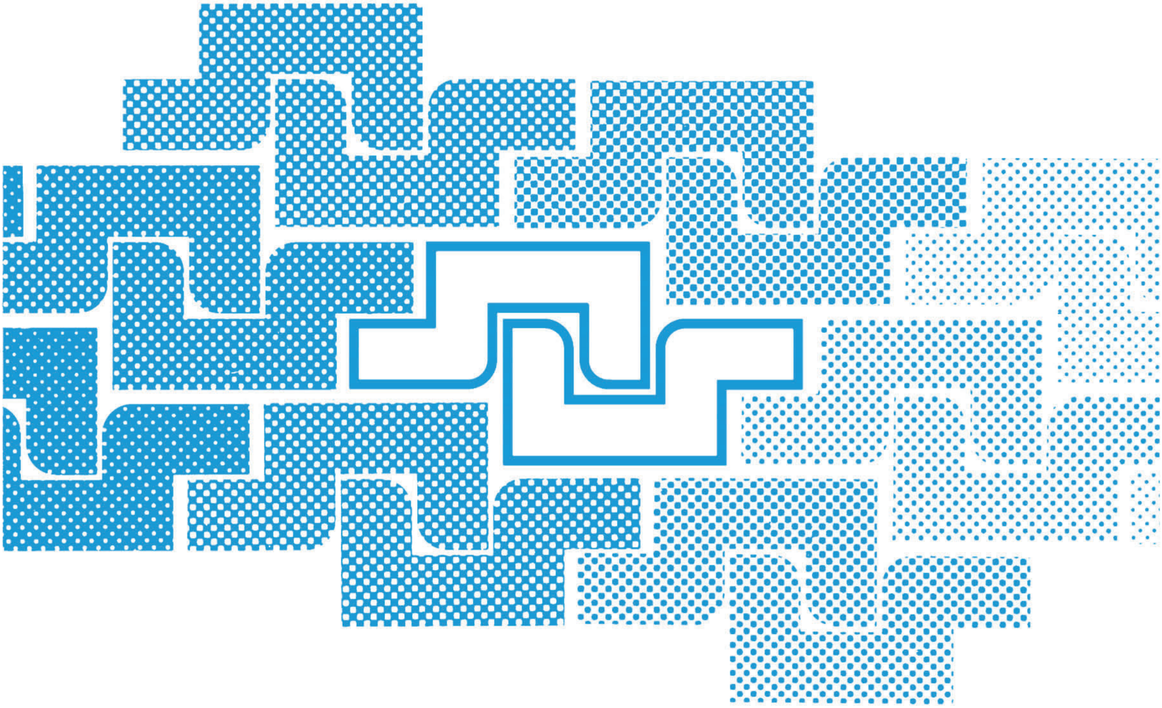


Dynamics of Complex Interconnected Systems: Networks and Bioprocesses

Edited by

Arne T. Skjeltorp and
Alexander V. Belushkin

NATO Science Series

II. Mathematics, Physics and Chemistry – Vol. 232

Dynamics of Complex Interconnected Systems: Networks and Bioprocesses

NATO Science Series

A Series presenting the results of scientific meetings supported under the NATO Science Programme.

The Series is published by IOS Press, Amsterdam, and Springer in conjunction with the NATO Public Diplomacy Division

Sub-Series

I. Life and Behavioural Sciences	IOS Press
II. Mathematics, Physics and Chemistry	Springer
III. Computer and Systems Science	IOS Press
IV. Earth and Environmental Sciences	Springer

The NATO Science Series continues the series of books published formerly as the NATO ASI Series.

The NATO Science Programme offers support for collaboration in civil science between scientists of countries of the Euro-Atlantic Partnership Council. The types of scientific meeting generally supported are “Advanced Study Institutes” and “Advanced Research Workshops”, and the NATO Science Series collects together the results of these meetings. The meetings are co-organized bij scientists from NATO countries and scientists from NATO’s Partner countries – countries of the CIS and Central and Eastern Europe.

Advanced Study Institutes are high-level tutorial courses offering in-depth study of latest advances in a field.
Advanced Research Workshops are expert meetings aimed at critical assessment of a field, and identification of directions for future action.

As a consequence of the restructuring of the NATO Science Programme in 1999, the NATO Science Series was re-organised to the four sub-series noted above. Please consult the following web sites for information on previous volumes published in the Series.

http://www.nato.int/science
http://www.springer.com
http://www.iospress.nl

Series II: Mathematics, Physics and Chemistry – Vol. 232

Dynamics of Complex Interconnected Systems: Networks and Bioprocesses

edited by

Arne T. Skjeltorp
Institute for Energy Technology,
Kjeller, Norway

and

Alexander V. Belushkin
Frank Laboratory of Neutron Physics,
Dubna, Russia

Springer

Published in cooperation with NATO Public Diplomacy Division

Proceedings of the NATO Advanced Study Institute on
Dynamics of Complex Interconnected Biosensor Systems: Networks and Bioprocesses
Geilo, Norway
11–21 April 2005

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-5029-1 (PB)
ISBN-13 978-1-4020-5029-9 (PB)
ISBN-10 1-4020-5028-3 (HB)
ISBN-13 978-1-4020-5028-2 (HB)
ISBN-10 1-4020-5030-5 (e-book)
ISBN-13 978-1-4020-5030-5 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

CONTENTS

Preface vii

Organizing committee and participants ix

Structure and Communication in Complex Networks 1
Kim Sneppen (invited)

Effects of Community Structure on Search and Ranking in Complex Networks 29
Huafeng Xie, Koon-Kiu Yan and Sergei Maslov (invited)

The SOS response of Bacteria to DNA damage 39
Joel Stavans (invited)

Self-Affine Scaling during Interfacial Crack Front Propagation 49
Stéphane Santucci (invited), Knut Jørgen Måløy, Renaud Toussaint and Jean Schmittbuhl

Diffusion, Fragmentation and Merging Processes in Ice Crystals, Alpha Helices and other Systems 61
Jesper Ferkinghoff-Borg, Mogens H. Jensen (invited), Poul Olesen and Joachim Mathiesen

Molecular Mechanisms in Biosignalling: Visual Reception 71
Mikhail A. Ostrovsky (invited)

The Architecture of Complexity: From WWW to Cellular Metabolism 107
Eivind Almaas and Albert-Lázló Barabási (invited)

Mahematical Modeling of Neural Activity 127
Gaute T. Einevoll (invited)

Braided Space-Time Particle Networks 147
Kai de Lange Kristiansen, Geir Helgesen and Arne T. Skjeltorp (invited)

Combining Optical Tweezers and Micropipettes for DNA Stretching: Elasticity of Micropipette crucial 163
Thomas Møller Hansen, Nader Reihani and Lene Oddershede (invited)

Universal Networks and Processes in Soft and Complex Matter: From Nano to Macro 175
Jon Otto Fossum (seminar)

What Economists should Learn from Econophysics 191
Joseph McCauley (seminar), Gemunu H. Gunaratne and Kevin E. Bassler

The Minority Game: Statistical Physics of Collective Behaviour of Adaptive Agents in a competitive Market 203
David Sherrington (seminar)

Index 211

PREFACE

This volume comprises the proceedings of a NATO Advanced Study Institute (ASI) held at Geilo, Norway, 11-21 April 2005, the eighteenth ASI in a series held every two years since 1971. The objective of this ASI was to identify and discuss areas where synergism between modern physics and biology may be most fruitfully applied to the study of bioprocesses for molecular recognition, and of networks for converting molecular reactions into usable signals and appropriate responses.

Many fields of research are confronted with networks. Genetic and metabolic networks describe how proteins, substrates and genes interact in a cell; social networks quantify the interactions between people in the society; the Internet is a complex web of computers; ecological systems are best described as a web of species. In many cases, the interacting networks manifest so-called emergent properties that are not possessed by any of the individual components. This means that the detailed knowledge of the components is insufficient to describe the whole system. Recent work has indicated that networks in nature have so-called scale-free characteristics, and the associated dynamic network modelling shows unexpected results such as an amazing robustness against accidental failures, a property that is rooted in their inhomogeneous topology. Understanding these phenomena and turning them to use in chemical and biological threat detection and response will require exploring a wide range of network structures as well. Questions related to error and attack tolerance of complex networks and their robustness in particular, and the dynamics of networks in general also have to be addressed. Modelling the signal transduction networks in bioprocesses as in living cells is a challenging interdisciplinary research area. It is now realized that the many features of molecular interaction networks within a cell are shared to a large degree by the other complex systems mentioned above, such as the Internet, computer chips and society. Thus knowledge gained from the study of complex non-biological systems can be applied to the intricate braided relationships that govern cellular functions. Bio-inspired processes provide an attractive option for sensing chemical and biological (CB) agents because nature has solved many of the problems inherent to the sense-and-respond task. For example, many biological responses such as blood clotting, gene expression and the activation of enzymes require enormous amplification of signals carried by as few as a single molecule or ion. Adaptability to local environments, atomic level control of self-assembled structures, benign processing, combinatorial synthesis and complex computation are other features of biological systems that are likely to prove useful in CB sensor development.

The starting point, and the underlying theme throughout the ASI, was a thorough discussion of general network theory. The next focus was on genetic networks and bioprocesses. Finally, focus was placed on the possible universality of network structures and how this knowledge can be combined to attack the urgent problem of rapid detection and diagnosis of CB agents.

The NATO ASI format in Geilo proved very efficient in getting researchers in different areas together and focus on the underlying theme that was common to all of them – that of networks and bioprocesses. In this manner a rapid communication was possible because a common vocabulary was developed during the ASI.

The scientific content of the school was timely and these proceedings should provide a useful definition of the current status. The Institute brought together many lecturers, students and active researchers in the field from a wide range of countries, both NATO and NATO Partner Countries. The lectures fulfilled the aim of the Study Institute in creating a learning environment and a forum for discussion on the topics stated above. They were supplemented by a few contributed seminars and a large number of poster presentations. These seminars and posters were collected in extended abstract form and issued as an open report available at the Institute for Energy Technology, Kjeller, Norway (Report IFE/KR/E-2005/005, ISBN 82-7017-535-8).

Financial support was principally from the NATO Scientific Affairs Division, but also from the Institute for Energy Technology and the nationally coordinated research team COMPLEX in Norway.

The editors are most grateful to M.H. Jensen, J.L. McCauley, R. Pynn and H. Thomas who helped them plan the programme and G. Helgesen for helping with many practical details. Finally, we would like to express our deep gratitude to Trine Løkseth of the Institute for Energy Technology, for all her work and care for all the practical organization before, during, and after the school, including the preparation of these proceedings.

December 2005

Arne T. Skjeltorp Alexander V. Belushkin

LIST OF PARTICIPANTS

Organizing Committee:

Skieltorp, Arne T., Director
Institute for Energy Technology, POB 40, N-2027 Kjeller, Norway

Belushkin, Alexander V, Co-director
Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research, 141980
Dubna, Moscow region, Russia

Helgesen, Geir, Technical assistant
Institute for Energy Technology, POB 40, N-2027 Kjeller, Norway

Løkseth, Trine, Secretary
Institute for Energy Technology, POB 40, N-2027 Kjeller, Norway

Participants:

Andersson, Anna
Hillerodsgade 69, 2200 Copenhagen, Denmark

Amatori, Andrea
Gruppo di fisica del nucleo Dipartimento di fisica, via celoria 16, 20133 Milano, Italy

Angheluta, Luiza
Peder Skrams Gade, 26, 4tv., 1054 Copenhagen

Apostol, Simona
Physics Department, Faculty of Sciences, Valahia University, 24 Bd Unirii, 0200 Targoviste, jud Dimbovita, Romania

Bakaeva, Zulfiya
28, Katartal street, 700 135, Tashkent, Uzbekistan

Barabasi, Albert
Department of Physics, 225 Nieuwl and Science Hall, University of Notre Dame, Notre Dame, IN 46556, USA

Beck, Peter
Sicherheit und Risiko, Austrian Research Centers GmbH – ARC, A-2444
Seibersdorf, Austria

Bernhardsson, Sebastian
c/o J. Hansen, Osterbrogade 95, 2100 Copenhagen, Denmark

Bock Axelsen, Jacob
Niels Bohr Institutet, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Cernak, Jozef
University of P.J. Safarik, Department of Biophysics, Jesenna 5,
SK-04000 Kosice, Slovak Republic

Christophorov, Leonid
Bogoliubov Institute for Theoretical Physics, NAS Ukraine, 14 B
Metrologichna Str, Kiev 03143, Ukraine

da Silva, Geraldo José
Instituto de Fisica, Universidade de Brasilia, CP 04525 70919-970
Brasilia DF, Brazil

Dmitrochenko, Oleg
Department of Mechanical Design Engineering, RIMT, CAE Lab, Pusan
National University, 30 Jangjeon-Dong, Geumjeong-Gu, Busan 609-735,
South Korea

Duman, Memed
Institude of Graduate and Applied Sciences, Bioengineering Division,
Hacettepe University, Beytepe, 06800 Ankara, Turkey

Eisler, Viktor
Institute for Theoretical Physics, Eotvos University, 1117, Pazmany
setany 1/A, Hungary

Einevoll, Gaute T.
Department of Mathematical Sciences and Technology
Norwegian University of Life Sciences, P.O.Box 5003, N-1432 Aas,
Norway

Elgsaeter, Arnljot
Department of Physics, NTNU, Høgskoleringen 5, N-7491 Trondheim,
Norway

Fossum, Jon Otto
NTNU Department of Physics, Høgskoleringen 5, N-7034 Trondheim, Norway

Garipcan, Bora
Institute of Graduate and Applied Sciences, Bioengineering Division
Hacettepe University, Beytepe, 06800 Ankara, Turkey

Giaever, Ivar
Institute of Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Goksu, Huseyin
Istiklal Mh. Fatih Sk. 21/5, Isparta, 32300, Turkey

Grigoryan, Arsen
Departmetnt of Molecular Physics Faculty of Physics, Yerevan State University, Al. Manoogian Str. 1, Yerevan 375025, Annenia

Hamedi, Mahiar
Inst. för fysik och mätteknik, Linköpings Univesitet/Tekniska Högskolan, 581 83 Linköping, Sweden

Hansen, Alex
Institutt for fysikk, NTNU Gløshaugen, N-7491 Trondheim, Norway

Heiberg-Andersen, Henning
Institute for Energy Technology, POB 40, NO-2027 Kjeller, Norway

Hervoches, Charles
Institute for Energy Technology, POB 40, NO-2027 Kjeller, Norway

Hritz, Jozef
Department of Biophysics, PF-UPJS, Jesenna 5, 04154 Kosice, Slovak Republic

Høgh Jensen, Mogens
Niels Bohr Institute, Blegdarnsvej 27, DK-2100 KØBENHAVN Ø, Denmark

Karlik, Bekir
Department of Computer Engineering, Engineering Faculty, Halic University, Molla Gurani cad. No. 16-18 Findikzacle, 34280-Istanbul, Turkey

Kharlamov, Alexey
3 Krjijanovskiy str., 03142 Kiev, Ukraine

Kharlamova, Ganna
14, Glushkov str., aprt.89., 03187 Kiev, Ukraine

Khizanishvili, Ana
E.Andronikashvili Institute of Physics, 6 Tamarashvili str., Tbilisi, 0177, Georgia

Knudsen, Kenneth D.
Institute for Energy Technology, POB 40, NO-2027 Kjeller, Norway

Kristiansen, Kai de Lange
Fysisk Institutt, University of Oslo, Box 1048 Blindern, 0316 OSLO, Norway

Korraa, Soheir
3 Ahmed El Zomour St. 8th Sector P.O. Box 29 Nasr City, Cairo, Egypt

Krishna, Sandeep
Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen O, Denmark

Libchaber, Albert
Laboratory of Experimental Condensed Matter Physics, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

Lunika, Michael
58 Petropavlovskaja Str., Sumy 40030, Ukraine

Marconi, Daniela
c/o Dipartimento di fisica Campaninini Renato, Viale Berti Pichat 6/2, 40100 Bologna, Italy

Maslov, Sergei
Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, USA

Mathiesen, Joachim
Department of Physics, NTNU, Hogskoleringen 5, 7491 Trondheim, Norway

McCauley, Joseph L.
Tiefetfeld Str. 14, A6632 Ehrwald, Austria

Méheust, Yves
Department of Physics, NTNU, Hogskoleringen 5, 7491 Trondheim, Norway

Micheelsen, Mille Ankerstjeme
The Niels Bohr Institute, University of Copenhagen Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Måløy, Knut Jørgen
Fysisk Institutt, University of Oslo, Box 1048 Blindem, 0316 Oslo, Norway

Muller, Jiri
Institute for Energy Technology, POB 40, NO-2027 Kjeller, Norway

Nawrocki, Waldemar
ul. Piotrowo 3A, PL-60965 Poman, Poland

Nepusz, Tamas
Department of Biophysics, Research Institute for Particle and Nuclear Physics, Konkoly – Thege Miklos u. 29-33, 1121 Budapest, Hungary

Oddershede, Lene
Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

Oliveira, Fernando A.
International Center for Condensed Matter Physics, Universidade de Brasilia, CP 04513, 70919-970 Brasilia DF, Brazil

Ostrovsky, Mikhail
Institute of Bio-Chemical Physics, Russian Academy of Sciences, Kosygin St 4, 117334 Moscow, Russia

Podolskaya, Olga
Heat Physics Department, Uzbekistan Academy of Sciences, Katartal 28, Tashkent 700135, Uzbekistan

Priezzhev, Vyacheslav, B
Office 203, Sector 14, BLTPJINR, 141980 Dubna, Moscow Region, Russia

Pynn, Roger
Materials Research Lab., UCSB, Santa Barbara, CA 93106-5130, USA

Ramos, Osvanny
Department of Physics, P.O. Box 1048 Blindern, N-0316 Oslo, Norway

Rosvall, Martin
Institutionen för fysikk, Umeä universitet, 90187 Umeä, Sweden

Santucci, Stephane
Department of Physics, P.O.Box 1048 Blindem, N-0316 Oslo, Norway

Shantsev, Daniel
Department of Physics, University of Oslo, POB 1048 Blindern, NO-0316 Oslo, Norway

Sherrington, David
University of Oxford, Theoretical Physics, I Keble Rd. Oxford OXI 3NP, United Kingdom

Sibanf, Paolo
Fysisk Institut, SDU, Campusvej 55, DK5230 Odense M, Denmark
Skau, Karl Isak
Department of Mathematics and Statistics, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

Skoge, Monica
Grad College – Princeton Univ., 88 College Rd., West Princeton, NJ 08544, USA

Skotheim, Jan
Cambridge University, Queens College, Cambridge CB3 9ET, United Kingdom

Stan, Mariana
Sat Straosti nr. 99, Comuna Dragodana, Judet Dambovita, Cod 0184, Romania

Sneppen, Kim
Nordita, Blegdamsvej 17, 2100 Copenhagen OE, Denmark

Stavans, Joel
Department of Physics of Complex Systems, Weimann Institute of Science, PO Box 26, Rehovot 76100, Israel

Steinsvoll, Olav
Institute for Energy Technology, POB 40, NO-2027 Kjeller, Norway

Svåsand, Eldrid
Institute for Energy Technology, POB 40, NO-2027 Kjeller, Norway

Tchesskaya, Tatyana
Department of Information Technology, Odessa State Environmental University, 15 Lvovskaya str., 65016 Odessa, Ukraine

Thomas, Harry
Dept. of Physics, University of Basel, CH-4056 Basel, Switzerland

Tourleigh, Yegor
Leninskiye Gory, 1/12, Dept. of Biophysics, Moscow State University, Russia

Tropin, Timur
FLNP, JINR, 141980 Dubna, Moscow reg, Russia

STRUCTURE AND COMMUNICATION IN COMPLEX NETWORKS

KIM SNEPPEN
Center for Models of Life
Niels Bohr Institute, Blegdamsvej 17,
2100 Copenhagen

1. Introduction

Networks is tool to describe systems composed of many different units which each typically interact with a few of the other units. Networks thus used to quantify complex systems from the intricate interactions of proteins inside a living cell, to ecosystems, social systems and computer networks. In most cases the network quantify communication channels in the system. Thus directly connected nodes communicates easy, while more distant nodes only obtain exchange information through a number of intermediate steps. In fact already in 1982 a detailed study of social networks within university departments revealed that mutual information of one member about another one was decaying exponentially with their distance, and increased linearly with number of common friends (degenerate paths). We will take this viewpoint and consider a Network as a description of who get direct information from who, and which parts that has to resort to second hand or even more inaccurate information: networks quantify the extent to which complex systems operate under the constraints of a limited information horizon.

In the text we will first review a few facts about real world networks, including broad degree distributions and degree correlations and some simple models for how such broad degre distributions may self organize. Secondly we introduce a few measures to characterize the topology of network: Correlation profiles, network motifs, topological hierarchy, search information and a measure for order in topology of the networks. Thirdly we discuss self organization of networks under various degrees of limited information. Finally we discuss biological networks, with focuss on basic physics of gene regulation.

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 1–28.

2. Basic Network Concepts

In Figure 1-4 we define a few basic quantities in graph theory. For a pair of nodes one define their distance as the length of the shortest path between them. For a connected graph we can define its diameter as the maximum distance between pairs. Examples of graphs include lattices in any dimension.

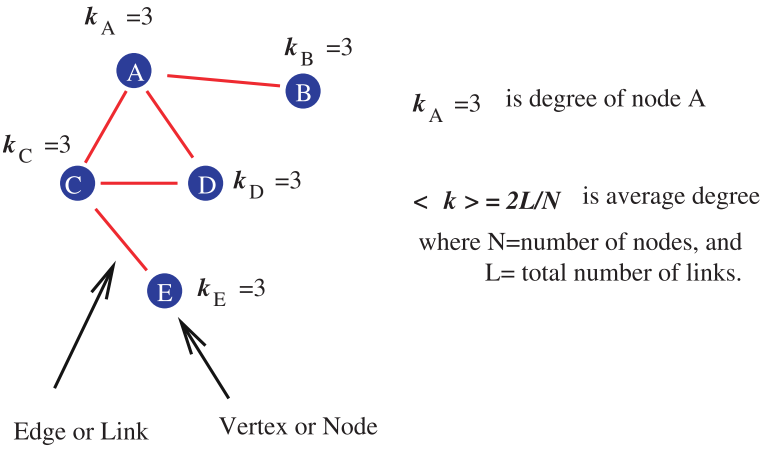


Figure 1. A graph, with nodes and edges.

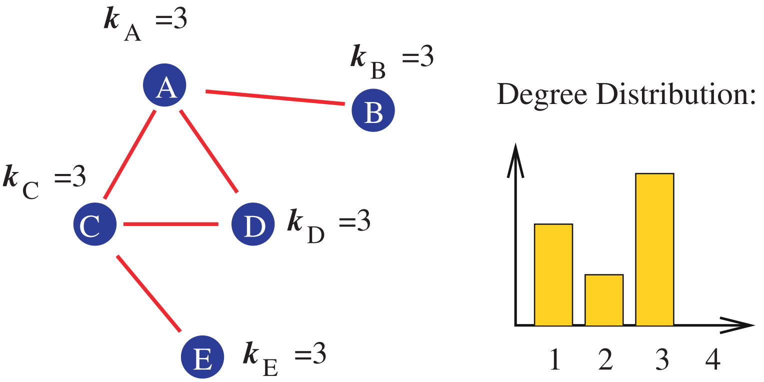


Figure 2. Degree distribution, here for undirected graph.

We now want to consider the simplest possible model for a random graph, the Erdos-Reynei Graph. In this type of graph one have N nodes, each pair connected to each other with probability p. The average (expected number of edges would be $L = p \cdot N(N-1)/2$. The degree distribution (probability that a given node have k links to the remaining N-1 nodes) will be

$$P(k) = C^k_{N-1} p^k (1-p)^k \tag{1}$$

where $C^k_{N-1} = (N-1)!/(k!(N-1-k)!)$. For a random graph one gain the neighborhood of any node increases by factor $\langle k \rangle$ for each step out in the graph (for large graphs, as long as there is essentially zero probability to reach same nodes through two different paths). Thus the network is traversed when

$$\sum_{i=0}^{diam} \langle k \rangle^i \approx N \tag{2}$$

that is when $diam = log(N)/log(\langle k \rangle)$, i.e. the distance between any two points in a random graph grows very slowly with its overall size.

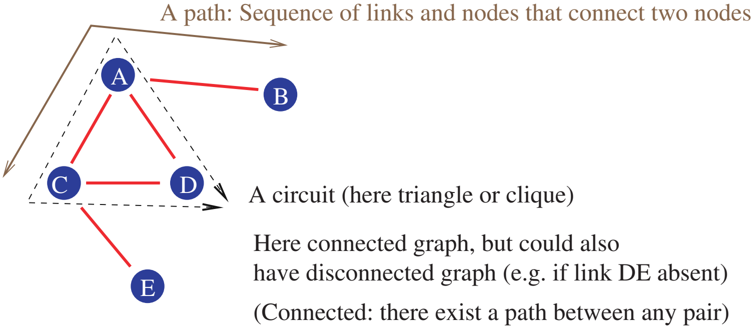
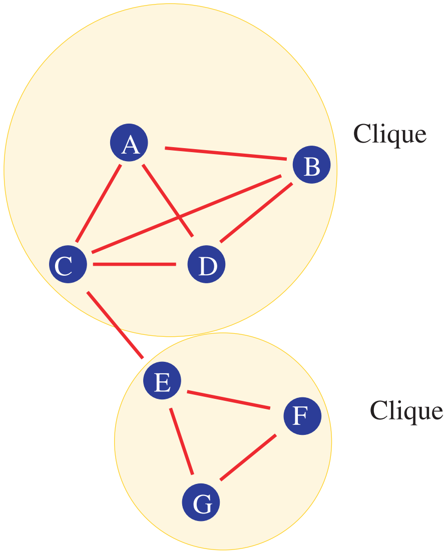


Figure 3. Path's in graphs.

Another quantity which often is used to characterize networks is their cliquishness. For each node this is defined as the fraction shown in Fig 4, meaning that a large cliquishness indicate large locality in the network everybody knows each other locally).

A network is defined as having small world property, when it is having a relatively large cliquishness, while still having a diameter of order $log(N)/log(k)$. Many networks are indeed found to have this interplay between global accessibility and local connectedness.

Finally we would like to stress that a network also have a unique matrix representation, A_{ij} where a link from node i to node j implies that $A_{ij} = 1$. A undirected network, even a symmetric matrix, $A_{ij} = A_{ji}$. The matrix representation opens for some simple manipulation where one for example can calculate the number of triangles in a network $= trace A^3/6$, where the factor $6 = 3 \cdot 2$ comes from going two ways around each triangle, and from starting the triangle a any of the three nodes.



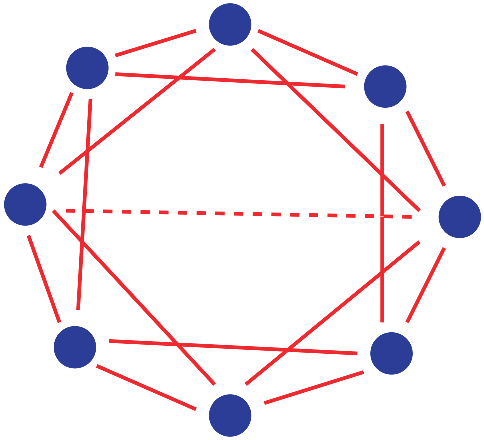
$$C_A = \frac{\text{no. of cliques}}{\text{max no of cliques}}$$

$$= \frac{\text{no. of cliques}}{k_A(k_A - 1)/2}$$

$$= \frac{3}{3} = 1$$

$$C_C = \frac{3}{6}$$

Figure 4. Cliquishness definition: the fraction of possible triangles that a node participate in, given its degree. If all its friends are friends, its cliquishness is 1.



Network with rater large cliquishness

C = 3/6 (except ––––

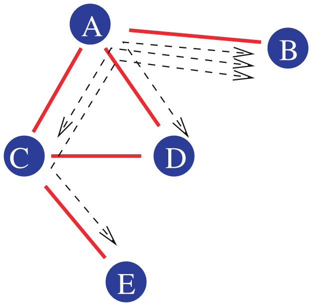
and with few additional links (–––), it also have rather small diameter.

Then SMALL WORLD

Figure 5. Small world network (Watts and Strogatz).

3. Real Networks

From the genetic blueprint in our DNA to the world-wide Internet, information and its dynamic counterpart communication sets the stage for further action. However, we live under the limited information horizon, in the sense that information is often imperfect and communication is always finite. Many complex systems are associated with information transfer. This includes neural networks with their synaptic rewiring, molecular networks



Betweenness:

The betweenness of a node is
number of shortest paths
that goes through the node

β_A = no. path through A = 3

similar definition of link betweenness

(Freeman, Sociometry 1977)

Figure 6. Betweenness (load) on a node, assuming that all transport goes through shortest paths.

evolved to modulate the protein activity in living cells, and social networks exemplified by the Internet. Below we show respectively the worldwide internet, a social network and some examples of molecular networks inside living cells. In all these cases the purpose of the network is to serve as a scaffold for information transfer.

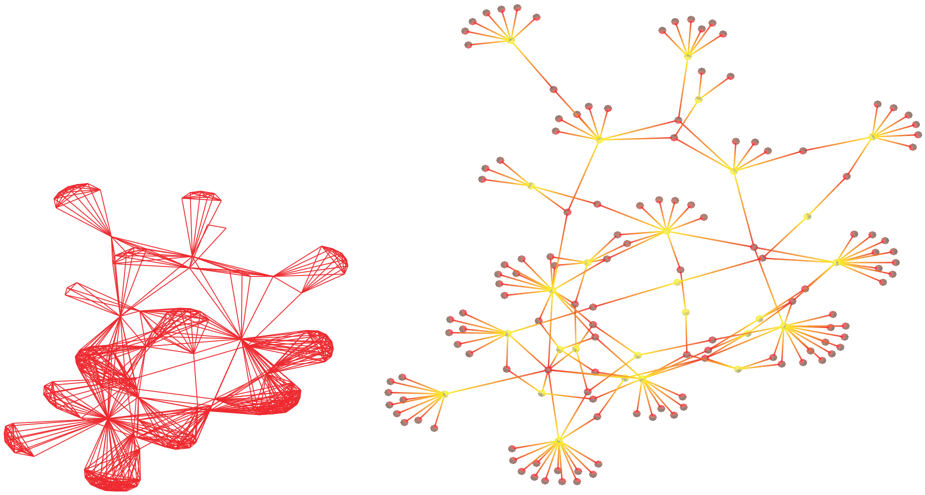


Figure 7. Directors in US: Left representation, connected by link if two nodes in same board. Right representation: Bipartite network representation consist of respective board nodes and member nodes, defining two distinct classes of nodes. In a bipartite graph there are only links from nodes in one class to a node in the other class. The shown network is a small sub-part of the full network of all CEO’s in USA.

For many real world networks, the distribution of proteins with a given number of neighbors (connectivity) K may (very crudely) be approximated by a power law

$$N(K) \propto 1/K^{\gamma} \tag{3}$$

with exponent $\gamma \sim 2.5 \pm 0.5$ (Jeong et al. (2001)) for protein-protein binding networks, and exponent $\gamma \sim 1.5 \pm 0.5$ for "out-degree" distribution of transcription regulators. Notice that the distribution of the number of proteins a given protein regulate, the "out-degree", differs from the much more narrow distribution of "in-degrees". We would now like to discuss features and possible reasons for why life may have chosen to organize its signaling in this way.

One aspect of a wide distribution of connectivity, is the possible amplification of signals in the network. Consider a signal that enters a node, and assume that it is transmitted along all exit links (unspecific broadcasting). Thus it is amplified by a factor K_{out}. However, not all nodes has equal chance to amplify signals. The probability to enter a node is proportional to K_{in}. Thus one in average visit nodes with probability $\propto K_{in}$ and the weighted average amplification factor in a *directed* network (Newman et al. 2001):

$$\mathcal{A} = \frac{\langle K_{in} K_{out}(given \;\; K_{in}) \rangle}{\langle K_{in} \rangle} = \frac{\langle K_{in} K_{out} \rangle}{\langle K_{in} \rangle} \tag{4}$$

The first equality assumes that there is no correlations between degree of a node and the degree of its neighbors. The second equality assumes that there is no correlations between a given proteins "in" and "out" degree. For undirected random network the amplification would be:

$$\mathcal{A} = \frac{\langle K(K-1) \rangle}{\langle K \rangle} \tag{5}$$

If all nodes has close to the same connectivity we recover the simple result that when $\langle K \rangle = 2$ then $\mathcal{A} = 1$. Thus to have marginal transmission, one input signal in average should lead to one output signal through a new exit. When $\mathcal{A} > 1$ signals tends to be exponentially amplified, and thus most signals will influence signaling over the entire network. For broad connectivity distributions $\mathcal{A}$ typically depend on the node with highest connectivity. To see this, assume that the number degree distribution is power law distributed (eq. 3). Then:

$$\mathcal{A} = \frac{\int_1^N \frac{K^2 dK}{K^{\gamma}}}{\int_1^N \frac{K dK}{K^{\gamma}}} - 1 \sim N^{3-\gamma} \tag{6}$$

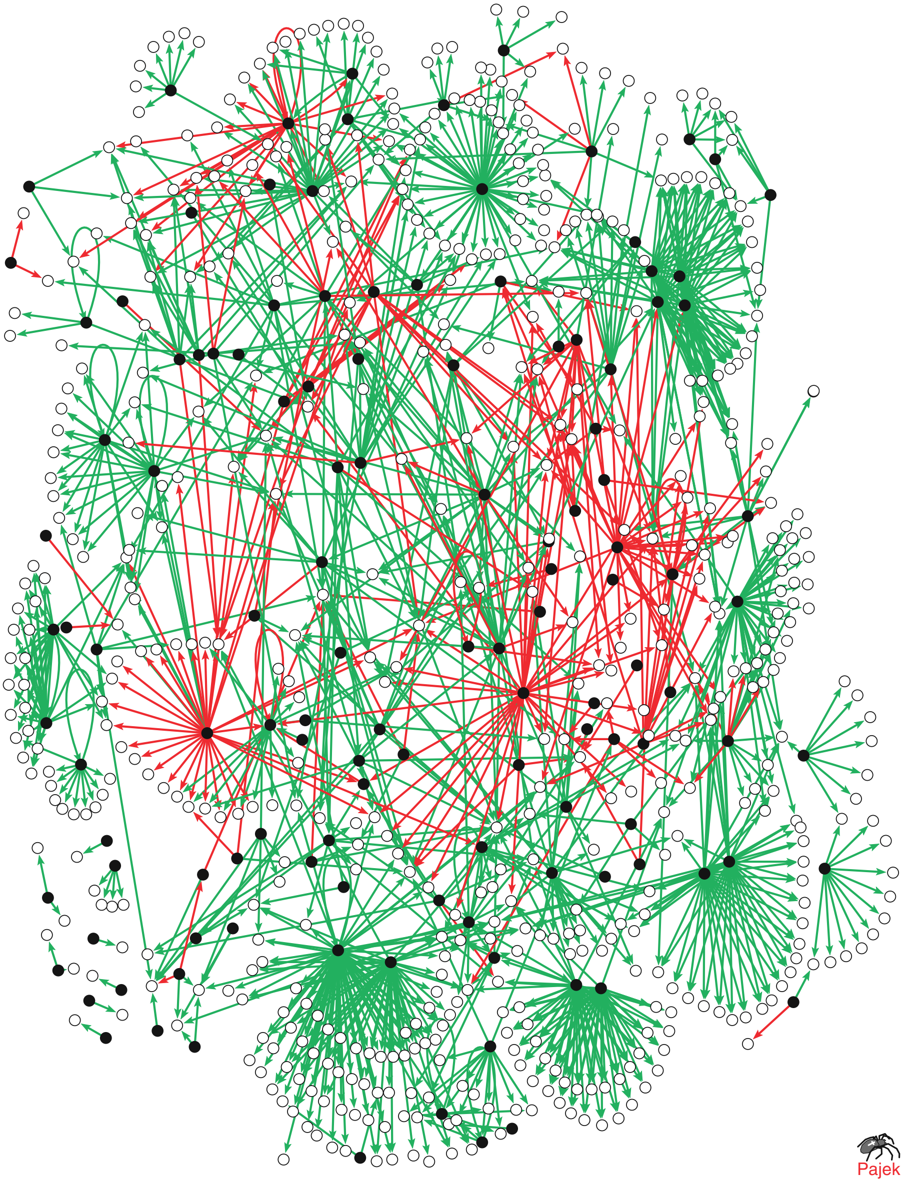


Figure 8. Presently known transcription regulations in Yeast (Sacromyces Cerevisia). Transcription regulation describes how one protein regulate the production of another protein. The network is directed, with both positive and negative regulations (see later for more detailed explanation).

for $\gamma > 2$. That is, in case of $\gamma > 2$ the denominator becomes independent of N in the limit of large N. One notice that for $\gamma < 3$ then $\mathcal{A}$ is dependent of the upper cut off in the integral, which represent the node (protein) with highest connectivity. Most real world networks are fitted with exponents between 2 and 3. We stress that the estimate in eq. 6 only is valid when the network consist of nodes which are randomly connected to each other. Further we are analyzing the networks as if all signals are going anywhere where there is connections.

