H7-1	24	26	26	0	0	76	in room
H7-2	15	15	16	15	15	76	dispersed
H7-3	0	0	0	0	76	76	concentrated
H8-1	27	28	28	0	0	83	in room
H8-2	0	0	0	0	83	83	concentrated

The testees are cadets of seaman colleges, 19 or 20 years old, male or female. They are instructed not to run, not to push each other, to walk as usual normally and to reach the refuge place as quickly as possible.

About twenty video cameras take video pictures of testees passed. Walking speed, maximum flux of human flow and evacuation time to the refuge place are afterwards measured from video films in research office. Fig. 7 is the example.



Fig. 7. View of Demonstration.

### Comparison Between Result of Demonstration and Calculation

The comparison as a result of an experiment and calculation is shown in Fig. 8. It is left-hand side as a result of an experiment, and right-hand side is as a result of

calculation. Parameters are the selected route numbers and those sum totals. Right and left are in agreement about all cases. The cases of (A), (D), and (G) which are room arrangement, and the cases of (B), (C), and (E) which are dispersedly arrangement specially show good agreement.

Problem of dispersed distribution is different psychology between men who have get out from bottom room E and men who are originally on the second deck A, B and D.

Problem of concentrated distribution is found in case of (F) and (H) that real men would foresee congestion in front of the stair which many persons are gathering to, but calculation has not programmed it.

These are small discrepancy, so approximation method should be adopted for simple treatment, that is, the hypothesis of psychologically single group and no foresight. So the human model is verified.

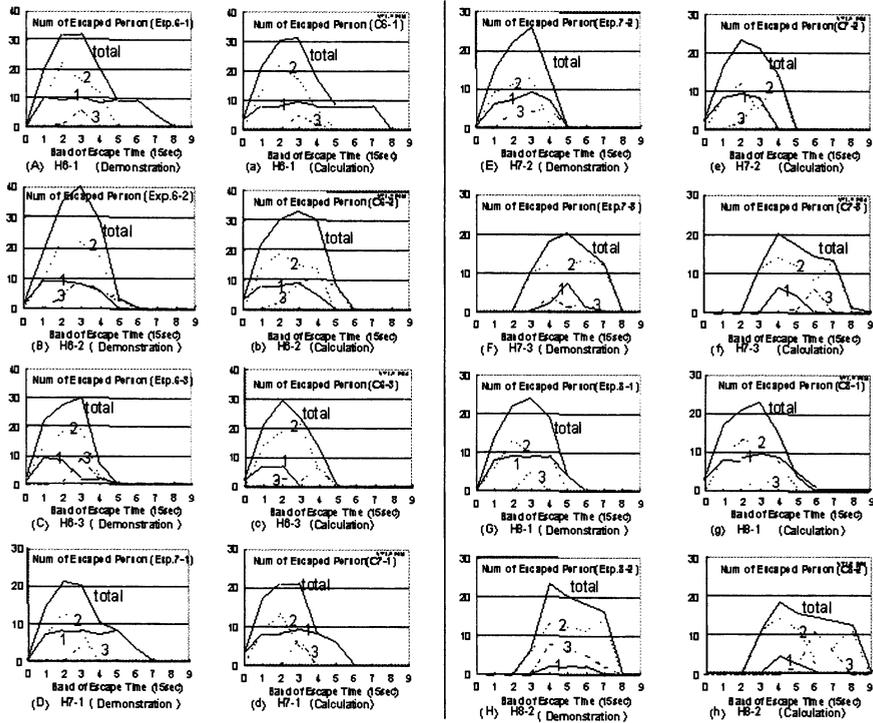


Fig. 8. Comparison of Demonstration and Calculation.

### Influence of Quantity of Psychology

The influence of an amount of psychology called imagination distance is to be investigated. Model ship is used for calculation. With six stairs and three decks there

are two essentially independent and significant parameters of imaginary distance.

Fig. 9 shows the influence of two imaginary distances. There are three plateaus, which are the cases of all men gathering one stair. 3-dimensional graph is seen as if at Colorado Valley three rivers flow between cliffs and the junction of three rivers is deepest. Z-axis is evacuation time so this deepest point is optimum case. Fig. 10 is evacuation time curve in the optimum case. Every curve finished at the same time. This means persons optimally distributed to every stairs. This psychological condition is realized by staff guidance in emergency. Fig. 11 is the information of optimum evacuation routes.

This evacuation safety analysis is applied to two large domestic RORO ferry and a ocean cruiser. These trials are succeeded.

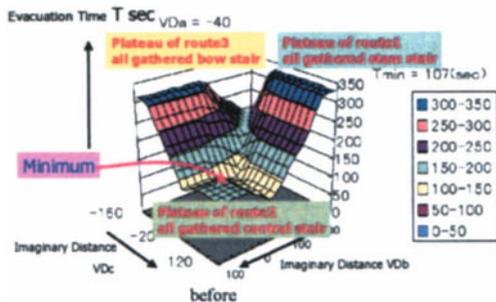


Fig. 9. Influence of Imaginary Distances.

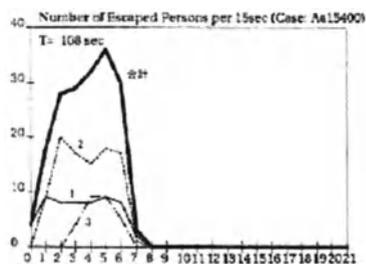


Fig. 10. Evacuation Time Graph in Optimum Case.

## Conclusions

Multi-agent simulation of evacuation on ship using model of group psychology is verified by real ship demonstration, and is applied to optimum design of evacuation route.

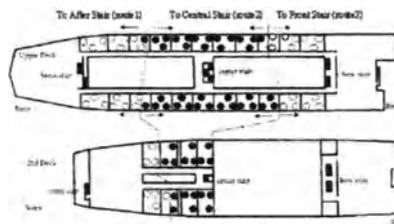


Fig. 11. Information of Optimum Evacuation Routes.

## References

1. IMO Document Resolution A.757(18), 4 November, 1993.
2. Daniel M. Alvord, Status Report of Escape and Rescue Model, NBS-GCR-83-4.32, Juni, 1983.
3. M. Katuhara et al., Simulations and Demonstrations of Human Evacuee Escape on Board, RINA Conference (with The Nautical Institute), "Escape, Evacuation and Rescue", November 1996, London, UK.

# Comparison of an Evacuation Exercise in a Primary School to Simulation Results

H. Klüpfel<sup>1</sup>, T. Meyer-König<sup>2</sup>, and M. Schreckenberg<sup>1</sup>

<sup>1</sup> Physics of Transport and Traffic, Gerhard-Mercator-University Duisburg,  
Lotharstr. 1, 47048 Duisburg, Germany

<sup>2</sup> TraffGo GmbH, Grabenstr. 132, 47057 Duisburg, Germany

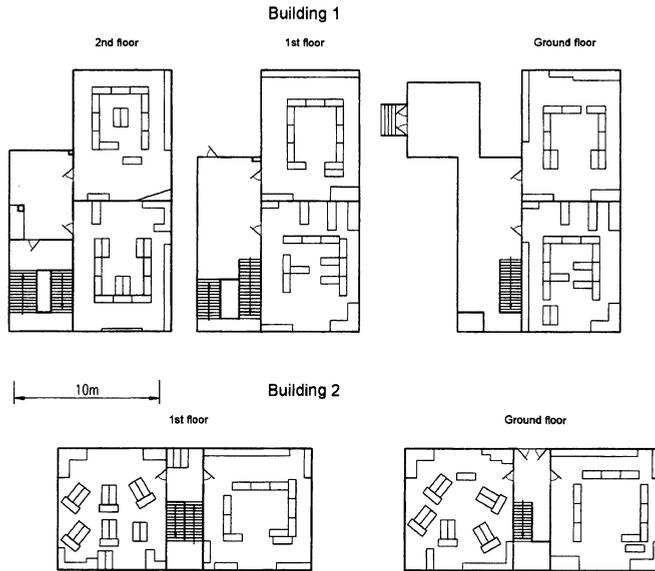
**Abstract.** The modeling of pedestrian movement has received growing interest over the last decades. This is due to the potential applications in facility design and especially evacuation simulation as well as the fascination of its fundamental properties. Empirical data plays a particular role with respect to both aspects. The key challenge in modeling and simulating crowd movement is to validate the model assumptions on the one hand and the simulation results on the other hand. In this paper we present empirical data on an evacuation exercise in a primary school. About two hundred pupils (and their teachers) took part in the drill. Three drills were carried out. The premises are divided into two separate buildings, the larger one containing 6 classrooms with about 120 pupils. The results of the exercise are reported. Additionally, the time measured is compared to the time distribution gained from simulations.

## 1 Modeling and Simulating Evacuation Processes

Pedestrian dynamics is a vital and growing interdisciplinary field of research [18,19]. Especially the subject of evacuation simulation has attracted increasing interest. Overviews over the different simulation and optimization methods available can be found in [9,10,14]. They can be roughly classified into:

1. regression models (hydraulic/hydrodynamic),
2. queuing models,
3. route-choice models (mesoscopic), and
4. microscopic models (social force model) [11] and Cellular-Automaton-Models (CA-Models) [16,2,3,5].

Regardless of the type of model used, calibration of the parameters and validation of simulation results is of major concern. This holds for normal crowd movement as well as for the evacuation or emergency case. This distinction is in a sense reflected in the one between empirical observations and experiments (active preparation of the conditions). Experiments are usually not carried out for a large number of persons and do therefore rather investigate individual behavior and movement than the characteristics of crowd motion. The latter has mainly been accessed via observations. Calibration of input parameters can therefore be based on experimental results, the validation of full scale simulations is mainly done via observations or data from fire drills.

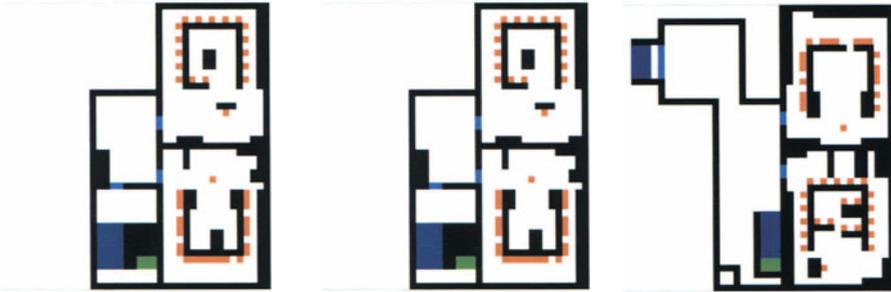


**Fig. 1.** Layout of the school building. It is separated into two independent parts, building 1 (top) and building 2 (bottom). Building 1 has three floors (from right to left), building 2 two floors. The students gather on the playground just in front of each building. The drawing is in scale: the doors of the rooms are 1m wide, the main exit on the ground floor is 2.2m wide.