4. Models for scale free networks

The scale free nature of the connectivity distribution itself is not that convincing for biological network as for other real networks. More convincing examples are found in for example the Internet (Faloustas et al. (1999), Barabasi & Alberts (1999) and Broder et al. 2000). The close to scale free features of many networks have initiated some proposals for mechanisms, which we review because emphasize certain potential evolutionary and topological features of networks.

- **Preferential attachment/cumulative advantage** One mechanism to obtain scale free networks is through growth models, where nodes are subsequently added to the network, with links attached preferentially to proteins that are highly connected (Price (1976), Barabasi & Alberts (1999)). It is a growth model based on minimal information in the sense that each new link is attached to the end of a randomly selected old link. Thus one connect new nodes with a probability of connecting proportional to the degree of the older nodes. Highly connected nodes therefore grow faster. It pays to be "popular". As a result the degree distribution becomes

  $$\frac{dn}{dk} \Rightarrow n(k) \propto \frac{1}{k^\gamma} \tag{7}$$

  with an exponent γ that can be tuned to values between 2 and 3, by adding links also without adding new nodes.

  The preferential growth model was originally proposed in an entirely different contexts, relating to modeling of human behavior quantified by the Zipf law (1949): That "law" states the empirical observation that incomes, or assets, or number of times a particular words are used, all tends to be distributed with power laws of type $1/s^2$. H. Simon (1955) suggested that this reflected the human tendency to preferentially give to what already has. For networks, a feature of this history dependent model is that the most connected nodes also are the oldest.

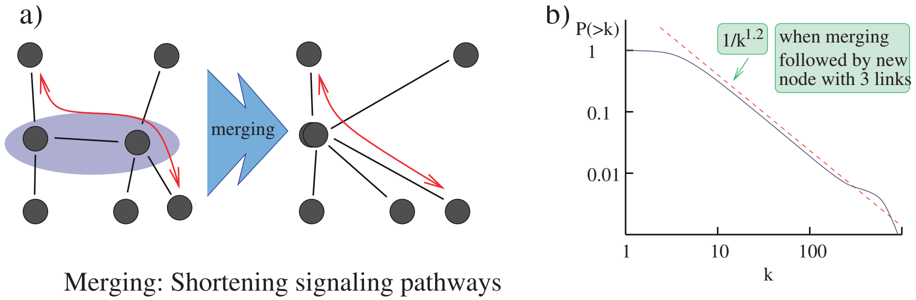
- **Threshold networks:** Another model for generating networks with power law distributed connectivity is the threshold networks considered by Calderelli et al.(2002). In this model each protein i is assigned an overall binding strength G_i selected from an exponential distribution, $P(G) \propto exp(-G)$. Then one assign a link to all proteins pairs i, j where $G_i + G_j$ is larger than a fixed detection threshold Θ. Thereby a network with scaling in some limited range are generated.

  Analytically the scaling comes about because a given protein is assigned a G with probability $exp(-G)$, and thereby a number of binding partners equal the number of proteins with $G' > \Theta - G$. This number is proportional to $exp(+G)$. Therefore there is probability $P(> G) = exp(-G)$ for having $N \propto exp(+G)$ partners:

  $$P(> N) \;=\; 1/N \quad \Rightarrow \quad \frac{dP}{dN} \propto 1/N^2 \tag{8}$$

  Thus threshold networks also generate scale free networks (when threshold Θ is rather large). In contrast to the preferential attachment model, the threshold model is history independent. But in the above formulation it is very non-specific: good binders bind to all reasonably strong binders. Presumably real networks have a specific reason to have a broad connectivity distribution. Life would would definitely not favor a prion like mechanism where some proteins aggregate many other proteins into one giant connected clump.

- **Merging and creation (/duplication):** A third scenario for generating scale free network is the merging and creation model introduced by Kim et al. (2003). In this one generate a scale free network by merging nodes (proteins) and generate new proteins. In detail, one time step of this algorithm consist of selecting a random node, and one of its neighbors. These are merged into one node, see Fig. 9a). Subsequently one add a new node to the network and link it to a few randomly selected nodes. As a result one may generate a nearly scale-free network with degree exponent $\gamma \sim 2.2$, see Fig. 9b). The justification for this algorithm for protein networks would be merging in order to shorten pathways, and creation in order to generate new functions. In contrast to the preferential attachment model, the merging/creation scenario does not demand persistent growth. Instead it suggests an ongoing dynamics of an evolving network which at any time has a very broad degree distribution.



Merging: Shortening signaling pathways

Figure 9. Merging and creation model of Kim et al. (2003). In addition to the shown merging step, a steady state network demands addition of a node for each merging. After a transient this evolutionary algorithm generate networks with scale free degree distribution, as illustrated in right panel.

> The merging and creation model has its correspondence in physics, where it original was suggested in the form of aggregation and injection model for dust (Fields & Saslow (1965)). The mechanism is probably important in the creation of larger aggregates in the interstellar vacuum. For protein interaction networks an on going merging process may seems hopeless. Rather a potential application may be in economy, where merging of companies is a widespread phenomena.

As a summary, we stress that any of the above models only presents some possible evolutionary elements in obtaining a broad degree distribution. As we will subsequently see, then maybe the relative positioning of highly connected proteins relative to each other may provide us with a more functional view.

5. Basic analysis of network topologies

We now want to discuss how to identify non trivial topological features of networks. That is, we want to go beyond the single node property defined by the degree distribution, and thus deal with the networks as objects that are indeed connected to each other. The hope is that this in the end may help us to understand function-topology relationship of various types of network. The key idea in this analysis is to compare the network at hand with a properly randomized version of it. As we want to go beyond degree distributions we want to compare with random networks with exactly the same degree distribution as the real network we are analyzing. One way to generate such random networks is shown in Fig 10. Technically the significance of any pattern is measured by its Z score:

$$Z(pattern) = \frac{N(pattern) - \langle N_{random}(pattern)\rangle}{\sigma_{random}(pattern)} \tag{9}$$

where $N_{random}(pattern)$ is the number of times the pattern occurs in the randomized network.

$$\sigma^2_{random}(pattern) = \langle N_{random}(pattern)^2\rangle - \langle N_{random}(pattern)\rangle^2 \tag{10}$$

is the variance of this number among the random networks. Considering patterns of links between proteins with various degrees, Maslov & Sneppen (2002) reported a significant suppression of links between hubs, both for regulatory networks and for the protein-protein interaction networks in yeast, see Fig. 11. Similarly, by defining higher order occurrences of various local patters of control, Shen-Orr et al. found some very frequent motifs in gene regulation networks. These are illustrated in Fig. 12.

We stress that one should be careful when judging higher order correlation patterns, because evaluation of these patterns are very sensitive to the null model against which they are judged. At least one should maintain the in and out degree of all nodes. But even when maintaining the degree distributions, for example loops will easily appear to be hugely over represented when comparing to a randomly reshuffled network. This is because a simple randomization does not take into account that proteins with similar functions tend to interact with each other (see question 1). This locality in itself give more loops. In fact one often use the number of loops as a measure of locality, quantified in terms of the so called cliquishness (Watts & Strogatz (1998)).

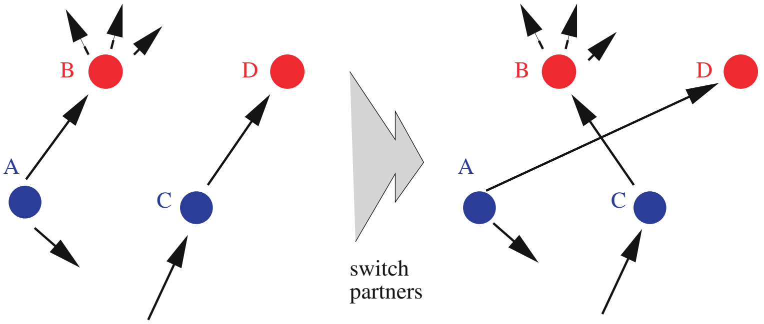


Figure 10. One step of the local rewiring algorithm (see Maslov et al. (2002)). A pair of directed edges A→B and C→D. The two edges then switches connections in such a way that A becomes linked to D, while C to B, provided that none of these edges already exist in the network, in which case the move is aborted and a new pair of edges is selected. An independent random network is obtained when this is repeated a large number of time, exceeding the total number of edges in the system. This algorithm conserves both the in- and out- connectivity of each individual node.

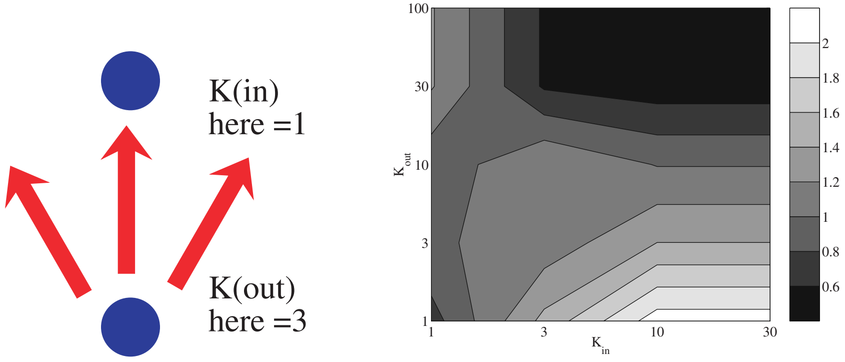
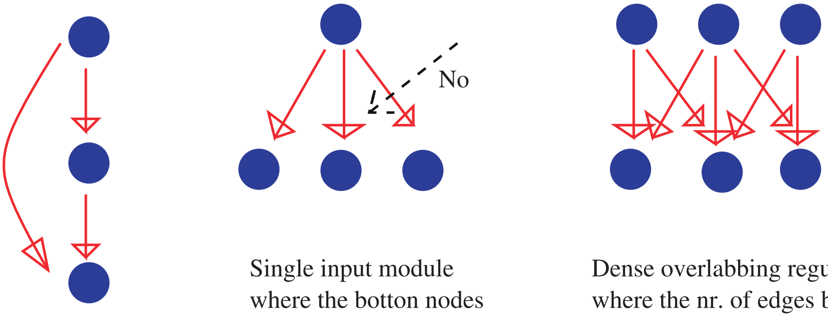


Figure 11. Correlation profile (Maslov et al. 2000) showing correlation between connected proteins in the regulatory network of yeast, quantified in terms of Z scores. The single output modules is reflected in the abundance of high K_{out} controlling single K_{in} proteins. The dense overlapping regulons correspond to the abundance of connections between $K_{out} \sim 10$ with $K_{in} \sim 3$ proteins. Finally the suppression of connections between highly connected proteins show that these tend to be on the periphery of the network.



Feed forward

Single input module
where the botton nodes
only are controled by one.

Dense overlabbing regulons
where the nr. of edges between
the two layers are large

Figure 12. Genetic control motifs which are found to be over represented in regulatory networks of both E.coli and Yeast (Shenn-Orr et al. 2002). Shenn-Orr et al. suggest that the feed-forward loop may act as a low pass filter, see question 1.

The tendency of highly connected proteins to be at the periphery of regulatory networks may teach us something about the origin of broad connectivity distributions: Maybe the hub proteins that gives orders to the many tend to give the same order to everyone below them.

One may speculate that the broad degree distribution of molecular networks is not an artifact of some particularly evolutionary dynamics (gene duplication, merging etc.), but rather reflect the broad distribution of number of proteins needed to do the different tasks required in a liven cell (Maslov et al. (2004)). Thus the broad connectivity distributions in signaling and regulative molecular networks may reflects the widely different needs associated to widely different functions that a living cell needs to cope with.

6. A Hierarchy measure of Networks

The topology of a network can be further characterized by its topological hierarchy, defined as the extent to which the network is hierarchical, when one assume that rank correlates with degree.

In fact in a number of systems nodes with higher degrees are on average *more important* than their lower degree counterparts. For example, for the Internet the number of hardwired connections a given Autonomous System serves as a proxy of its importance with the most connected nodes being global Internet Service Providers. For WWW the in-degree of a web page can serve as a measure of its popularity and hence importance; highly connected hub-airports of airline networks typically located in large cities, etc. In what follows we use the degree of a node as a proxy for its rank in the hierarchy based on the relative importance of nodes.

We quantify the hierarchical topology of a network is using the concept of a *hierarchical path*: a path between two nodes in a network is called hierarchical if it consists of one or two parts: either just an "up path" - where one is allowed to step from node i to node j only if their degrees k_i, k_j satisfy $k_i \leq k_j$, or alternatively an "up path" followed by a "down path" - where only steps to nodes of lower or equal degree are allowed. Either up or down path is allowed to have zero length.

Thus,

- If path first from low to high degrees,
- and then from high to low
- The two nodes are connected hierarchically

Example: Consider the internet, moving from local provider, up to district provider, state provider .... and the down again until another local provider is reached.

This definition of a hierarchical path follows the above mentioned trajectory of a request which is first forwarded up and then descends down the levels of a hierarchy quantified by k_i. It is also similar to the definition proposed in by L. Gao, (*Proc. IEEE INFOCOM*, November (2000)). The length of the shortest hierarchical path between a given pair of nodes can be either: 1) equal to the length of the overall shortest path; 2) longer than it; 3) not exist at all if these two nodes cannot be connected by any hierarchical path. The fraction of pairs in the first of these three categories is denoted as $\mathcal{F}$. Thus the hierarchical fraction $\mathcal{F}$ is the fraction of shortest paths in the network that are hierarchical.

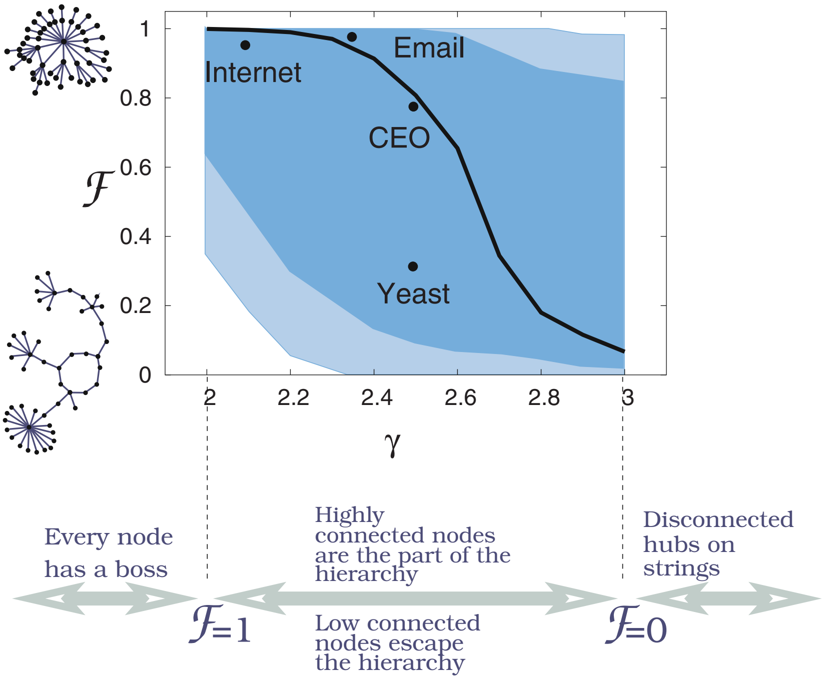


Figure 13. Different ways to position low and high degree nodes: Left part of figure illustrate Maximally hierarchical (top), random (middle) and anti-hierarchical (bottom) networks of size $N = 50$ nodes and node degree distribution $f(k) \propto 1/k^{2.5}$. The Main figure show how $\mathcal{F}$ depend on degree distribution in random scale free networks with distribution $P(k) \propto 1/k^{\gamma}$. One see that as degree distribution narrows, the hubs tend to separate. For $\gamma > 3$ the network effectively behaves as an Erdos-Reynei network.

In the figure 13 we illustrate how $\mathcal{F}$ behaves as function of γ for random networks where all nodes are connected by at least one path. We also compare with a few real world networks.

The overall behavior that $\mathcal{F}$ decrease with γ for random network can be understood from the following simple local consideration. Let us first calculate the probability that a given edge is attached to a node with degree larger than k

$$\begin{aligned} P_{edge}(\geq k) \propto & \int_k^K k' f(k') dk' \\ \propto & \begin{cases} 1 - \left(\frac{k}{K}\right)^{2-\gamma}, & \text{for } \gamma < 2 \\ k^{2-\gamma}, & \text{for } \gamma > 2 \end{cases} \end{aligned} \tag{11}$$

Here for $\gamma < 2$ one can only have a scale-free distributions below an upper cutoff K. For $\gamma > 2$ we can ignore the upper cutoff N. (We assume thermodynamic limit and do not restrict ourself to connected network). Thus in the absence of degree-degree correlations the probability that a node of degree k has at least one neighbor of degree higher than itself is given by

$$\begin{aligned} P(k_{neighbor} \geq k) &\propto (1 - P_{edge}(\geq k))^k \\ &\propto \begin{cases} 1 - \left(\frac{k}{K}\right)^{(2-\gamma)k}, & \text{for } \gamma < 2 \\ 1 - (1 - k^{2-\gamma})^k, & \text{for } \gamma > 2 \end{cases} \end{aligned} \tag{12}$$

For $\gamma \leq 2$ both low and high degree nodes always have a higher connected neighbor. and for $\gamma > 3$ the high k nodes nearly never have a boss. For $2 < \gamma < 3$, low connected nodes often have no higher connected neighbors, but as $P(k_{neighbor} > k) \rightarrow 1$ for increasing k there is a hierarchical core of highly connected nodes. In popular terms, at these intermediate values of γ many low degree nodes escape the hierarchy, while medium and highly connected nodes have bosses. Above $\gamma = 3$, $P(k_{neighbor} > k)$ decreases to zero with degree. Thus for these high values of γ a network becomes modular with each of the modules centered around a local hub.

7. A communication perspective on network topology

A key feature of molecular as well as most other networks is that they define the channels along which information flows in a system. Thus, in a typical complex system one may say that the underlying network constrain the information horizon that each node in the network experience. This view of networks can be formalized in terms of information measures that quantify how easy it would be for a node to send a signal to other specific nodes in the rest of the network (Sneppen et al. (2004)). To do this one count the number of bits of information required to *transmit* a message to a specific remote part of the network, or reversely, to *predict* from where a message is received (see Fig. 14).

In practice, imagine that you at node i want to send a message to node b in a given network (left panel in Fig. 14). Assume that the message follow the shortest path. That is, as we are only interested in specific signals we limit ourselves to consider only this direct communication. If the signal deviate from the shortest path, it is assumed to be lost. If there are several degenerate shortest paths, the message can be sent along any of them. For each shortest path we calculate the probability to follow this path, see Fig. 14. Assume that without possessing information one would chose any new link at each node along the path with equal probability. Then:

$$P\{p(i,b)\} = \frac{1}{k_i} \prod_{j \in p(i,b)} \frac{1}{k_j - 1}, \tag{13}$$

where j count all nodes on the path from a node i to the last node before the target node b is reached. The factor $k_j - 1$ instead of k_j takes into account the

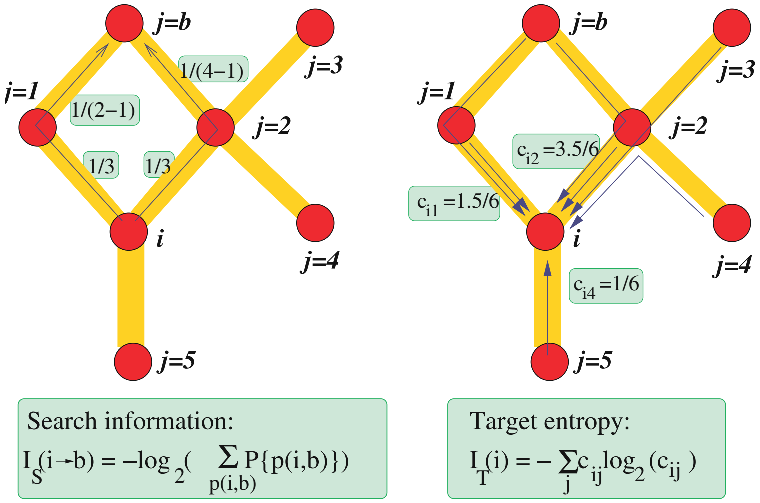


Figure 14. Information measures on network topology: Left panel: *Search information* $S(i \to b)$ measures your ability to locate node b from node i. $S(i \to b)$ is the number of yes/no questions needed to locate any of the shortest paths between node i and node b. For each such path $P\{p(i,b)\} = \frac{1}{k_i} \prod_j \frac{1}{k_j - 1}$, with j counting nodes on the path $p(i,b)$ until the last node before b. Right panel: *Target entropy* T_i measures predictability of traffic to you located at node i. c_{ij} is the fraction of the messages targeted to i that passed through neighbor node j.

information we gain by following the path, and therefore reduce the number of exit links by one. In Fig 14 we show the subsequent factors in going along any of the two shortest path from node i to node b. The total information needed to identify one of all the degenerate paths between i and b defines the "search information"

$$I_S(i \to b) = -log_2 \left(\sum_{p(i,b)} P\{p(i,b)\} \right), \tag{14}$$

where the sum runs over all degenerate paths that connect i with b. A large $I_S(i \to b)$ means that one needs many yes/no questions to locate b. The existence of many degenerate paths will be reflected in a small I_S and consequently in easy goal finding.

The value of $I_S(i \to b)$ teaches us how easy it is to transmit a specific message from node i to node b. To characterize a node, or a protein in a network, one may ask how easy is it in average to send a specific message from one node to another in the net:

$$\mathcal{A}(i) = \sum_b I_S(i \to b) \tag{15}$$

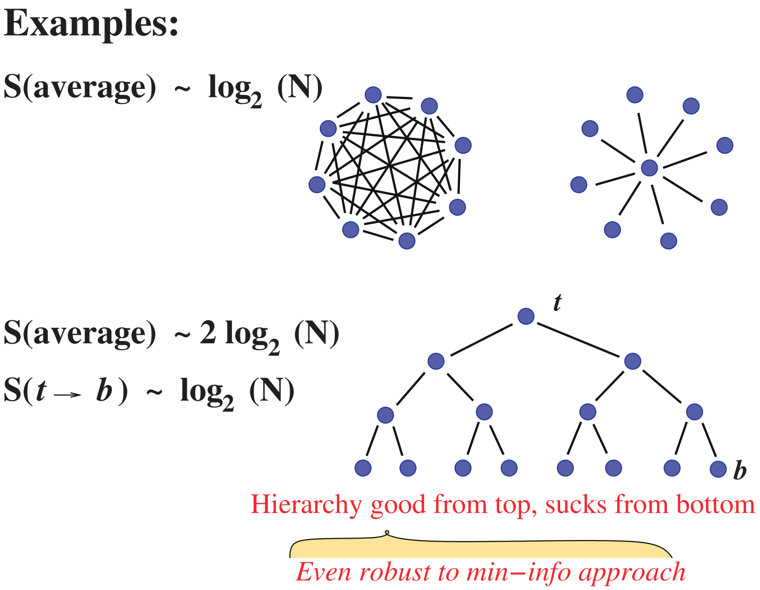


Figure 15. Examples of goal finding in different simple networks. Here S is short for the I_S used in text. The twoupper networks is networks where it is particularly easy to locate a given target node. In contrast, the hierarchy in bottom is a network where it is only easy to locate a target node, if you start at the top.

The overall ability for specific communication

$$\mathcal{I}_S = \sum_i \mathcal{A}(i) = \sum_{i,b} I_S(i \rightarrow b) \tag{16}$$

In a figure we investigate a few networks, and illustrate what high and low I_S means. In terms of degree distribution, then

$$\mathcal{I}_S \approx \sum_{i,b} \beta_i log(k_i) \tag{17}$$

where β is the betweenness of a node, counting the number of times it is used in paths between other nodes. As betweenness of high degree nodes typically are large, we get

$$\mathcal{I}_S(scale - free) > \mathcal{I}_S(Erdos - Reynei) \tag{18}$$

We will now see than real networks further tend to place the hubs, such that I_S gets even larger. Thus we compare with the value $\mathcal{I}_S(random)$ obtained for a randomized network, maintaining degree distribution. In Fig. 16 we plot the Z score defined as

$$Z = \frac{\mathcal{I}_S - \langle \mathcal{I}_S(random) \rangle}{\sqrt{\langle \mathcal{I}_S(random)^2 \rangle - \langle \mathcal{I}_S(random) \rangle^2}} \tag{19}$$

for the protein-protein network for both yeast (*Sacromyces Cerevisia*) (Uetz et al. 2000 and Ito et al. 2001) and fly (*Drosophilia*) (Giot et al. 2003) as well as for the hardwired Internet and a human network of governance (CEO) defined by company executives in USA where two CEO's was connected by a link if they are members of the same board. One sees that $\mathcal{I}_S > \mathcal{I}_S(random)$ for most networks, except for the fly network. Thus most networks have a topology that tends to hide nodes.

In Fig. 16 we also show another quantity, the ability to predict from which of your neighbors the next message to you will arrive from. This quantity measures predictability, or alternatively the order/disorder of the traffic around a given node i. The predictability based on the orders that are targeted *to* a given node i is

$$I_T(i) \;=\; -\sum_{j=1}^{k_i} c_{ij} log_2(c_{ij}), \tag{20}$$

where $j = 1, 2..., k_i$ denotes the links from node i to its immediate neighbors j and c_{ij} is the fraction of the messages targeted to i that passed through node j.

Notice that I_T is a measure of order in the network. In analogy with the global search information I_S one may also define overall predictability of a network

$$\mathcal{I}_T = \sum_i I_T(i) \tag{21}$$

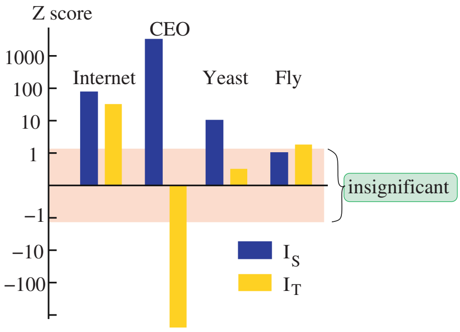


Figure 16. Measure of communication ability of various networks (Rosvall et al. (2004)). A high Z-score implies relatively high entropy. In all cases we show $Z = (\mathcal{I} - \mathcal{I}_r)/\sigma_r$ for $\mathcal{I} = \mathcal{I}_S$ and $\mathcal{I}_T$, by comparing with $\mathcal{I}_r$ for randomized networks with preserved degree distribution. σ_r is the standard deviation of the corresponding $\mathcal{I}_r$, sampled over 100 realizations. Results within the shaded area of two standard deviations are insignificant. All networks have a relatively high search information $\mathcal{I}_S$. The network of governance CEO show a distinct communication structure characterized by local predictability, low $\mathcal{I}_T$, and global inefficiency, high $\mathcal{I}_S$.

This quantity is large for Erdos-Reynei networks, and smaller for scale free networks, thus presenting an intuitively right measure for order in network topologies. For a given network $\mathcal{I}_T$ may subsequently be compared with its random counterpart (keeping degree distribution). In general as the organization of a network gets more disorganized $\mathcal{I}_T$ increases and the number of hubs with disordered traffic increases. Also, as one consider networks with increasing value of $\mathcal{I}_T$, nodes of low degree tend to be positioned between the hubs.

In summa, networks are coupled to specific communication and their topology should reflect this. The optimal topology for information transfer relies on a system-specific balance between effective communication (*search*) and not having the individual parts being unnecessarily disturbed (*hide*).

8. Communication and Structure in Social Networks

In a society the information horizon is set by each individual's social contacts, which in turn is a part of the global network of human communication. One simple goal for individuals is to be central. Thus we here investigate a model for a society where players try to be as close as possible to everybody else by moving their social connections. Local communication gives rise to global organization. Communication and not correctness appears as a success-strategy for individuals.

In other words we explore the local dynamic origin of global network organization by modeling response to information transfer in a simplified social system. The scenario is a set of players, that each tries to be as close as possible to everybody else. The players adjust their social connections to achieve this goal, guided by a limited knowledge about the individual players' positions in the network. The finite information is in turn obtained by local communication. When local communication is weak, the system disorganizes into a highly dynamic and chaotic network where no single player is dominating the system, see Fig. 17. In network language, the degree distribution

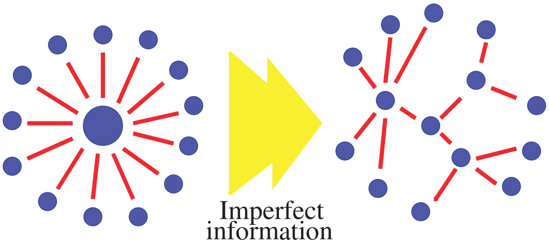


Figure 17. In a perfect world, a single vertex that can differentiate all exit edges from each other might distribute all tasks and information efficiently. In real world networks, no perfect "distributor" exists: even when every vertex "tries" to minimize its distances to all other vertices, typical vertices tend to connect through more than one intermediate. Imperfect information destabilize the central hub.

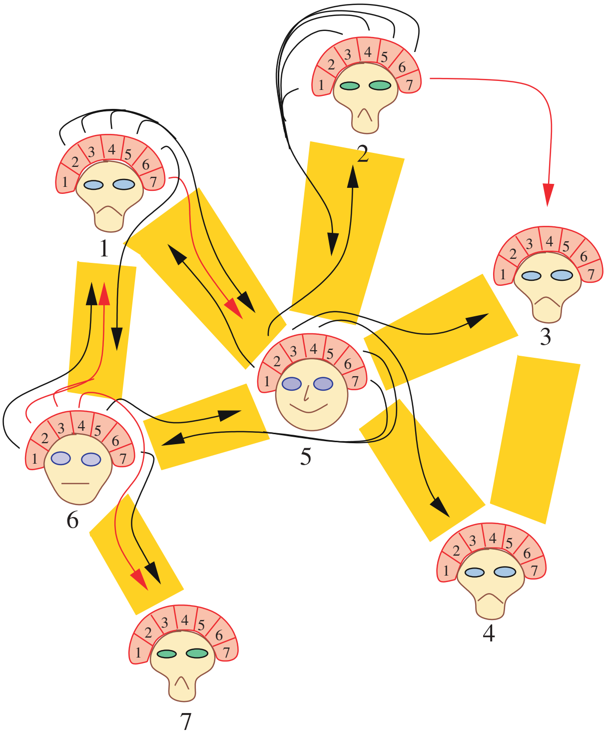


Figure 18. Illustration of social model, where agents optimize their position based on information obtained through communication with other agents. Versions of the model can be studied in java applets on cmol.nbi.dk.

is narrow, or in technical words exponential. On the other hand, when local communication is strong, the system organizes into a coherent structure dominated by a central hub that remains indefinitely frozen. In between, there is a critical transition in the dynamics where no hubs take over for ever, and where at the same time the network has players with all types of connectivities. The network is scale-free and furthermore hierarchical, in a way that resembles the Internet and often social and biological networks.

Every agent has a memory that corresponds to a rough picture of the network, see Fig. 18. The memory consists of an estimated shortest path length to any other agent in the network and the direction of the path in the form of the nearest neighbor on the corresponding path. By successive rewiring attempts the agents try to optimize their positions, that is, minimizing the distances to other agents. After a successful rewiring the agents in the local neighborhood of the rewiring (3 agents are involved) are allowed to “chat” with each other to update their information of the network. The conversation corresponds to a comparison of a fraction S of the two nearest neighbors’ memories. If a neighbor provides shorter paths to some agents, the agents adapt the new paths. The correctness of the information the agents have about the system can accordingly be adjusted by the parameter S (0 ¡ S ¡ 1). A small S results in bad information and the rewiring is close to random and no agent can survive as a hub with many links for a long time. A high S results in good information, the agents' memories give a good picture of the network, and an agent can survive as a hub for a long time since this topology minimizes the path lengths for agents in the network.

Given the fixed number of edges, the hub structure is the core of any network where distances is truly minimized. However, in practice each vertex may have only limited information about the location of other vertices. When changing their neighbors by moving links from one vertex to another, they may make mistakes, e.g. because of their limited local information. This will destabilize the centralized hub, and may lead to a distributed network as shown in right panel of Fig. 14.

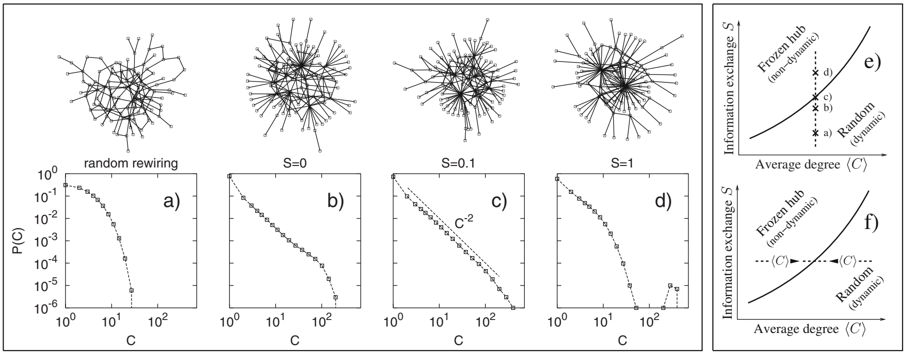


Figure 19. Vertex degree distribution of the evolved network with 4 levels of information exchange: no information between i and k in (iii,iv), full information in (iii) with respectively and $S = 0$, $S = 0.1$ and $S = 1.0$ in (iv). In all lower cases we sample dynamics of an $N = 1000$ vertex system with $E = 1500$ edges ($\langle C \rangle = 3$). The plots show average of many samples. The upper graphs show the corresponding networks for size $N = 100$.

In detail, the agent based model consists of different agents have different and adjustable memories. Every agent, named by a number $i = 1, 2, 3 \ldots n$, is a vertex in a connected network that consists of N vertices and E edges. Agent i has a memory

$$M_i = \left\{ \begin{array}{l} D_i(l) \\ P_i(l) \end{array} \right. , \; l = 1, 2, \ldots, i-1, i+1, \ldots N,$$

with $N - 1$ *distances* D and *pointers* P to the other agents in the network. The distance $D_i(l)$ is agent i's estimated shortest path length to l. The pointer $P_i(l)$ is agent i's nearest neighbor on the estimated shortest path to l. Thus M_i may be seen as a simplified version of the gateway protocol used by the autonomous systems to direct transmission of E-mails across the hardwired internet. Here, however, the memory will be used to rewire edges in the network.

Initially the network is a hub of the $N-1$ agents connected to a center agent by $N-1$ edges (as in Fig. 1, left) plus $E-N+1$ randomly placed edges on the periphery of the "hub". The rewiring consists of three steps:

> (i) An agent i and one of its neighbors j is chosen randomly.
>
> (ii) An agent $l \neq i,\ j$ is chosen randomly and if $D_i(l) > D_j(l)$ then the edge between i and j is rewired to an edge between i and $k = P_j(l)$. If l did not satisfy the above criteria a new l is randomly chosen. If no such l exists the rewiring is aborted.
>
> (iii) The information i has lost by disconnecting j is replaced by information from k. Further, there is full exchange of information between i and k: If agent k list a shorter path to some other agents, then i adopts this path with a pointer to k. Similarly for k, if agent i list a shorter path then k adopts this path through i. The information j has lost by disconnecting i is replaced by letting agent j change all its previous pointers toward i to pointers toward k and add 1 to the corresponding distances.

The model defines an update of both the network (ii) and the information (iii) that agents in the network have about each other's locations. The step (ii) represents local optimization where agent i rewire from j to k with a probability given by the fraction of the network which is estimated to becomes closer. We stress that only a small part of the system is informed about a changed geometry, and that decisions on moves may be based on outdated information. When repeated many times the model leads to a break down of the central hub into a steady state ensemble of networks with a broad distribution of vertex degrees.

Figure 19 show that the degree distribution for vertices in the network is very broad, in fact close to the Zipf law $1/C^2$ reported for some real world networks as well as for the size distributions of industrial companies. However there is correction to scaling at intermediate and large vertex degrees. This limitation of the model can be removed by increasing the information between agents during the rewiring. In particular we introduced a parameter S for information exchange between agent j and agent k in the above algorithm by adding a fourth step:

(iv) j consider a fraction S of the information it has stored with a pointer toward k. For this fraction it is checked whether k list a shorter path than j. For each path where this is the case, the memory of k is used to update the memory of j.

The number of links in the network play a similar role as the information exchange rate S. With many links in the system the agents suffer from their limited information horizon and get messed up by all the links. With added links an increased S is accordingly necessary to obtain a similar topology as for a network with less links. On a transition line between the chaotic, highly dynamic and "confused" state (typically low S and many links) and the ordered and one-hub dominated frozen state (typically high S and few links) the degree distribution is broad, in fact of scale-free form.

Scale free networks here appear at critical conditions between a frozen (hub dominated) and a chaotic (Erdos-Reynei) stage.

9. Networks in molecular biology

Cells are controlled by the action of molecules upon molecules. Receptor proteins in the outer cell membrane, sense the environment and may subsequently induce changes in the states of specific proteins inside the cell. These proteins then again interact and convey the signal further to other proteins, and so forth, until some appropriate action is taken. The states of a protein may for example be methylation status, phosphorylation or allosteric conformation as well as sub-cellular localization. The final action may be transcription regulation, thereby making more of some kinds of proteins, it may be chemical, or it may be dynamical.

Molecular networks are typically directed, consisting of proteins that regulate activity or production of other proteins. In Fig. 20 we illustrate two basic regulatory mechanisms in the genetic regulation of protein production, respective a positive (activation) and a negative (repression) one.

As for social or computer networks, the molecular networks are the backbone for information across complex systems. The interactions in molecular networks are obviously different from interactions in the other networks, i.s. they are constrained to the basic rule of binding and unbinding reactions in chemistry. When understanding molecular networks it is thus important to quantify the dynamics associated to possible reactions:

- A negative regulation may be mere inhibitive for the production, it may be blocking activity through binding or it may direct the degradation of the regulated protein.
- A positive regulation may activate production, it may change the protein property through f.ex. phosphorylation and thus instantly activate hidden

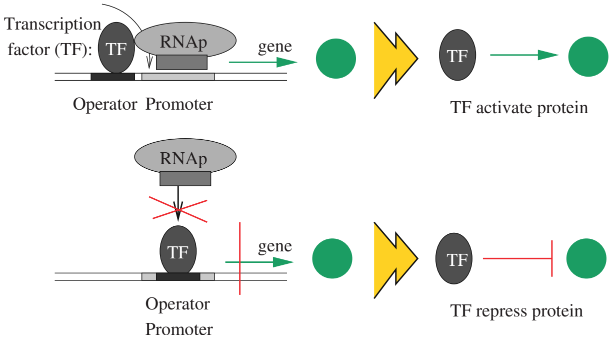


Figure 20. Positive and negative regulation by a transcription factor (TF).The positive regulation is through a binding between the TF and the RNAp, that increases the chance that RNAp binds to the promoter that is shown as medium dark region on the DNA strand. Negative regulation occur when the operator is placed such that the bound TF prevents the RNAp from binding to the promoter. On the rightmost panels we show how one typically draw the elementary regulation in as an arrow in a regulatory network.

> reservoir of passive proteins. Also on fast timescale, it may activate a hidden reservoir of passive mRNA and thereby lead to a sudden burst in protein production.

In Fig. 21 we show different strategies for regulating the concentration of a protein. In all cases the external regulation is taking place through a change in **c**, that typically will represent either a protein or a change in binding constant. The shown strategies may be combined, for example may a given protein be both positively regulated by another protein, and negatively auto-regulated by itself.

Comparing the regulations in Fig. 21 one sees that protein degradation (for negative), and translation control (for positive regulation) provides the most dramatic change in protein concentration. In accordance with this observation, it is typically these types of interactions on find in relation to stress responses, as f.es. the heat shock (model see Arnvig et al (2000) or the the unfolded protein response i yeast (see Cox & Walter (1996), Sidrauski et al. (1997), and for model see Bock & Sneppen 2004).

Simple regulation may be combined to provide intricate complete organism networks, as indeed illustrated by the yeast network in beginning of this section. A particularly simple organism is the 186 phage or the lambda phage, see Fig. 22. Such networks exhibit many of the features found in the larger network as for example the presence of hubs. In addition they provide a possibility to address the reason for the hubs, which in case of the phages

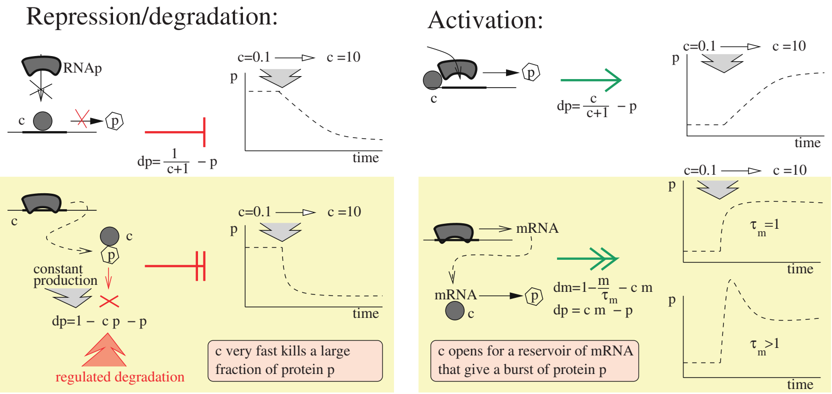


Figure 21. Some ways to regulate the concentration of a protein (**p**), through respective negative and positive control, and through respective transcription regulation and something faster (from Bock & Sneppen (2004)). The upper panels is the simplest regulation option, where *c* simply regulate transcription of mRNA for **p**. In this simplest case the time to obtain new steady state is set by degradation rate of **p**. Fast regulation and eventual overshooting is obtained when one use active degradation (lower left), respective translational control (lower right). **m** in lower right panel is inactive mRNA for protein **p**, and it has a regulated conversion time to an active but fast decaying active form.

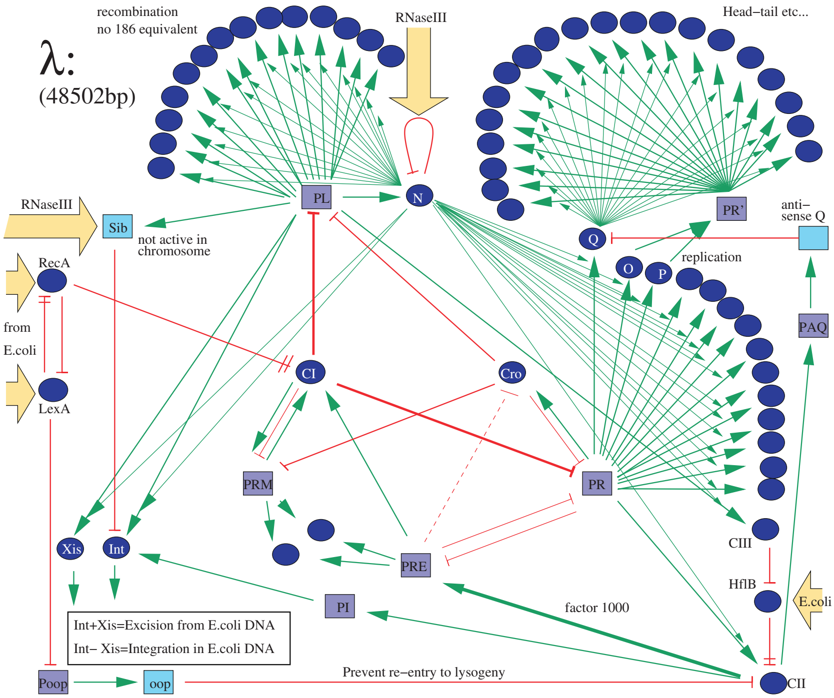


Figure 22. Regulatory network of λ phage. Squares represent promoters, circles represent proteins. RecA, LexA, Lon, HflB and RNAaseIII belongs to the host cell, which is *E.coli*. Thus whatever goes on in the network, depends on signaling from these input proteins.

clearly are associated to proteins which are associated to the same function (e.g. building the capsule for the phage).

Further information on structure-function-evolution relationships for networks can be found on our homepage on cmol.nbi.dk. Here we also have java applets that visualize details of molecular mechanisms in the lambda phage and elsewhere.

Acknowledgment

I thank for support from Danmarks Grundforskningsfond through the center "Models of Life".

References

K.B. Arnvig, S. Pedersen and K. Sneppen (2000). *"Thermodynamics of Heat Shock Response"*. Phys. Rev. Lett. **84** 3005 (cond-mat/9912402)

E. Aurell, K. Sneppen (2002). *Epigenetics as a first passage problem*, Phys. Rev. Letters. **88**, 048101-1.

E. Aurell, S. Brown and K. Sneppen (2002). *Stability puzzles in phage* λ, Physical Review E **65**, 051914.

A.-L. Barabasi, R. Albert (1999). *Emergence of scaling in random networks*, Science, **286**, 509.

J. Bock and K. Sneppen (2004). *Quantifying the Benefits of Translation Regulation in the Unfolded Protein Response.* To appear in Physical Biology.

S. Bornholdt & K. H.Ebel (2001). *World Wide Web scaling from Simon's 1955 model*, Phys. Rev. E. 035104(R).

S. Bornholdt and K.Sneppen (2000). "*Robustness as an evolutionary principle*", Proc Roy. Soc. London, B **267** 2281.

A. Broder *et al.* (2000). *Graph Structure in the Web*, Computer Networks **33**, 309-320.

K.T. Bæk, S. Svenningsen, H. Eisen, K. Sneppen & S. Brown (2003). *"Single-cell analysis of lambda immunity regulation*", J. Mol. Biol. **334** 363.

G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muoz (2002) *Scale-Free Networks from Varying Vertex Intrinsic Fitness.* Phys. Rev. Lett. **89** 258702.

J.S. Cox and P. Walter. (1996). *A Novel Mechanism for Regulating Activity of a Transcription Factor That Controls the Unfolded Protein Response. Cell.* 87:391-404.

Davidson et al. (2002). *A Genomic Regulatory Network for Development.* Science **295** (5560): 1669-2002.

E.H. Davidson, (2001). *Genomic Regulatory Systems. Development and Evolution.* Academic Press, San Diego.

S.F. Elena & R.E. Lenski (1999). *Test of synergetic interactions among deleterious mutations in bacteria Nature* **390**, 395.

P. Erdös & A. Rényi (1960). *On the evolution of random graphs.* Publ. Math. Inst. Hung. Acad. Sci. **5**, 1760.

J.E. Ferrell Jr. (2000). *What do Scaffold Proteins Really Do*, Science STKE, www.stke.sciencemag.org/cgi/content/full/sigtrans;2000/52/pe1

G.B. Field and W.C. Saslow (1965). Astrophys. J. **142** 568.

A.-C. Gavin, *et al.* (2002). *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature **415**, 141-147.

L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray, (1999). *From molecular to modular cell biology*, Nature **402** (6761 Suppl), C47.

A.M. Huerta, H. Salgado, D. Thieffry & J. Collado-Vides (1998). *R*egulonDB: *A database on transcriptional regulation in Escherichia coli* Nucleic Acid Res. **26**, 55. (see also *http:// www.smi.stanford.edu/ projects/ helix/ psb98/ thieffry.pdf*)

D. Hughes, M. Paczuski, R.O. Dendy, P. Helander and K.G. McClements. Physical Rev. Letters **90** 131101 (2003).
M. Paczuski and D. Hughes, cond-mat/0311304 (2003).

T. Ito, *et al.* (2001). *A comprehensive two-hybrid analysis to explore the yeast protein interactome*, Proc. Natl. Acad. Sci. USA **98**, 4569.

H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai & A.-L. Barabasi (2000). *The large scale organization of metabolic networks.* Nature, **407**, 651-654.

H. Jeong, S. Mason, A.-L. Barabasi, Z.N. Oltvai (2001). *Centrality and lethality of protein networks.* Nature **411**, 41 -42.

B.J. Kim, A. Trusina, P. Minhagen and K. Sneppen (2003). *Scale free networks from merging and creation.* Submitted to Physical Review Letters.

B.J. Kim, A. Trusina, P. Minnhagen and K. Sneppen (2004). *Self Organized Scale-Free Networks from Merging and Regeneration.* nlin.AO/0403006.

D.E. Koshland, Jr., A. Goldbeter, and J.B. Stock (1978). *Amplification and adaptation in regulatory and sensory systems.* Science. **217** 220

S. Maslov and K. Sneppen (2002). *Specificity and stability in topology of protein networks*, Science **296** 910.

S. Maslov and K. Sneppen (2002). "*Pattern Detection in Complex Networks: Correlation Profile of the Internet*" Submitted to Phys. Rev. Lett., cond-mat/0205379

S. Maslov and K. Sneppen (2004). *Computational Architecture of the Yeast regulatory network.* Preprint.

P. Minnhagen, M. Rosvall, K. Sneppen, A. Trusina (2004). *Self-organization of structures and networks from merging and small-scale fluctuations.* cond-mat/-0406752.

M. E. J. Newman, S. H. Strogatz, and D. J. Watts (2001). *Random graphs with arbitrary degree distributions and their applications*, Phys. Rev. E, **64**, 026118, 1.

E.van Nimwegen (2003). *Scaling Laws in the functional content of genomes.* Trends in Genetics bf 19 479

K.J. Peterson and E.H. Davidson, *Regulatory evolution and the origin of the bilaterians.* Proc. Natl. Acad. Sci. USA 97, 4430-4433, 2000

D.J. de S. Price (1965). *Networks of scientic papers.* Science **149**, 510.

D.J. de S. Price (1976). *A general theory of bibliometric and other cumulative advantage processes.* J. Amer. Soc. Inform. Sci. **27** 292.

M. Ptashne & A. Gann (1997). Nature **386** 569.

N. Rosenfeld, M. Elowitz & U. Alon (2002). *Negative Autoregulation Speeds the Response Times of Transcription Networks.* JMB, **323** 785-793.

M. Rosvall & K. Sneppen (2003). *Modeling Dynamics of Information Networks.* cond-mat/0308399.

M. Rosvall, A. Trusina, P. Minnhagen and K. Sneppen (2004). *Networks and Cities: An Information Perspective.* cond-mat/0407054.

S.S. Shen-Orr, R. Milo, S. Mangan and U. Alon (2002). *Network motifs in the transcriptional regulation of Escherichia coli* Nature Genetics, Published online: 22 April 2002, DOI:10. 1038/ng881

C. Sidrauski, J. Cox, & P. Walter. 1997. *tRNA Ligase Is Required for Regulated mRNA Splicing in the Unfolded Protein Response. Cell* **87** 405-413.

H. Simon (1955). Biometrika **42** (1955) 425.

K. Sneppen, M. Rosvall, A. Trusina and P. Minhagen (2004). *A simple model for self-organization of bipartite networks* Europhysics letters **67** 349.

K. Sneppen, M. Rosvall, A. Trusina, P. Minnhagen. *A simple model for self organization of bipartite networks.* Europhys. Lett. **67** (2004) 349

K. Sneppen, A. Trusina and M. Rosvall (2004). *Hide and seek on complex networks.* cond-mat/0407055

C.K.Stover et al. (2000), *Complete genome sequence of Pseudomonas Aeruginosa PA01, an opportunistic pathogen Nature* **406**, 959.

G. Tiana, M. H. Jensen, K. Sneppen (2002). *Time delay as a key to Apoptosis Induction in the p53 Network.* Eur. Phys. J. B **29** 135.

A. Trusina, S. Maslov, P. Minnhagen & K. Sneppen (2003). *Hierarchy and Anti-Hierarchy in Real and Scale Free networks.* cond-mat/0308339.

A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani (2001). *Modelling of protein interaction networks* arXiv:cond-mat/0108043.

P. Uetz, *et al.* (2000). *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisia.* Nature **403**, 623.

D. J. Watts and S. H. Strogatz (1998). *Collective dynamics of “small-world” networks.* Nature, **393** 440.

EFFECTS OF COMMUNITY STRUCTURE ON SEARCH AND RANKING IN COMPLEX NETWORKS

HUAFENG XIE
New Media Lab, The Graduate Center, CUNY New York, NY 10016, USA
and
Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, USA

KOON-KIU YAN
Department of Physics and Astronomy, Stony Brook University,
Stony Brook, NY 11794, USA
and
Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, USA

SERGEI MASLOV
Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, USA
(`maslov@bnl.gov`*)*

Abstract. The community structure in complex networks has been a popular topic in recent literature. It is present in all types of complex networks ranging from bio-molecular networks, where it reflects functional associations between proteins to information networks such as the The World Wide Web (WWW). The World Wide Web – a quintessential large complex network – presents formidable challenge for the efficient information retrieval and ranking. Google has reached its current position as the world's most popular search engine by efficient and ingenious utilization of topological properties of this WWW network for ranking of individual webpages. The topological structure of the WWW network is characterized by a strong community structure in which groups of webpages (e.g. those devoted to a common topic) are densely interconnected by hyperlinks. We study how such network architecture affects the average Google ranking of individual webpages in the community. We demonstrate that the Google rank of community webpages could either increase or decrease with the density of inter-community links depending on the exact balance between average in- and out-degrees in the community. The magnitude of this effect is described by a simple analytical formula

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 29–37.

and subsequently verified by numerical simulations of random scale-free networks with a desired level of the community structure. A new algorithm allowing for generation of such networks is proposed and studied. The number of inter-community links in such networks is controlled by a temperature-like parameter with the strongest community structure realized in "low-temperature" networks.

Networks have recently emerged as a unifying theme in complex systems research. It is in fact no coincidence that networks and complexity are so heavily intertwined. Any future definition of a complex system should reflect the fact that such systems consist of many mutually interacting components. These components are far from being identical as say electrons in systems studied by condensed matter physics. In a truly complex system each of them has a unique identity allowing one to separate it from the others. The very first question one may ask about such a system is which other components a given component interacts with? This information system-wide can be visualized as a graph, whose nodes correspond to individual components of the complex system in question and edges to their mutual interactions. Such a network can be thought of as a backbone of the complex system. Of course, system's dynamics depends not only on the topology of an underlying network but also on the exact form of interaction of components with each other, which can be very different in various complex systems. However, the underlying network may contain clues about the basic design principles and/or evolutionary history of the complex system in question.

The World Wide Web (WWW) – a very large ($\sim 10^{10}$ nodes) network consisting of webpages connected by hyperlinks – presents a challenge for the efficient information retrieval and ranking. Apart from the contents of webpages, the topology of the network itself can be a rich source of information about their relative importance and relevance to the search query. It is the effective utilization of this topological information [1] which advanced the Google search engine to its present position of the most popular tool on the WWW and a profitable company with a current market capitalization over 100 billion dollars . To rank the importance of webpages Google simulates the behavior of a large number of "random surfers" who just follow a randomly selected hyperlink on each page they visit. The number of hits a given page gets in the course of such simulated process determines its ranking. It is intuitively clear that the larger is the number of hyperlinks pointing to a given webpage (its in-degree in the network) the higher are the chances of a random surfer to click on one of them and, therefore, the higher would be the resulting Google rank of this webpage. However, the algorithm goes beyond just ranking nodes based on their in-degrees. Indeed, the traffic directed to a given webpage along a particular incoming hyperlink is proportional to the popularity of the webpage containing this link. Therefore, the Google rank of a node is given by the weighted in-degree where the weight of each neighboring webpage reflects its importance and is determined self-consistently.

The WWW is a very heterogeneous collection of webpages which can be grouped based on their textual contents, language in which they are written, the Internet Service Provider (ISP) where they are hosted, etc. Therefore, it should come as no surprise that the WWW has a strong community structure [2] in which similar pages are more likely to contain hyperlinks to each other than to the outside world. Formally a web community can be defined as a collection of webpages characterized by a higher than average density of links connecting them to each other. In this letter we are going to address the question: how the community structure affects the Google rank of webpages inside the community. One might naively expect that the community structure always boosts the Google rank of its webpages as it tends to "trap" the random surfer inside the community for a longer time. However, it turned out that it is not generally true. In fact the Google rank of community webpages could either increase or decrease with the density of inter-community links depending on the exact balance between average in- and out-degrees in the community. In the heart of the Google search engine lies the PageRank algorithm determining the global "importance" of every web page based on the link structure of the WWW network around it. While the details of the algorithm have undoubtedly changed since its introduction in 1997, the central "random surfer" idea first described in [1] remained essentially the same. To a physicist the algorithm behind the PageRank just simulates an auxiliary diffusion process taking place on the network in question.

Similar diffusion algorithms have been recently applied to study citation and metabolic networks [4] and the modularity of the Internet on the "hardware level" represented by an undirected network of interconnections between Autonomous Systems [5]. A large number of random walkers are initially randomly distributed on the network and are allowed to move along its directed links. As in principle some nodes in the network could have zero out-degree but non-zero in-degree and would thus "trap" random walkers, the authors of the algorithm introduced a finite probability α for a surfer to randomly select a page in the network and directly jump there without following any hyperlinks. This leaves the probability $1 - \alpha$ for a surfer to randomly select and follow one of the hyperlinks of the current webpage. According to [3] the original PageRank algorithm used $\alpha = 0.15$. The PageRank then simulates this diffusion process until it converges to a stationary distribution. The Google rank (PageRank) $G(i)$ of a node i is proportional to the number of random walkers at this node in such steady state. We chose to normalize it so that $\sum_i G(i) = N$, where N is the total number of nodes in the network, but in general the normalization factor does not matter as ranking relies on relative values of $G(i)$ for different webpages. When one enters a search keyword such as e.g. "statistical physics" on the Google website the search engine first localizes the subset of webpages containing this keyword and then simply presents them in the descending order based on their PageRank values. The main equation determining the PageRank values $G(i)$ for all webpages in the WWW is

$$G(i) = \alpha + \sum_{j\to i}(1-\alpha)\frac{G(j)}{K_{out}(j)} \cdot \tag{1}$$

Here $K_{out}(j)$ denotes the the number of hyperlinks (the out-degree) in the node j and the summation goes over all nodes j that have a hyperlink pointing to the node i. In the matrix formalism the PageRank values are given by the components of the principal eigenvector of an asymmetric positive matrix related to the adjacency matrix of the network. Such eigenvector could be easily found using a simple iterative algorithm [3]. The fast convergence of this algorithm is ensured by the fact that the adjacency matrix of the network is sparse. We first consider the effect of the community structure on Google ranking in the simplest and most physically transparent case of $\alpha = 0$. In order for the algorithm to properly converge in this case we need to assume that $K_{out}(i) > 0$ for all nodes in the network.

Consider a network in which N_c nodes form a community characterized by higher than average density of edges linking these nodes to each other. Let E_{cw} denote the total number of hyperlinks pointing from nodes in the *community* to the outside *world*, while E_{wc} - the total number of hyperlinks pointing in the opposite direction (See Fig. 1 for an illustration). Similarly E_{cc} and E_{ww} denote the total number of links connecting nodes within the community and, respectively, the outside world. The total number of hyperlinks pointing to nodes inside the community is given by $E_{cc} + E_{wc} = N_c\langle K_{in}\rangle_c$ where $\langle K_{in}\rangle_c$ is the average in-degree of community nodes. Similarly, $E_{cc} + E_{cw} = N_c\langle K_{out}\rangle_c$,

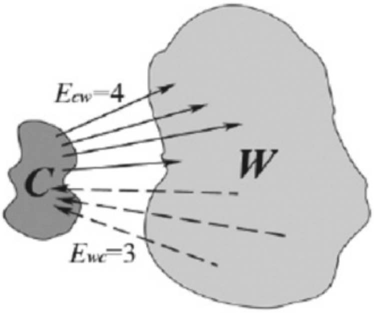


Figure 1. The illustration of hyperlink connections between the community C and the outside world W. E_{cw} and E_{wc} are numbers of links from the community to the outside world and from the outside world to the community, respectively.

where $\langle K_{out} \rangle_c$ is the average out-degree in the community, gives the total number of hyperlinks originating on community nodes. The Google rank is computed in the steady state of the diffusion process where the average number of random surfers currently visiting any given webpage does not change with time. This means that the total current of surfers J_{cw} leaving the community for the outside world must be precisely balanced by the current J_{wc} entering the community during the same time interval. Let $G_c = \langle G(i) \rangle_{i \in C}$ denote the average Google rank inside the community given by the average number of random surfers on its nodes. If edges pointing away from the community to the outside world start at an unbiased selection of nodes in the community the average current flowing along any of those edges would be given by $G_c / \langle K_{out} \rangle_c$ while the total current leaving the community $J_{cw} = E_{cw} G_c / \langle K_{out} \rangle_c$. Similar analysis gives $J_{wc} = E_{wc} G_w / \langle K_{out} \rangle_w$, where $\langle K_{out} \rangle_w$ is the average out-degrees of nodes in the world outside the community. Balancing these two currents one gets:

$$\frac{G_c}{G_w} = \frac{E_{wc}}{E_{cw}} \cdot \frac{\langle K_{out} \rangle_c}{\langle K_{out} \rangle_w} \ . \tag{2}$$

The Eq. 2 is based on a mean-field assumption that average values of the Google rank and the out-degree on those community nodes that actually send links to the outside world are equal to their overall average values inside the community [6]. It is tempting to assume that higher than average density of hyperlinks connecting nodes in the community is beneficial for the Google rank of its nodes as it "traps" random surfers to spend more time within the community. It turned out that this naive argument is not necessarily true. In fact one is equally likely to observe an opposite effect: an excess of intra-community links could lead to a lower than average Google rank of its nodes. To see it explicitly one should replace E_{wc} and E_{cw} in Eq. 2 with identical expressions $\langle K_{in} \rangle_c N_c - E_{cc}$ and $\langle K_{out} \rangle_c N_c - E_{cc}$ respectively:

$$\frac{G_c}{G_w} = \left(\frac{\langle K_{in} \rangle_c N_c - E_{cc}}{\langle K_{out} \rangle_c N_c - E_{cc}} \right) \cdot \frac{\langle K_{out} \rangle_c}{\langle K_{out} \rangle_w} \ . \tag{3}$$

From this equation it follows that enhancing the community structure (increasing E_{cc}) while keeping other parameters such as $\langle K_{in} \rangle_c$,$\langle K_{out} \rangle_c$ and$\langle K_{out} \rangle_w$ fixed can be both good and bad for the average Google rank of the community webpages. It depends on $\langle K_{in} \rangle_c / \langle K_{out} \rangle_c$ – the ratio between average in- and out-degrees of community nodes. If the ratio is less than 1 the increase in E_{cc} leads to a further decrease of G_c / G_w below one. If the community constitutes just a small fraction of the whole network one could safely assume that G_w remains approximately constant so that the average Google rank of the community, G_c, has to decrease. Similarly if the ratio is larger than 1, G_c grows with the number of inter-community links E_{cc} (see Fig. 2 for an illustration of both cases).

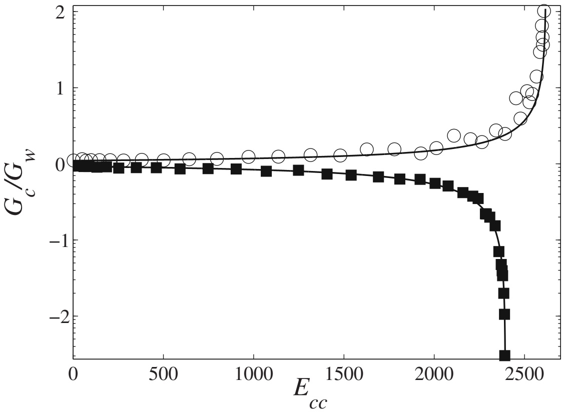


Figure 2. The ratio of average Google ranks in the community and the outside world G_c/G_w as a function of E_{cc} – the number of intra-community links – in two series of model networks with varying degree of community structure. Open circles correspond to a beneficial effect of the community structure on Google ranking in a scale-free network with $\langle K_{out}\rangle_c = 5.24 < \langle K_{in}\rangle_c = 5.9$. On the other hand, filled squares show a detrimental effect in another series of networks where $\langle K_{out}\rangle_c = 5.6 > \langle K_{in}\rangle_c = 4.8$. Solid lines are fits with the Eq. 3 with a given set of parameters for each of the networks. All networks with $10,000$ nodes have a community of 500 nodes were generated by the Metropolis rewiring algorithm described later on in the text.

The real-life Google algorithm uses a non-zero value of $\alpha \simeq 0.15$. In this case one needs to consider the contribution to currents J_{cw} and J_{wc} due to surfers' random jumps that do not follow the existing hyperlinks. The total number of random walkers residing on the nodes inside the community is G_cN_c and the probability of them to randomly jump to a node in the outside world is $N_w/(N_c + N_w)$. So the contribution to the outgoing current due to such jumps is given by $\alpha G_cN_cN_w/(N_c + N_w)$ which for $N_c \ll N_w$ can be simplified as αG_cN_c. The total outgoing current then can then be written as $J_{cw} = (1 - \alpha)G_cE_{cw}/\langle K_{out}\rangle_c + \alpha G_cN_c$. Similarly the incoming current J_{wc} is given by $(1 - \alpha)G_wE_{wc}/\langle K_{out}\rangle_w + \alpha G_wN_c$. The Eq. 2 remains valid for $\alpha > 0$ if one replaces E_{wc} and E_{cw} with "effective" numbers of edges E^*_{wc} and E^*_{cw} given by

$$\begin{aligned} E^*_{cw} &= E_{cw}(1 - \alpha) + N_c\langle K_{out}\rangle_c\alpha \ ; \\ E^*_{wc} &= E_{wc}(1 - \alpha) + N_c\langle K_{out}\rangle_w\alpha \ . \end{aligned} \tag{4}$$

These effective numbers take into account contributions to both currents due to random jumps.

For numerical studies of networks with a community structure we propose a particular version of the Metropolis network rewiring algorithm [8]. It allows one to generate an ensemble of random networks with user-defined in- and out-degree distributions as well as with any desired density of links

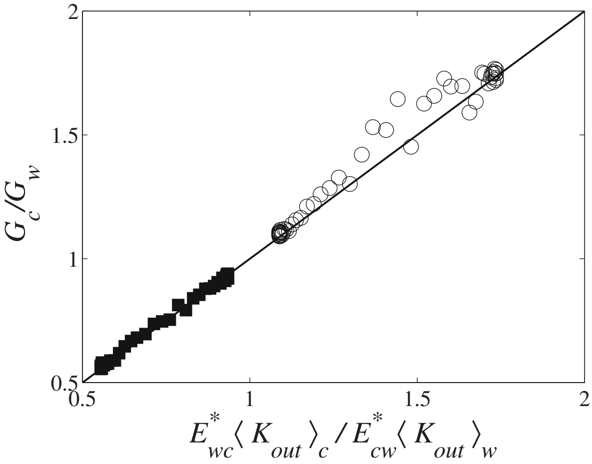


Figure 3. The ratio of average Google ranks in the community and the outside world G_c/G_w as a function of the ratio of effective numbers of links E^*_{wc}/E^*_{cw}. As predicted by the Eq. 2 these two ratios are basically equal to each other. Different symbols correspond to series of networks described in Fig. 2.

between pre-selected N_c "artificial community" nodes. It starts from a seed network with the preferred (scale-free in our case) distributions of in- and out-degrees [9] and proceeds by a sequence of edge-swapping steps changing a pair of randomly selected edges A→B and C→D into A→D and C→B correspondingly. The decision on whether to accept or reject an elementary step depends on changes in the energy function H and the inverse temperature β. For our purposes of generating networks with community structure we choose $H = -E_{cc}$, where E_{cc} is the number of edges connecting pre-selected community nodes to each other.

Figure 3 shows the results of a numerical test of Eqs. 2, 4on modeldirected networks with scale-free distributions of in- and out-degrees: $P(K_{in}) \sim K_{in}^{-2.1}$ and $P(K_{out}) \sim K_{out}^{-2.5}$ correspondingly. The exponents were selected to be identical to their values in the actual WWW network [2, 7] and a community structure was generated by the Metropolis algorithm described above.

The reciprocal temperature β used in the Metropolis algorithm indirectly determines the number of links within the community. A network without any community structure is realized at an infinite temperature ($\beta = 0$), while the algorithm run at zero temperature ($\beta = \infty$) produces a network with the largest possible number of links within the community. One could also run the algorithm with $\beta < 0$. Negative values of β generate networks with an anti-community structure in which the number of intra-community links is lower than that in a random network. The relation between E_{cc} and β for both positive and negative values of β is shown in Fig. 4.

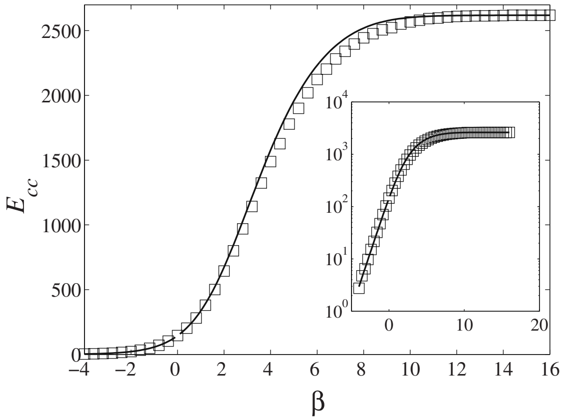


Figure 4. The number of intra-community links E_{cc} in networks generated by the rewiring algorithm as a function of the inverse temperature β. Negative values of β correspond to networks with anti-community structure and are generated by changing the sign in front of the Hamiltonian H. The solid line is the fit with the analytical expression obtained by solving the Eq. 5 for E_{cc}. The inset shows the same plot with a logarithmic scale of the Y-axis.

To analytically derive such relation we consider the detailed balance in the steady state of the Metropolis rewiring algorithm, in which the probabilities of an increase and a decrease in E_{cc} are equal to each other. It results in the equation

$$E_{cw}E_{wc} = E_{cc}E_{ww}e^{-\beta} \ . \tag{5}$$

Additional constraints (i) $E_{cc} + E_{wc} = \langle K_{in}\rangle_c N_c$ (the sum of in-degrees of all nodes within the community), (ii) $E_{cc} + E_{cw} = \langle K_{out}\rangle_c N_c$ (the sum of out-degrees of all nodes within the community) and (iii) $E_{cc} + E_{cw} + E_{wc} = E$ (the total number of edges in the network) plugged into the Eq. (5) result in a quadratic equation for E_{cc} as a function of $\langle K_{in}\rangle_c$, $\langle K_{out}\rangle_c$, E, and β – the parameters strictly conserved in our rewiring algorithm. The Fig. 4 compares the analytical expression for $E_{cc}(\beta)$ obtained by solving the Eq. 5 with numerical simulations for different values of β. Clearly, E_{cc} increases with β in general accord with the Eq. 5. When β is sufficiently large, E_{cc} exponentially approaches a limiting value equal to $\min(\langle K_{in}\rangle_c, \langle K_{out}\rangle_c)N_c$ – the minimal number of links within a community given the set of in- and out-degrees of its nodes. The deviations between the analytical formula and numerical results visible for large values of β could be attributed to the “no multiple edges” restriction in networks generated by our rewiring algorithm. As the density of inter-community links increases more and more of rewiring steps leading to an increase of E_{cc} are aborted as the new link they are attempting to create within a community already exists. This situation is more appropriately described by the following equation: $E_{cw}E_{wc}(1 - E_{cc}/E)(1 - E_{ww}/E) = E_{cc}E_{ww}(1 - E_{cw}/E)(1 - E_{wc}/E)e^{-\beta}$, reminiscent of the detailed balance equation in two-fermion scattering (see also [11]).

In summary, we investigated the effect of web-communities have on Google rank of their constituent webpages. Contrary to naive arguments, we found that the density of links within the community only indirectly affects the average Google ranking of its constituent webpages. The sign of this effect can be negative as well as positive depending on the exact balance of links leading from the community to the outside world and, conversely, from the outside world to the community. An important practical question is: how community members could boost the average Google rank of their webpages? Within our framework there is one solution easily available to the community members: they could collectively reduce the total number of links leading from their pages to the outside world (here we assume that they have no direct control over the number of links going in the opposite direction). It is interesting to note that this prescription does not work on the level of individual webpages: the Google ranking of a webpage is generally independent on the number of hyperlinks it contains. We are currently investigating whether this simple collective strategy was empirically discovered in the real WWW and what other strategies WWW communities actually use to boost their ranking [12].