The data presented in this paper are not truly experimental, since they are based on an evacuation exercise and not on laboratory experiments. However, there is a well defined scenario and the situation is controlled to some extent. Therefore, it mainly addresses the question of evacuation time and whether it can be predicted by the simulation. Please note that the case described should be considered ideal (optimal movement and behavior), since there are no hazards present and the participants were aware of the fact that it was an exercise.

Phenomena like lane formation and oscillation at bottlenecks [12], flow-density relationships [13,20,17], or the level of service concept [7] provide further methods for checking and calibrating simulation results. However, they can usually not be observed in this type of scenario, since there is no movement into the building (no counterflow), which is characteristic for most of these phenomena but not for evacuation drills.

The model that is used here to simulate the egress time (cf. eq. 1) is described in [16,21]. Du to the limited space, we do not describe the model here in detail. It is a CA model (floor-plan=square grid, cell= $0.4\text{m} \times 0.4\text{m}$ ) with  $v_{\max} = 5\text{cells}/\Delta t$ ,  $\Delta t = 1\text{s}$ , and each cell being occupied by at most one



**Fig. 2.** Initial distribution (grey cells) in the simulation for the three floors of building 1 (cf. fig. 1, same order). Each small grey square corresponds to one person. The initial number of persons is therefore (from top left to right, top to bottom): 21, 21, 22, 22, 18, 25.

person. The total evacuation time is obtained by the following equation [8]:

$$t_{\text{evac}} = t_{\text{process}} + t_{\text{react}} + t_{\text{egress}}, \quad (1)$$

where  $t_{\text{process}}$  is the time for realizing the situation (e.g., the time it takes to trigger the alarm),  $t_{\text{react}}$  is the reaction time, and  $t_{\text{egress}}$  the time for the movement out of the building.  $t_{\text{process}}$  can be set to zero, since in an exercise the time starts with triggering the alarm. What is then measured is  $t_{\text{react}} + t_{\text{egress}}$ . Those two could be distinguished by recording information about the events within the classroom. This was not done, however. This is justified by the fact that the time until the first person reached the door of the room was in the order of a few seconds, e.g., only  $t_{\text{egress}}$  is relevant and  $t_{\text{evac}} \approx t_{\text{egress}}$ . A more detailed report about the exercise can be found in [15].

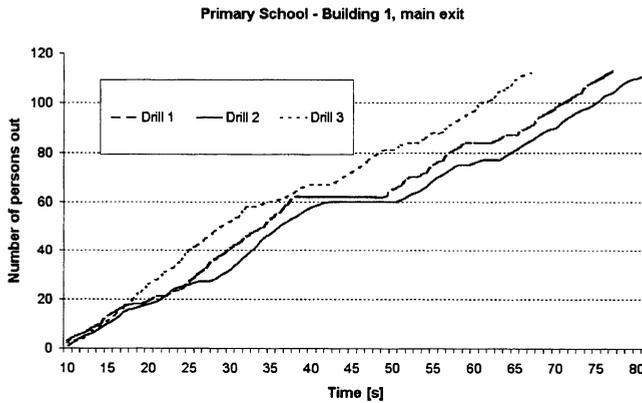
## 2 The Evacuation Exercise

The building consists of two separated parts. It houses a primary school with about 200 pupils. The geometrical details are shown in fig. 1.

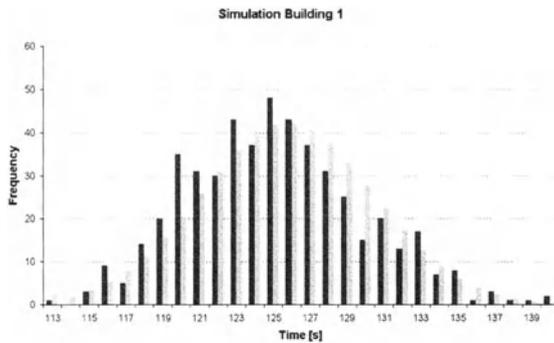
The procedure for the evacuation exercise was the following:

1. The alarm siren was triggered.
2. The persons started evacuating.
3. A person was considered evacuated when she reached the outside, e.g., had left the building via its main exit.

The initial distribution for the simulation is taken from the statistical records of the headmaster (class sizes and rooms, not taking into account absences). It is shown in fig. 2. Therefore, there might be slight deviations in the number of persons between the simulation and the exercise (see figs. 2 and 3).



**Fig. 3.** Number of persons out vs. time for building 1 (cf. fig. 1). Drill 1 to 3 are three different runs of the same scenario, with the same population and initial conditions. Drill 3 proceeded slightly faster due to learning effects.



**Fig. 4.** Distribution of the simulated evacuation times for building 1 (500 simulation runs). The different times result from two factors: the statistical distribution of the personal parameters to the persons and the influence of stochastic parameters.

The participants were all children of the age 6 to 10 (first to fourth grade). Demographic data can be taken into account by the parameter settings. One aim of this endeavor was to check the validity of the parameter settings for such a population (cf. the next section).

### 3 Results

The evaluation is based on the videotapes taken during the exercise. Three different drills with the same initial conditions were performed. Figure 3 shows the number of persons having left building 1 via the main exit on the ground floor vs. time (evacuation or egress curve). It can be seen that there

is a small learning effect which leads to a smaller egress time for the repeated drills.

The simulated egress times (for 500 simulation runs in each case) are:

	Exercise	Simulation
Building 1:	$(125 \pm 4.7)s$	80s
Building 2:	$(80 \pm 4.7)s$	58s

Due to lack of space, the egress curve is only shown for building 1. However, the detailed analysis of building 2 is comparable and would not provide new insights.

## 4 Summary and Conclusion

We have reported results on an evacuation exercise performed in a primary school. There is a deviation between the egress time predicted by the simulation and the actual time it took to evacuate the building. Actually, the simulated time is too high by a factor of 1.6 (building 1), resp. 1.4 (building 2). This is mainly due to the fact that the movement was extremely orderly and organized by well-trained teachers. This fact corresponds the absence of route choice or orientation problems. The population is naturally divided into groups (classes). And the pupils are physically fit, probably even more than adults, very familiar with the building, and used to follow the advice of their teachers. The simulation was based on a standard scenario which corresponds to a less optimal situation. It can therefore be concluded that it gives a more conservative estimate of the egress time than the drill. E.g., walking speed was set to 1.2–2m/s in the simulation; in the drill walking speeds up to 5m/s were observed. Another remarkable result is the reduction of the evacuation time for a repeated drill, even though the motivation decreased. However, this influence was more than compensated by harmonization of the flow.

## Acknowledgment

We would like to thank the primary school in Duisburg Rahm, especially Mr. Tehbas, the caretaker and Mrs. Dommers, the headmistress, for their support, and last but not least all the kids. Our special thanks goes to Roland Chrobok for assisting us in videotaping the exercise. We are grateful to the German Ministry of Education and Research (bmb+f) for funding the BYPASS research project (Assessment and Analysis of Evacuation Processes on board Passenger Ships by Microscopic Simulation), the framework for conducting this research.

## References

1. S. Bandini and T. Worsch, (Eds.), *Theoretical and Practical Issues on Cellular Automata*, (London, Springer, 2000).
2. V. Blue and J. Adler, Cellular automata microsimulation for modeling bi-directional pedestrian walkways, *Transpn. Res. B* **35**, 293–312 (2001).
3. C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz, Simulation of pedestrian dynamics using a 2-dimensional cellular automaton, *Physica A* **295**, 507–525 (2001).
4. D. Canter, (Ed.), *Fires and Human Behaviour*, (David Fulton Publishers, London, 2nd edition, 1990).
5. J. Dijkstra, H. Timmermans, and A. Jessurun, A multi-agent cellular automata system for visualising simulated pedestrian activity, in: Bandini and Worsch [1], pages 29–36.
6. P. DiNenno, (Ed.), *SFPE Handbook of Fire Protection Engineering*, (National Fire Protection Association, 2nd edition, 1995).
7. J. Fruin, *Pedestrian Planning and Design*, (Metropolitan Association of Urban Designers and Environmental Planners, New York, 1971).
8. E. Graat, C. Midden, and P. Bockholts, Complex evacuation; effects of motivation level and slope of stairs on emergency egress time in a sports stadium, *Safety Science* **31**, 127–141 (1999).
9. H. Hamacher and S. Tjandra, Mathematical modelling of evacuation problems – a state of the art, in: Schreckenberg and Sharma [18], pages 227–266.
10. D. Helbing, I. Farkas, P. Molnar, and T. Vicsek, Simulation of pedestrian crowds in normal and evacuation situations, in: Schreckenberg and Sharma [18], pages 21–58.
11. D. Helbing, I. Farkas, and T. Vicsek, Freezing by heating in a driven mesoscopic system, *Phys. Rev. Lett.* **84**, 1240–1243 (2000).
12. D. Helbing, P. Molnár, I. Farkas, and K. Bolay, Self-organizing pedestrian movement, *Environment and Planning B: Planning and Design* **28**, 361–383 (2001).
13. L. Henderson, The statistics of crowd fluids, *Nature* **229**, 381–383 (1971).
14. S. Hoogendoorn, P. Bovy, and W. Daamen, Microscopic pedestrian wayfinding and dynamics modelling, in: Schreckenberg and Sharma [18], pages 123–154.
15. H. Klüpfel, T. Meyer-König, and M. Schreckenberg, Experimental results on an evacuation exercise in a primary school, *Fire Safety Journal*, 2002, submitted.
16. H. Klüpfel, T. Meyer-König, J. Wahle, and M. Schreckenberg, Microscopic simulation of evacuation processes on passenger ships, in: *Proc. Fourth Int. Conf. on Cellular Automata for Research and Industry*, pages 63–71, (London, Springer, 2000).
17. J. Pauls, Movement of people, in: DiNenno [6].
18. M. Schreckenberg and S. Sharma, (Eds.), *Pedestrian and Evacuation Dynamics (PED)*, (Berlin, Springer, 2002).
19. R. Smith and J. Dickie, (Eds.), *Engineering for Crowd Safety*, (Elsevier, Amsterdam, 1993).
20. Transportation Research Board, Washington, D.C. *Highway Capacity Manual*, 1994.
21. Available for download at [www.traffic.uni-duisburg.de](http://www.traffic.uni-duisburg.de). (category publications).