Work at Brookhaven National Laboratory was carried out under Contract No. DE-AC02-98CH10886, Division of Material Science, U.S. Department of Energy.

References

1. S. Brin and L. Page, Computer Networks and ISDN Systems, **30**, 107 (1998).
2. R. Kumar, P. Raghavan, S. Rajagopalan, and A.Tomkins, Computer Networks **31**, 11 (1999).
3. L. Page, S. Brin, R. Motwani and T. Winograd, Stanford Digital Library Technologies Project (1998).
4. S. Bilke and C. Peterson Physical Review E **64**, 036106 (2001).
5. K. A. Eriksen, I. Simonsen, S. Maslov and K. Sneppen, Phys. Rev. Lett. **90**, 148701 (2003).
6. Strictly speaking the selection of nodes sending links to the outside world is biased to those with higher values of K_{out} and thus might be not representative of the whole community. This is not a problem in random scale-free networks where we can prove the absence of correlations between the out-degree of a node and its Google ranking. It remains to be seen to what extent the same statement holds in the real WWW [12].
7. A. Albert, H. Jeong, and A.-L. Barabasi, Nature, **401**, 130, (1999).
8. S. Maslov, K. Sneppen, A. Zaliznyak, Physica A, **333**, 529 (2004).
9. Such a seed network can be created e.g. using a stub reconnection procedure described in M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
10. S. Maslov and K. Sneppen, Science, **296**, 910 (2002).
11. J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
12. H. Xie, K.-K. Yan, S. Maslov, in preparation.

THE SOS RESPONSE OF BACTERIA TO DNA DAMAGE

A Single Cell Study

JOEL STAVANS
Department of Physics of Complex Systems
Weizmann Institute of Science
Rehovot 76100
Israel

A variety of sources such as radiation, chemical mutagens and products of metabolism induce damage to the genomes of organisms very often, from bacteria to man. Damage can be fatal for the organism since it can prevent DNA replication, and thus cell division. Evolution has given rise to elaborate mechanisms to either repair or bypass this damage. Upon encountering damage to their genomes, bacteria such as *E. coli* respond by activating the SOS network, consisting of about forty genes whose task is to repair/bypass the DNA damage, in order to enable DNA replication. The SOS genetic network deploys a variety of specific functions such as detecting damage, repairing it correctly by nucleotide excision (the NER mechanism) or by recombination, and if these functions do not succeed, bypassing damage by mutagenesis. The activation of all these functions requires a high degree of coordination and regulation, whose understanding is poor in spite of decades of study. I will survey recent findings in which the execution of the response was followed at the level of individual cells. These findings illuminate certain aspects of the concerted response, which are inacccessible to techniques in which large cell ensembles are interrogated. In particular, the findings show that the response exhibits highly precise modulations in the activation of a number of gene promoters, modulations which posess a digital character. Importantly, the precise timing mechanism responsible for the modulations is independent of the cell cycle, the main built-in clock of the cell. Genes responsible for the precision are identified. I will also highlight the importance of this network as one of the main forces driving the evolution of bacterial genomes.

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 39–47.

1. Introduction

DNA, the repository of all the information needed to encode an organism and the proteins that enable it to live is very often under attack. Damage to DNA can come from a variety of sources such as UV light the organism is exposed to, an assortment of chemicals that the organism ingests or breaths, and even products of the organism's own metabolism such as oxygen radicals. Lesions on the DNA may be lethal to a cell: they halt the DNA replication machinery, preventing the duplication of genetic information and cell division. In higher organisms, lesions may also cause cancer.

In order to survive, evolution has selected for mechanisms that help organisms to overcome DNA damage, by either repairing it, or by bypassing it. In the case of bacteria such as *Escherichia coli*, the set of genes comprising a genetic network in charge of this is known as the SOS response [1]. The SOS network consists of nearly forty genes documented today, not all of whose function has been completely elucidated [2]. There are also recent hints that RNA may control certain aspects of the response. A considerable amount of control of the response is also carried out at the level of protein-protein interactions. The proteins encoded by these genes carry out a variety of functions, serving as detectors of damage, fixing the damage exactly and if this cannot be done, then replicating DNA over the lesion, in order to enable DNA replication to restart, thus enabling the cell to divide. There are also mechanisms which prevent cell division until repair/bypass have been accomplished, and mechanisms to shut off the response. In these set of lectures I will provide glimpses of how the SOS network works, illustrating the use of following gene expression at the level of single cells. A full account of the experiments will be published elsewhere [3]. More complete reviews of the SOS response are available [1]. These experiments reveal aspects of the network's performance that are not accessible to techniques in which large ensembles of cells are interrogated. One must not forget that ultimately, proteins act at the molecular level. I will therefore also illustrate the complexity of the response at this level, by presenting a birds eye view of recent findings about a protein complex playing a key role in the network: the RuvAB complex [4].

1.1. DETECTION OF DNA DAMAGE AND THE EVENTS THAT ENSUE

Under normal conditions, when a cell is healthy and its DNA is intact, the cell does not need to engage upon wasteful synthesis of proteins needed for repairing or bypassing lesions. The genes encoding these proteins are therefore downregulated, by a master repressor protein called LexA. Things change when damage occurs. Damage is detected during DNA replication, when the accurate DNA polymerase encounters a lesion. Lesions are detected fast: a DNA polymerase complex translocates at a rate of about 1000 bp/sec. A concerted chain of events then ensues, that allows repair genes to be expressed and carry out their function. The story goes more or less like this.

A variety of agents such as UV or high energy radiation may give rise to lesions on the DNA such as dimerization of neighbouring pyrimidines. Upon encountering such a lesion, the highly accurate replication machinery of the cell at a replication fork comes to a halt. One important consequence of halting is that a tract of single-stranded DNA is exposed without being replicated. Single-stranded DNA is prone to attack and the cell quickly covers it with an important protein called RecA, which polymerizes along the single-stranded DNA substrate. This polymerization enables RecA to function as a coprotease: it greatly increases the inherent tendency of the LexA repressor to auto cleave. As we will see later, the RecA-DNA filament used to trigger the SOS response is also endowed with other important functions [5]. If the LexA pool decreases enough, then repair genes begin to be expressed [6]. Genes in the SOS network have LexA boxes at their promoters, with a consensus sequence. There are slight variations in the sequence of these boxes, with a consequent difference in the affinity of LexA. Genes or operons with weak boxes are expressed earlier than those with strong boxes. I will now sketch some of the important functions in repair and damage bypass, and refer the reader to more extensive reviews of the SOS response in the literature.

1.2. DIFFERENT MECHANISMS FOR REPAIR

One of the earliest expressed set of genes are the *uvr* genes, which encode for proteins which carry out nucleotide excision repair (NER). These proteins detect the damage on one strand, and excise it together with neighbouring bases on the same strand. The gap is quickly filled by DNA polymerases which leave the DNA intact, as if damage had not occurred. NER is by far the dominant repair mechanism. About 85% of lesions are repaired this way. Some lesions either escape NER, and another repair mechanism then sets in: recombination. Recombination is also carried out by RecA. The RecA on the RecA-DNA filament formed during the initial stages of the SOS response catalyses first the pairing of the single-stranded DNA in the filament with duplex DNA, one strand of which bears homology to the single-stranded DNA. After the homologous tract has been found, RecA catalysis strand exchange.

There are situations however when both NER and recombination fail to repair lesions on the DNA, and as a last resort, replication is restarted at a stalled fork, but at a price: mutagenesis. A special DNA polymerase, PolV, encoded by the processed products of the *umuDC* operon, carries out replication past the lesion, but in an inaccurate fashion [7,8]. It may introduce wrong nucleotides, and therefore mutations. PolV is formed when the UmuD protein undergoes cleavage, again by the RecA filament, to form UmuD'. Two of these molecules associate with UmuC, thus forming PolV. After a few nucleotides have been added PolV is replaced by the accurate PolIII DNA polymerase and the replication fork continues advancing until the next lesion is encountered.

The events described above must occur before cell division. A healthy *E. coli* cell divides every ~25 min. Naturally, dealing with lesions takes time, and cell division must be prevented before repair/bypass have been accomplished succesfully. This is carried out by another SOS protein, SulA, which inhibits the function of proteins involved in the mechanics of cell division.

Lastly, after successful repair of DNA damage, the LexA concentration must be restored in order to shut off the SOS network. There are a number of proteins suspected of playing roles in the shutoff. One of them is DinI, whose structure bears some resemblance with DNA [9]. It is believed that this structure enables it to interact with the RecA filament, inhibiting its function both as a protease and a recombinase. However there are other proteins which may also play a role, such as RecX [10], whose functions have not been fully elucidated.

The picture that emerges from the above description of the SOS response is that provided enough lesions on the DNA exist to activate it, a promoter of the network would be expected to exhibit a singly-peaked activation response.

2. Study of the SOS network by single cell experiments

While a considerable amount of information has been amassed over the years on the SOS network and the biochemistry of its components, not much is known about the coordination of its onset, progression through the different repair mechanisms and final shutoff. One problem that has impeded advance in this direction is that experiments are typically carried out with large ensembles of cells, in which one measures only an average response. Experiments monitoring gene expression in single cells may provide a wealth of new information and a new window at how the network operates, by enabling the study of deviations from the average response.

Our experiments on the SOS network were performed using fluorescence techniques [3]. In brief, plasmids encoding for the green fluorescent protein (GFP), fused to one of the SOS promoters was introduced into *E. coli* cells, with about 10 plasmids per cell. Expression of GFP allowed to obtain the activity of the particular promoter on the plasmid, by the following procedure: the total fluorescence intensity I(t) at a given time t emitted by a cell, as well as its length L(t) are measured. The activity of the promoter is then given by:

$$PA(t) = \frac{1}{L(t)} \frac{d}{dt} \left[I(t) L(t) \right]$$

Cells were then irradiated in situ by UV light in order to induce DNA damage in a controllable fashion, and observed with a home-made microscope held at 37 ^{o}C, both by dark field and fluorescence techniques.

3. Results

Typical curves of promoter activity of pRecA over time in individual cells are shown in Fig. 1, for UV doses of 10 and 20 J/m^2. At 10 J/m^2 the response is induced in many cells but not in all. Cells in which the response is induced fail to divide when their length is about four times their diameter, as healthy cells do. Instead they form long filaments due to the action of the SulA protein, that prevents cell division until repair has been accomplished. As the damage dose increases, less and less cells reach the stage when they are able to divide again. As in the case of Fig. 1 (left), most of the cells in which the response is induced exhibit a single peak of RecA activation, but a few exhibit a second peak as well. In contrast, the 20 J/m^2 dose is high enough to activate the SOS response in practically all cells, but low enough not to lead to cell death in the large majority of cells. The activity is noticeably modulated and three peaks can be discerned, appearing at times T_1, T_2 and T_3. The activity at each peak is plotted as a function of the time of its appearance for many cells in Fig. 2 for both doses.

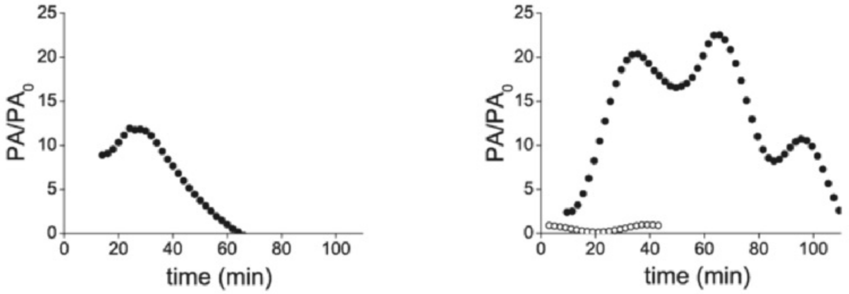


Figure 1. Normalized promoter activity PA/PA_0 as a function of time from the *recA* promoter, measured in two individual cells irradiated at 10 J/m^2 (left) 20 J/m^2 (right).

The data form three well-defined clusters, whose amplitude variation is larger than the temporal one. Part of the amplitude variation is due to the 20% variation in the plasmid copy number per cell. Note that cells with slightly higher T_1, have a corresponding higher T_2 and T_3. A full statistical analysis of data obtained at different doses shows that as the damage dose is increased, the modulations appear at longer times [3]. For example, at 35 J/m^2 the first peak appears after 40 min instead of at 29 min at 20 J/m^2. One can also monitor the rate at which the cell size grows. The data also shows that this rate slows down considerably with increasing dose: while irradiated cells divide every 30 min, cells irradiated at 50 J/m^2 show a doubling time above 100 min. Hence DNA damage slows down all the cell's metabolism.

Importantly, the cell cycle of the cells was not synchronized before the cells were irradiated (for about 15 sec). Thus there is a highly precise timing device as part of the SOS genetic network, which is independent of the cell cycle. In spite of this precision, it is interesting to note that the modulations in promoter activity have not been observed in bulk experiments probing a large ensemble of cells. The varying amplitude of the peaks between cells, the spread in the timing of their appearance account partly for this. However, other factors such as inhomogeneous induction due to absorption of UV light when cells are irradiated in bulk may also lead to considerable spread in timing, averaging out the peaks.

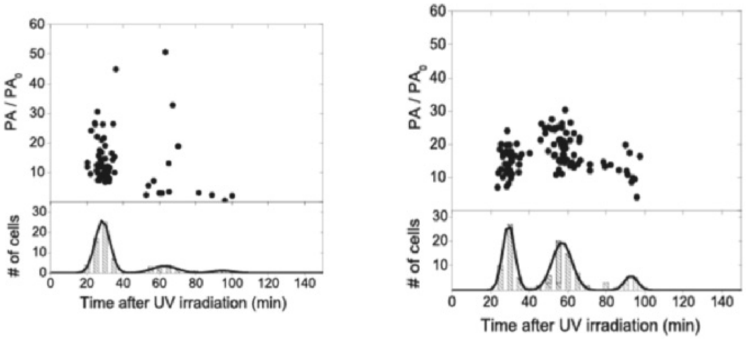


Figure 2. Quantitative analysis of the oscillatory behavior. Normalized amplitudes of the peaks in promoter activity of *recA* are plotted as function of peak time for individual cells irradiated with a UV dose of 10 J/m^2 (left) and 10 J/m^2 (right).

The behavior described above is not limited to the pRecA promoter. Indeed, experiments carried out with the pLexA and pUmu promoters exhibit similar modulations in promoter activity (those of the pUmu promoter appear at somewhat longer times). These facts suggest that modulations in promoter activity are due to modulations in the concentration of LexA. An analysis of the data at different damage doses yields the following picture. When cells are irradiated with a lower dose, 10 J/m^2, most cells exhibit a first peak, few exhibit a second peak and almost none a third one. Conversely, when cells are irradiated with a higher dose, 20 J/m^2, practically all cells exhibit three peaks. Note also that the amplitude of the three modulations rapidly becomes saturated with UV dose. Hence the SOS network was designed by evolution to respond "digitally": the number of modulations but not their amplitude increases with increasing DNA damage.

How is time counted in the cell, how can one explain the high precision of the timing device, and which elements in the SOS network play a role?. It has been proposed by G. Walker and coworkers [11] that the products of the *umuDC* operon may act as a prokaryotic DNA damage checkpoint effecting a timed pause in DNA replication, in addition to their role as an error-prone DNA polymerase (PolV). Evidence has shown that UmuD, in its uncleaved form may delay the resumption of DNA replication, providing more time for NER to repair accurately as many lesions as possible. The inhibition of replication may occur for example by the direct interaction of UmuD with Pol III. The cleavage of UmuD by the RecA filament into UmuD' removes on one hand the replication block and by forming PolV, it helps in the resumption of replication by translesion synthesis.

The findings by Walker suggested to us that the products of the *umuDC* operon may play a role in the modulations we observe. To test this hypothesis, we tampered with the *umuDC* operon. Our results with Δ*umuDC*, a mutant backround in which the *umuDB* operon was deleted, as well as with an uncleavable mutant of the UmuD protein (K97A) proved very revealing. Three salient features were observed: first, the second peak almost disappeared in both cases. Second, both the temporal variation in the first and third peaks was considerably larger than in the wild type. Third, the amplitude of the first peak depended on damage dose, contrary to the independence in the wild type. Hence the products of the *umuDC* operon play important roles in setting the precision of the timing device, and the digitality of the response.

4. Summary and outlook

DNA damage in organisms is very common, and is due both to environmental sources as well as internal ones such as products of metabolism. Genomic damage poses a big threat to life, and therefore evolution and natural selection have given rise to elaborate mechnisms for coping with it. In the case of bacteria, the SOS network is in charge of dealing with DNA damage. The experiments described above, which probe the response of the network at the level of individual cells, reveal a high level of regulation which is not limited to transcriptional control. An extensive set of interactions at the protein-protein level controls with high precision the temporal execution of the network's response, in a cell cycle-independent fashion. Modulation in the activity of a number of promoters indicates that the activity is controlled by changes in the concentration of a global regulator, the LexA repressor. The experiments show that modulations are highly precise from cell to cell, in spite of noise sources such as fluctuations in the number of regulatory proteins. While the experiments did not identify each modulation with a particular function, they revealed that the products of the *umuDC* operon, which carry out various functions, are involved in the temporal control of the modulations, and in the presence of the second peak of activity. Interestingly, the cell responds "digitally" by increasing the number of modulations instead of increasing the amplitude of the response. Further investigations will study the effects of other genes, how is the response shut off, what is the origin of the modulations and many other questions. One can speculate as to the advantages of having digital behavior versus an analog one. For example, avoding too large a response amplitude may avoid an overshoot and make it easier for the response to shut off.

It is known that under adverse conditions, such as in the absence of nutrients, cells may enter a stationary phase, in which growth is very limited. Under these conditions, many physiological changes occur, and in some of the cells, the SOS response is induced. As a result of this, DNA polymerases other than the highly accurate PolII enter the fray. We've met one: PolV. But also PolII and PolIV are induced. These replicate DNA introducing errors, thereby allowing the cell to try new mutations that will enhance its survival under the new conditions. The study of mutagenesis during stationary phase is being currently carried out by many groups (see for example [12]), and it is expected that it will have a strong impact on the understanding of the evolution of microorganisms.

References

1. Little, J.W. (1996) The SOS regulatory system. In Lin, E.C.C. and Simon, A. (eds.), *Regulation of Gene Expression in E. coli*. R. G. Landes Company, pp. 453-479.
2. Courcelle, J., Khodursky, A., Peter, B., Brown, P.O. and Hanawalt, P.C. (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. *Genetics*, **158**, 41-64.
3. Friedman, N., Vardi, S., Ronen, M., Alon, U. and Stavans, J. (2005) To be published in PLoS.
4. Amit, R. and Stavans, J. (2004) Direct observation of RuvAB-catalyzed branch migration of single Holliday junctions. *Proc. Natl. Acad. Sci. USA*, **101**, 11605-11610.
5. Sassanfar, M. and Roberts, J.W. (1990) Nature of the SOS-Inducing Signal in Escherichia-Coli - the Involvement of DNA-Replication. *J. Mol. Biol.*, **212**, 79-96.
6. Ronen, M., Rosenberg, R., Shraiman, B.I. and Alon, U. (2002) Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 10555-10560.
7. Tang, M.J., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R. and Goodman, M.F. (1999) UmuD ' C-2 is an error-prone DNA polymerase, Escherichia coli pol V. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 8919-8924.
8. Reuven, N.B., Arad, G., Maor-Shoshani, A. and Livneh, Z. (1999) The mutagenesis protein UmuC is a DNA polymerase activated by UmuD ', RecA, and SSB and is specialized for translesion replication. *J. Biol. Chem.*, **274**, 31763-31766.
9. Yasuda, T., Morimatsu, K., Horii, T., Nagata, T. and Ohmori, H. (1998) Inhibition of Escherichia coli RecA coprotease activities by DinI. *EMBO J.*, **17**, 3207-3216.
10. Stohl, E.A., Brockman, J.P., Burkle, K.L., Morimatsu, K., Kowalczykowski, S.C. and Siefert, H.S. (2003) Escherichia coli RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo. *J. Biol. Chem.*, **278**, 2278-2285.
11. Opperman, T., Murli, S., Smith, B.T. and Walker, G.C. (1999) A model for a umuDC-dependent prokaryotic DNA damage checkpoint. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 9218-9223.
12. Bjedov, I., Tenaillon, O., Gerard, B., Souza, V., Denamur, E., Radman, M., Taddei, F. and Matic, I. (2003) Stress-induced mutagenesis in bacteria. *Science*, **300**, 1404-1409.

SELF-AFFINE SCALING DURING INTERFACIAL CRACK FRONT PROPAGATION

STÉPHANE SANTUCCI, KNUT JØRGEN MÅLØY,
Fysisk Institutt, Universitetet i Oslo,
P. O. Boks 1048 Blindern, N-0316 Oslo 3, Norway

RENAUD TOUSSAINT AND JEAN SCHMITTBUHL
Institut de Physique du Globe de Strasbourg - UMR 7516,
5, rue René Descartes, 67084 Strasbourg Cedex, France

Abstract. We have performed an experimental study of slow crack front propagation through a weak plane of a transparent Plexiglas block. Spatial random toughness fluctuations along the weak interface generate a rough crack line in pinning locally the crack front, and leads to an intermittent dynamics of the crack front line. Using a high speed and high resolution camera we are able to capture the features of this complex dynamics.

A new analysis procedure is proposed in order to measure the waiting time fluctuations, and study the local burst dynamics and structure along the crack front during its propagation. First, we confirm previous results [1]: the fracture front dynamics is governed by local and irregular avalanches with very large size and velocity fluctuations, and can be described in terms of a Family-Vicsek scaling with a roughness exponent $\zeta \simeq 0.6$ and a dynamic exponent $\kappa \simeq 1.2$. Then, focusing in particular on the avalanches structure, we show that the system exhibits self-affine scaling with the same roughness exponent ζ for the local burst and the fracture front line itself.

Keywords: Interfacial fracture, depinning transition, self-affinity, dynamic scaling.

1. Introduction

Since the pioneering work of Mandelbrot, Passoja and Paullay [2], it is now well established that crack surfaces are self-affine objects. The scaling properties of the morphology of cracks manifest themselves through self-affine long range correlations [2, 3, 4, 5, 6, 7, 8, 9] with a roughness exponent which is found to be very robust for different materials and a broad range of length scales. However, the physical role played by the heterogeneities which lead to these self-affine long range correlations is not well understood.

In recent years front propagation in disordered media has become a challenging problem trying to describe the dynamics of interfaces in many different physical systems such as crack fronts [9], magnetic domain walls [10],

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 49–59.

or wetting [11]. Actually there exists few experimental data describing crack front propagation through heterogeneous material, essentially due to the difficulty of making direct observation and following the crack front line. Indeed, the crack front line growing in a 3d heterogeneous medium has itself a 3-d shape with different in-plane and out-of-plane roughnesses, respectively $\zeta_{\|}$ and $\zeta_{\perp}$. Therefore, the interfacial crack front problem simplifies the 3-d original one, both experimentally [12, 13] and theoretically [14]. Since the crack front is constrained geometrically to lie in the plane where the motion is driven by the stress transmitted through the two elastic plates, it is possible to perform direct visualization and to follow the fracture front line. So far most experiments on fracture front lines have been focused on the fracture front line morphology leading to the estimated roughness exponent $\zeta = 0.55 \pm 0.03$ [12], followed up by a longer study leading to $\zeta = 0.63 \pm 0.03$ [13]. Recently the interfacial crack front propagation has started to be investigated [1]. This study has shown that the fracture front line dynamics is intermittent -the depinning on asperities triggers local instabilities- and can be described in terms of a Family-Vicsek scaling [15] with a roughness exponent $\zeta = 0.6$ and a dynamic exponent $\kappa = 1.2$. In contrast to earlier numerical and theoretical studies [8, 16, 17, 18], recent numerical simulations interpreted as a stress-weighted percolation problem [19] give consistent results on the experimental roughness and dynamic exponents.

In this work, we went further on in the investigation of the local dynamics and this study appears as the continuation of the experimental work initiated by Måløy and Schmittbuhl [1]. In the first part, we will recall the experimental set-up and the sample preparation that permits us the direct observation of an in-plane crack front which propagates into the annealing plane of two transparent polymethylmethacrylate (PMMA) plates [12]. It is important to underline that now, using a really powerful high speed and high resolution camera (Photron Ultima), we are able actually to capture the details of the complex crack front dynamics. Then, in order to analyze the local burst dynamics and in particular to extract the local waiting time fluctuations, we propose in the second section a new analysis procedure. Both this analysis and the fast video recording confirm the previous observations and results, showing that the dynamics of the fracture front is driven by local and irregular avalanches with very large size and velocity fluctuations. In this paper we would like to focus on the scaling of the crack front line. We will first confirm that the development of the crack roughness follows a Family-Vicsek scaling and then examining in details the structure of the local avalanches we will show that the system exhibits self-affine scaling with the same roughness exponent ζ for the whole set of local bursts, and for the fracture front line itself. More details concerning the dynamics and in particular the velocity and waiting time fluctuations will be given elsewhere [20].

2. Experimental procedure

2.1. SAMPLE PREPARATION

Two Plexiglas plates ($32\times 14\times 1cm$ & $34\times 12\times 0.4cm$) are annealed together at $205°C$ during 30 *mn* under several bars of normal pressure, in order to create a single block with a weak interface. Before the annealing procedure both plates are sand-blasted on one side with $50\mu m$ steel particles or $100\mu m$ glass beads. Sand-blasting introduces a random topography which induces local toughness fluctuations during the annealing process. In order to estimate the characteristic size of the local heterogeneities arising from the sand-blasting process, we have measured the profile of a sand-blasted Plexiglas surface, using a white light interferometry technique (performed at SINTEF laboratory). We found that these local irregularities have an upper cut-off size estimated as $18 \pm 2\mu m$.

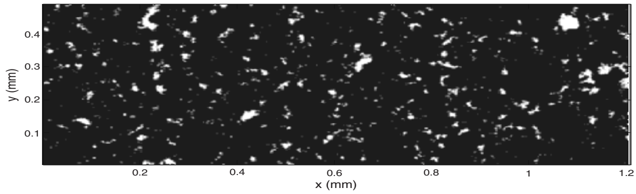


Figure 1. Two dimensional map of a sand blasted PMMA surface: white patches correspond to asperities higher than $0.85\mu m$, which corresponds to the standard deviation of the height of this profile.

The two key points of the procedure are *the transparency* of the material allowing *the direct observation* of the fracture front, and *the random toughness* introduced along the interface which generates a rough crack line in pinning the crack front.

2.2. MECHANICAL AND OPTICAL SET-UP

While the upper Plexiglas plate is clamped to a stiff aluminium frame, a press applies a normal displacement to the lower one (1 cm thick) corresponding to a crack opening in mode *I* configuration (pure tensile mode) at a low and constant rate $\sim 10\mu m.s^{-1}$.

A high speed and high resolution camera (Photron Ultima) mounted on a microscope allows us to follow the slow crack front propagation. Using this camera at a spatial resolution of 1024×512 pixels, and an acquisition rate of 1000 f.p.s. we can follow the crack front during $\simeq 12s$ (obtaining $\simeq 12000$ images). In this work, we will focus in particular on a given experiment with

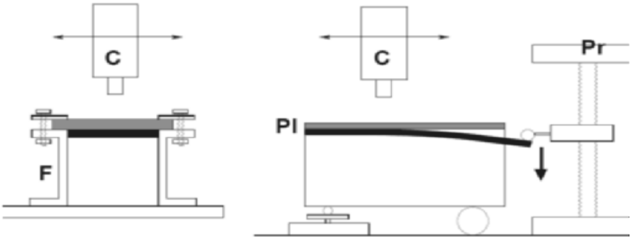


Figure 2. Sketch of the experimental setup: a thick PMMA plate (PL) is clamped into a rigid frame (F). A normal displacement is imposed by a press (Pr) on the thin plate with a cylindrical rod. A high speed and high resolution camera (C) mounted on a microscope follows the crack front propagation.

an average crack front speed $\langle v \rangle = 28.1\mu m/s$ an a pixel size $a = 3.5\mu m$. It is important to notice that the pixel size is smaller than the characteristic scale of these local heterogeneities, estimated as $18 \pm 2\mu m$ arising from the sand-blasting process. In a forthcoming paper [20], we will present using an extended set of data that all of these preliminary results are stable while changing the experimental conditions (varying both the average front line speed and the pixel size).

3. Analysis procedure and results

3.1. SCALING ANALYSIS OF THE CRACK FRONT

Image analysis is performed to extract the crack front by computing the gradient of the gray levels. A typical result is shown on figure 3. The front position being defined as the contrast interface between the uncracked (in black) and cracked parts is given by $y = h(x, t)$.

The power spectrum $P(k)$ of the deviations from the mean front position $h(x) - \langle h(x) \rangle$, averaged over all the crack front position detected during an experiment (in this case 9000 fronts), as function of the wave number k, is shown on figure 4. The slope of the best fit $P(k) \propto k^{-(1+2\zeta)}$ gives an estimation of the roughness exponent $\zeta = 0.54 \pm 0.06$. This value for the roughness exponent is consistent with previous careful estimations $\zeta = 0.55 \pm 0.05$ [12] and $\zeta = 0.63 \pm 0.03$ [13], where it has been extensively checked for fronts at rest over a much larger range of scales (around 3.5 decades), and using several techniques.

Moreover, we confirm that the development of the crack roughness follows rather well the Family-Vicsek scaling ansatz with a roughness exponent $\zeta = 0.6$ and a dynamic exponent $\kappa = 1.2 \pm 0.2$. Considering the power spectrum of the relative position $\Delta h(x, t) = h(x, t) - h_i$, where $h_i = h(x, t_i)$

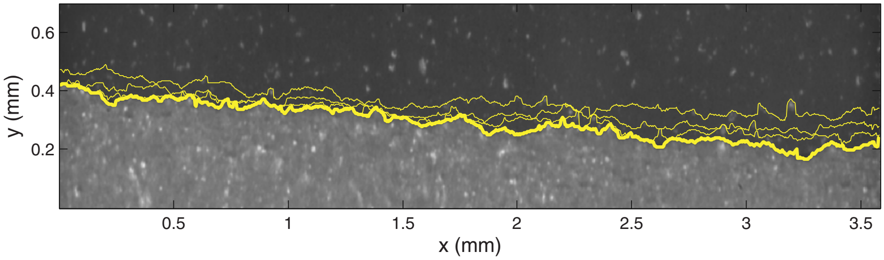


Figure 3. Typical example of a picture recorded by the high speed camera (Photron Ultima) during an experiment with an average crack front speed $\langle v \rangle = 28.1 \mu m.s^{-1}$. The pixel size is $a = 3.5 \mu m$. During this experiment, the camera recorded in total 9000 frames with a time delay of $1ms$ for each picture. The crack front propagates from bottom to top. The thick solid line represents the interface separating the uncracked (in black) and cracked parts extracted after image analysis. We superimpose 3 fracture front positions for later times (0.5, 1 and $2s$), suggesting the crack pinning and the burst activity.

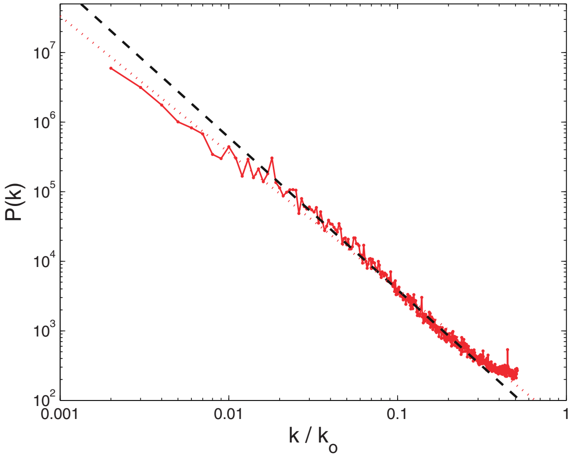


Figure 4. Power spectrum of the deviations from the mean front position $h(x) - \langle h(x) \rangle$ as function of the reduced wave number k/k_o with $k_0 = 2\pi/(3.5\mu m)$ averaged over 9000 crack front positions detected. The lines correspond respectively to the best fits $P(k) \propto k^{-(1+2\zeta)}$ of all the spectrum (dotted line), and excluding small and high wave number, $0.25 \geq k/k_o \geq 10^{-2}$ (dashed line). An average over the slopes of the fits gives an estimation of the roughness exponent: $\zeta = 0.54 \pm 0.06$. The reduced quantity is $k_0 = 2\pi/(3.5\mu m)$. The vertical axis is arbitrary.

is the initial front, the Family-Vicsek scaling ansatz can be written in the following way.

$$P(k, \Delta t) = \Delta t^{(1+2\zeta)/\kappa} G(k \Delta t^{1/\kappa}) \qquad \text{with} \qquad G(x) \propto \begin{cases} b & x \ll c \\ x^{-(1+2\zeta)} & x \gg c \end{cases}$$

where $\Delta t = t - t_i$ is the time delay between the analyzed images, b and c are characteristic constants.

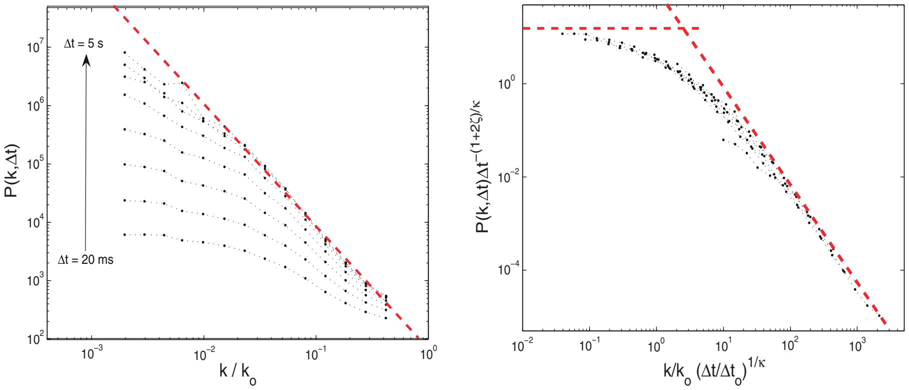


Figure 5. Left: Power spectrum of the relative position $\Delta h(x,t)$ as function of the reduced wave number k/k_o with $k_0 = 2\pi/(3.5\mu m)$ for logarithmically increasing time delays $\Delta t = t - t_i$, from $\Delta t = 20ms$ to $\Delta t = 5s$. Right: Data collapse for the scaling function $G(k\Delta t^{1/\kappa}) = P(k,\Delta t)\Delta t^{-(1+2\zeta)/\kappa}$, showing a dynamic Family-Vicsek scaling with a roughness exponent $\zeta = 0.54$ and a dynamic exponent $\kappa = 1$. The power spectra of the relative position $\Delta h(x,\Delta t) = h(x,t) - h(x,t_i)$ have been averaged over 4000 different initial front $h_i = h(x,t_i)$. The dashed lines are guides for the eyes and have a slope of -2.1. The horizontal axis correspond to reduced dimensionless quantities k/k_0, t/t_0, with $k_0 = 2\pi/(3.5\mu m)$, and $\Delta t_0 = 1ms$. The vertical axis are arbitrary.

On Fig. 5 to the left, we show the power spectrum of the relative position $\Delta h(x,t)$ for logarithmically increasing time delays Δt, from $\Delta t = 20ms$ to $\Delta t = 5s$, averaged over 4000 different initial front $h(x,t_i)$. When increasing the time delays Δt, we observe a crossover behaviour from a flat spectrum - indicating that no spatial correlations are present at small time delays Δt -, towards a power law behaviour at larger times Δt, consistent with the self-affine long range correlations previously observed $P(k) \propto k^{-(1+2\zeta)}$ with a roughness exponent $\zeta = 0.54$. Then, we plot on figure 5 to the right, the scaling function $G(k\Delta t^{1/\kappa}) = P(k,\Delta t)\Delta t^{-(1+2\zeta)/\kappa}$ as a function of $k\Delta t^{1/\kappa}$ with a roughness exponent $\zeta = 0.54$ previously measured. A satisfying data collapse is obtained for a dynamic exponent $\kappa = 1$. Performing the same procedure for different experiments leads to the following estimation for the dynamic exponent $\kappa = 1.2 \pm 0.2$ with a roughness exponent of $\zeta = 0.6$, consistent with previous experimental results [1].