# **Biology**

# Simulations of Evacuation Using Small World Network

N. Ohi, M. Ikai, and K. Nishinari

Department of Applied Mathematics and Informatics, Ryukoku University, Seta, Ohtsu 520-2194, Japan

**Abstract.** A stochastic cellular automaton(CA) model using Small World Network [5] has been proposed to simulate evacuation of persons from a room with plural exits. The number of persons escaping from each door is studied numerically in several kinds of rooms. Numerical results of them are compared with those calculated according to the law of "Building Standard Art" (BSA) in Japan. The calculations in the BSA are based on a few experimental facts and a simple queue theory, thus we will check how they coincide with our direct simulations. It is found that on average the results of simulations and of the BSA show good agreement though their variations are quite large. The effect of the SWN is seen in the total time of evacuation from the room, which becomes the maximum value in this case. Since the transmission speed of information is high in the SWN, it is seen that many people will rush into the door simultaneously and they cannot move freely due to congestion.

## 1 Introduction

Recently, behaviors of pedestrians have been extensively studied by physicists, and various models have been proposed to research collective effects, such as the spontaneous lane formation and the arching phenomena in front of an exit [1–4]. In this paper, we propose a stochastic cellular automaton(CA) model using Small World Network(SWN) [5] in order to simulate evacuation of persons from a room with plural exits. We consider the number of persons escaping from each door, and the total time of evacuation from the room. These values are quite important when we design and evaluate buildings. There is the law called "Building Standard Art" (BSA) in Japan, in which a simple method is proposed experimentally to estimate the above quantities. One of the main purpose of the calculations in the BSA is to check whether all the people in a building can evacuate safely within the permissible time or not in case of a fire. We will compare numerical results of these quantities in our simulations with those calculated according to the BSA in detail.

## 2 Calculations According to the BSA

The calculations in the BSA are based on a few experimental facts and a simple queue theory. There are many tables of formulae of calculations according to various types of buildings, such as office buildings and theaters. For example, the density of people in a typical office room is defined as  $0.25(\text{person}/\text{m}^2)$  and

the flow rate escaping from a door per 1m width is 1.5(person/second-m). In the case of a room with plural exits, each person is assumed to use his nearest door in the BSA. Thus to obtain the number of people using each door, we first divide the floor of a room into subareas considering which door is the nearest one. The floor is discretized to a fine lattice and the distance from each cell to each door is all checked, then we select the door of the minimum distance at each cell. Finally we calculate the area of each part and divide the total number of people into each door according to their ratio. This division rule is one of the simplest ones, and is expected to hold in average in the case of evacuation from a room. We have made a program of calculating the above division in an arbitrary kind of room by using the Dijkstra method [6], and we will compare the division with direct simulations given in the next section.

### 3 A CA Model Using SWN

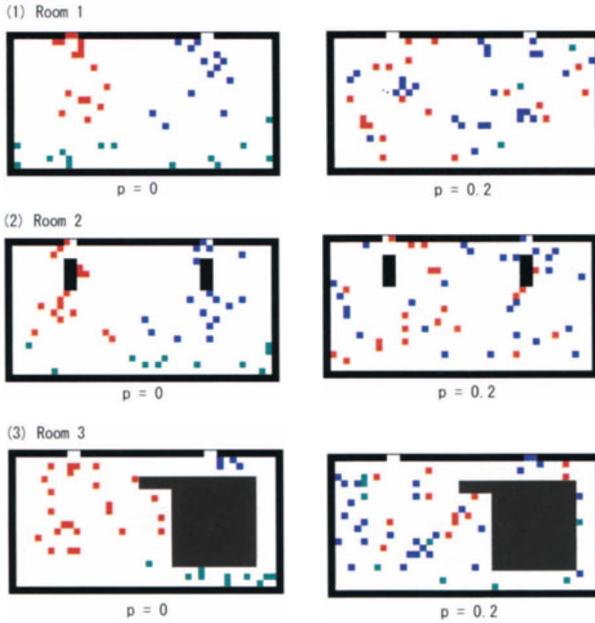
The behavior of evacuating persons is complex in reality. There are continuous fluctuations of judging which direction to escape in case of a fire. We simulate these situations by using the combination of cellular automaton (CA) and the SWN. Originally, the SWN is proposed as intermediate networks between order ( $p = 0$ ) and randomness ( $p = 1$ ), where  $p$  represents the randomness parameter [5]. We can quantify the structural properties of the SWN by a characteristic path length  $L(p)$  and clustering coefficient  $C(p)$ .  $L(p)$  measures the typical separation between two vertices (a global property), whereas  $C(p)$  measures the cliquishness of typical neighborhood (a local property). Regular networks have large  $L$  and  $C$  and random networks have small  $L$  and  $C$ . The SWN has small  $L$  and large  $C$ , and  $p$  is between 0.1 and 0.3. There are numerous researches on the SWN up to now, and we have found that transmission speed of information becomes high in the SWN when the probability is near  $p = 0.2$  [7].

Our model is based on the two-dimensional CA, i.e., the floor is divided into a regular square lattice and the size of each cell is  $0.5\text{m} \times 0.5\text{m}$ , which is the area of one person. Time is discretized with the interval 0.5s and the velocity of person is assumed as 1.0 m/s. Each person moves one cell per unit step with probability if the target cell is not occupied. The rule for the nearest interaction of persons is determined on their Moore neighborhoods as follows. Each person want to move to the direction of one of the doors, which is determined by the SWN explained later, so the weights are set to the neighboring 8 cells so as to be proportional to the distance to the door. The nearest cell to the door has the highest weight of probability. If there are persons in some of the 8 cells, then we set the weight of the occupied cells as zero. Then one of the neighboring cells is chosen according to the probability and the person moves there at one time step. The update rule is parallel and if more than one person want to move the same cell, then we choose one among them with equal probability. Moreover, we use the SWN to consider a kind of remote information. If a fire occurs, then we assume that the virtual information of each door propagates and expands to the room. If a

cell receives the information from one of the doors, then the person on the cell will begin to move to the door. The information of each cell is fixed to the first information it receives, and it rejects the information that come later. Every cell that has obtained information will send it only once to other cells at the next step. If more than one information comes simultaneously into one cell, then we choose one of them with equal probability. The propagation of information is like the SWN. It propagates from every door to nearest cells at each time step like usual wave propagation, but also we choose  $s$  ( $1 \leq s \leq 8$ ) remote cells randomly and send the information to those cells. Thus we choose  $8 - s$  cells of the Moore neighborhoods and  $s$  remote cells at each time. This correspond to the SWN with  $p = s/8$ . Initially all the people in the room stand still and distributes randomly. At a time the information from doors will start to propagate and if people receive the information, then they begin to escape from the room. Persons will go to the door of the first information they receive, and do not change the target door until they are surrounded by other people with different information. If more than two persons which have the same different information exist in his Moore neighborhoods, then he will change his information to the same as the other persons and move to the same door with the neighboring persons. We have taken account of the fact that people will follow the other persons in a panic situation by using this changing rule.

## 4 Simulations and Results

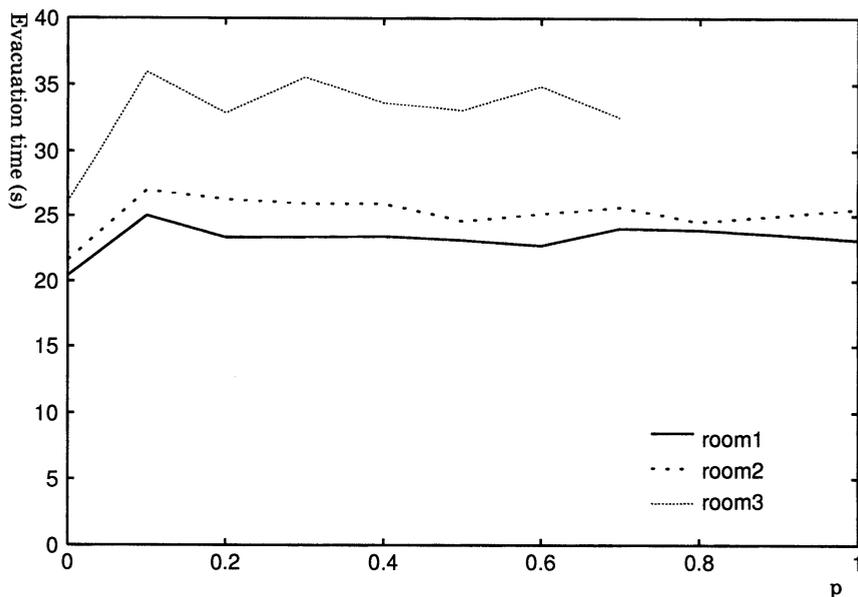
In our simulations, we assume that a fire occurs in a typical office building and we set the density of people according to the BSA. The number of people is 50 and the size of room is  $10\text{m} \times 20\text{m}$ . Two doors exist and their widths are both  $1\text{m}$  (2 cells). We simulate the evacuation with three kinds of rooms (fig.1.1-1.3). There is a large obstacle in an asymmetrical location to each door in the room 3. The parameter  $p$  is changed from 0 to 1 in this simulations in order to study the effect of remote information. Typical snapshots of the simulations of each room when  $p = 0$  and  $p = 0.2$  are given in fig.1. We see that the information of each door expands regularly in the case of  $p = 0$ , while random spreading is observed in  $p = 0.2$ . We have obtained the division of people to each door as listed in the table 1. From table 1, it is found that on average the results of the division and of the BSA show good agreement in the case of  $p = 0$  though their variations are quite large. In the case of large  $p$ , the difference of the two values becomes large. Thus the division rule in the BSA corresponds to the regular propagation of information from each door. In reality, people will avoid congestion near doors and the division is expected to be similar like the case  $p > 0$ . The flow rate is between 1.2 and 1.4 on average in all simulations, which are slightly smaller than that of the BSA. The total evacuation time is given in fig. 2. It is found that in the case of  $p = 0$ , i.e., regular network, the value becomes minimum. In the case of the SWN of  $p = 0.1$ , it becomes maximum. Since the transmission speed of information is high in the SWN, it is seen that many people will rush into the door simultaneously and they cannot move freely due to congestion. As