3.2. WAITING TIME MATRIX **W** AND LOCAL FRONT VELOCITY MATRIX **V**

In order to study the local burst dynamics during the slow crack propagation we have computed a waiting time matrix $\mathbf{W(x,t)}$. The fracture front lines extracted from image analysis of the digital pictures (see Fig. 3) were added to obtain a waiting time matrix $\mathbf{W}$. This matrix has the dimension of the original image and an initial value equal to zero. We add the value 1 to the matrix element w corresponding to each pixel of the front line position detected. This procedure has been performed for all frames of a given experiment in order to obtain the final waiting time matrix $\mathbf{W}$ for each experiment. Then, a local normal speed of the interface $\mathbf{V}(\mathbf{x},\mathbf{t})$ can be deduced by computing the matrix of the inverse waiting time w times the ratio of the pixel size a on the typical time between two images δt. Therefore, we can associate to each pixel corresponding to the crack line in each image, a local front velocity $v = \frac{a}{w\delta t}$.

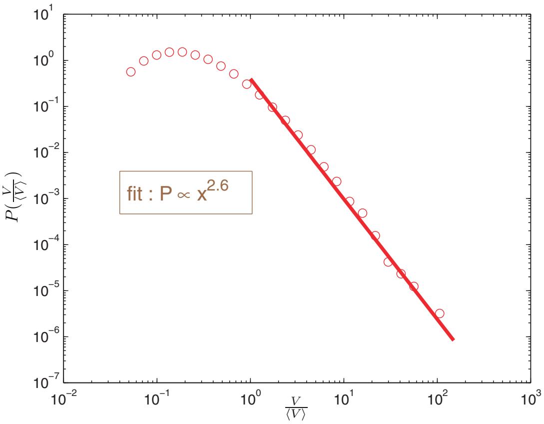


Figure 6. Distribution of the local front velocity normalized by the average crack speed $\frac{v}{\langle v\rangle}$ for an experiment with an average crack front speed $\langle v\rangle = 28.1\mu m/s$. The solid line represents a fit for velocities 2 times larger than the average crack front speed $\langle v\rangle$ and has a slope of −2.6.

Finally, we can obtain the probability distribution functions of the local waiting time w and the local front velocity v, in estimating the occurrence number of each measured waiting time or velocity on all pixels in all fracture front line images. A typical example of a distribution $P(\frac{v}{\langle v\rangle})$ of the local front velocity v, obtained for an experiment with an average crack front speed $\langle v\rangle = 28.1\mu m/s$ is shown on Fig. 6, in log-log scale. For velocities larger than the average crack front speed $\langle v\rangle$, we clearly observe that the local front velocities are power law distributed $P(v/\langle v\rangle) \propto (v/\langle v\rangle)^{-\eta}$ with an exponent $\eta \simeq 2.6$. This power law behaviour for the velocity distribution reveals a rich and non trivial underlying dynamic as one can observe on a fast video recording that the crack front is growing through irregular avalanches on all length scales. Since in this study we would like to focus on the local burst dynamic, more details concerning the velocity and waiting time fluctuations will be given elsewhere [20], and now we are going to examine in particular the structure of the local avalanches.

3.3. LOCAL BURST STRUCTURE

In order to analyze the local burst activity, let us consider the velocity matrix **V**. We can generate a clipped velocity matrix from **V** by setting the matrix elements v equal to one for $v > C \cdot \langle v \rangle$ and zero elsewhere. We present on Fig. 7 the spatial distribution of clusters of different sizes S obtained from the clipped matrix for a clip level $C = 8$. The white clusters correspond to velocities eight times larger than the average front speed $\frac{v}{\langle v \rangle} > C = 8$, which was for this experiment $\langle v \rangle = 28.1\mu m.s^{-1}$. The clusters connected to the first and the last front, and thus belonging to the upper and lower white parts are excluded from the analysis.

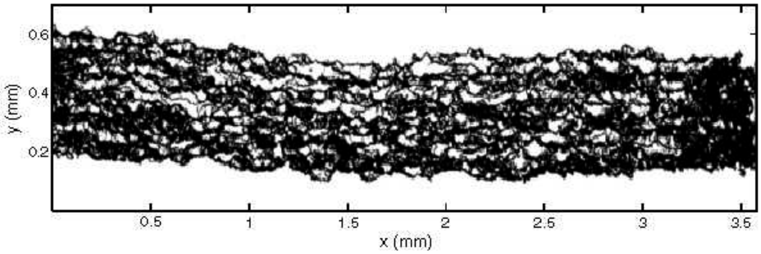


Figure 7. Spatial distribution of clusters size S. White clusters correspond to velocities $v > C \cdot \langle v \rangle$ with a clip level $C = 8$, during an experiment with an average crack front speed $\langle v \rangle = 28.1\mu m.s^{-1}$. The pixel size is $a = 3.5\mu m$. The crack front propagates from bottom to top.

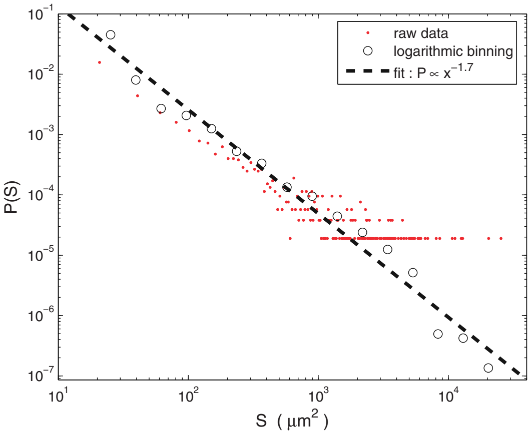


Figure 8. Distribution of the burst sizes S for an experiment with an average crack front speed $\langle v \rangle = 28.1\mu m.s^{-1}$ and a pixel size $a = 3.5\mu m$. The clip level used here $\frac{v}{\langle v \rangle} > C = 8$. Notice that the logarithmic binning allows to extract the scaling law underlying the large statistics, over a large range of event sizes. The dashed line represents the best fit $P(S) \propto S^{-\gamma}$ and gives an exponent $\gamma = 1.7$

In Fig. 8, the cluster sizes S distribution $P(S)$ is given in a double logarithmic scale. We clearly observe a power law behaviour $P(S) \propto S^{-\gamma}$ with an exponent $\gamma = 1.7$ proving that the burst dynamics occurs on all length scales. We have checked that this critical behaviour, and in particular the exponent γ, is really stable: normalizing by the average burst size $\langle S \rangle$, we can rescale all the different distributions corresponding to diverse experimental conditions and a wide range of clip level values (see [20] for more details).

We expect a correlation between the burst structure on small scales and the self-affine scaling of the crack front line on larger scales. In order to investigate the spatial scaling on small scales in detail, we have for each cluster S chosen the smallest bounding box enclosing it. The size of the bounding box gives the length scale L_y of the clusters along the growth direction and the length scale L_x of the clusters in the direction of the average fracture frontline.

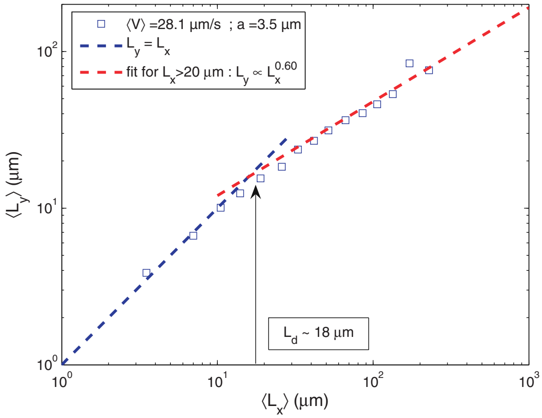


Figure 9. Average length scale $\langle L_y \rangle$ in function of the average length scale $\langle L_x \rangle$. The dashed line is a fit to the data points for $L_x > 20$ μm and has a slope 0.60 consistent with the roughness exponent of the fracture front line. The dotted line represents the curve $y = x$ and serves as a guide for the eye.

Figure 9 shows the dependence of the average length scale $\langle L_y \rangle$ as function of $\langle L_x \rangle$ in a double logarithmic plot for the same previous experiment ($\langle v \rangle = 28.1 \mu m.s^{-1}$, and $a = 3.5 \mu m$), for a given clip value $C = 8$. We clearly see that the avalanche clusters become anisotropic above a characteristic length scale $L_d \sim 18 \mu m$. This typical size can be interpreted as a correlation length for the disorder introduced by the sand-blasting technique as we have observed previously on Fig. 1. Below L_d the local toughness is marked by the same individual asperity and as a result the clipped velocity bursts appear isotropic. A linear fit of the data points for $L_x > 20 \mu m$ gives a slope 0.60 consistent with the roughness exponent $\zeta = 0.63 \pm 0.03$ of the fracture front line itself. This result shows that the system exhibits self-affine scaling with the same roughness exponent ζ for the local dynamic bursts as for the fracture front line itself, and brings a new confirmation of the roughness exponent determined in such interfacial crack experiments. It is important to note that most present theoretical and numerical works predict a lower value for this roughness exponent [8, 16, 17, 18].

4. Conclusion

We have performed an experimental study of slow interfacial crack front propagation. This work confirms previous experimentals results [1]: the fracture frontline dynamics can be described in terms of a Family-Vicsek scaling with a roughness exponent $\zeta \simeq 0.6$ and a dynamic exponent $\kappa \simeq 1.2$. Moreover, we have observed that this dynamics is driven by local and irregular avalanches whose size and velocity are power law distributed. We show that the dynamics and the structure of the local bursts are playing a crucial role for the scaling and the dynamics of the crack front itself. In particular, above a typical size $L_d \sim 20\mu m$, the burst size scales differently in the direction parallel and normal to the fracture front with an exponent consistent with the roughness exponent of the fracture front $\zeta \simeq 0.6$. However, if it appears consistent to interpret this result as finite size effect related to the disorder, more experimental work is needed to control and change the typical disorder size, in order to be conclusive and affirm that this cut-off length scale really corresponds to the quenched disorder correlation length introduced by the sand-blasting procedure.

Recent simulations interpreted as a stress-weighted percolation problem [19] give consistent results with the experimental roughness and dynamic exponent measured, in contrast to earlier numerical and theoretical studies [8, 16, 17, 18]. However so far, no theory or simulations have examined the dynamics and in particular the local burst activity of the crack line, and therefore it appears of central importance to develop these numerical studies.

Acknowledgements

We acknowledge E.L. Hinrichsen, M. Fleissner, and SINTEF laboratories in Oslo for their hospitality, permitting us to perform the white light interferometry measurements. We also thank L. Vanel, O. Ramos, A. Hansen, E.G. Flekkøy, and J.P. Vilotte for fruitful discussions. This work was supported by the CNRS/NFR PICS program, and the NFR Petromax and SUP program.

References

1. K. J. Måløy and Jean Schmittbuhl. *Phys. Rev. Lett.*, **87**, 105502 (2001)
2. B. B. Mandelbrot, D. E. Passoja, and A. J. Paullay. *Nature*, **308**, 721 (1984);
3. S. R. Brown and C. H. Scholz. *J. Geophys. Res.*, **90**,12575 (1985);
4. E. Bouchaud, G. Lapasset, and J. Planés. *Europhys. Lett.*, **13**, 73 (1990);
5. J. Schmittbuhl, S. Gentier, and S. Roux. *Geophys. Res. Lett.*, **20**, 639 (1993);
6. B. L. Cox and J. S. Y. Wang. *Fractals*, **1**, 87 (1993);
7. K. J. Måløy, A. Hansen, E. L. Hinrichsen, and S. Roux. *Phys. Rev. Lett.*, **68**, 213 (1992);
8. J. Schmittbuhl, F. Schmitt, and C. H. Scholz. *J. Geophys. Res.*, **100**, 5953 (1995).
9. E. Bouchaud. *J. Phys*, **9**, 4319 (1997).
10. S. Lemerle, J. Ferr, C. Chappert, V. Mathet, T. Giamarchi and P. Le Doussal *Phys. Rev. Lett.*, **80**, 2849, (1998).
11. E. Rolley, C. Guthmann, R. Gombrovisch and V. Repain, *Phys. Rev. Lett.*, **80**, 2865, (1998).
12. J. Schmittbuhl and K. J. Måløy. *Phys. Rev. Lett*, **78**, 3888 (1997);
13. A. Delaplace, J. Schmittbuhl, and K. J. Måløy. *Phys. Rev. E*, **60**, 1337 (1999); J. Schmittbuhl, A. Delaplace, and K. J. Måløy. *In Physical Aspects of Fracture.* (eds E. Bouchaud), (Kluwer Academic Publishers, 2001)
14. J. Schmittbuhl, S. Roux, J. P. Vilotte and K. J. Måløy.*Phys. Rev. Lett.*, **74**, 1787 (1995).
15. F. Family and T. Vicsek. *J. Phys. A*, **18**,L75 (1985).
16. J. P. Bouchaud, E. Bouchaud, G. Lapaset, and J. Planes. *Phys. Rev. Lett.*, **71**, 2240 (1993).
17. S. Roux and A. Hansen. *J. Phys. I*, **4**, 515, (1994).
18. S. Ramanathan and D. Fisher. *Phys. Rev. B*, **58**, 6026 (1998);
19. J. Schmittbuhl, A. Hansen and G. Batrouni. *Phys. Rev. Lett.*, **90**, 045505, (2003)
20. K. J. Måløy, S.Santucci, R.Toussaint and J. Schmittbuhl. submitted to *Phys. Rev. Lett.*

DIFFUSION, FRAGMENTATION AND MERGING PROCESSES IN ICE CRYSTALS, ALPHA HELICES AND OTHER SYSTEMS

JESPER FERKINGHOFF-BORG, MOGENS H. JENSEN
(`mhjensen@nbi.dk`),
POUL OLESEN
Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen ØDenmark
JOACHIM MATHIESEN
NTNU, Institutt for fysikk, N-7491 Trondheim

Abstract. We investigate systems of nature driven by combinations of diffusive growth, size fragmentation and fragment coagulation. In particular we derive and solve analytically rate equations for the size distribution of fragments and demonstrate the applicability of our models in very different systems of nature, ranging from the distribution of ice crystal sizes from the Greenland ice sheet to the length distribution of α-helices in proteins. Initially, we consider processes where coagulation is absent. In this case the diffusion-fragmentation equation can be solved exactly in terms of Bessel functions. Introducing the coagulation term, the full non-linear model can be mapped exactly onto a Riccati equation that has various asymptotic solutions for the distribution function. In particular, we find a standard exponential decay, $\exp(-x)$, for large x, and observe a crossover from the Bessel function for intermediate values of x.

1. Dynamics of Ice Crystal Formation

The North Greenland Ice Core Project (NorthGRIP) ice core provides paleoclimatic information back to about 120 kyr before present [1]. Each year, precipitation on the ice core covers it with a new layer of snow, which gradually transforms into ice crystals as the layer sinks into the ice sheet. We study the dynamics of the ice crystal distributions and use data previously obtained from the upper 880 m of the NorthGRIP ice core [2]. The data covers a time span of approximately 5300 years and going backwards in time, towards the depth of 880m, the distributions evolve to a universal curve. The universality indicates a common underlying physical process in the formation of crystals. We identify this process as an interplay between fragmentation of the crystals and diffusion of their grain boundaries. The process is described by the two-parameter diffusion-fragmentation equation equation to which we obtain the exact solution.

The crystal size distributions are obtained from fifteen vertical thin sections of ice evenly distributed in the depth interval 115 - 880 m. [2]. The thin

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 61–70.

sections have dimensions of 20 cm × 10 cm (height × width) and a thickness of 0.4±0.1 mm. Where the height is close to the present annual snow accumulation in the area which on the average, at the considered depths, becomes equivalent to 19.5 cm ice. Therefore each sample roughly covers a full year and averages over the seasonal variations. Digital images of ice thin sections placed between crossed linear polarizers are used to map the dimensions of individual ice crystals in the sample. In our analysis we define the crystal size as the vertical extent of the smallest vertically aligned rectangle enclosing the crystal., see Fig. 3.

In Fig. 1 we show the distributions of the crystals at selected depths. Overall the average crystal size is increasing until some depth corresponding to 2500 years back in time where it seems to saturate. Effectively, the process behind the grain growth can be described in terms of boundary diffusion between the ice crystals [3]. In this respect, it is the grain boundaries rather than the grains themselves that have physical significance. The grains, characterized by small radii of curvature, tend to be eliminated by motions toward their centers of curvature as time progresses. This causes smaller crystals to be incorporated into the larger ones leading to an overall growth of the mean

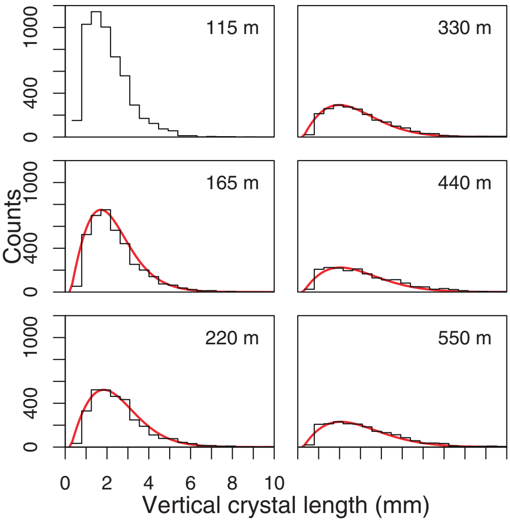


Figure 1. Distributions of ice crystal sizes at depths 115m, 165m, 220m, 330m, 440m and 550m. The crystal size is defined as the vertical extension of the individual crystals. The ragged curves are the measured histograms and the smooth lines are the temporal evolution predicted by eq. (4) starting from the initial distribution at 115m. The total counts of ice crystals decreases with depth (due to the overall increase of sizes) until the steady state is reached. Note that the distributions vanish at a small, finite value to account for the experimental bias; see [6].

crystal size [4, 5]. This approximative description is captured by the widely used Normal Grain Growth law. However, the diffusive growth never saturates, in clear contradistinction to the experimental observations. Moreover, the Normal Grain Growth law explains little about the actual shape of the distributions which recently have been proposed to resemble log-normals. Log-normal distributions are typically explained in terms of a multiplicative process, and in fact such a process could be important in the crystal dynamics if the individual crystals succesively did fragment. On the other hand, this process would lead to an overall decrease in the crystal sizes.

Recently, we have proposed a new framework predicting to great accuracy the evolution of the crystal size distributions and the observed distributions [6, 7]. Our framework combines the process of grain bounday diffusion with that of an occasional fragmentation, and demonstrates that the distributions more likely are to be described by Bessel functions.

Our proposed model is formulated as a rate equation in the quantity $N(x, t)$, which is the density of ice crystals of length x at time t measured before present. At a given time, $N(x, t)$ can be increased or decreased by diffusion with a diffusion constant D. It can receive fragments of size x from fragmentations of larger crystals and it can decrease by its own fragmentation. The fragmentation is defined as a rate f in length and time, i.e. for a given time step dt the average number of fragmentation events over a length L is $fLdt$. We therefore end up with an integro-differential equation of the form:

$$\frac{\partial N(x,t)}{\partial t} = D\frac{\partial^2 N(x,t)}{\partial x^2} - fxN(x,t) + 2f\int_x^\infty N(x',t)dx' \tag{1}$$

Here, the first term on the right hand side is a diffusion term that corresponds to the grain boundary diffusion of the Normal Grain Growth law. The second term in Eq.(1) is the rate at which crystals of size x fragment and the last term is the contribution from fragmentation of larger crystals into crystals of size x. Combining the two ways a crystal of size $x' > x$ can produce a fragment of size x with the assumption that there is a uniform probability, $1/x'$ for where the crystal breaks, we get

$$fx'N(x') \cdot \frac{2}{x'}, \tag{2}$$

where $fx'N(x')$ is the number of crystals of size x' that fragments per time and $2/x'$ is the probability for generating a fragment of size x. If we integrate (2) over all crystals larger than x we achieve the last term in (1). The model has been generalized to over fragmentation kernels in [9].

The integro-differential equation has analytical solutions, $B((x+\lambda)/a^{1/3}$, which are explicitly given in terms of the Bessel-function $K1/3$ and eigenvalues λ (see [8] for details),

$$B\left(\frac{x+\lambda}{a^{1/3}}\right) = \frac{x^{3/2}}{\sqrt{3}} K_{1/3}\left[\frac{2}{3}\left(\frac{x+\lambda}{a^{1/3}}\right)^{3/2}\right]. \tag{3}$$

Here, the parameter $a = D/f$ is the ratio of the diffusion constant and fragmentation rate and $a^{1/3}$ sets the natural length scale of the ice crystals. The Bessel function distribution has a characteristic stretched exponential tail $\exp(-x^{3/2})$, distinctively different from a log-normal tail $\exp(\exp(-x^2))$. Using the boundary condition that no new crystals are nucleated meaning that $N(0,t) = 0$ for all times [6, 7], implies that only a discrete set of non-positive values for λ is allowed. They are found by solving the equation $B(\lambda/a^{1/3}) = 0$. The largest eigenvalues are $\lambda_0/a^{1/3} = 0$, $\lambda_1/a^{1/3} = -2.338\ldots$, $\lambda_2/a^{1/3} = -4.088\ldots, \ldots$. The general solution can then be expressed as a linear combination of the eigenfunctions B in the form,

$$N(x,t) = \sum_{n=0}^{\infty} c_n B\left(\frac{x+\lambda_n}{a^{1/3}}\right) e^{\lambda_n f(t-t_0)}, \tag{4}$$

where the coefficients c_n is determined by matching $N(x, t = t_0)$ with the intial distribution at time $t = t_0$. The fact that all eigenvalues are negative except $\lambda_0 = 0$ guarantees that the dynamics converges to a unique steady state solution, $N(x,t) \sim B(x/a^{1/3})$, at high ages ($t \to \infty$). The characteristic time τ of the exponential growth towards this steady steady state is found from $lamda_1$;

$$\tau = -\frac{1}{\lambda_1 f} \approx \frac{1}{2.338 f a^{1/3}}.$$

In Fig. 2 we show the mean vertical size of the ice crystals, $\langle x \rangle(t)$ as function of time. The dots are the experimental values and the solid line is an exponential fit corresponding to the two leading terms in the solution,

$$\frac{\langle x \rangle_\infty}{1 + \left(\frac{\langle x \rangle_\infty}{\langle x \rangle_0} - 1\right) e^{-(t-t_0)/\tau}},$$

where $\langle x \rangle_0$ is the observed average length at time $t_0 = 500$ years and $\langle x \rangle_\infty$ is the mean vertical size in the steady state, $\langle x \rangle_\infty = 3^{2/3}\frac{\Gamma(4/3)}{\Gamma(3/2)} a^{1/3}$. From the figure, we estimate the characteristic time, $\tau = 600 \pm 100$ years and the average length in the steady state $\langle x \rangle_\infty = 2.9 \pm 0.1$ mm. The two parameters correspond to an effective fragmentation rate and a diffusion constant of respectively $f = (3.6 \pm 0.2) \cdot 10^{-4}$ mm$^{-1}\cdot$yr^{-1} and $D = (2.8 \pm 0.4) \cdot 10^{-3}$ mm^{2}yr^{-1}. Using these estimates, we can predict the time evolution of crystal

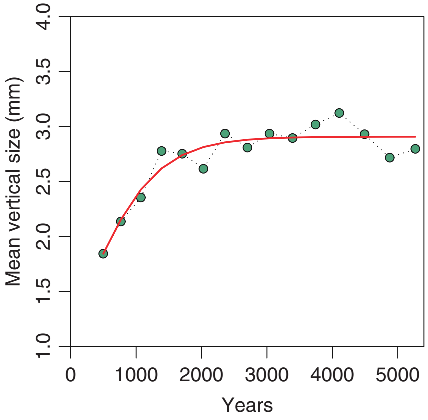


Figure 2. The mean vertical size of the ice crystals shown versus their age in years B.P. The red line shows the best fit predicted from our dynamical description of ice crystal growth. From the fit we read off the diffusion constant, $D \approx 2.8 \cdot 10^{-3}\mathrm{mm}^2 \cdot \mathrm{yr}^{-1}$, and fragmentation rate, $f \approx 3.6 \cdot 10^{-4}\mathrm{yr}^{-1} \cdot \mathrm{mm}^{-1}$. The time scale is taken from ref [1].

sizes from any initial distribution. The solid lines in Fig. 1 show the time evolution of the distribution observed at time $t_0 = 500$ years (depth 115m), in excellent agreement with the experimental results.

We can be generalize our diffusion-fragmentation model by changing the fragmentation rate to be proportional to x^β to account for energy/surface arguments. In this case the stationary distribution can also be found, where e.g. for $\beta = 2$ the solution will be $B(x) \sim \sqrt{x} K_{1/4}(x^2/2)$ instead of eq.(3). It turns out that the expression for $\beta = 2$ provides a poor fit to the experimental data for the ice crystals.

2. The evolution of α-helices in proteins

The α-helix is the most abundant structural motif of proteins, covering about one-third of the amino acid residues [10]. The α-helix is a stable structure in which the $C = 0$ group of residue i makes a hydrogen bond with the NH of residue $i + 4$.

The length distribution of α-helices in a representative set of 299 high resolution structures with low homology [11] exhibits a clear maximum around a length of 7 amino acids followed by a long tail, see Fig. 3. The data do not fit well any simple statistical distribution and a polynomial fit requires at least a fourth order expression to become acceptable [11]. In the following we will consider this ensemble of α-helices as the result of an ’infinitely’ long evolutionary process where diffusive growth competes with occational

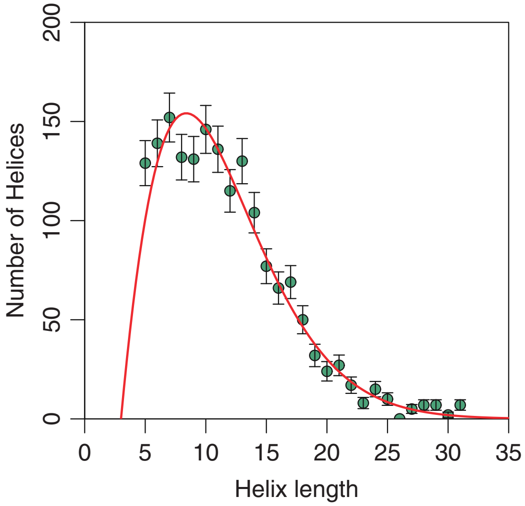


Figure 3. The black dots represent the length distribution of α-helices from 299 high resolution protein structures with low homology [11]. The error bars are estimated as the square root of the numbers. The full curve is the stationary distribution (i.e. infinite time limit) obtained from eq. (1) plotted versus the helix length measured in terms the number of amino acids. The ratio between the diffusion constant D and the fragmentation rate f resulting in the best fit to the data is $(D/f)^{1/3} \approx 6.1$ number of amino acids. Note that $N(x)$ vanishes for a helix length equal to 3, as a α-helix needs one turn to be identified.

fragmentation. Here, the diffusive growth reflects the small structural changes a given protein can accomodate without changing the arrangement of its structural elements (its tertiary structure), whereas a ‘fragmentation’ process reflects the breaking of a helix required to maintain an overall globular shape. By considering the dynamics of the ensemble of α-helices, we effectively randomize the particular evolutionary advantages of growing or shrinking any particular α-helix in any particular protein. However, we believe that the result of such ensemble average will reveal the main determinants of α-helix formation in proteins, much in the same way as binding energies between amino acids tend to influence the frequencies of neighboring relations between them in an ensemble of proteins [12].

The stationary solution to eq. (1) is defined by one free parameter (the ratio D/f) and it nicely reproduces the overall rise and decline of the observed length dependence of helix frequencies with $D/f \approx 6.1$, see Fig. 3, Simple energy arguments suggest that a point mutation in the middle of the helix to a residue not forming hydrogen bonds has a statistical weight of $\approx 1/400$ relative to the weight of growing/shrinking the helix [13]. This corresponds to the ratio $(D/f)^{1/3} \approx 7.4$ in reasonable agreement with our result. We take the difference to indicate the presence of other determinants of helical formation, such as stability constraints of the overall protein structure.

3. Diffusion, Fragmentation and Coagulation

Since the pioneering work of Smoluchowski [14, 15] dating back to the beginning of the last century, the literature on coagulation and fragmentation processes has grown considerably. Smoluchowski original coagulation equation [14, 15] provides a mean field description of clusters that coalesce by binary collisions with a constant rate. Scaling theory and exactly solvable models in the kinetics of irreversible aggregation has recently been reviewed in [16].

Fragmentation and coagulation was first considered as combined processes in [17], and mean field type of coagulation-fragmentation models have subsequently been used in a diverse range of applications, including polymer kinetics [18], aerosols [19], cluster formation in astrophysics [20] and animal grouping in biology [21, 22]. We refer to [23, 24] (and references therein) for a survey of the progress in the study of coagulation-fragmentation process.

We now present an extension to the diffusion and fragmentation model introduced above to account for coagulation processes as well. The extension can be considered as a generalization of the Smoluchowski coagulation-fragmentation equation to include processes where size diffusion is important. We emphasize that our approach differs from the approach in e.g. [25], where the clusters represent particles immersed in a gas or liquid type of medium and the diffusive term added to the coagulation-fragmentation equation represents the random movement of the center of mass of each cluster. Rather, the diffusion considered here is equivalent to the grain boundary diffusion.

In this section we shall discuss an extenstion to Eq. (1), with constant diffusion and a fragmentation probability growing linearly with the fragment size. We add a nonlinear coagulation process where two random fragments merge at a rate $\beta > 0$ and independent of their sizes. The basic equation for the steady state solution is thus

$$D\,\partial_x^2 N(x) - f x N(x) + 2f \int_x^\infty dx' N(x') + \beta/2 \int_0^x dx' N(x') N(x-x') - 2\tilde{\beta} N(x) = 0, \tag{5}$$

where the second last term on the left side is the feed to fragments of size x from all smaller fragments and the last term is the drain from merging events with all larger fragments. We have used the definition

$$\tilde{\beta} = \beta/2 \int_0^\infty dx N(x). \tag{6}$$

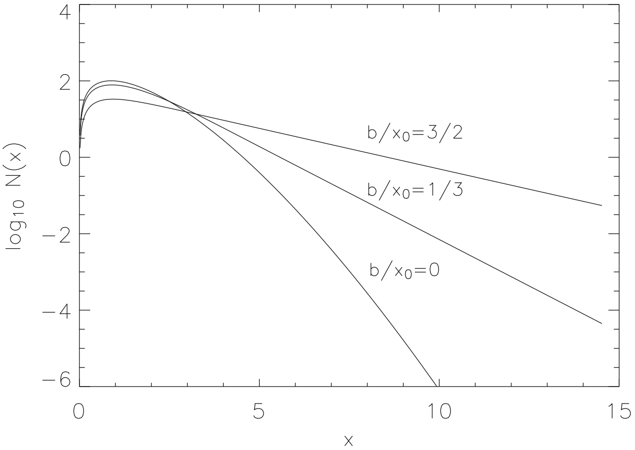


Figure 4. The numerical steady-state distributions $N(x)$ as function of x (both quantities dimensionless) for different coagulation-to-fragmentation ratios b/x_0. Note the crossover from a a bessel-like tail $\sim \exp(-2/3x^{3/2})$ to a pure exponential tail.

In addition to the characteristic length scale

$$x_0 = (D/f)^{1/3} \tag{7}$$

in the pure diffusion-fragmentation equation we now have a new length scale in the presence of coagulation

$$b = \tilde{\beta}/f. \tag{8}$$

In general we would like to solve equation (5) with the boundary conditions that $N(0) = 0$ and $N(x) \to 0$ for $x \to \infty$. Furthermore, we need of course also to invoke the condition that $N(x)$ is positive. In [8] we map this non-linear equation to a Riccatti type equation, which again can be mapped to a linear second order differential equation doing a proper substitution. In [8] we do a saddle point approximation to this linear equation and find a stable saddle point leading to an approximation of $N(x)$ for large x,

$$N(x) \approx \text{const.}\ e^{\tau_0\ x}, \tag{9}$$

where the relevant τ_0, under the condition that $N(x)$ is finite and positive, becomes

$$\tau_0 = \frac{i}{< x > -b/x_0} \left(\sqrt{\frac{2 < x >}{b/x_0} - 1} - 1 \right). \tag{10}$$

In Fig. 4 we show $N(x) = \tilde{N}(\tilde{x})$ as function of $x = \tilde{x}/x_0$ for different values of b/x_0. For small values of b/x_0, one indeed observes a cross-over from the bessel-like function in Eq. (3) at intermediate values of x to a pure exponential for large values of x. At larger values of b/x_0, only a single exponential form is observed for $x > 1$.

4. Conclusion

We have proposed a general dynamical scheme which involves the three fundamental physical processes diffusive growth, size fragmentation and fragment merging. In the case where merging is absent, we have applied the scheme successfully to the crystal size distribution of the ice core of the Greenland ice and to alpha-helices in proteins.We believe that the full dynamical scheme should also be relevant in many physical situations, such as for solutions of macromolecules like polymers, proteins and micelles. In fact, from measured distributions of fragment sizes we suggest that one might be able to identify how important the merging process is compared to the fragmentation process.

References

1. D. Dahl-Jensen, N. Gundestrup, H. Miller, O. Watanabe, S.J. Johnsen, J.P Steffensen, H.B. Clausen, A. Svensson, and L.B. Larsen, *Annals of Glaciology* **35**, 1-5, 2002.
2. A. Svensson, K.G. Schmidt, D. Dahl-Jensen, S.J. Johnsen, Y. Wang, J. Kipfstuhl, and T. Thorsteinsson, Annals of Glaciology **37**,113-118 (2003).
3. N. P. Louat, Acta Metallurgica **22**, 721 (1974).
4. R. B. Alley, J. H. Perepezko and C. R. Bentley, J. Glaciol. **32**(112), 425-433 (1986).
5. Paterson, W.S.B. *The Physics of Glaciers*, 480 pp., Pergamon, New York (1994).
6. J. Mathiesen, J. Ferkinghoff-Borg, M.H. Jensen, M. Levinsen, P. Olesen, D. Dahl-Jensen and A. Svensson, J. Glaciol. **50**, 325 (2004).
7. J. Ferkinghoff-Borg, M. H. Jensen, J. Mathiesen, P. Olesen and K. Sneppen, Phys. Rev. Lett. **91**, 266103 (2003).
8. P. Olesen, J. Ferkinghoff-Borg, M. H. Jensen and J. Mathiesen. Phys. Rev. E in print, cond-mat/0411514.
9. Laurençot P., *Steady states for a fragmentation equation with size diffusion*, to appear in a volume of Banach Center Publications.
10. A. Fersht, *Structure and Mechanism in Protein Science*, W.H Freeman and Company (1999).
11. Penel, S., Morisson, R. G., Mortishire-Smith, R. J. and Doig, A. J. J. Mol. Biol., 293, 1211-1219 (1999).
12. S. Miyazawa and R.L. Jernigen, Macromolecules **18**, 534 (1985); S. Miyazawa and R.L. Jernigen, J. Mol. Biol. **256**, 623 (1996)
13. A. J. Doig, A. Chakrabartty, T. M. Klingler and R. Baldwin, Biochemistry **33**, 3369 (1994).
14. M. von Smoluchowski, Phys. Z. **17**, 557, 585 (1916).

15. M. von Smoluchowski, Z. Physik. Chem. **92**, 129 (1917).
16. F. Leyvraz, Phys. Rep. **383**, 95 (2003).
17. Z. A. Melzak, Trans. Amer. Math. Soc **85** 547 (1957).
18. R. M. Ziff, J. Stat. Phys. **23**, 241 (1980).
19. S. K. Frielander. *Smoke, dust and haze: fundamentals of aerosol dynamics*, Wiley, New York (1977).
20. J. Silk and S. D. White, Astrophysical J. **223** L59 (1978).
21. A. Okubo, Adv. Biophys. **22**, 1 (1986).
22. S. Gueron and S. A. Levin, Math Biosci. **128**, 243 (1995).
23. R. L. Drake, *A general mathematical survey of the coagulation equation*. In G.M. Hidy and J.R. Brock, editors, Topics in Current Aerosol Research (Part 2), pages 201-376. Pergamon Press, Oxford, 1972.
24. P. Laurençot and S. Mischler, *On coalescence equations and related models* in "Modeling and computational methods for kinetic equations", Editors Degond, Pierre; Pareschi, Lorenzo and Russo, Giovanni, in the Series Modeling and Simulation in Science, Engineering and Technology (MSSET), Birkhauser (2004)
25. H. Amann, H., Arch. Rat. Mech. Anal. **151**, 339 (2000); H. Amann. and Weber, *Adv. Math. Sci. Appl.* **11**, 227-263 (2001); P. Laurençot and S. Mischler, Arch. Rat. Mech. Anal. **162**(1), 45 (2002).