**Fig. 1.** Snapshots of simulations of evacuation from three kinds of rooms. There are obstacles in the Room 2 and 3. The red and blue squares represent persons who move to the left and right door, respectively. The green squares are persons who do not move yet. The light-pink and light-blue area are the information from left and right door. We see random spreading of the information when  $p = 0.2$

Room	1				2				3			
$p$	0		0.2		0		0.2		0		0.2	
Door	L	R	L	R	L	R	L	R	L	R	L	R
data1	22	28	38	12	22	28	22	28	41	9	29	21
data2	24	26	31	19	20	30	20	30	38	12	33	17
data3	27	23	18	32	27	23	33	17	32	18	32	18
data4	24	26	30	20	28	22	27	23	38	12	44	6
data5	23	27	32	18	30	20	27	23	37	13	42	8
data6	25	25	21	29	31	19	31	19	40	10	36	14
data7	23	27	28	22	25	25	31	19	34	16	37	13
data8	25	25	15	35	22	28	22	28	40	10	18	32
data9	29	21	18	32	22	28	37	13	43	7	41	9
average	1:1.03		1.05:1		1.02:1		1.25:1		3.21:1		2.26:1	
BSA	1:1				1:1				3.25:1			

**Table 1.** The division of the number of people to each door is shown both in the case of simulations and BSA. Both of them agree well on average in the case of  $p = 0$ .



**Fig. 2.** The time of evacuation of all the people in each room is shown. It becomes the maximum when  $p = 0.1$  and the minimum when  $p = 0$  in all cases. In the room 3, some of the people cannot receive the information from doors if  $p > 0.7$ .

seen in fig. 2, the existence of obstacles will make the evacuation time longer in these simulations. It is also found that if  $p$  exceeds 0.7, then information is not transmitted to all the people in the room 3. Therefore, the random spread of information causes bad effects to the evacuation from a room.

## References

1. D. Helbing, *Complex Systems* **6**, 391 (1992).
2. D. Helbing, I. Farkas, and T. Vicsek, *Nature* **407**, 487 (2000).
3. M. Fukui and Y. Ishibashi, *J. Phys. Soc. Jpn.*, **68**, 2861 (1999).
4. C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz, accepted for publication in *Physica A*.
5. D.J. Watts and S.H. Strogatz, *Nature* **393**, 440 (1998).
6. M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational geometry* (Springer, Germany, 1997).
7. T. Maeshiro, N. Ohi, and K. Shimohara, *Proc. Sixth Int. Symp. on Artificial Life and Robotics*, 27 (2001).

# Oscillation Patterns in Cytoplasmic Networks of the *Physarum* Plasmodium

H. Yamada<sup>1</sup> and T. Nakagaki<sup>2,3</sup>

<sup>1</sup> Department of Physics, Nagoya University, Nagoya 464-8602, Japan

<sup>2</sup> Research Institute for Electronic Science, Hokkaido University, Sapporo 060-0812, Japan

<sup>3</sup> The Institute of Physical and Chemical Research (RIKEN), Wako 351-0198, Japan

**Abstract.** The plasmodium of a true slime mould is an amoeboid organism with vein networks. We regard this organism as the system of biological oscillators on a graph and study collective behaviour on it.

## 1 Introduction

In ecological or social population dynamics, individual elements behave as if they have some shared aim although they have no leader but weakly and locally interact each other. Such collective phenomena are often affected by the topological structure of interactions [1]. For instance, green algae on coral reefs spawn synchronously with time lags between closely related species [2], and mycorrhizal fungi associated with the roots of plants transfer nutrients and determine plant biodiversity [3]. Complex network systems exhibiting diverse functions exist not only in population system but in an organism. Neural networks govern a wide range of information processes, and intra- and inter-cellular metabolic networks regulate homeostasis and control various rhythmic cycles. For lower organisms without neural systems, metabolic networks are properly organized to keep alive in a fluctuating environment. A unicellular amoeboid protozoan, *Physarum polycephalum*, is one of the remarkable organisms which show visible cytoplasmic networks.

The plasmodium of *Physarum* is a giant amoeboid cell and behaves as individual with dynamical changes of its cell shape. The plasmodium shows contraction and relaxation cycles everywhere within the organism, which is generated by oscillatory mechanochemical reactions among various chemicals, actin, intracellular organelles, and so forth. The rhythmic contractions regulate intracellular transport of endoplasm including nuclei and mitochondria through the vein structure. The plasmodium can crawl when the endoplasmic flow is organized throughout the cell by synchronous oscillations [4].

The rhythmic contraction of the plasmodium is almost synchronous, but synchrony has some spatio-temporal variation. This synchronous oscillation pattern depends on the cell shape with vein structures. After various types of stimulation are applied to the organism, the plasmodium shows very different vein structures accompanying with changes of the contraction pattern and mixing

rate of the protoplasm [5]. By means of the systematic control of cellular configuration, the plasmodium is regarded as a set of biological coupled oscillators [6]. The vein-network morphology of the plasmodium and its formation process have been studied under the path-finding task in a maze [7].

To get a clue to mechanisms of the plasmodial behaviour induced by rhythmic contraction based on the network of vein structure, we study collective motions of coupled oscillators on a graph. However, we soon met some difficulties known by most of the researchers of networks [1]: the structural complexity of networks and how to represent the phenomena on them. Statistical approaches to regular, random or small-world networks are the most fruitful strategies in this area. But we study more simple and symmetric graphs because the vein network of the plasmodium is almost planar and the degree of each node is small. In the present report, we only deal with trees having almost homogeneous degrees. Introducing pace makers on the graphs as the stimulation points, we study the collective phenomena called phase locked/drift solutions and the flow induced by these solutions.

## 2 Coupled Oscillators on Graphs

Although the plasmodium always changes its cell shape, we focus on the oscillation phenomena with the static network architecture on the proper time scale. The vein network is represented as a graph, where nodes of the graph are nonlinear dynamical systems (oscillators) and edges are interactions among the nodal oscillators. We adopt the following oscillatory systems governed by the coupled complex Ginzburg-Landau equations,

$$\begin{aligned} \dot{w}_j &= (i\omega_j + \mu)w_j + b|w_j|^2w_j + \sum_{k \sim j} c(w_k - w_j), \\ w_j &= u_j + iv_j, \quad b = b_1 + ib_2, \quad c = c_1 + ic_2, \end{aligned} \quad (1)$$

where  $k \sim j$  means the node  $k$  adjacent to the node  $j$ , and thus the interaction terms correspond to the Laplacian generated from the adjacency matrix and the diagonal degree matrix (not normalized by the degree of nodes). We assume nearly identical oscillators and homogeneous couplings, and thereby parameters  $\mu$ ,  $b$ , and  $c$  have the same values for each node. This system of equations is the normal form around the Hopf bifurcation point of the generic coupled oscillators.

In the following, we assume that the zero amplitude solution is unstable ( $\mu > 0$ ) and the amplitude of each oscillator is saturated by the nonlinear term ( $b_1 < 0$ ) for coupling-free oscillators. We also assume that the diffusion coefficients are positive ( $c_1 > 0$ ). In the numerical calculations, we have set the parameters as the homogeneous rotation ( $b_2 = 0$ ) and no rotational couplings ( $c_2 = 0$ ).

For the network architecture, we concentrate our analyses on simple and symmetric tree structures: binary trees with the height  $n$  and  $2^n$  leaves, and finite height Cayley trees. Leaves of trees are coupling only their parent oscillators, and this means the boundary conditions are set the Neumann type (flux-free).

Pace makers, oscillators having the fast natural frequency (large  $\omega_j$ ), are put on the root or leaves (boundaries) and they are regarded as stimulus points or information sources for the organisms. In the case that the root oscillator is the pace maker, we pass through successive branches from the root to leaves. In other case that the leaves are the pace makers, we meet junctions from leaves to root. We aim to find out some effects of such a topological difference on the dynamics of (1).

### 3 Phase Locked Solutions

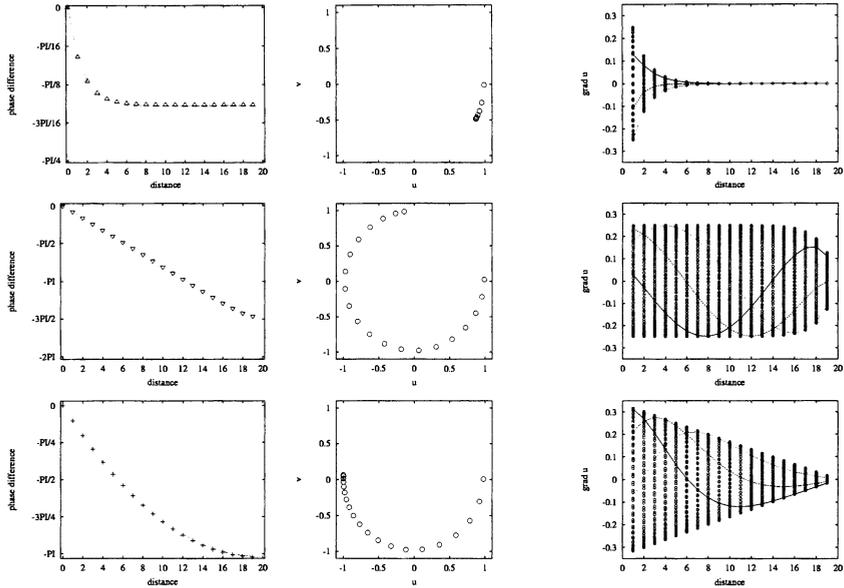
When the positive diffusion coefficients are large in (1) on the symmetric binary tree with identical oscillators, a phase locked solution with in phase (phase differences are 0) is an attractor for proper parameter regions. If we replace some oscillators with pace makers, then a phase locked solution can exist for some conditions. This phase locked solution is no longer in phase, but it has non-zero phase lag between neighbouring oscillators. In the case that the root or all leaves are pace makers, the system (1) has a phase locked solution whose phases are the same values in each level of the symmetric tree, and it is based on the symmetry of exchanges between two sub-trees in the same levels.

Setting parameters in (1) as

$$\begin{aligned}\omega_j &= 1.2 \text{ (pace makers), } 1.0 \text{ (otherwise),} \\ \mu &= 1.0, \quad b = -1.0, \quad c = 0.4 \text{ (binary trees), } 0.6 \text{ (linear chains),}\end{aligned}$$

we obtain the stable phase locked solutions with remarkable phase lags by numerical calculations. Here, we show three cases: (i) the symmetric binary tree with the root being the pace maker, (ii) one with all leaves being the pace makers, and (iii) the linear chain of oscillators with one end being the pace maker. The last case (iii) is arranged as the finite one-dimensional system.

The left column in Fig. 1 shows phase differences from the pace maker for these three cases. The angle of the phase is defined as the argument in the complex plane ( $u_j, v_j$ ) (see the middle column in Fig. 1). In the case of the fast root (i), the phase difference exponentially depends on the distance from the pace maker, and thus the oscillators are almost in-phase near the leaves. Each leaf can follow the neighbouring oscillation since it is only coupling with its parent. Oscillators near the root, however, have long-term lags: heavy sub-trees are hanging in their back. The case of the fast leaves (ii) is remarkably contrastive to (i): the phase difference linearly depends on the distance, that is, the phase differences between neighbouring oscillators are almost the same values. To compare these two cases, we have also calculated the intermediate case between them: the finite linear chain (iii). In this case, the phase difference has parabolic dependence on the distance. These three types of dependence are analytically estimated by means of homogeneous amplitude approximation. According to this approximation, phase locked solutions are the solutions of one-dimensional phase equations with asymmetric couplings, and the phase differences are determined by coupling coefficients of the upward and downward [8].



**Fig. 1.** Phase locked solutions. The upper, middle, and lower rows are respectively the cases of the tree with the root being the pace maker, one with all leaves being the pace makers, and the linear chain with one end being the pace maker. The left column shows relative phase difference plotted against the level of each tree by numerical calculations. Dotted lines are obtained by means of the homogeneous amplitude approximation. The middle column shows a snapshot of oscillators in the complex plane. In the right column, the gradients of  $u$  are plotted in some time intervals. Lines are the profiles of snapshots.

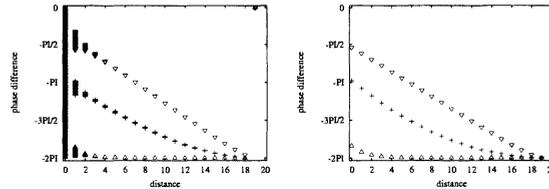
In the right column of Fig. 1, we show the difference (gradient) of  $u_j$  between neighbouring oscillators in some time intervals. Each profile of envelopes represents the dependence of the phase lag between neighbouring oscillators on the distance. Suppose the flow or transportation variables are proportional to the gradient, as the endoplasmic flow is induced by the gradient of contraction pressure in the plasmodium. In such a view, exchanges of the substance, information or something are all over the the networks in the fast leaves case (ii), while they are only near the pace maker in the fast root case (i).

### 4 Phase Drift Solutions

Decreasing the positive diffusion coefficients in (1), we get phase drift (phase slip) solutions: not all of the oscillators have the same averaged frequency. We study the phase drift solutions whose phases are the same value in each level as the phase locked solutions in the previous section. Setting parameters in (1) as

$$\begin{aligned} \omega_j &= 1.2 \text{ (pace makers), } 1.0 \text{ (otherwise),} \\ \mu &= 1.0, \quad b = -1.0, \quad c = 0.04 \text{ (binary trees), } 0.06 \text{ (linear chains),} \end{aligned}$$

we get the phase drift solutions. The only pace maker is drifting when the diffusion coefficients are small enough. It is possible that drifting solutions have more drifting oscillators for larger coupling coefficients.

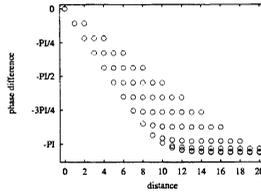


**Fig. 2.** The relative phase differences of phase drift and locked solutions. The left plot shows relative phase difference of phase drift solutions obtained by numerical calculations. Triangles, inverted triangles, and crosses respectively correspond to the cases of the tree with the root being the pace maker, one with all leaves being the pace makers, and the linear chain with one end being the pace maker. Only pace makers are drifting in all cases. The right one is the similar plotting of phase locked solutions combined three parts in the left column of Fig. 1.

Figure 2 shows phase differences. The profile of phase differences is almost the same as the case of phase locked solutions except for the pace makers (drifting oscillators). Because the drifting oscillators pass other oscillators in some long period, the phase differences of drifting ones take all values in the phase angle, and the amplitude of gradient of  $u$  becomes large between the drifting oscillators and others. The homogeneous amplitude approximation is no longer valid for the phase drift solutions, since they vary their amplitude dynamically. But the approximation can estimate the length of the cluster of drifting oscillators with some modifications (cf. [8]).

## 5 Centres and Leaves of a Tree

The Cayley tree is the most symmetric tree with homogeneous degrees. We deal with the finite height Cayley tree with all degrees of nodes being three except for leaves. Only one pace maker is put on one of the leaves and system parameters are set similar values as above but  $c = 0.6$ . Fig. 3 shows the phase difference from the pace maker for the phase locked solutions. In this figure, we notice two remarkable points: the linear dependence of the phase difference on the distance from the pace maker, and branches with almost constant (may be exponential) phase differences from this linear arrangement. In other words, the phase lags between the neighbouring oscillators are constant on the path from the pace maker to the centre of the tree, but, in contrast, the phase lags decrease to zero as far from the pace maker and the centre. These observations indicate that the centre and the boundaries are landmarks to illustrate the oscillation dynamics on trees.



**Fig. 3.** The relative phase differences of phase locked solutions of the finite Cayley tree with one leaf being the pace maker.

## 6 Conclusion

We studied synchronization phenomena on trees and showed branching structures have some effects on the profile of phase locked solutions. We deal with the simple and symmetric trees, the binary trees with the height  $n$  and  $2^n$  leaves. In the case that the root oscillator is the pace maker, the phase difference exponentially depends on the distance from the pace maker, and neighbouring sites have very small phase lags except for near the root. In other case that all the leaves are the pace makers, the phase difference linearly depends on the distance, and therefore the neighbouring phase lags are about the same value all over the tree. The profile of these solutions reflect the network architecture.

It is known that some of the rhythmic behaviour like gaits of animal locomotion and the wave-like movement of swimming fish are generated by neural networks called central pattern generators [9]. In the case of migration and information process of the plasmodium, such mechanisms must be based on visible networks of vein structures. The local information like the phase lag is useful to acquire the whole state of the system as shown above. However, we need more data on the behaviour of the plasmodium for manifestations of biological functions generated by the network architecture.

## References

1. S.H. Strogatz, *Nature* **410**, 268 (2001).
2. K.E. Clifton, *Science* **275**, 1116 (1997); M. Hay, *ibid.* 1080.
3. M.G.A. van der Heijden *et al.*, *Nature* **396**, 69 (1998); D. Read, *ibid.* 22.
4. N. Kamiya, *Protoplasmatol.* **8**, 1 (1959); K. Matsumoto, T. Ueda, and Y. Kobatake, *J. Theor. Phys.* **122**, 399 (1986); Y. Miyake *et al.*, *J. Theor. Biol.* **178**, 341 (1996); T. Nakagaki and T. Ueda, *J. Theor. Biol.* **179**, 261 (1996).
5. T. Nakagaki, H. Yamada, and T. Ueda, *Biophys. Chem.* **84**, 195 (2000).
6. A. Takamatsu, T. Fujii, and I. Endo, *Phys. Rev. Lett.* **85**, 2026 (2000); A. Takamatsu *et al.*, *Phys. Rev. Lett.* **87**, 078102 (2001).
7. T. Nakagaki, H. Yamada, and Á. Tóth, *Nature* **407**, 470 (2000); *Biophys. Chem* **92**, 47 (2001).
8. H. Yamada, Phase-locked and phase drift solutions of phase oscillators with asymmetric coupling strengths (2002) [arXiv:nlin.PS/0203032](https://arxiv.org/abs/nlin.PS/0203032).
9. J.D. Murray, *Mathematical Biology*, 2nd edn., (Springer, Berlin, 1993) and cited therein.

# Jamming Bacterial Traffic: Bioconvection

I.M. Jánosi<sup>1</sup>, A. Czirók<sup>2</sup>, D. Silhavy<sup>3</sup>, and A. Holczinger<sup>4</sup>

<sup>1</sup> Department of Physics of Complex Systems, Eötvös University

<sup>2</sup> Department of Biological Physics, Eötvös University  
P.O. Box 32, H-1518 Budapest, Hungary

<sup>3</sup> Agricultural Biotechnology Center, P.O. Box 411, H-2100 Gödöllő, Hungary

<sup>4</sup> Department of Biotechnology and Molecular Genetics, Szent István University,  
H-2103 Gödöllő, Hungary

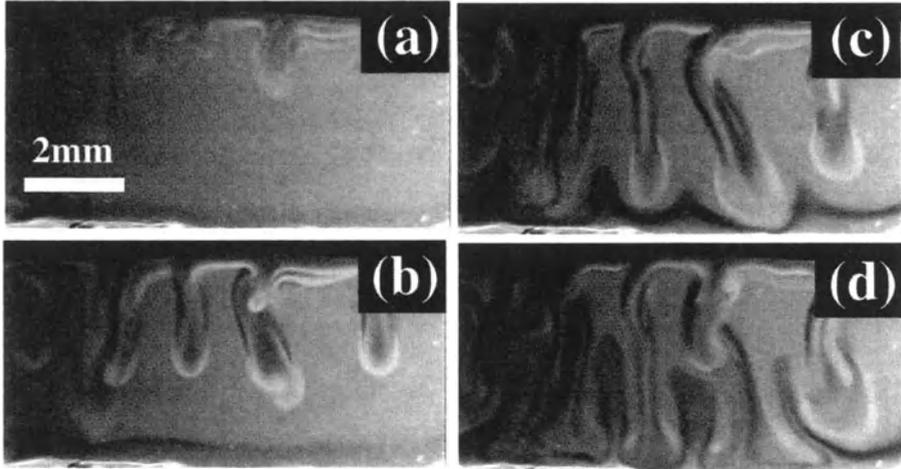
**Abstract.** Bioconvection is a fascinating pattern forming phenomenon driven by the swimming activity of microbes. There is a common belief that bioconvection has a positive effect on the whole microbial population by aerating deep layers in the suspension. In order to detect such a biological benefit, we performed experiments with several strains of *Bacillus subtilis* and *Bacillus licheniformis* of different swimming capabilities. Bioconvection is a robust phenomenon, we observed it at numerous strains in different growth phases. Nevertheless data evaluation has not revealed a similarly robust positive effect on population growth.