MOLECULAR MECHANISMS IN BIOSIGNALLING: VISUAL RECEPTION

MIKHAIL A. OSTROVSKY
Institute of Biochemical Physics, Russian Academy of Sciences, Moscow
Joint Institute for Nuclear Research (JINR), Dubna

1. Main chapter

The main general idea, main topic of my lecture today is photobiological paradox of vision. The core of the paradox is: light in vision is not only a carrier of information, but also a risk factor of damage to the eye structures, first of all to the retina and to the layer of cell behind the retina – so called retinal pigment epithelium. In fact, the same paradox is real to other photobiological process – to photosynthesis. In this case light is used not only to convert and accumulate energy but also light can damage the photosynthetic molecular machinery. In both cases complex photoprotective systems, along with physiological system of visual reception or photosynthesis, have been developed in the course of evolution. In both cases the sophisticated photoprotective systems are able to solve the paradox of both photobiological processes. The impairment of these systems can lead in case of vision to human retina diseases or play role in progression of eye diseases like age-related macula degeneration, and in case of photosynthesis to destruction of plant cells or photosynthetic microorganisms.

That is why speaking about vision two functional systems simultaneously and close relating to each other have been developed in the course of visual system evolution. I mean systems of photoreception and photoprotection. The first one includes on the level of the retina photoreceptor cells physiological processes of phototransduction, light and dark adaptation, and also molecular basis of colour discrimination. The second one consists of at least three lines of protection: permanent renewal of rod and cone outer segment, power and complex antioxidant system and optical media of the eye as cut-off filters.

Let me start from the first system of photoreception, other words from the molecular mechanisms of visual biosignalling, and then I will consider the molecular mechanisms of potential light damage to the retina and retinal pigment epithelium and the complex, multipart system of photoprotection.

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 71–106.

In order to present you all I am going to present, I should first of all to describe the eye structures – from eye itself to molecules, where all these physiological and pathological processes take place.

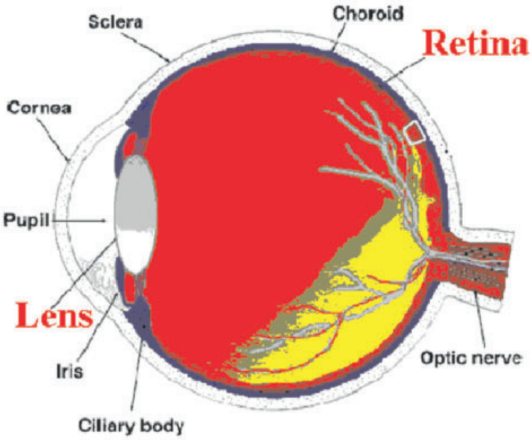


Figure 1. Eye and Retina.

The **figs 1 and 2** show you the general schemes of eye and retina. Light penetrate into the eye, goes through the retina, since the retina in the vertebrate and human eye is inverted, and absorbs partly (no more than about 15% of incoming light) by photoreceptor cells – rods and cones and the rest incoming light absorbs by retinal pigment epithelium cells that are placed behind the retina.

Ramon-y-Cajal
1852-1934

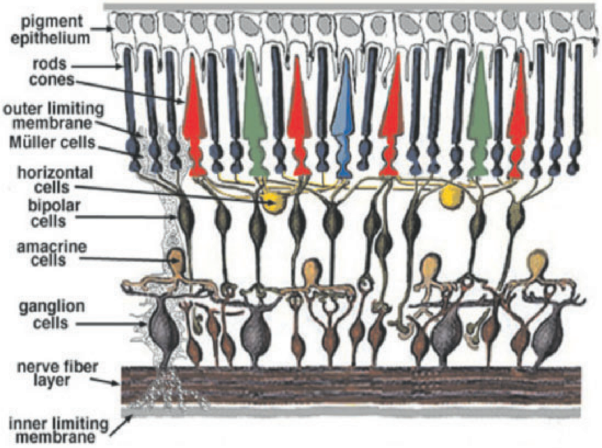
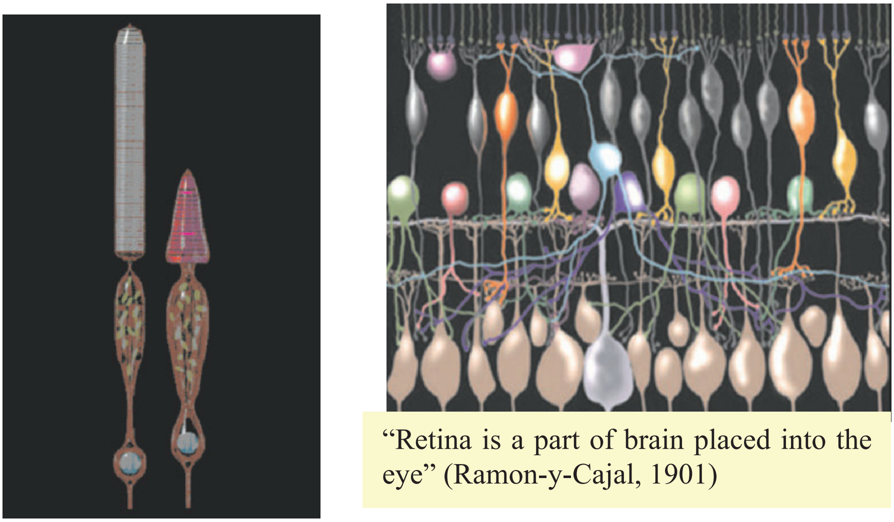


Figure 2. Simple diagram of the organization of the retina.

Usually, two parts of retina is distinguished – the layer of light sensitive photoreceptor cells and neural retina. The **fig. 3 (left)** shows you drawing of a rod and a cone. The rod- and cone-shape outer segment of both cells is the place where photons are absorbed and trigger the visual process. The **fig. 3 (right)** shows you drawing of the neuronal part of retina. As famous Spanish histologist Ramon-y-Cajal (his picture on the **Fig. 2)** on the very beginning of XX century, said: "Retina is a part of brain placed into the eye". This is absolutely true, and neurophysiology of retina is an essential part modern neuroscience. Also, retina is a very convenient object for electrophysiological studies not only in retina, but in brain neurophysiology as well. The retina is a layered structure that processes information in two stages. The outer synaptic or so called plexiform layer (the upper part of drawing of the neuronal part of retina) comprises the first stage and is where photoreceptors, bipolar cells, and horizontal cells interact synaptically. This is the synaptic layer where ON and OFF responses to light are formed, as well as the site where receptive field centre and surround organization is first thought to occur. The inner synaptic or plexiform layer (the middle part of drawing of the neuronal part of retina) is where the second stage of synaptic interactions occurs. This synaptic layer is where subsequent visual processing occurs that may contribute to the formation of transient responses, which may underlie motion and direction sensitivity. In addition, synaptic interactions in the inner synaptic layer may also contribute to the classical ganglion cells receptive field properties. The underside of the drawing represents the layer of ganglion cells, which long processes - axons leave eye and penetrate into brain.



"Retina is a part of brain placed into the eye" (Ramon-y-Cajal, 1901)

Figure 3. Drawing of Rod and Cone(left). Drawing of neuronal part of retina (right).

Now, let me focus on the rod and cone photoreceptor cells and light-sensitive visual pigment molecule situated in the outer segment of the cells.

The fig. 4 shows nice electron microscopy images both human rods and cones. One the left you can see the 3D images of the cells, please, pay your attention to the large rod outer segment. The suspension of these segments can be rather easy isolated and studied *in vitro*. One the right you can see the 2D electron microscopy images of the cells, and what is most important it is clear to see (upper right image) the inner structure of rod outer segment that contains hundreds photoreceptor flatted discs surrounding by plasma membrane of the cell. Each disc is build up with two membranes linked or locked on the edges; as a result there is a narrow cleft or space inside of disc - about 10A /angstrom/ – between the membranes.

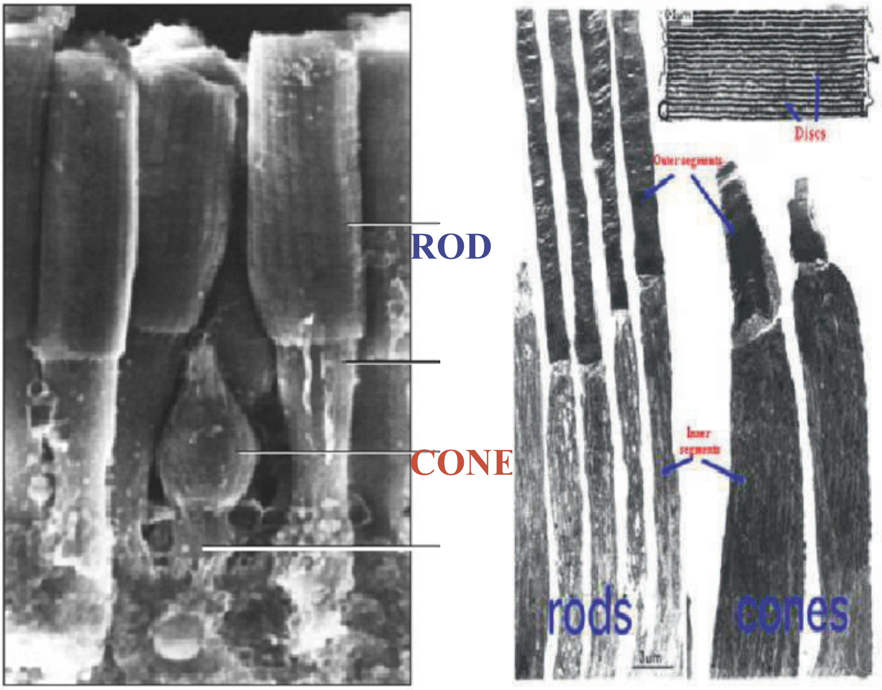


Figure 4. Rods and cones: electron microscopy.

Next fig. 5 (left) shows you a drawing of rod cell, disc and visual pigment rhodopsin – typical membrane protein incorporated into the membrane. Within the core of the protein part of rhodopsin (this part named as opsin) one can see the chromophore group that is covalently attached to the opsin. The chromophore group of all known visual pigments as vertebrates as invertebrates animals as well as human is 11-cis isomer of retinal$_1$ or retinal$_2$ that are aldehydes of vitamin A$_1$ or A$_2$. The 11-cis isomer is the only isomer that is served as a visual pigments chromophore group among sixteen possible isomers of retinal. Other words, there is very good conformation correspondence between 11-cis retinal and nearby protein (opsin) surrounding in the chromophore domain of rhodopsin. It is clear to see on the **fig. 5 (right),** where 11-cis retinal (red) is situated in the opsin (green). This picture is not drawing one, but the real result of recently published X-ray analysis 3D structure of crystallized rhodopsin molecule.

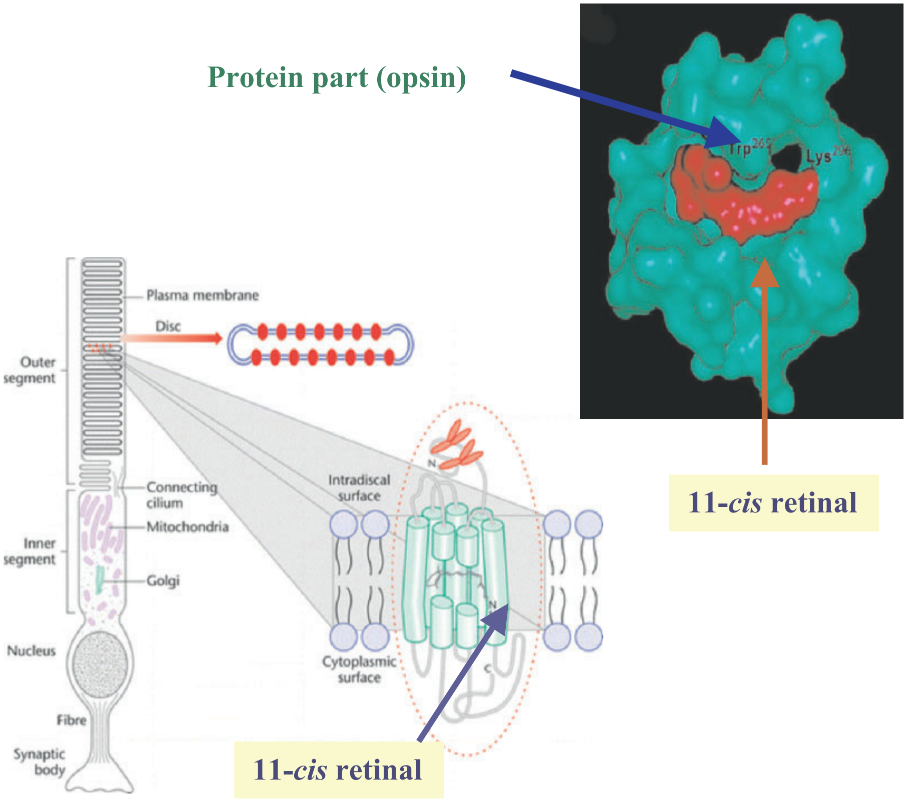
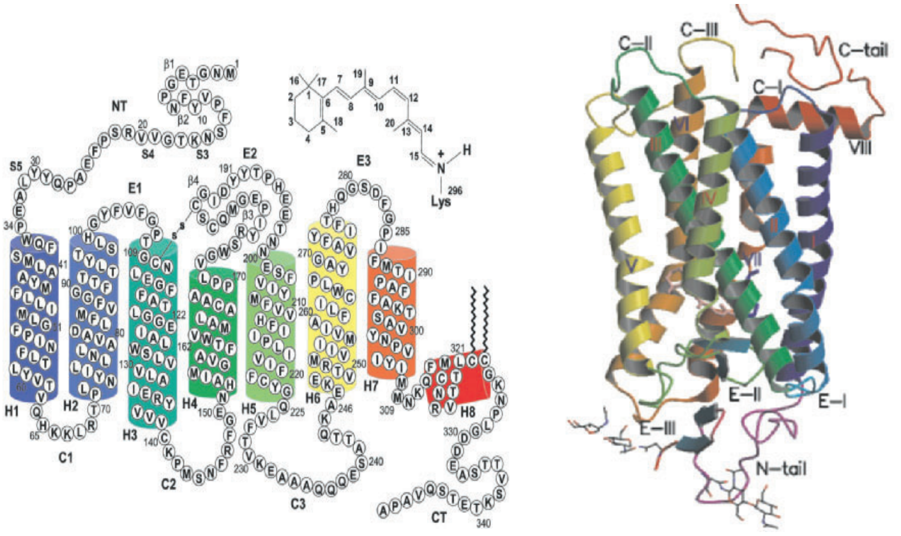


Figure 5. Rod, disc and visual pigment rodopsin (left) and Rhodopsin (upper right).

In reality, rhodopsin was the very first animal membrane protein which primary structure and topography in the membrane have been estimated. It was done almost simultaneously by two groups— in Russia by Yuri Ovchinnikov and his co-workers and in the US by Paul A. Hargrave co-workers.

As you can see on the **Fig. 6 (left)** rhodopsin is a heptahelical transmembrane protein where seven transmembrane alpha-helical parts ("helical bundles") linked by six extramembrane fragments ("loops"). In fact, rhodopsin is a protypical member of a large G-protein-coupled receptor (GPCR) family that plays a key role in all regulatory processes of living organisms. The signalling pathways regulated by these receptor proteins determine numerous crucial biological processes, including sensory reception, endocrine regulation and synaptic transmission. Approximately 5% of the human genome (above 600 genes) contains information about of these proteins. Over 40% of the currently used drugs aim at G-protein-coupled receptors as the main target. The visual pigment rhodopsin localized in the photoreceptor membrane of the photoreceptor cell rod outer segment is a perfect model for studies of structure and functions of G-protein-coupled receptors.



(Ovchinnikov et al., 1982) **(Palczewski et al., 2000)**

Figure 6. Rhodopsin - Two- and Three-Dimensional organization for rhodopsin in the disk membrane.

On the **Fig. 6 (right)** the 3D structure of rhodopsin is presented. It takes near 20 years to find way to crystallize membrane protein rhodopsin. Mainly postdoc Okada in Japan did it. After that Palczewski and co-workers in America carried out the X-ray analysis. The picture shows you the highly organized structure in the extracellular region forms a basis for the arrangement of the seven-helix transmembrane motif

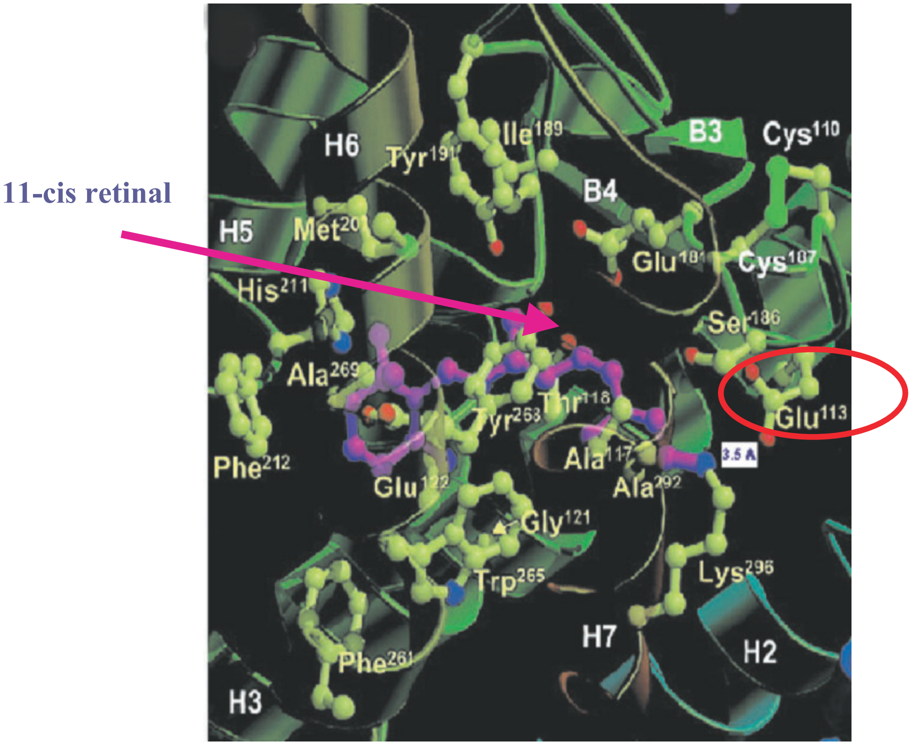
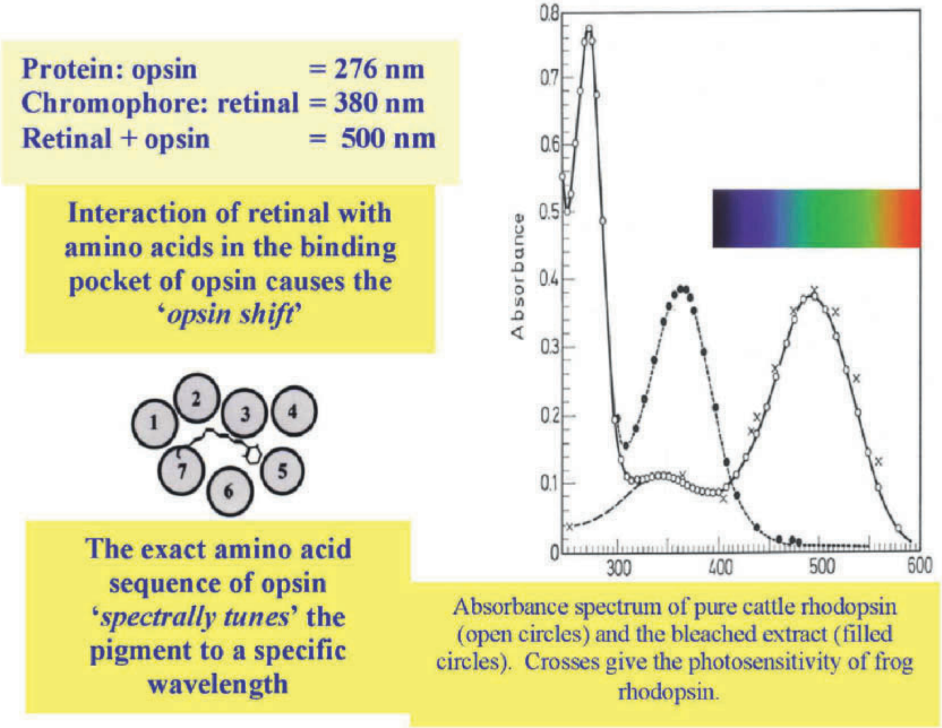


Figure 7. Chromophore site of rhodopsin (spectral tuning, photoisomerization).

Next **Fig. 7** represents the 3D structure of chromophore site of rhodopsin that is responsible for spectral tuning, 11-cis retinal photoisomerization and keeping rhodopsin as a G-protein-coupled receptor in the dark "silent" state. Other words, the ground-state chromophore, 11-cis-retinal, holds the transmembrane region of the protein in the inactive conformation. Interactions of the chromophore with a cluster of key residues determine the wavelength of the maximum absorption. Changes in these interactions among rhodopsins facilitate colour discrimination. Identification of a set of residues that mediate interactions between the transmembrane helices and the cytoplasmic surface, where G-protein activation occurs, also suggests a possible structural change upon photoactivation.

As far as of the wavelength of the maximum absorption and colour discrimination — this is the **Fig. 8** — it is well known that the interaction of retinal with amino acids in the binding pocket of opsin causes the so called “opsin shift”. It means that if opsin itself absorbs at near 280 nm and free retinal – at 380 nm, the chromoprotein rhodopsin absorbs at 500 nm.



Protein: opsin = 276 nm
Chromophore: retinal = 380 nm
Retinal + opsin = 500 nm

Interaction of retinal with amino acids in the binding pocket of opsin causes the ‘*opsin shift*’

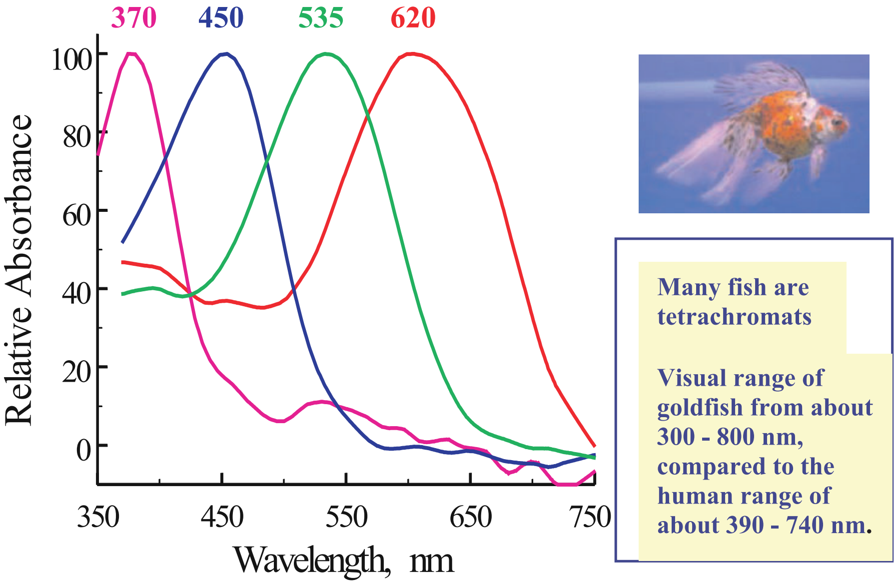
The exact amino acid sequence of opsin ‘*spectrally tunes*’ the pigment to a specific wavelength

Absorbance spectrum of pure cattle rhodopsin (open circles) and the bleached extract (filled circles). Crosses give the photosensitivity of frog rhodopsin.

Figure 8. Absorbance spectrum (spectral sensitivity) of rhodopsin – spectral tuning.

The broad diversity of the absorption maximum position of all known visual pigments in animals and human in all types of visual cells is determined by the exact amino acid sequence of opsin that “spectrally tunes” the pigment to a specific wavelength.

The goldfish four types cones visual pigments are the excellent example. As you can see on the **Fig. 9** the visual range of goldfish is occupied the near UV and all visible part of optical spectrum from about 300 - 800 nm. The visual range of human is narrower.



Many fish are tetrachromats

Visual range of goldfish from about 300 - 800 nm, compared to the human range of about 390 - 740 nm.

Figure 9. Goldfish, Carassius auratus: cone visual pigments - spectral tuning.

Before I will go to rhodopsin photochemistry, I should show you a photo of George Wald — **fig. 10 —** who was a great scientist in vision and awarded the Nobel Prize in physiology or medicine in 1967 for his monumental contributions to our understanding of the molecular basis of vision. He not only showed that the visual pigment molecules consist of a protein (termed opsin) to which is bound a derivative of vitamin A (vitamin A aldehyde, termed retinal). Wald and his co-workers, including his closest friend and co-workers - his wife Ruth Hubbard, discovered that only 11-cis retinal serves as chromophore for visual pigments and that the only photochemical reaction in vision is photo isomerization of 11-*cis* form of retinal to the all-*trans* form.

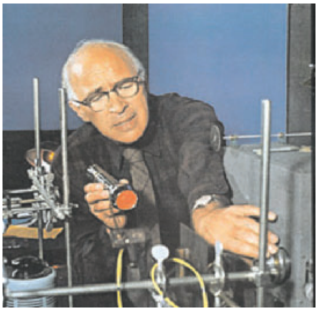


Figure 10. George Wald (1906- 1972).

Next **fig 11. (right part)** shows you the George Wald's scheme of the sequence of events that occurs following the absorption of a quantum of light by the rod visual pigment, rhodopsin that I took from the obituary of his pupil and also very famous scientist John E. Dowling.

The scheme shows how light initiates the conversion of rhodopsin to retinal and opsin through a series of metarhodopsin intermediates. Metarhodopsin II is the active intermediate leading to excitation of the photoreceptor cell. Eventually, the chromophore of rhodopsin, retinal, separates from the protein opsin and is reduced to vitamin A (retinol). For the resynthesis or regeneration of rhodopsin, the vitamin A must be isomerized from the all-*trans* to the 11-*cis* form, and this isomerization takes place in the pigment epithelium overlying the receptors.

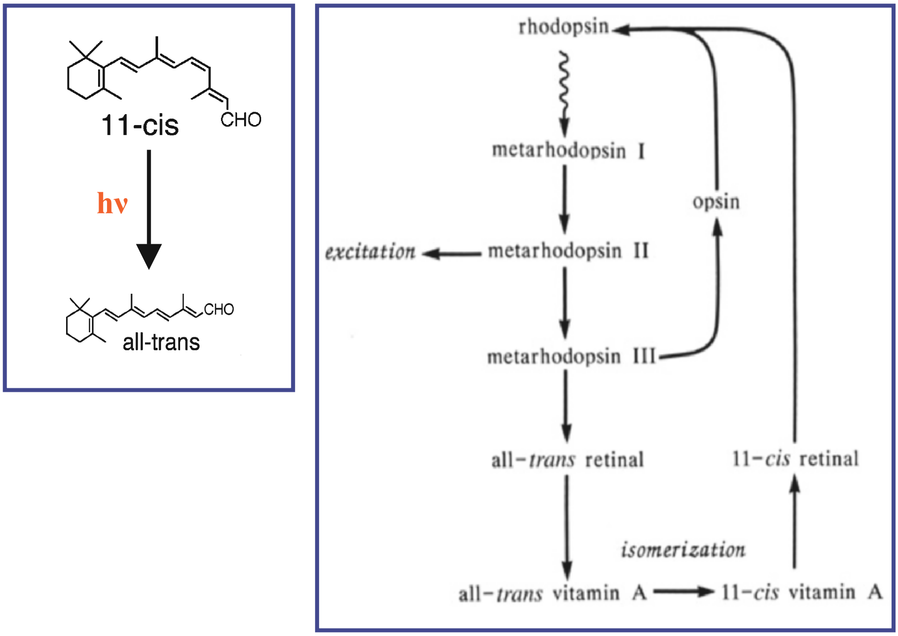


Figure 11.

The fig.11 (left part) shows the photochemical event itself, namely how absorbed photon change the isomeric configuration of 11-*cis* retinal to its to all-*trans* form. Now days it is known that to be in the chromophore site of rhodopsin the 11-*cis*-retinal, on the contrary to free retinal state, exists in a twisted nonplanar state and that the β-ionone ring is additionally turned by about 65° relative to the polyene chain. It is also believed that the twisting of the polyene chain plays an important role in retinal isomerization upon absorbing a photon.

Now, let me show you the modern scheme of rhodopsin photochemistry or rhodopsin photolysis —**fig. 12.**

I have to stress on this picture three key events of this photocycle.

Number one event is photochemical one, namely photoisomerization of the 11-*cis* retinal chromophore to its distorted *trans* form. This is as I said the only light-dependent event in both vertebrate and invertebrate vision. Photoisomerization occurs on an ultrafast time scale with photorhodopsin as the photoproduct formed on a femtosecond time scale. Important photochemical properties of 11-*cis* retinal chromophore is a very high quantum efficiency (~0.67 versus ~0.20 for free retinal in solution) and an extremely low rate of thermal isomerization.

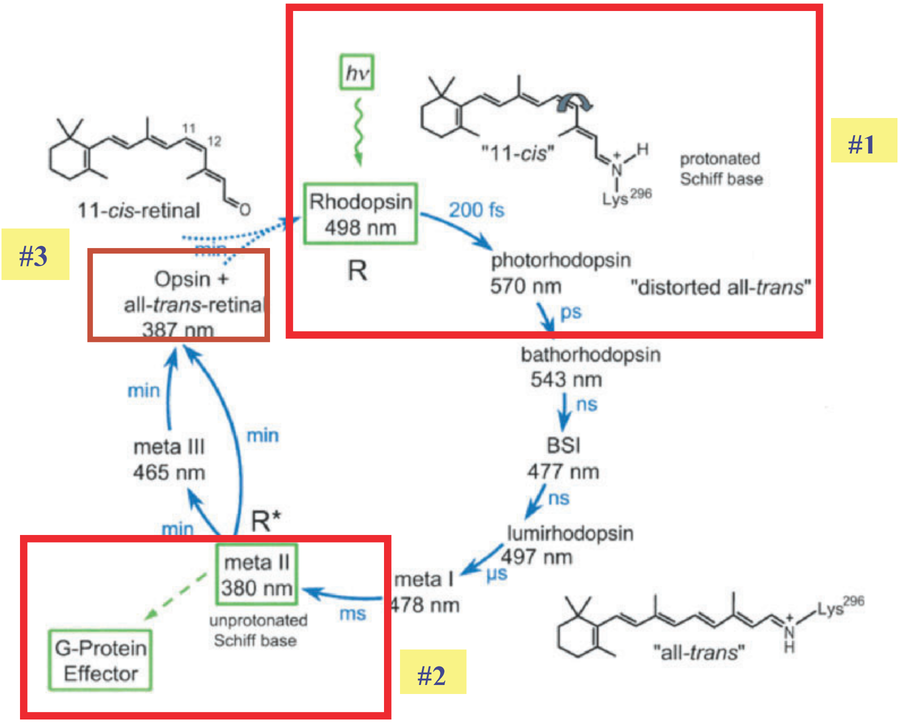


Figure 12. Rhodopsin photochemistry (photolysis).

After photoisomerization the rhodopsin molecule proceeds through a number of well-characterized spectral intermediates. Because the protein-chromophore interactions change after chromophore photoisomerization the distinct λ_{max} values of the intermediates are observed. Metarhodopsin-I is in a dynamic equilibrium under physiological conditions with metarhodopsin II (meta-II), which is characterized by an unprotonated Schiff base imine and a dramatically blue-shifted λ_{max} value (380 nm).

The number two key event is appearance of metarhodopsin II. Meta-II or so called physiologically activated rhodopsin is the photoproduct that is able to bind and activate G-protein. G-protein is named transducin in case of visual phototransduction. Meta-II eventually decays to free all-*trans* retinal and opsin apoprotein. In the course of Meta-I to Meta-II transition the retinal Schiff base is hydrolyzed and all-*trans*-retinal is released, and then is reduced to all-*trans*-retinol in the rod outer segment.

So, the formation of free all-*trans* retinal and opsin in the rhodopsin photocycle I can mark as the number three key event. This process is rather slow one, and it occurs after the phototransduction process in the rod outer segment is completed. What I should stress regarding this third key event is: the appearance of free all-*trans* retinal in the disk photoreceptor membrane creates a hazard of light damage as to the rod cell as to the retinal pigment epithelium cell. I am going to consider this subject that I call as a photobiological paradox of vision a little bit later.

But now let me stop briefly on the very first photochemical event of rhodopsin photocycle.

Visual phototransduction, the conversion of incoming light to an electrical signal, takes place in the outer segments of the rod and cone photoreceptor cells.

Light reduces the concentration of cGMP, which, in darkness, keeps open cationic channels present in the plasma membrane of the outer segment. Ca2+ plays an important role in phototransduction by modulating the cGMP-gated channels as well as cGMP synthesis and breakdown. Ca2+ is involved in a negative feedback that is essential for photoreceptor adaptation to background illumination.

Visual phototransduction represents one of the best-characterized G-protein-coupled receptor-mediated signaling pathways. Structural analyses of visual pigment rhodopsin, G protein, and several other phototransduction components have revealed common folds and motifs that are important for function.

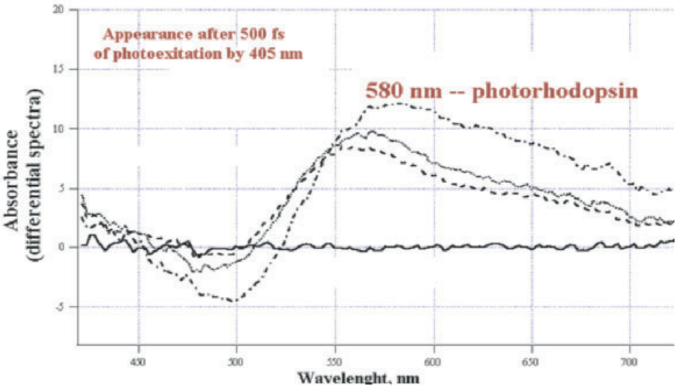
Given the accelerated pace of structure and functional determination, it is anticipated that visual phototransduction will be the first G-protein-coupled pathway for which a complete molecular description is ultimately available.

Because light in vision is not only a carrier of information, but also a risk factor of damage to the eye structures, the complex photoprotective system, along with physiological system of visual reception, has been developed in the course of evolution. The sophisticated photoprotective system is able to solve the photobiological paradox of vision. The impairment of the system can lead to human retina diseases or play role in progression of eye diseases like age-related macula degeneration.

Molecular mechanisms of potential light hazard to the eye structures, first of all to the retina, and main components of natural photoprotective system (permanent renewal of rod and cone outer segment, antioxidants and optical media of the eye) will be considered.

Now let me stop briefly on the very first photochemical event of rhodopsin photocycle.As it is marked on the **Fig. 13** rhodopsin is characterized by an unusually high rate of chromophore isomerization (faster than 200 femtoseconds). The quantum yield of this photoreaction is also high, namely 0.67 or in accordance with recently made estimations – 0.65, but not less. It should be stressed that the rate of 11-*cis*-retinal isomerization as a rhodopsin chromophore group is about two orders of magnitude higher than in solution. The crucial role of the protein environment that produces this enormously high rate of isomerization as well as high quantum yield is doubtless. However, the underlying molecular mechanisms have not yet been studied in full. The very first product rhodopsin photocycle in the ground state is photorhodopsin. There are several laboratories, including our own lab, where this very first photochemical event is careful studied. The Fig. 13 shows you the typical experimental results we have obtained recently in Moscow in collaboration with the laser physicists from Institute of Chemical Physics of the Russian Academy of Sciences. Using modern femtoseconds spectroscopy technique we were able to record the appearance of photorhodopsin. On the upper part of the picture you can see the differential spectrum of the product that has absorption maximum around 580 nm and that is recorded after 500 fs after rhodopsin solution photoexcitation by 70 fs laser flash. The wavelength of the flash in our case is 405 nm, namely in the blue part of alpha-band of rhodopsin absorption spectrum. In other words, the 405 nm flash is absorbed by rhodopsin chromophore group 11-*cis* retinal itself.