## 1 Introduction

Physical interaction might be significant between swimming microorganisms living in aquatic habitats. In natural ecosystems (seas, lakes, rivers, etc.), the strong coupling between the liquid medium and the environment generates usually large-scale flows and perturbations, which determine crucially the motion of the microorganisms: They are advected more or less passively. In a quiescent fluid, however, even their slow motion (typically a couple of meters per day) can result in considerable spatial rearrangements, dynamic pattern formation, or bioconvection [1,2].

Bioconvection is a fluid instability which arises spontaneously in aqueous suspensions of heavy, swimming biological organisms such as algae, bacteria or ciliated protozoa [3–5]. The cell density of microbial species exhibiting bioconvection is larger than that of the water by typically 5-15%. Convective pattern formation in a shallow layer occurs when the concentration of organisms is sufficiently high, and when the average direction of their swimming is upward in response to some external stimulus; e.g., gravity, light or oxygen. The mechanism of one instability is that the continuous upward swimming results in mass inversion, i.e., the surface layer of the suspension becomes denser than below. Subsequently the heavy top layer starts to sink, generating a fluid turnover and pattern formation (see Fig. 1).

At least two major classes of bioconvective systems can be distinguished, based on the taxis mechanism of the cells. (i) Swimming microorganisms can be bottom heavy and as a result of this, tend to swim upwards (geotaxis). (ii) The oxygen concentration gradient, generated by the oxygen consumption of the



**Fig. 1.** Side view of pattern formation produced by *Bacillus subtilis* in a thin upright container. Snapshots are taken with a time lag of 1 minute.

cells, and by supply from the air interface can induce upward swimming, towards regions of higher oxygen concentration (aerotaxis, or in general, chemotaxis).

Continuum theories for bioconvection must include three sets of coupled phenomena: (a) The continuity and Navier-Stokes equations driven by variations in organism concentration, i.e. local fluid density, (b) the equation for conservation of organism concentration including advection and directional swimming (e.g. geotaxis or aerotaxis), and (c) the balance equation for the molecular species that the organisms consume, including diffusive supply and advection. This set of coupled nonlinear equations and the methods for modeling the part of the organism flux due to swimming have been discussed in [5,6]. There has not been much quantitative analysis of experiments available for testing theoretical developments. Recent quantitative studies concerned the geotactic patterns of the algae *Chlamydomonas nivalis* [7], and the onset time of aerotaxis driven bioconvection in cultures of *Bacillus subtilis* [8]. Macroscopic differences were identified by quantitative pattern analysis for the geotactic and aerotactic classes of bioconvective systems [9].

## 2 Motivation

There is a common belief that bioconvection has a positive effect on the whole microbial population: Parcels of surface fluid dragged down with sinking plumes may carry a rich load of oxygen deep into the culture at a rate much faster than diffusion. Even analogies with more developed circulatory-respiratory systems are drawn [10].

This idea, however, is often criticized as unlikely to bear relevance in any natural context [11]. Indeed, laboratory studies are usually performed in qui-

escent, shallow, nutrient rich suspensions at very high cell concentrations of a clean microbial strain.

Further criticism may be based on consideration of transport efficiency. For example, the measured rate of oxygen consumption for *Bacillus subtilis* is in the range of  $0.47 - 1.86 \times 10^6$  molecules/s/cell (depending on the growth phase) [12]. Oxygen concentration at the top layer of the suspension can be estimated as  $8.2 \text{ mg/l} \approx 10^{17}$  molecules/cm<sup>3</sup> (solubility limit at room temperature). Bioconvection usually develops when the cell concentration is in the range of  $10^9$  cells/cm<sup>3</sup>, thus an estimated oxygen depletion time is approximately 1 minute (note that bacteria cease swimming activity at a low but finite oxygen concentration). The observed settling velocity for plumes in thin vertical containers is 3-10 mm/min (see Fig. 1) which means that net oxygen transport is possible only in shallow (1-2 mm) layers only. It is possible that the fall velocity of plumes is smaller in thin containers with typical wall-to-wall distance of 1 mm, thus mixing is more effective without side wall constraints. Nevertheless observations suggest that deep layer plumes can not stir the whole fluid [13].

If we accept that bioconvection is beneficial for the whole population, some interesting questions can be raised on evolutionary connections. Convecting suspension represents an organized structure on a much larger scale than the size of microbes. Although chemical signaling related to population density — the so called quorum-sensing — has been recognized recently for several bacteria [14], it remains difficult to imagine what kind of behavioral response could organize the independent and autonomous cells to enhance or even trigger off convective pattern formation, not to mention its genetic origin.

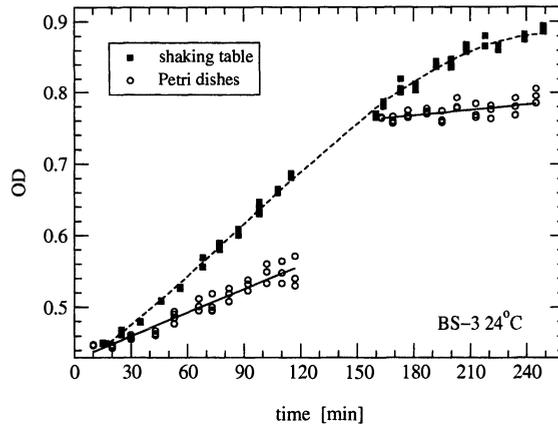
As for experiments, it would be easy to test biological benefits by “switching on and off” bioconvection in the same culture. This seems feasible e.g. by increasing the density of the surrounding medium [15]. Related experiments by Bees [16] were inconclusive: The added unreactive chemical at higher concentrations unwantedly increased the viscosity, and most of the algal cells were perished at the end of the assays.

### 3 Experiments

The main question is whether bioconvection exhibits a clear benefit for the population, or the phenomenon should be considered as a side effect of overcrowding in laboratory cultures. As a first step in a systematic study, we worked out a simple assay to detect a difference in population growth rate at pattern-forming and non-pattern-forming cultures.

We performed experiments with 8 strains of *Bacillus subtilis* and 13 strains of *Bacillus licheniformis* with different swimming capabilities (details will be published elsewhere). Both species are gram-positive, rod-shaped, flagellated common soil bacteria, some of the genetic variants swim fast ( $20-40 \mu\text{m/s}$ ). In general, the cells consume oxygen, weak anaerobe growth is possible in the presence of nitrate.

The essential steps of our experiments were the following:

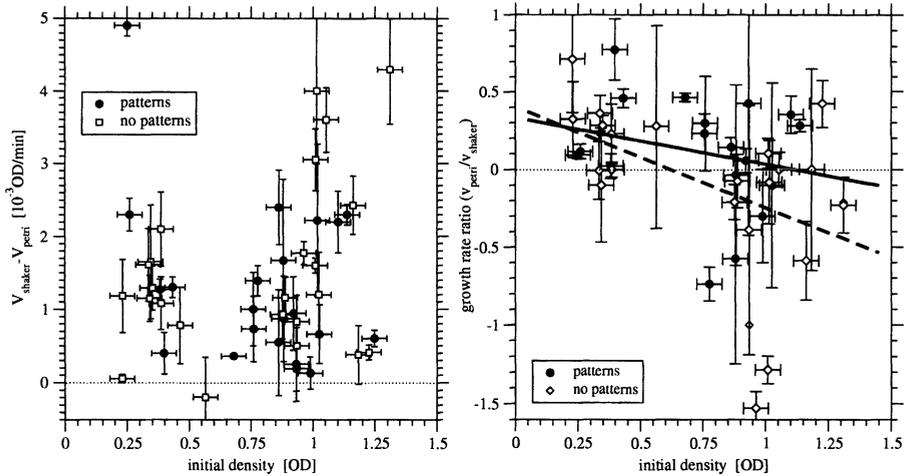


**Fig. 2.** Time evolution of the optical density in a split *Bacillus subtilis* culture. 0.85 OD corresponds to  $6 \pm 2 \times 10^8$  cells/cm<sup>3</sup>.

1. Overnight liquid cultures were proliferated in standard Luria-Bertani (LB) broth aerobically at 37°C.
2. Starter cultures were diluted (1 part suspension + 1 part LB + 6 parts distilled water) to initiate fast swimming, if possible. These cultures were bred on a shaker at room temperature (24°C) for 2-6 hours in order to minimize possible heat-shock effects.
3. At the beginning of the measurements, the cultures were split. One half of the suspension was distributed into Petri dishes (16 ml each, nominal layer depth 2.8 mm), the dishes were *kept in rest* on a solid table. The rest of the culture was put back onto the shaker. Population growth was followed by measuring the optical density (OD) at 600 nm (Phillips PU 8625) at timesteps  $t_0, t_1, t_2 \dots$ . We emphasize that resting Petri dishes were not touched until the OD measurement, after that they were discarded from the process. Pattern formation was detected by visual inspection (the Petri dishes were put on black photo-paperboard).

Typical results are shown in Fig. 2. In general, bacterial growth rate decreased strongly in Petri dishes after splitting the cultures, both for pattern-forming and non-pattern-forming cases. Growth rates were estimated by linear regression, in case of apparently nonlinear growth curves (see Fig. 2) the minimal and the maximal rates around the average contributed to the error bars (see Fig. 3).

Inspections under light microscope revealed that pattern formation is promoted by small active cells (2-5  $\mu\text{m}$ ) at high enough cell concentrations, however we observed patterns already at OD 0.25 in a few cases. In old cultures, where sporulation was clearly visible, the patterns were weak if any. Some strains simply do not swim, they form a dense sediment at the bottom of Petri dishes, or clump together forming macroscopic flakes. We did not observe convective



**Fig. 3.** Growth rate difference (left) and growth rate ratio (right) as a function of cell concentration (OD). Heavy circles and empty squares denote pattern-forming and non-pattern-forming cultures. The error bars reflect nonlinear growth curves, instrumental errors and scatter for individual Petri dishes. Fitted lines on the right (solid for pattern-forming and dashed for non-pattern-forming subsets) have different slopes, but the difference is statistically not significant (see text).

patterns in such populations, note however, that the growth rate was positive in some of these cultures as well.

## 4 Results and Discussion

The measured quantities were the growth rate on the shaker and in the Petri dishes  $v_s$  and  $v_P$ , the initial density  $OD_0$ , and the pattern forming time  $T$ . We performed several statistical tests in order to detect differences between the pattern-forming (22 cases) and non-pattern-forming (24 cases) subgroups. Fig. 3 shows a representative example, where the growth rate drop ( $v_s - v_P$ ) and the growth rate ratio ( $v_P/v_s$ ) are plotted as the function of initial optical density. The difference between the two subgroups is rather small. Quantitative methods, such as the Kolmogorov-Smirnov test [17], indicated that the distinctions reflected e.g. by the linear fits in Fig. 3 are at the edge of statistical significance, which is not very convincing.

We think that the main difference between cultures on a shaker and in Petri dishes is the difference in oxygen supply. The populations at the beginning were the same, grown in the same broth, adapted to room temperature growth, and genetically clean. If bioconvection was an effective aerating process, we expect that it should be reflected by higher relative growth rates compared to non-pattern-forming populations. Note that we do not compare absolute rates belonging to different strains, we compare growth rates for the two halves of a split culture.

Our results can not prove that bioconvection does not have a positive effect on a whole microbial population. We can claim, however, that the detection of biological benefits is difficult, it is far from being such robust as the phenomenon itself occurring in a wide range of cell concentrations in many cultures of different micro-organisms. The opposite possibility is that an overall positive effect does not exist at all, in this case bioconvection can be interpreted as a hydrodynamic phenomenon generated by jamming in microbial traffic at the surface of the fluid medium.

This work was supported by the Hungarian National Science Foundation (OTKA) under Grants No. F026645 and T032423.

## References

1. J.R. Platt, *Science* **133**, 1766 (1961).
2. M.S. Plesset and H. Winet, *Nature* **248**, 441 (1974).
3. J.O. Kessler, *Contemp. Phys.* **26**, 147 (1985).
4. T.J. Pedley and J.O. Kessler, *Sci. Progress* **76**, 105 (1992).
5. J.O. Kessler and M.F. Wojciechowski, Collective behavior and dynamics of swimming bacteria, in: *Bacteria as Multicellular Organisms*, J.A. Shapir and M. Dworkin, (Eds.), pp. 417–450 (Oxford University Press, New York 1997).
6. J.O. Kessler, G.D. Burnett, and K.E. Remick, Mutual Dynamics of Swimming Microorganisms and Their Fluid Habitat, in: *Nonlinear Science at the Dawn of the 21st Century*, P.L. Christensen, M.P. Soerensen, and A.C. Scott, (Eds.), pp. 409–426 (Springer, Heidelberg 2000).
7. M.A. Bees and N.A. Hill, *J. Exp. Biol.* **200**, 1515 (1997).
8. I.M. Jánosi, J.O. Kessler, and V.K. Horváth, *Phys. Rev. E* **58**, 4793 (1998).
9. A. Cziráok, I.M. Jánosi, and J.O. Kessler, *J. Exp. Biol.* **203**, 3345 (2000).
10. J.S. Turner, *The Extended Organism: The Physiology of Animal-Built Structures*, Chapter 4, (Harvard University Press, Cambridge, 2000).
11. M. LaBarbera, *Science* **289**, 1882 (2000).
12. L.S. Wong et al., *J. Bacteriol.* **177**, 3985 (1995).
13. J.O. Kessler and N.A. Hill, Microbial Consumption Patterns, in: *Spatio-Temporal Patterns*, P.E. Cladis and P. Palffy-Muhoray, (Eds.), (Addison-Wesley, New York 1995).
14. G.M. Dunny and B.A.B. Leonard, *Annu. Rev. Microbiol.* **51**, 527 (1997); S.I. Aizawa, C.S. Harwood, and R.J. Kadner, *J. Bacteriol.* **182**, 1459 (2000).
15. M. Lebert and D.P. Häder, *Nature* **379**, 590 (1996).
16. M.A. Bees, Non-Linear Pattern Generation by Swimming Micro-Organisms. PhD Thesis, University of Leeds, Leeds (1998).
17. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes*, 2nd ed., (Cambridge University Press, Cambridge 1992).

# Molecular Dynamics Simulation of Lubricating Films

T. Murakami, S. Yukawa, and N. Ito

Department of Applied Physics, School of Engineering, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

**Abstract.** Lubricating films are studied using a simple microscopic model via molecular dynamics simulation. This model consists of discrete hard-core particles sandwiched between two hard flat walls and one wall moves with constant velocity. In the steady state of this system, it is observed that strong shear induces “freezing” and the interface between solid and fluid is formed autonomously. Regardless of existence of interface, it is confirmed in this simple model that the profiles of velocity and temperature are described by macroscopic continuum description and the properties of kinetic friction is well-reproduced.

## 1 Introduction

Control in nanoscopic level is turned to practical use in various kinds of fields. Even in nanoscopic case, control of friction is essential in the same way as macroscopic control. This is called nanotribology. But dominant factor characterizing the properties of object changes stage by stage as the object becomes smaller. It is not self-evident for continuum description to be applicable to the control of friction in nanoscopic level. Therefore, understanding of nanotribology on the atomic and molecular level is needed.

We find that our fundamental understanding from microscopic level is very good in equilibrium, but it is pretty poor in nonequilibrium. But now, molecular dynamics (MD) simulation provides an alternative approach for understanding from microscopic level in not only equilibrium but also nonequilibrium state [1,2], because computer has been developed ever-progressingly. For tribology, this approach is also applied in many cases [3–6]. As pioneering work of the present paper, we showed hard-core particles model using MD that the interface between solid and fluid is realized steadily and strong shear induces freezing [7]. Using this model, we study the lubricating films that have the possibility of partly induced freezing.

In the present paper, we describe first, the simple model that is investigated as lubricating film. Then we show the profiles and kinetic frictional properties obtained by MD comparing with continuum description. The reason, why macroscopic continuum description is used, is that frictional properties of lubricating film (not thin) is understood well using this description called hydrodynamic lubrication theory.

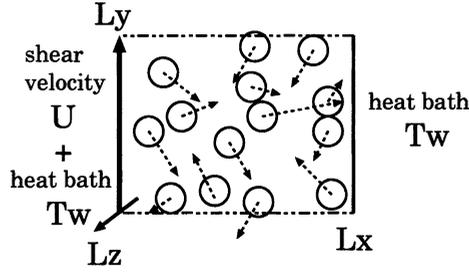


Fig. 1. Image of the model.

## 2 Model

The objective system we study consists of hard-core (Young's modulus is  $\sim 7.9 \times 10^5$ ) particles in a rectangular parallelepiped ( $90 \times 43 \times 16$ ). The particles with uniform radius ( $\sigma = 0.9377$ ) and mass ( $m = 1$ ), collide elastically with each other. The number of particles is 8400 and the density over the entire system is about 0.63, relative to the closed packing. This is below freezing density. Periodic boundary conditions are imposed along  $y$  and  $z$  directions. At the both ends of the  $x$  direction, we put flat walls. When a particle collides with these walls, we impose a stochastic boundary condition that the outgoing velocity is randomly chosen according to the thermal equilibrium distribution with temperature  $T_w (= 1.0)$ :

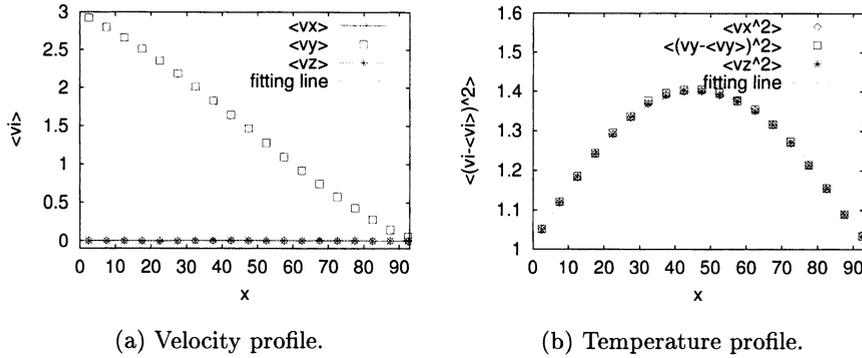
$$f(v_x, v_y, v_z) = \frac{(2\pi T_w)^{-\frac{3}{2}}}{T_w} |v_x| \exp\left(-\frac{v_x^2 + (v_y - \langle v_y \rangle)^2 + v_z^2}{2T_w}\right), \quad (1)$$

independent of the incoming velocity. Here Boltzmann constant is set equal to 1 and  $\langle v_y \rangle$  denotes the constant velocity of each wall in  $y$  direction.  $\langle v_y \rangle$  is set equal to  $U$  at left wall and 0 at right wall, respectively. An image of this model is shown in Fig.1. This model is simulated via molecular dynamics.

## 3 Results

Using this simple model of a lubricating film sandwiched between a moving wall and a fixed one, we investigate the shear velocity dependence of kinetic friction as  $U$  is changed from 0.1 to 7.5 in nonequilibrium steady state. To begin with, profiles for  $x$  direction of two observables are calculated: velocity and temperature. These observables are measured locally at each subsystem whose shape is a rectangular parallelepiped ( $5.0 \times L_y \times L_z$ ). In this case, it is easy to predict that shear induced from the left wall produces the gradients of these observables.

Results of profiles are shown at  $U = 3.0$  as a typical example. Velocity and temperature profiles are shown in Fig.2. Each velocity of no shear direction ( $v_x$



**Fig. 2.** Velocity and Temperature profiles at  $U=3.0$ .

and  $v_z$ ) is 0 within statistical error at any place and velocity profile of shear direction has constant linear gradient almost all over the regions. This gradient is in good agreement with that of Couette flow of hydrodynamic phenomena.