Absorption spectra of primary products of rhodopsin photolysis at increasing delay of probing pulse. The very first product is photorhodopsin with 580 nm absorption maximum.



Kinetics of optical absorbance increasing at different wavelength:
500, 540 and 580 nm

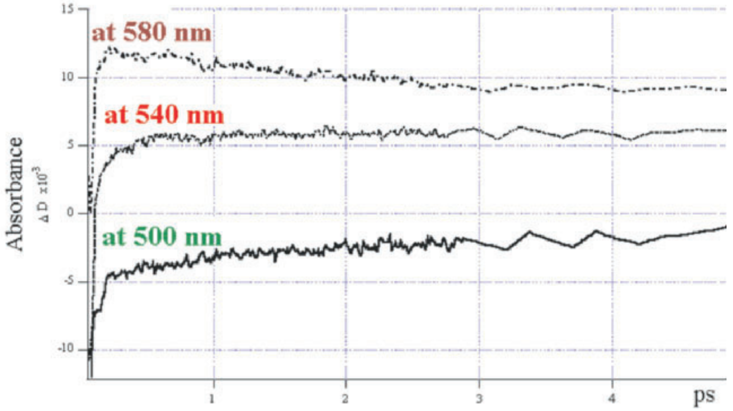


Figure 13. Femtoseconds spectroscopy of rhodopsin (Ostrovsky et al., submitted).

Then, the photorhodopsin due to following conformation changes within the chromophore site of rhodopsin molecule turns in the absence of light into the next product, so called batorhodopsin. The absorption maximum of batorhodopsin is shifted to the more short-wavelength part of visible spectrum in comparison with photorhodopsin. On the lower part of the picture three kinetics curves of the rise of optical absorbance at different wavelength is presented. It is clear to see that the absorption maximum at 580 nm that is the absorption maximum of photorhodopsin, reaches its highest value during about 200 - 250 fs. The physical importance of such high rate of rhodopsin chromophore group photoisomerization is clear: the energy of absorbed photon should be used as much as possible for this useful photochemical reaction that triggers the visual process, and not to be loosed via other concurrent reactions like worming up or fluorescence. The same situation takes place in case of other crucial photobiological process — photosynthesis, where the biological useful charge separation also occurs in the femtoseconds time scale.

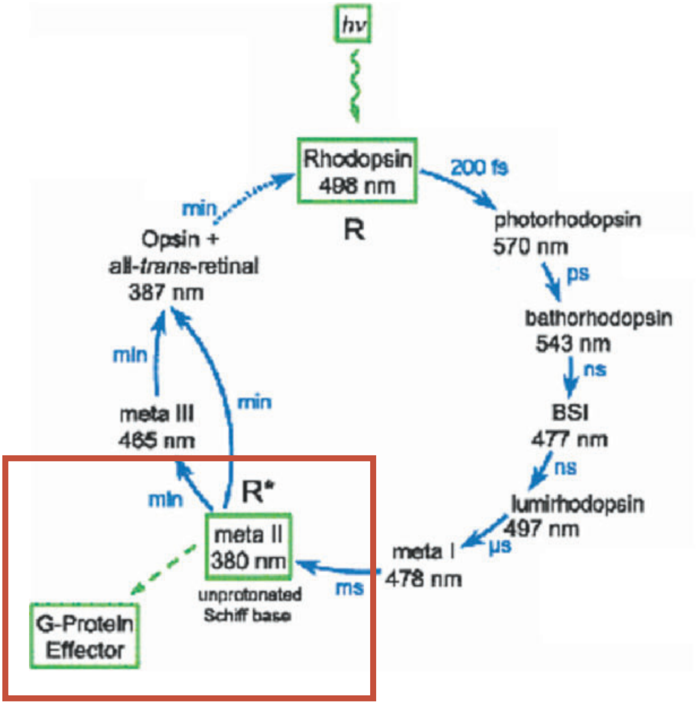


Figure 14a. Site-directed spin labeling of Cys140, 316 and pair Cys 322, 323 and electron paramagnetic resonance spectroscopy (Ostrovsky et al. 1985).

The second biologically most important event in the rhodopsin photocycle is appearance of the transient product, called metarhodopsin II. This product and this stage of photocycle are presented on the left part of **Fig. 14a.** A little bit later I will consider the visual pigment rhodopsin is a prototypical member of a large G-protein-coupled receptor family in more detail. The only I should say now is: these membrane protein receptors play a key role in very many regulatory processes of living organisms. But now let me stress that a unique feature of rhodopsin as a G-protein-coupled receptor is that its incorporated ligand, 11-*cis*-retinal chromophore, acts as a highly efficient antagonist that enables to maintain rhodopsin in its dark state inactive. This is extremely important for physiology of visual process to keep rhodopsin salient, to prevent its spontaneous interaction with G-protein in the dark adaptation state, in order to diminish as much a possible dark noise of photoreceptor cell. As a result of photon absorption and 11-*cis* to all-*trans* retinal isomerization, the all-*trans*-retinal turns into a highly-efficient agonist. On the stage of metarhodopsin II all-*trans*-retinal is still covalently bound up with the protein part of rhodopsin molecule — opsin. In spite of rhodopsin in its dark state, the metarhodopsin II, which general conformation state is now rather different from the initial dark rhodopsin conformation state, is capable to interact very efficiently with G-protein ; in case of vision (phototransduction) G-protein named transducin.

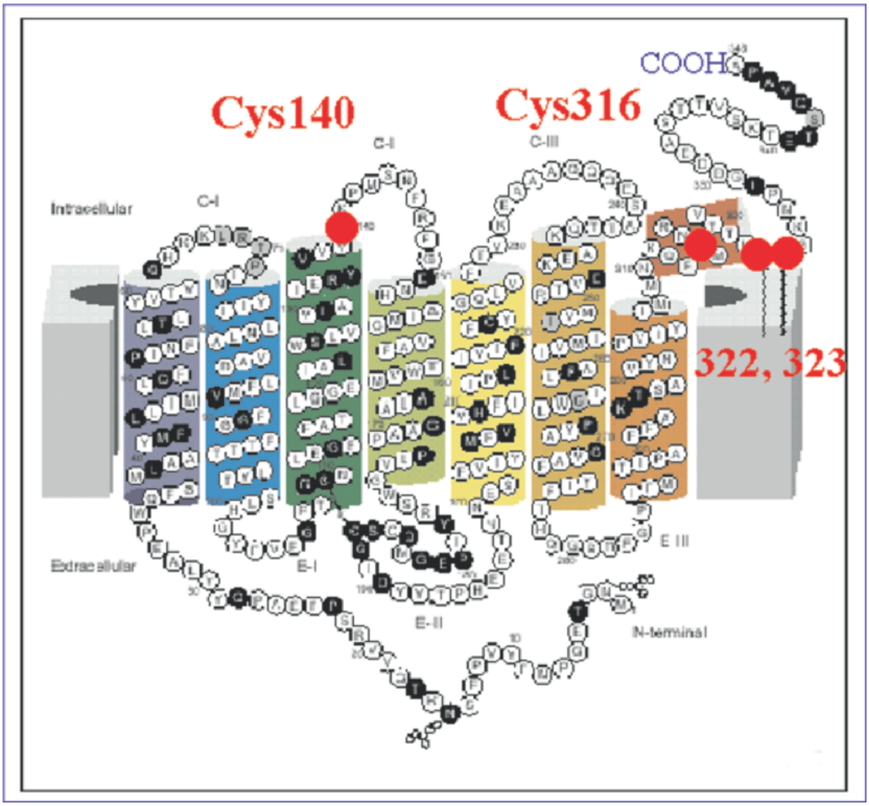


Figure 14b.

At present there are numerous experimental data obtained by different experimental techniques that have clear shown in all details the conformation changes of rhodopsin molecule that lead to appearance of metarhodopsin I and then metarhodopsin II, which can bind G-protein and activate it. One of the powerful approach is site-directed spin labelling of rhodopsin and by means of electron paramagnetic resonance spectroscopy studying its light-induced conformation changes. We were one of the first who have applied this technique and demonstrated how the binding sites of opsin can be open on the stage of metarhodopsin II, in order to be prepared to bind and interact with G-protein. On **fig. 14b** you can see the two-dimension model of rhodopsin molecule in the membrane and the sites, namely Cys140, 316 and pair Cys 322, 323 situated in the hydrophilic cytoplasm part of molecule, where the spin labels were attached.

Figure 15. Light induced mobility of spin labels attached to SH-groups. Effects of transducinadding. (Ostrovsky et al.,1985. Pogozheva et al., 1989).

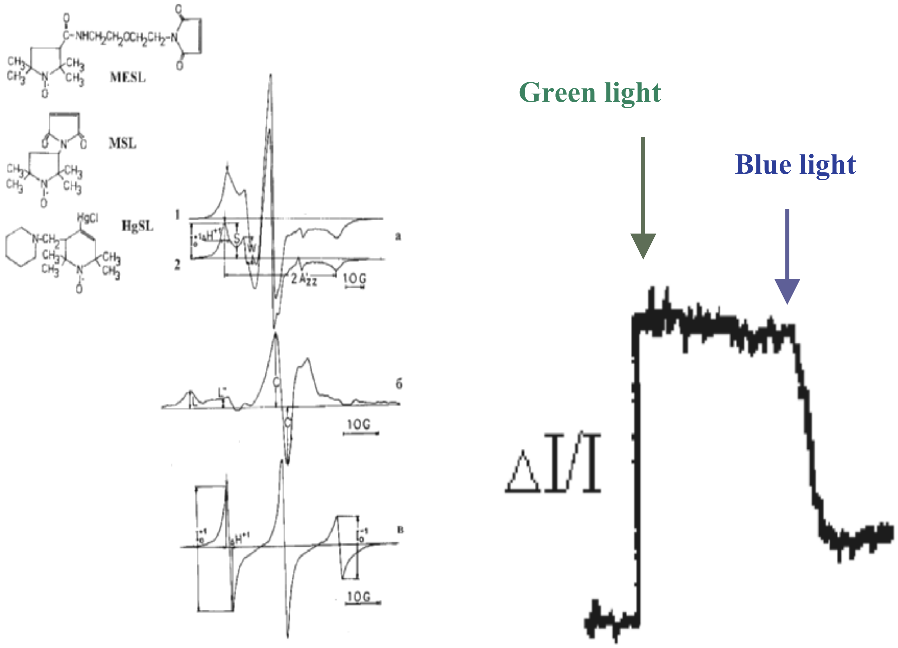


Figure 15a. Spin labels and EPR spectra.

Figure 15b. Increase of light-induced mobility of spin label attached to Cys140.

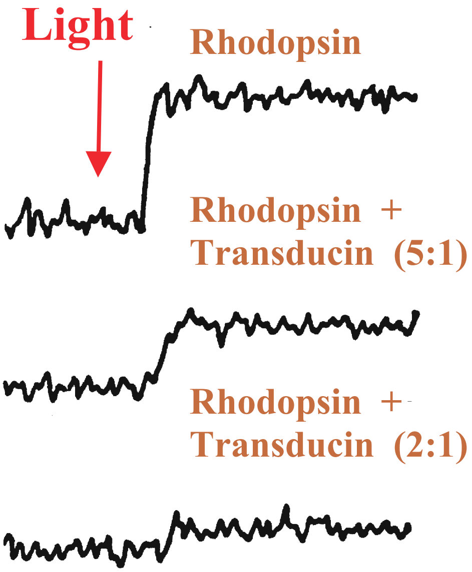
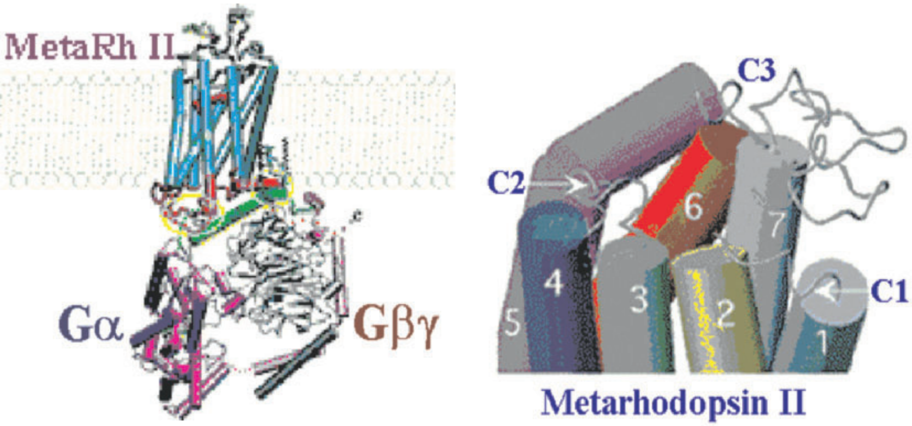


Figure 15c. Decrease of light-induced mobility of Cys140 spin label after transducin adding.

On the next **Fig. 15** the main results are presented. On figure 15a the structure of spin labels molecules and electron paramagnetic resonance (EPR) spectra are presented. On figure 15b you can see the main phenomenon: green light that is absorbed by rhodopsin induces increase of spin label mobility that attached to SH-group (in this case to Cys140), and blue light that is absorbed mainly by metarhodopsin II leads to decrease of spin label mobility. It means that the conformation of photoregenerated product, first of all so called metarhodopsin III, returns, at least partly, to initial state. On figure 15c it is clear to see that adding of transducin to the photoreceptor membrane suspension decrease or completely prevents the light-induced mobility of spin label attached to Cys140. This is one of the first results demonstrated how light is able to open the hydrophilic loops of rhodopsin in the membrane during its transition to metarhodopsin II.

The next slide **(Fig. 16)** shows you the modern models based on the plentiful data obtained by different techniques. On the right three-dimension model of binding sites of metarhodopsin II is presented: it is clear to see that the hydrophilic loops C2 and C3, and nonloop cytoplasmic regions C1 are accessible to interaction with G-protein, namely with its alpha-subunit. On the right side the drawing of alpha- and beta-gamma G-protein subunits binding to cytoplasmic surface of metarhodopsin II is presented. You can see that the area of metarhodopsin II to bind both alpha- and beta-gamma G-protein subunits is not enough. That is why a new model of rhodopsin molecules topography within the photoreceptor membrane as dimer, but not monomer, has been recently proposed. There are now several publication regarding the functional crosstalk between G protein-coupled receptors, first of all rhodopsin itself, in a homo- or heterodimeric assembly however the nature of this interface is not yet established

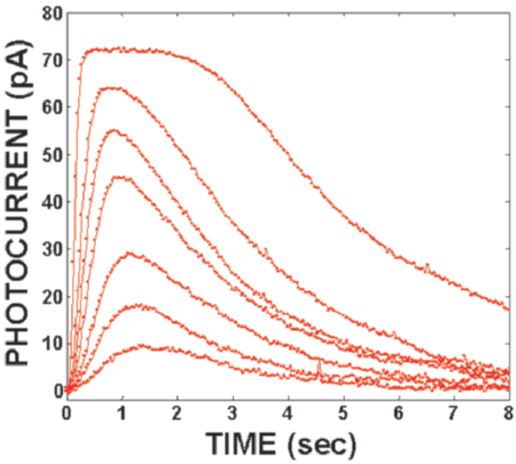


The loops C2 and C3, and nonloop cytoplasmic regions C1 are involved on the metarhodopsin II stage in the specific interaction with alpha-subunit of G-protein.

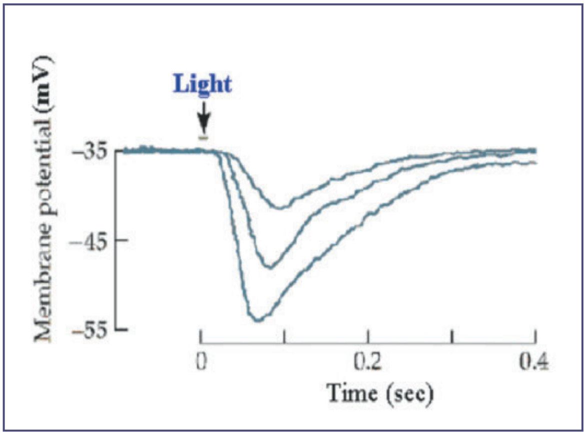
Figure 16. Three-dimension model of binding sites of R* (Metarhodopsin II) with the α-subunite of the G-protein (Choi et al, 2002).

Thus, the visual pigment rhodopsin is a seven trans membrane protein and is from the family of G protein coupled receptors (GPCRs). These trans membrane proteins switch from an inactive to an active form on the binding of a ligand. The activated receptor can trigger an intracellular signal cascade, in case of photoreceptor cell called phototranduction.

Visual phototransduction, the conversion of incoming light to an electrical signal, takes place in the outer segments of the rod and cone photoreceptor cells. This process represents one of the best-characterized G-protein-coupled receptor-mediated signalling pathways. Structural analyses of visual pigment rhodopsin, G protein, and several other phototransduction components have revealed common folds and motifs that are important for function process, by which light, captured by a visual pigment molecule, generates a physiological response, namely an electrical signal called photoreceptor potential.



Photocurrent (decrease of "dark current")



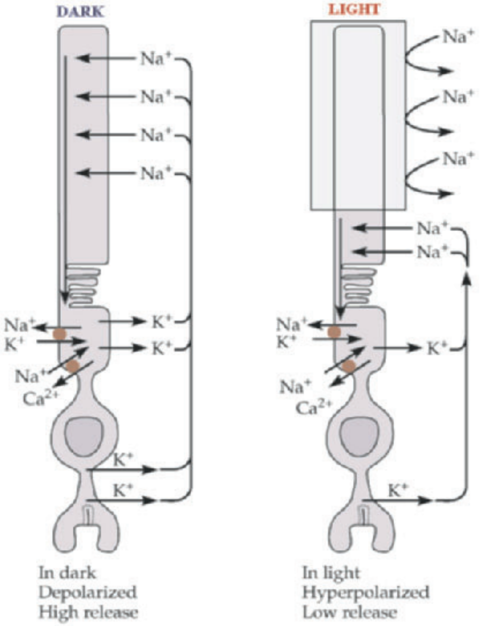
Hyperpolarization

Figure 17 a.

On **Fig. 17a** the final electrophysiological step of complex phototransduction process is presented.

It was rather surprising discovery made in the second part of 60th actually at the same time by Japanese neurophysiologist Tsuneo Tomita and Russian neurophysiologist Yuri Trifomov that vertebrate photoreceptor cell, in contrast to in vertebrate one, generates not depolarisation, but hyperpolarisation electrical response to light stimulus. On figure 17a, the lower part, you can see three intracellular recorded potentials in response to three light stimuli with increasing intensity. Later Dennis Baylor from

Stanford University and Trevor Lamb from Cambridge (UK) have elaborated the technique using suction pipette as an electrode to measure electrical current through the rod and cone outer segment plasmatic membrane. On the left side above you can see series of photocurrents responses recorded for 20-ms light flashes with increasing intensities — from dim-flash to saturating levels. In fact, the photocurrent response reflects the diminishing of the constant dark current. It is easy to see not only rise of responses amplitude, but also the progression of time-to-peak from ~1.5 s to ~0.6 s as it is increased from very dim-flash to saturating one.

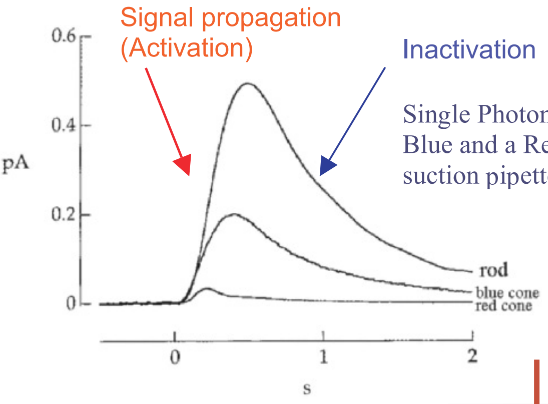


Circulating current (“dark current”) in a Rod

Figure 17b.

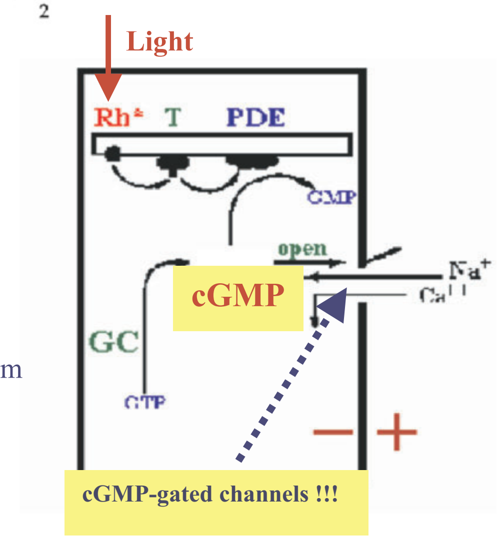
On figure 17b the circulating current (“dark current”) in a rod as a very simple scheme is depicted. In the dark sodium ions leak into the outer segments of the rods. This produces depolarizing potential that keep calcium channels open in their synaptic endings. This results in continuous neurotransmitter release, namely glutamate, by the rods at their synapses. When light triggers rhodopsin breakdown the sodium permeability of the outer segment membrane decreases rapidly. The effect of light turns off the sodium entry causing the rod cells to develop a hyperpolarizing receptor potential that inhibits their release of neurotransmitter. The inhibition or even prevention of neurotransmitter release serves as a signal to the second order neurons in the retina – bipolar and horizontal cells that visual pigment molecules absorbed light.

The next **Fig. 18a** shows you single photon photocurrents responses in a rod, and in a blue and red cones recorded with suction pipette.



Single Photon Responses in a Rod, and in a Blue and a Red Cones recorded with suction pipette (Mendez et al., 2000)

Figure 18a.



The simplest schematic diagram of phototransduction

Figure 18b.

It is clear to see that the phototransduction amplification cascade is enough powerful in rod, in order to govern the rate of neurotransmitter release. At the same time, amplification cascade in cones need much more than one absorbed photon to initiate the phototransduction process. That is why, as it is well and long time known, the cone diurnal vision is about tow orders less sensitive than rod scotopic vision.

What is also should be stressed is: the rising, going up, upward phase of the electrical response is the result of signal propagation or activation steps of amplification cascade, and descending, downward phase is the consequence of inactivation reactions of the cascade.The simplest schematic diagram of phototransduction process is presented on the **Fig. 18b.** Photoactivated rhodopsin (R*), namely metarhodopsin II, activates in the photoreceptor disk membrane the second protein — GTP-binding protein transducuin (T). A single active rhodopsin R is able to cause hundreds of transducins to exchange bound GDP for GTP, forming active transducin-GTP complexes. A greatly amplified signal now passes to a third protein, cGMP phosphodiesterase (PDE, phosphodiesterase), which is activated by transducin-GTP complexes. Before phosphodiesterase activation the cGMP kept in the dark the cationic channels in the cell membrane in the open state. Activated PDE hydrolyses phosphodiesterase to 5$'$-GMP, which now cannot open the cationic channels in the cell membrane. So, with cGMP removed, channels close, interrupting a steady inward currents of Na$^+$ and Ca$^+$, thus hyperpolarizing the cell. The next step is inactivation of cascade. We will consider both activation and inactivation steps a little bit later in more details. Now, we only should note that, finally, the cGMP concentration is restored to the dark level by cGMP synthesis, which is mediated by guanylat cyclase (GC).

Fig. 19 shows the rod phototransduction cascade in more details.
The picture show: the activation steps **(A**) those are responsible, as we said before, for the upward phase of the electrical response, and **B&C** that are responsible for its downward phase.

(A) As it clear to see, following absorption of a photon (hν), the activated rhodopsin (R* that is metarhodopsin II) repeatedly contacts molecules of the heterotrimeric G-protein, catalyzing the exchange of GDP for GTP, producing the active form G*α (=Gα–GTP). Two G*αsubunits bind to the two inhibitory γ- subunits of the phosphodiesterase (E), thereby activating the corresponding β and αcatalytic subunits, forming E* which then catalysis the hydrolysis of cGMP. The consequent reduction in the cytoplasmic concentration of cGMP leads to closure of the cyclic nucleotide gated channels, and blockage of the inward flux of Na$^+$ and Ca^{2+} (thereby reducing the circulating electrical current). The exchanger continues to pump Ca^{2+} out, so that the cytoplasmic Ca^{2+} concentration declines, activating three ‘calcium feedback’ mechanisms illustrated in this panel and in **(B).**

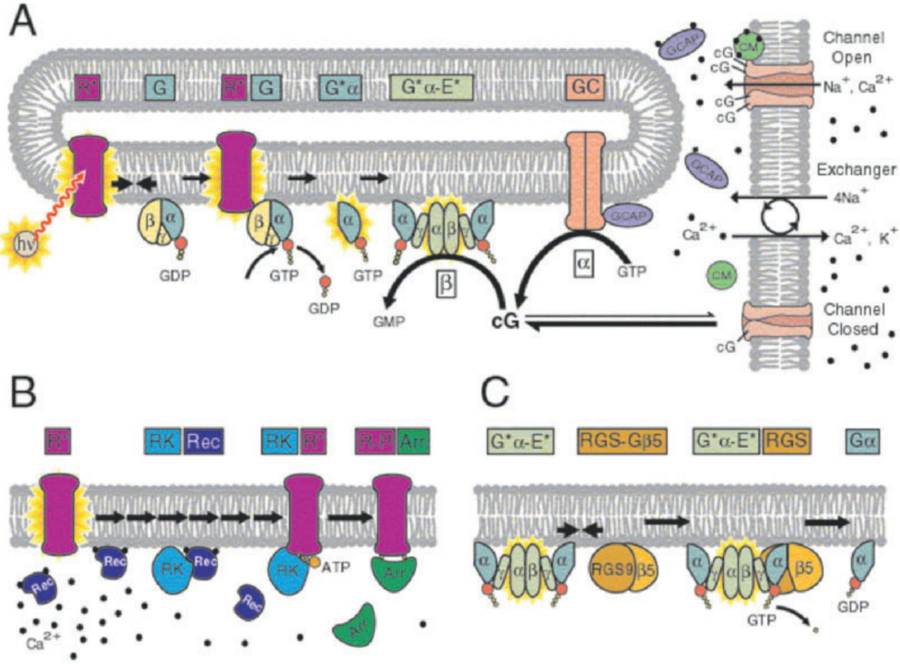


Figure 19. Schematic of the rod phototransduction cascade (from:http://dolphin.upenn.edu/~pugh/papers/PNL_99.html).

As far as ‘calcium feedback’ mechanisms I can say that loss of Ca^{2+} from guanylyl cyclase activating proteins (GCAPs) allows them to bind to a cytoplasmic domain of the guanylyl cyclase (GC), increasing its activity. Also, loss of Ca^{2+} from calmodulin (CM) causes it to dissociate from the channels, lowering the $K_{1/2}$ of the channel for cGMP.

(B) Now let we consider the first step of R* inactivation.

At the dark concentration of [Ca^{2+}]$_i$ (left side of diagram), most of the recoverin (Rec) that is specific Ca^{2+}- binding protein is in the Ca^{2+}-bound form at the membrane. In the dark Rec–2Ca forms a complex with rhodopsin kinase (RK), blocking its activity. Rhodopsin kinase is also a specific protein that is able to phosphoralate R*, if it is in the free state. Other words, at resting Ca^{2+} levels contact between R* and uncomplexed RK occurs infrequently. When [Ca^{2+}]$_i$ drops during the light response (arrows indicate progression of time), Ca^{2+} dissociates from Rec, which moves into solution. The concentration of free RK rapidly increases, increasing the frequency of interaction with R*, and leading to its rapid phosphorylation. Arrestin (Arr) that is the second protein involved in the R* inactivation then binds to the phosphoralated R*, substantially quenching the R* activity.

(C) The second step of phototransduction cascade inactivation is inactivation of G*–E* complex. G* or (Gα–GTP) is inactivated when the terminal phosphate of its bound GTP is hydrolyzed. Although the G protein has intrinsic GTPase activity, this capacity is only activated when the G* is bound to the γ-subunit of PDE (PDEγ) and when, in addition, the GTPase activating protein or GAP factor (RGS9–Gβ5) also binds. The resulting tetra-molecular complex, G*α –PDEγ–RGS9–Gβ5, rapidly hydrolyzes the GTP to GDP, returning the Gα-subunit to its inactive form, so that the E* and G*α are simultaneously inactivated.

As you can see, the molecular mechanism of cascade inactivation is even more complex that activation. Also, it should repeated that these mechanisms of cascade activation and inactivation are responsible for the upward and descending, downward phases of rod cell electrical response.

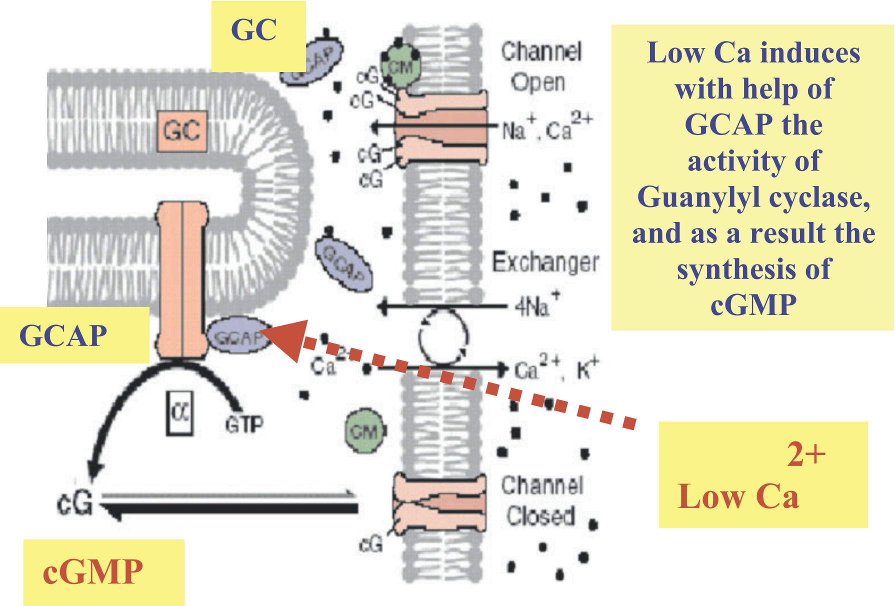


Figure 20. Guanylyl cyclase and synthesis of cGMP.

Figure 20 shows you a piece of the scheme Fig. 19A, in order to pay you attention to the final step of cascade activation and the very first step of cascade inactivation. The key protein of the final step of cascade is the cyclic nucleotide-gated (CNG) channel. These channels play a central role in the conversion of sensory information, such as light and odour, into primary receptor electrical signals. The cGMP-gated channel of photoreceptor outer segments is inserted solely into the plasma membrane. The channel is constantly kept open by cGMP und closes as cGMP phosphodiesterase hydrolyzes cGMP and reduces its concentration. The channel is nonselective for monovalent cations and discriminates bivalent cations for Ca^{2+}. This channel consists of a complex of two distinct subunits and one or more associated proteins. Calmodulin that is specific calcium–binding protein binds to the channel complex and modulates the sensitivity of the channel for cGMP in a Ca2+ dependent manner. The phenomenon of direct, reversible binding of cGMP to the channels was discovered in the middle of 80th simultaneously in Russia by Fensenko et al. and in the US by Yau and Nakatani. This allows for a steady influx of Na+ and Ca2+ into the outer segment and maintains the cell in a partially depolarized state.

The influx of Ca2+ through the channel is balanced by an efflux of Ca2+ by the Na+/Ca2+-K+ exchanger in the ROS plasma membrane, thereby maintaining the intracellular level of Ca2+ at a relatively constant level. Also, low intracellular Na+ concentration is maintained by the balanced extrusion of Na+ by Na+-K+ ATPase localized in the plasma membrane of the rod inner segment.

The light-induced closure of the channels to Ca2+ results in a reduction in intracellular Ca2+ since the Na+/Ca2+-K+ exchanger continues to extrude Ca2+ from the outer segment. The rod Na-Ca exchanger utilizes both the inward sodium gradient and the outward potassium gradient to drive calcium extrusion.

The reduction in Ca2+ is a key event in the recovery of the rod outer segment to its dark state and in light adaptation. This feedback mechanism is suggested to occur through the interaction of Ca2+- binding proteins with various proteins involved in the phototransduction process, and in particular, guanylate cyclase (GC). Indeed, low Ca induces with help of another Ca-binding protein – GCAP the activity of guanylyl cyclase, and as a result the synthesis of cGMP

Figure 21 shows you a possible scheme of timing of inactivation events. According to the scheme the timing of shut-off and recovery events for single photon response is: GC<RK (rhodopsin kinase) < Arr (arrestin) and GTPase. Generally speaking, increasing of guanylyl cyclase activity by low Ca as a result of cGMP-channels closing and cGMP synthesis is the first, earliest event in the mechanism of cascade inactivation.

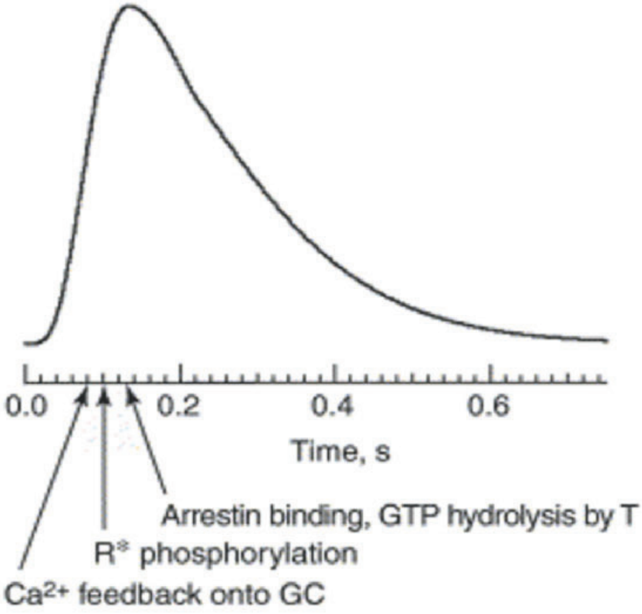


Figure 21. Timing of inactivation events.

Let me now to consider the last topic of my lecture, namely the so-called photobiological paradox of vision. The essence of the paradox is: light in vision is not only a carrier of information, but also a risk factor of damage to the eye structures. The most dangerous hazard of light damage is related to the eye retina, namely to the rod cells, and to the retinal pigment epithelium cells (RPE).

Why light is a risk factor to these cells? Because these cells are open t light and available to absorb light, and because they contain all three factors (components) that are necessary to initiate and develop the destructive free-radical photooxidation reaction. Firstly, these cells are full of coloured substances that are able to absorb light and act as photosensitizes. Secondly, they contain easy oxidized substrates of oxidation like lipids with polyunsaturated fatty acids and SH-groups contained proteins like rhodopsin. And finally, there is a high tension of oxygen in these cells. It should be stressed that all these compounds are crucial necessary for normal, physiological process of photoreception. That is why the complex protective system that is able to prevent light hazard to the retina and RPE has been developed in the course of evolution. The photoprotective system of the eye — as vertebrate as invertebrate animals — includes several lines of protection. The most obvious and important defence line is the permanent renewal of rod and cone outer segments. Two other lines are the powerful set of antioxidants, and optical media as cut-off filters. The lens plays a crucial role as a cut-off filter.

Both the photobiological system and also complex system of visual reception, as I said before, have been developed simultaneously in the course of evolution. The sophisticated photoprotective system is able to solve the unavoidable photobiological paradox that is appropriated to any photobiological processes like vision or photosynthesis. However, the impairment of the photoprotective system can lead to human retina diseases or play role in progression of eye diseases like age-related macula degeneration.

To understand the mechanism of the hazard of light damage let me now pay your attention on the so-called visual cycle. This cycle is a normal, physiological part of visual system goings-on. At the same time it is a potentially source of light hazard the retina and retinal pigment epithelium.

As it was clear from **Fig. 12**, the third key step in rhodopsin photolysis is hydrolysis of covalent link between isomerised all-*trans*-retinal and amino acid of opsin — Lys296, to which the chromophore is attached. In fact, the released free all-*trans*-retinal can be, if not processed properly, a cause of light damage.