Temperature profile, however, is not so easy to explain, because it is not consistent with linear gradient derived from heat conduction equation. This is because energy transport consists of not only heat conduction but also viscous flow in the bulk fluid. Before we interpret it using macroscopic description, we check that equipartition of energy is realized and velocity distribution function is approximately Gaussian. Therefore, it is valid to assume the local equilibrium, and temperature can be defined as equilibrium way. Energy conservation is expressed as

$$\frac{\partial}{\partial t} \left( \frac{1}{2} \rho v^2 + \rho \epsilon \right) = -\text{div} \left[ \rho \mathbf{v} \left( \frac{1}{2} v^2 + \omega \right) - \mathbf{v} \cdot \boldsymbol{\sigma}' - \kappa \text{grad} T \right], \quad (2)$$

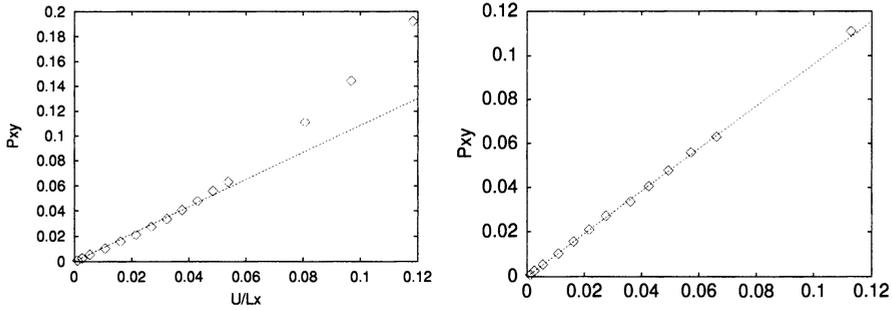
where  $\rho, \epsilon, \omega, \boldsymbol{\sigma}'$  and  $\kappa$  denote density, energy density, enthalpy density, viscous stress tensor and thermal conductivity, respectively. Assuming that pressure, thermal conductivity and shear viscosity  $\eta$  are constant at any position and the fluid is Newtonian, the solution of (2) becomes

$$T(x) = -Ax^2 + AL_x x + T_w, \quad (3)$$

$$A = \frac{\eta}{\kappa} \left( \frac{\partial v_y}{\partial x} \right), \quad (4)$$

As a result, the temperature profile obtained from our model is well-described by the single parameter parabolic function of (3). Therefore, continuum description within linear deviations from equilibrium can be applied to these profiles of this small model.

Now, we study kinetic frictional properties as lubricant of this model. Shear viscosity  $\eta$  is first of all, investigated, because viscous drag causes the increase of



(a) Velocity dependence of stress tensor.

(b) Shear rate dependence of stress tensor.  $-\frac{\partial v_y}{\partial x}$  is plotted along horizontal axis.

**Fig. 3.** Velocity dependence of stress tensor.

friction of oil lubricant, in the region of hydrodynamic lubrication we consider here.  $\eta$  is calculated from

$$P_{xy} = \eta \frac{\partial v_y}{\partial x} \quad (5a)$$

$$= \eta \frac{U}{L_x}, \quad (5b)$$

here  $P_{xi}$  is the stress tensor. In this study, it is measured as the momentum of  $i$  direction added from left wall per unit time and area. The velocity gradient can be regarded as  $U/L_x$  from the results of above-described velocity profiles. Assuming that shear viscosity is independent of shear velocity,  $P_{xy}$  should be proportional to  $U$  like (5b). Shear velocity dependence of  $P_{xy}$  is shown in Fig.3(a). This proportionality is realized in small shear velocity region, but appears to break down in large velocity region. This may be due to the change of shear viscosity. Supposing this possibility, property of velocity profiles should also depend on shear velocity. Therefore, we check the velocity profile at large  $U$ , for example  $U = 7.5$ . This is shown in Fig.4(a). This profile is much different from what is shown at  $U = 3.0$ . The region where  $v_y$  stays constant, is realized near each wall, but the other region has still Couette-type gradient. This result suggests that solid phase is organized at both ends and freezing is confirmed from investigating that local density in these regions are larger than solidified density. This transition is what we call shear-induced freezing [7]. This suggestion is also confirmed from linear gradients of temperature profile of Fig.4(b).

Due to freezing, the region where the lubricant behaves like fluid does not exist all over the system, and the effective shear rate  $-\frac{\partial v_y}{\partial x}$  is larger than  $U/L_x$ . Therefore, it seems to be appropriate to express  $P_{xy}$  not as (5b) but (5a) where  $-\frac{\partial v_y}{\partial x}$  is estimated from the velocity profile of the fluid region. This is shown in Fig.3(b). Above-described proportionality is got back and this simple model

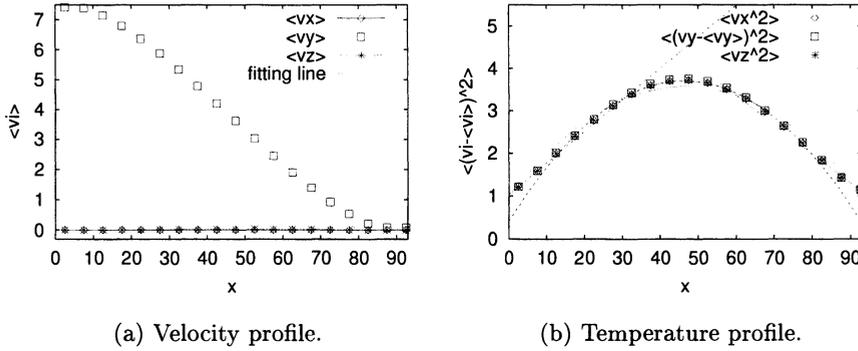


Fig. 4. Velocity and Temperature profiles at U=7.5.

of lubricating films is confirmed that shear viscosity is independent of shear velocity.

At last, we proceed to discuss what is aimed at. Two objects are considered: coefficient of kinetic friction  $\mu$  and frictional heat  $J$ .  $\mu$  and  $J$  are measured as

$$\mu = \frac{P_{xy}}{P_{xx}} \tag{6}$$

and the energy produced by left wall per unit time and area, respectively. From suggestion of above-described results, we estimate these two quantities using continuum description as

$$\mu = \frac{\eta}{P_{xx}} \frac{\partial v_y}{\partial x} \tag{7}$$

$$J(x) = -2\eta v_y \frac{\partial v_y}{\partial x} - \kappa \frac{\partial T}{\partial x} \tag{8}$$

The form of (7) is clear to consider the dependence of shear viscosity, shear rate and load. On the other hand, (8) is difficult to evaluate because thermal conductivity needs to be measured. Therefore, we use

$$J\left(\frac{L_x}{2}\right) = -2\eta v_y \frac{\partial v_y}{\partial x} \sim \eta \left(-U \frac{\partial v_y}{\partial x}\right) \tag{9}$$

instead of (9), here we assume  $\frac{\partial T}{\partial x} = 0$  at  $x = L_x/2$ . This assumption is proper in case existing just fluid phase like Fig.2(b) but is not clear in solid-fluid coexisting case. Temperature profile at  $U = 7.5$  is shown in Fig.4(b). This assumption is confirmed to be proper even in this case and furthermore, energy conservation of (2) describes this result very well for each phase. (In solid phase, energy flux consists of heat conduction only.) Shear velocity dependences of coefficient of kinetic friction and frictional heat are shown in Fig.5(a) and Fig.5(b), respectively. These results are consistent with macroscopic continuum description. Therefore

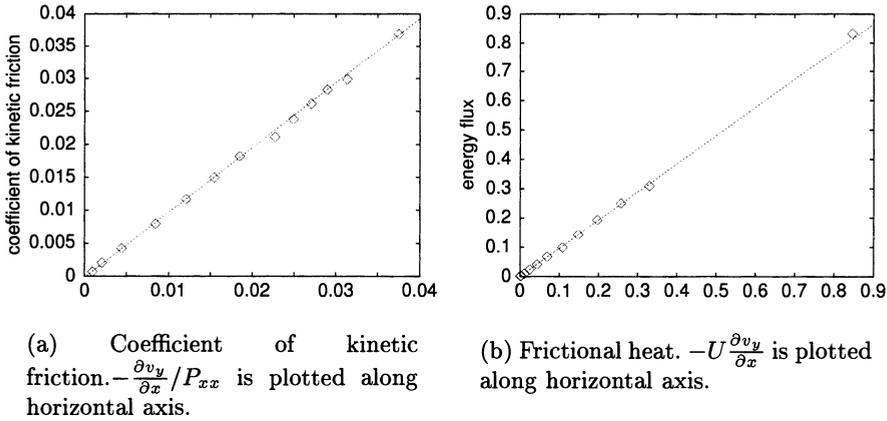


Fig. 5. Shear rate dependence of kinetic friction.

This small simple model represents the kinetic frictional properties of lubricants in spite of very small system.

## 4 Conclusion

We investigated the kinetic frictional properties of a simple model using MD and concluded that this small model consisting of discrete particles is in good accord with what macroscopic continuum description expects and shows the properties of lubricant regardless of existence of interface. In particular, it is very interesting that shear viscosity is estimated at the same value 0.96 within an accuracy of statistical error in three different ways: stress tensor, coefficient of kinetic friction and frictional heat. This study will contribute to understand phase transition accompanied with hydrodynamic flow. But the quantitative relation between this nonequilibrium molecular dynamics and linear response, and application to nano-machine, are the next challenges.

## References

1. D.C. Rapaport, Phys. Rev. A **36**, 3288 (1987).
2. M. Mareschal, M.M. Mansour, A. Puhl, and E. Kestemont, Phys. Rev. Lett. **61**, 2550 (1988).
3. P.A. Thompson and M.O. Robbins, Science **250**, 72 (1990).
4. M. Cieplak, J. Koplik, and J. Banavar, Phys. Rev. Lett. **86**, 803 (2001).
5. S. Nasuno, A. Kudrolli, and J.P. Gollub, Phys. Rev. Lett. **79**, 949 (1997).
6. T. Kawaguchi and H. Matsukawa, Phys. Rev. B **58**, 15866 (1998).
7. T. Murakami, S. Yukawa, and N. Ito: 'Non-equilibrium Energy Transport in Multi-phase System', in: *Recent Developements in Computer Simulation Studies in Condensed Matter Physics, February, 19-24, 2001*, D.P. Landau, S.P. Lewis, and H.-B. Schuttler, (Eds.), (Springer, Verlag), to appear.