Figure 22 shows you the scheme of visual cycle. As it was shown, rhodopsin in the rod outer segment undergoes photoactivation after photon absorption and finally the all-*trans*-retinal is released. To regenerate rhodopsin and maintain normal visual sensitivity, the all-*trans* isomer must be metabolised and reisomerized in the retinal pigment epithelium to produce the chromophore 11-cis-retinal again. I am not going to consider the rather complex biochemical steps that constitute the visual cycle and that involve both photoreceptor cells and retinal pigment epithelium. The only I should stress now is: a key step in the visual cycle is isomerization of all-*trans* retinol to 11-*cis*-retinol in the RPE.

In fact, there are two pathways to process the all-*trans*-retinal that is released from opsin. The first one is to enzymatically convert it to all-*trans*-retinol on the surface of the same disk membrane, where the rhodopsin molecule was bleached. This is a normal physiological way. It takes place, if retina was irradiated by visible light of physiological intensities or/if a special protein that situated in the edge of disk works properly. The physiological function of this protein is the active transport of all-*trans*-retinal from the disk. However, if too much rhodopsin molecules are bleached due to very bright light irradiation or if this special protein, so called ABCR protein, is deficient, the all-*trans*-retinal is not leave the disk membrane, but accumulated within the

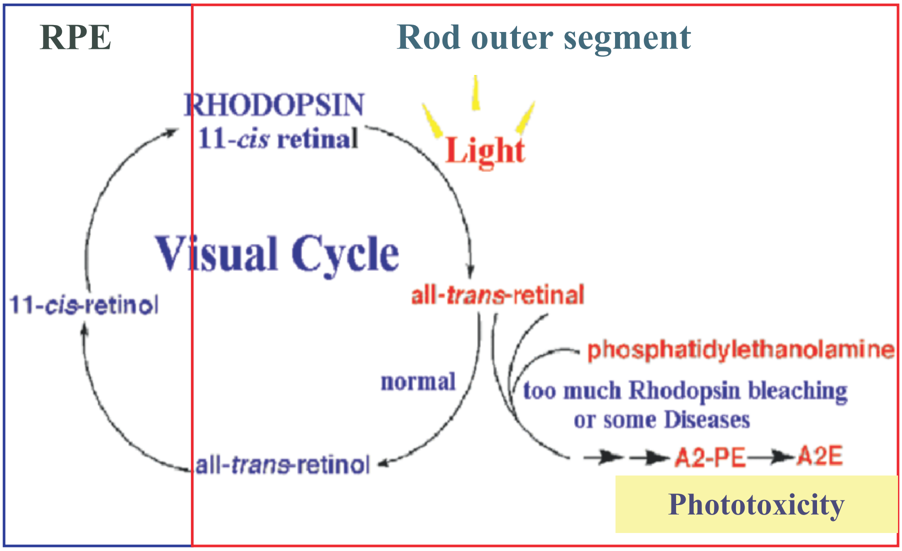


Figure 22. Visual Cycle and Light damage.

membrane. If so, one molecule of all-*trans*-retinal reacts very rapidly with one of membrane's phospholipids, namely phosphatidylethanolamine, and then the second molecule of all-*trans*-retinal reacts with the same molecule of phosphatidylethanolamine. As a result the complex and phototoxic compound, so called A2E, is appeared. A2E was originally named from the two equivalents of vitamin A aldehyde (*all-trans*-retinal) and one equivalent of ethanolamine. So, in the eye, A2E forms when an excess of *all-trans*-retinal leads to chemical condensation reactions between *all-trans*-retinal and phosphatidylethanolamine (2:1 ratio) in the photoreceptor disks. Once again, at normal physiological conditions most of *all-trans*-retinal molecules are reduced to *all-trans*-retinol by means of enzyme *all-trans*-retinoldehydrogenase that is located in the photoreceptor outer segment. Only *all-trans*-retinal that escapes normal conversion can enter the A2E biosynthetic pathway.

The right part of **Fig. 23** shows you the scheme of where the active carrier of *all-trans*-retinal — the membrane protein ABCR is situated and how it transports using the energy of ATP *all-trans*-retinal from the disk. The left part of **Fig. 23** shows you the drawing of rod outer segment apexes and adjoining RPE cells, and illustration of the well established fact that bright light or/and ABCR deficiency promote accumulation of lipofuscin granules or so called "age pigment". The lipofuscin granules are accumulated within the human

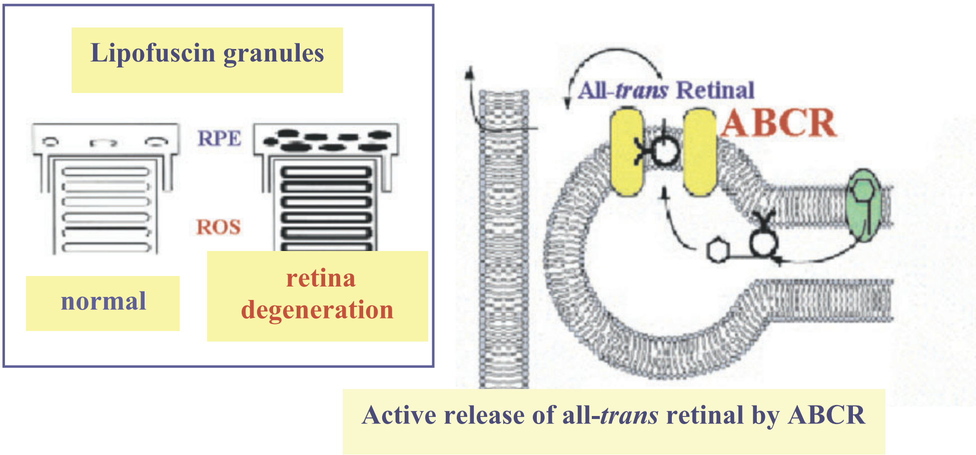
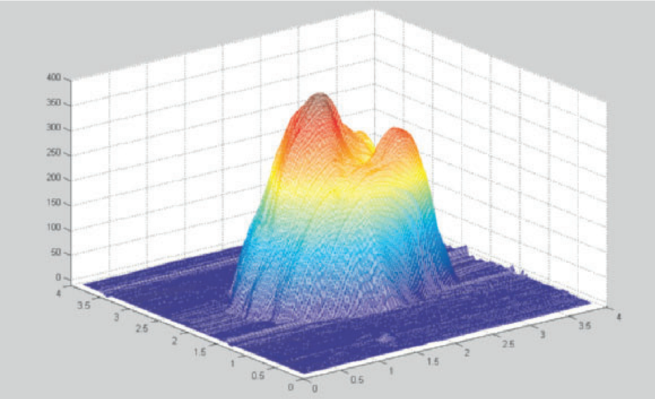


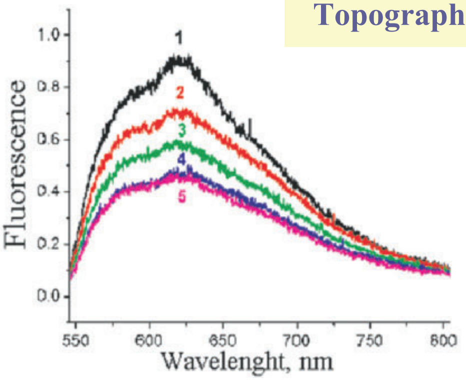
Figure 23. Bright light or/and ABCR deficiency promote accumulation of lipofuscin granules or "age pigment" in the RPE cells.

pigment epithelium cells in the course of life span and its build up is accelerated by too bright light and/ or deficiency of the special transport protein ABCR. The main fluorophore along with 10-12 other fluorophores within the lipofuscin granule is A2E.

These granules, as we and many other labs have established, are phototoxic and play an important role in the pathogenesis of retina degeneration deceases. Recently we have studied using atomic force and near-field microscopy techniques the ultrastructure organization and fluorescence properties of single lipofuscin granule isolated form human eye pigment epithelium. On the right part of the **figure 24** you can see the topography of a granule, which structure is not homogeneous, other words, not like a ball or bean. On the left part you can see the phenomenon we have observed: the decrease of fluorescence of the granule at 532 nm in the course of it irradiation by different energy of visible light (from 3 to 24 J/cm^2 – spectra 2 to 5); the spectrum 1 represents the fluorescence of a granule in the dark, before its irradiation. One can propose that this phenomenon is owing to breakdown (photooxidation) of main fluorophore A2E, and formation of its more short-wavelength toxic products, so called A2E-epoxides. It was shown recently that A2E-epoxides are even more toxic then A2E itself. The studies of origin, structure and photo- and biochemistry of lipofuscin granules are necessary for understanding the aggravating role of light in progression of retina degeneration deceases.



Topography of single lipofuscin granule

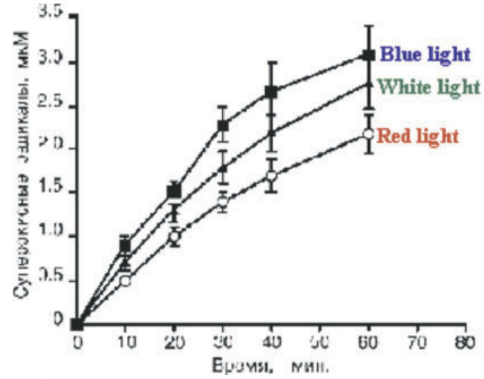


Decrease of fluorescence in the course of irradiation by visible light
Excitation at 420 nm, fluorescence at 625 nm

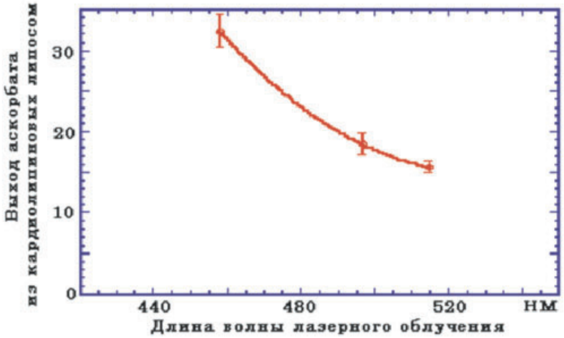
Figure 24. Structure and fluorescence of single lipofuscin granule: near field microscopy (Petrukhin et al., 2005).

The reason of lipofuscin granules phototoxicity is its ability to light-induced generation of toxic forms of oxygen.

On the **Fig. 25** is collected the main results related to light-induced reactions of lipofuscin granules. As you can see, light, first of all UV and blue light, absorbed by the granules is able to induce oxygen consume, to generate singlet oxygen and superoxide radicals, and damage biological membranes, in our case the liposome membranes.

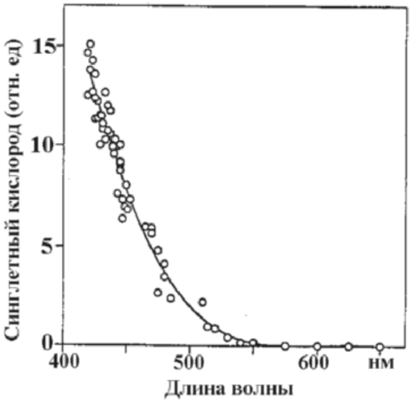


Superoxide radicals
(Boulton, Dontsov, Ostrovsky., 1993)



Light-induced damage to the liposome membrane

Singlet oxygen generation
(Rozanowska et al., 1998)



Oxygen consumption
(Pawak et al., 2002

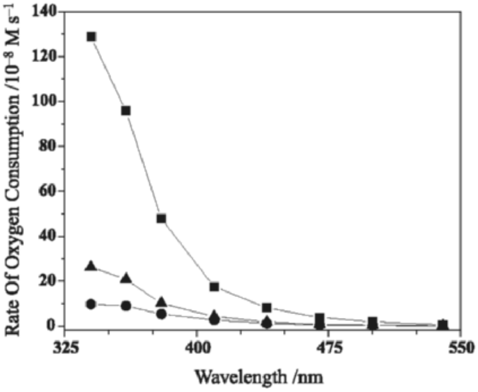


Figure 25. Wavelength dependence of light-induced reactions by lipofuscinranules.

The **Fig. 26** shows you the absorption spectra of all-*trans* retinal and lipofuscin that are phototoxic agents (photosensitizers) and absorption spectra of human rod and blue, green and red cones. It is clear to see that blue light is potentially dangerous due to these photosensitizers for rods and blue cones. If we will insert UV and blue light cut-off filters in front of retina, we can partly avoid the blue light hazard, but it could also partly destroy the blue colour perception. However, I believe to protect retina against the hazard of light damage, especially after cataract surgery, is more important than to destroy slightly the blue colour vision perception.

Now days we can say that the very possible molecular base of blue light hazard phenomenon that was discovered long time ago by American physiologist Ham is the phototoxicity of lipofuscin granules within the pigment epithelium cells.

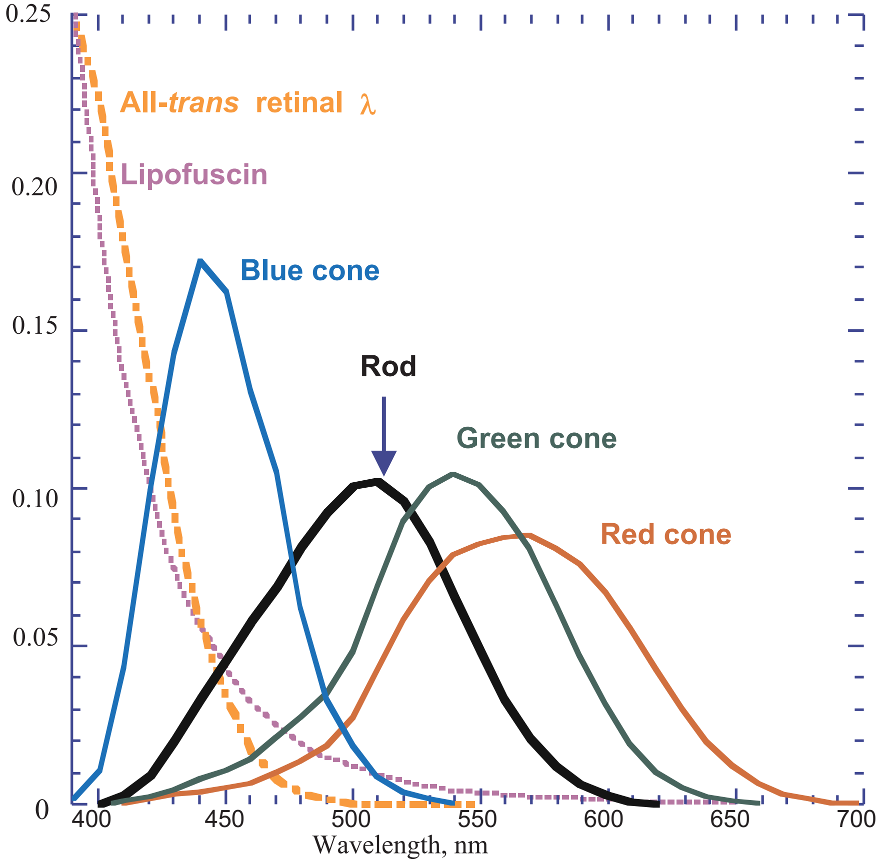


Figure 26. Absorption spectra of all-trans retinal and lipofuscin that are photosensitizers and Absorption spectra of human rod and blue, green and red cones.

The **Fig. 27** shows you the classical action spectrum of light damage (laser flashes) to monkey retina published by Ham, Meuller and Sliney in 1976.

As I talk about the complex physiological protective system before, the system that is able to prevent light hazard to the retina and retina pigment epithelium includes, at least, three lines of protection: permanent renewal of rod and cone outer segment, set of antioxidants, and optical media as cut-off filters, where lens plays a crucial role (**Fig. 28**, the table)

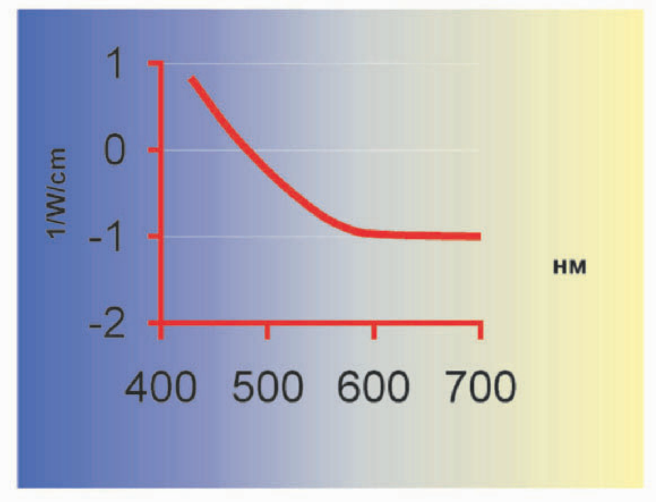
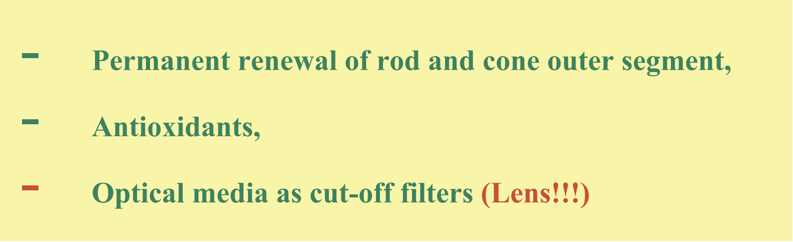


Figure 27. Action spectrum of light damage to monkey retina.(Ham, Meuller, Sliney, 1976).

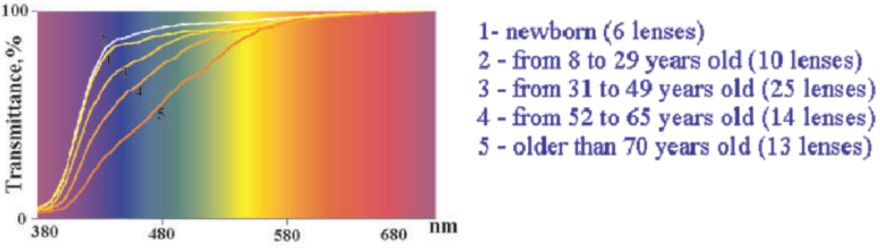


- **Permanent renewal of rod and cone outer segment,**
- **Antioxidants,**
- **Optical media as cut-off filters (Lens!!!)**

Figure 28. The complex protective system is able to prevent light hazard to the retina and RPE.

As far as human lens as a cut-off filter is concerned, it should be stressed that the well known physiological phenomenon of age-related yellowing of human lens is extremely important to protect retina and retina pigment epithelium against blue light hazard. Based on our knowledge regarding the absorption spectra of all-*trans* retinal as a photosensitizer and age-related yellowing of human lens (see **Fig. 29**), we have developed in the middle of 80th new yellow artificial lenses (intraocular lenses), which transmittance spectrum was similar to the transmittance spectrum of natural, non-cataractal human lens at age near 50. The idea was to cut-off the UV and blue light that could reach the retina and retina pigment epithelium, and if so, can be absorbed by all-*trans* retinal and lipofuscin granules and generates the toxic free radical of oxygen.

Lens as a cut-off filter
Age-related yellowing of human lens (Fedorovich, Zak, Ostrovsky, 1994)



Absorption spectra of photosensitizers, and transmittance spectra of human lens and yellow artificial lens "Spectrum"

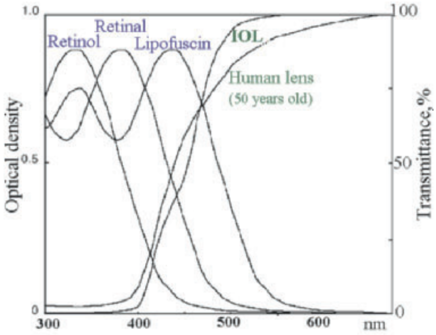


Figure 29.

Today about one million yellow intraolular lenses (IOLs "Spectrum") have been produced and implanted in Russia since 1986 **(Fig. 30)**. The comparative statistical analysis of yellow IOLs and colorless IOLs

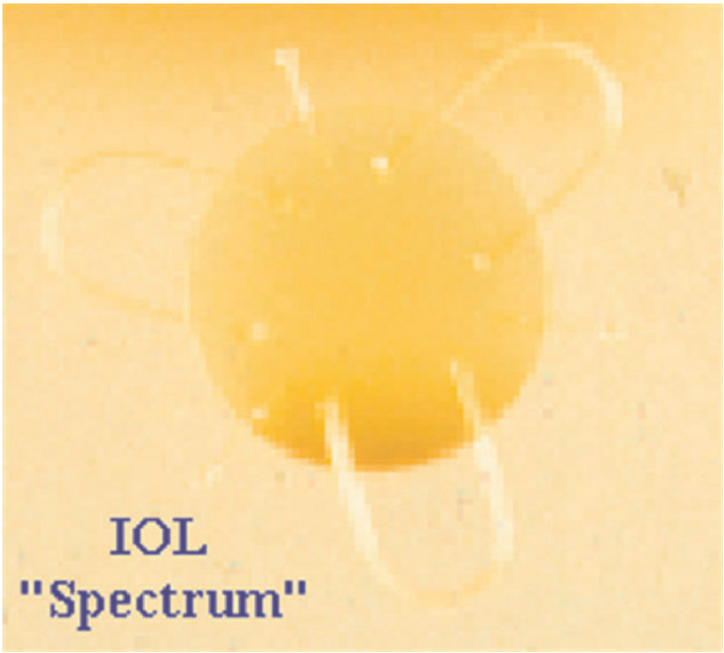


Figure 30. About one million yellowintraolular lenses IOLs “Spectrum” have been produced and implanted in Russia since 1986.

implantations has shown that so called macular edema – the adverse effect of cataract surgery – was observed about ten time less in case of yellow IOLs. Other words, yellow intraocular lenses are more protective than colorless ones. It is clear why: as it was shown in the recently published epidemiological study, the cataract surgery increases the amount of optical radiation that reaches the retina, and can promote retina diseases like age-related macula degeneration. About 2 or 3 years ago the American Alcon Company has developed soft yellow AcrySof® Natural intraocular lenses. The transmittance spectrum of these lenses is very similar to our hard IOLs “Spectrum”. Now days it is obviously that yellow IOLs that mimic the optical transmission of ageing crystalline lens and absorb both blue and UV radiation should be widely used. The yellow IOLs (“Spectrum”, Alcon AcrySof® Natural) can reduce the risk for or progression of age-related macular degeneration, and this is tremendously important. In fact, this statement is a direct result of our modern understanding of molecular mechanisms of the photobiological paradox of vision and its consequence.

THE ARCHITECTURE OF COMPLEXITY: FROM WWW TO CELLULAR METABOLISM

EIVIND ALMAAS AND ALBERT-LÁSZLÓ BARABÁSI*
Center for Complex Network Research and Department of Physics,
University of Notre Dame, Notre Dame, IN 46556

Abstract. Recent studies of complex systems indicate that real networks are far from random, instead having a highly robust, large-scale architecture that is governed by strict organizational principles. Here, we will focus on cellular networks, discussing their scale-free and hierarchical features. We will first discuss a few central network models, before illustrating the major network characteristics using examples primarily from bacterial metabolic networks. Additionally, as the interactions in real networks have unequal strengths, we discuss the interplay between network topology and reaction fluxes in cellular metabolic networks, as provided by the flux balance method. We find that the utilization of the metabolic networks is both globally and locally highly inhomogeneous, dominated by "hot-spots" that represent connected set of high-flux pathways.

Keywords: complex system; network; scale-free; cellular metabolism; flux balance analysis.

1. Introduction

The last century brought with it a dramatic progress in the natural sciences, and the majority of the scientific and engineering developments can be directly related to the reductionist approach: One presumes that the often complex behavior of a system can be predicted and understood from the detailed knowledge of the system's (often identical) elementary constituents. The last 30 years or so have clearly revealed that our ability to

---

*To whom correspondence should be addressed. Albert-László Barabási, Center for Complex Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA; e-mail: alb@nd.edu

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 107–125.

understand simple fundamental laws governing the individual "building blocks" is a far cry from being able to predict the overall behavior of a complex system (Anderson, 1972). Furthermore, for most complex systems there exists a considerable variation in the nature of both the elementary building blocks and their interactions requiring novel methods capable of analyzing and predicting their large-scale behavior.

In the last few years network approaches have shown great promise as a new tool to analyze and understand complex systems (Strogatz, 2001; Albert, 2002; Dorogovtsev, 2003; Bornholdt, 2003; Pastor-Satorras, 2004). For example, technological information systems like the internet and the world-wide web are naturally modeled as networks, where the nodes are routers (Faloutsos, 1999; Vázquez, 2002) or web-pages (Albert, 1999; Lawrence, 1999; Broder, 2000) and the links are physical wires or URL's respectively. Human society is also naturally described within the framework of network analysis, with people as nodes and the links between the nodes being either friendships (Milgram, 1967), collaborations (Kochen, 1989; Wasserman, 1994), sexual contacts (Liljeros, 2001), or co-authorship of scientific papers (Redner, 1998; Newman, 2001) to name just a few possibilities.

In the biological sciences we can represent systems as disparate as food webs in ecology and biochemical interactions in molecular biology as networks. In particular, the complex interactions of the various types of intracellular molecules offer a wide range of structures whose salient features are well captured by a network concept. Important examples include the many interactions between genes, proteins and metabolites. The development of high-throughput measurement tools in molecular biology during the last several years has made available a huge amount of genomic- and postgenomic data. For example, in the fields of transcriptomics and proteomics there is now a wealth of data on protein levels under various conditions, and genome wide analysis of gene expression at the mRNA level is now routine (Pandey, 2000; Caron, 2001; Burge, 2001).Furthermore, protein-protein interaction maps have been generated for a variety of organisms including viruses (Flajolet, 2000), prokaryotes, like *Helicobacter pylori* (Rain, 2001) and eukaryotes, like *Saccharomyces cerevisiae* (Ito, 2000; Ito 2001; Schwikowski 2000; Uetz 2000; Gavin 2002; Ho, 2002; Jeong, 2001) and *Caenorrhabditis elegans* (Walhout, 2000).

2. Tools in Network Characterization

We begin by discussing some of the tools of network analysis, and the key properties they reveal of real networks. It is important to realize that the representation of different systems as networks has uncovered surprising similarities, many of which are intimately tied to power laws. Although the details of the networks, e.g., the explicit nature of a node and the nature of its interactions with other nodes, are frequently quite unique, the overall statistical features of different networks can be very similar.

2.1. DEGREE DISTRIBUTION

The simplest statistical measure of a network property is the average number of nearest neighbors of a node, also called the average degree $\langle k \rangle$. A natural refinement of this property, revealing deeper insights into a network's organization, is the distribution of the number of nearest neighbors $P(k)$ of a randomly chosen node. The degree of a node is defined from the nearest neighbor matrix $\{a_{ij}\}$, which contains the description of the whole network, as $k_i = \sum a_{ij}$, where $a_{ij}=1$ if there is a link between nodes i and j and zero otherwise. For a surprisingly large number of networks, the degree distribution is best characterized by the power law functional form (Barabási, 1999) (Fig. 1a),

$$P(k) \sim k^{-\alpha}. \qquad (1)$$

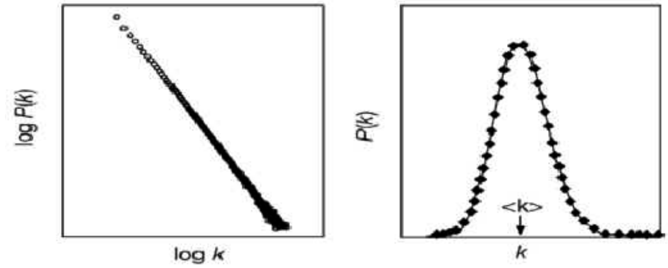


Figure 1. Characterizing degree distributions. For the power-law degree distribution **(a)**, there exists no typical node, while for distributions with a single peak (see **(b)**) most nodes are well represented by the degree $\langle k \rangle$ of the average (typical) node.

Examples include the Internet on both the domain and router level, the World-Wide Web, the network of movie actors, sexual network, co-author networks and food webs to mention a few (for a more complete listing together with the pertinent references, see Albert, 2002).

However, if the degree distribution is instead single-peaked (e.g., Poisson or Gaussian) as in Fig. 1b, the majority of the nodes would be well described by the average degree, and hence, the properties of a "typical" node. In contrast, for networks with a power-law degree distribution, the majority of the nodes have only one or two neighbors while coexisting with many nodes with hundreds and a few even with thousands of nearest neighbors. For the power law networks there exists no typical node, and thus they are often referred to as "scale-free". In Fig. 2a-c respectively, we show the degree distributions of the metabolic networks of the three notably disparate organisms of *Archaeoglobus fulgidus* (archae), *Escherichia coli* (bacterium) and *C. elegans* (eukaryote), all adhering to a power law functional form (Jeong, 2000). The hypothesis that the scale-free network structure probably is a universal feature of metabolic networks is further strengthened (Fig. 2d) by the presence of a clear power law when averaging over 43 diverse organisms (Jeong, 2000).

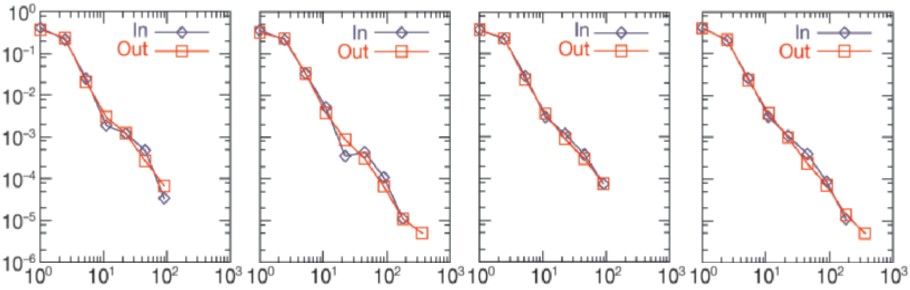


Figure 2. Degree distributions of metabolic networks. The degree distribution displays a power law in both the in- and the out degrees for **(a)** *A. fulgidus* (archae), **(b)** *E. coli* (bacterium), **(c)** *C. elegans* (eukaryote), and **(d)** when averaged over 43 organisms (Jeong, 2000).

2.2. CLUSTERING COEFFICIENT

In order to investigate the local network structure, we use the clustering of a node C_i, which measures the degree to which the neighborhood of a node resembles a complete subgraph (Watts, 1998). The clustering of a node *i* can also be thought of as the probability that two nodes with a common neighbor are also neighbors of each other (i.e. form a triangle). For a node *i* with degree k_i the clustering is defined as

$$C_i = \frac{2n_i}{k_i(k_i - 1)}, \tag{2}$$

representing the ratio of the number of actual connections n_i between the neighbors of node i to the number of possible connections. For a node that is part of a fully interlinked cluster (all the nodes are connected to each other) $C_i = 1$, while $C_i = 0$ for a node which acts as a bridge between different clusters or parts of the network. Accordingly, the overall clustering coefficient of a network with N nodes, given by

$$\langle C \rangle = \sum C_i / N, \tag{3}$$

gives a measure a network's potential modularity: The larger the value of $\langle C \rangle$, the more small, tightly connected small clusters exist in the network. One can also study the clustering of nodes with a given degree k, through which information about the actual modular organization of a network can be gleaned (Ravasz 2002; Ravasz 2003; Dorogovtsev, 2002; Vázquez, 2002): For all metabolic networks available, *C(k)* behaves like the power law

$$C(k) \sim k^{-\delta}, \tag{4}$$

suggesting the existence of a hierarchy of nodes with different degrees of modularity (as measured by the clustering coefficient) overlapping in an iterative manner (Ravasz, 2002). In Fig. 3a-c we show the clustering as function of k for the three single-celled organisms *Aquidex aeolicus* (archaea), *E. coli* (bacterium) and *S. cerevisiae* (eukaryote), respectively. In Fig. 3d *C(k)* is averaged over 43 organisms, displaying a robust power-law behavior.

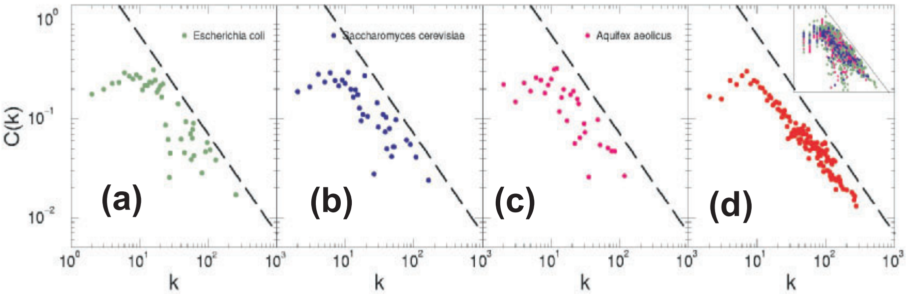


Figure 3. The clustering of metabolic networks. The average clustering as function of node degree k for **(a)** *A. aeolicus* (archaea), **(b)** *E. coli* (bacterium), **(c)** *S. cerevisiae* (eukaryote), and **(d)** averaged over 43 organisms, displays a power law behavior (Ravasz, 2002). The dashed lines represent *C(k)~1/k*. The inset in **(d)** displays all the 43 organisms together.

2.3. DEGREE CORRELATIONS

In many real networks, there exists non-trivial correlations in the node connectivities. In networks with assortative degree correlations, the high degree nodes tend to be connected to other nodes with high degree, while in dissassortative networks, high degree nodes tend to be connected to low degree ones. The type and level of degree correlations in any network is typically measured as a Pearson's correlation coefficient (Newman, 2002) or as the average connectivity $k_{nn}(k)$ of the nearest neighbors of nodes with degree k (Pastor-Satorras, 2001). Studies of real network correlations indicate that both technological and biological networks are dissassortative, while social networks tend to be assortative (Newman, 2003).

2.4. WEIGHTED NETWORKS

While the modelling of real networks has focused on understanding the structure of natural and man-made networks, it is also necessary to acknowledge the fact that the network links can be very different in their connection strengths or intensities, generally denoted link weights. The effect of this difference can be captured by generalizing the above-mentioned network tools to allow for varying link weights. For instance for a node i, the node strength s_i, is defined as

$$s_i = \sum_j w_{ij} a_{ij} , \tag{5}$$

which can be readily recognized as the generalization of the node degree k_i to weighted networks (Yook, 2001; Barrat, 2004). Comparison of the node strength distribution with the degree distribution, and studies of quantities like $s(k)$, can subsequently uncover non-trivial correlations between the link weights and the network topology (Barrat, 2004). In section 4 of this chapter, we will investigate in more detail the effects of link weights on bacterial metabolic networks.

3. Important Network Models

Several network models are currently available to describe many of the aforementioned aspects of real networks. In the following, we will limit ourselves to the description three such models, namely the random network model, the scale-free model and the hierarchical model. These models are then compared to the features actually observed in metabolic networks.

3.1. RANDOM NETWORK MODEL

Since the 1950's the properties of large networks with no apparent design principles were presumed to be well described by random graphs (Bollobas, 1985). The random graph model represents the simplest and most straightforward realization of a complex network. According to the Erdos-Renyi (ER) model of random networks (Erdos, 1960), we start with N nodes and connect every pair of nodes with probability p, creating a graph with approximately $pN(N-1)/2$ randomly distributed edges (Fig. 4a). For this model the degrees follow a Poisson distribution (Fig. 5a).

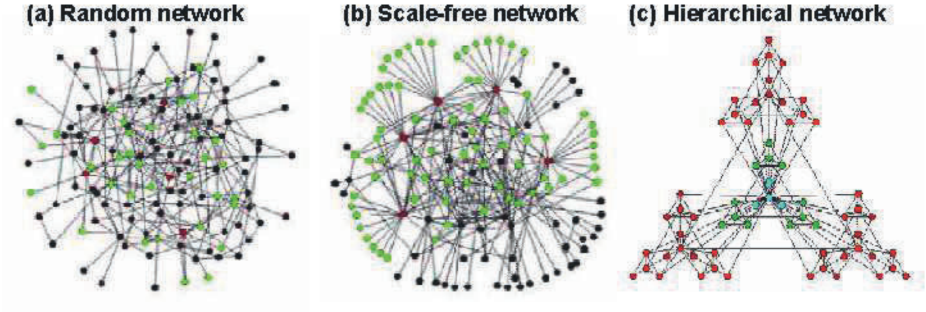


Figure 4. Graphical representation of three network models: **(a)** The ER (random) model, **(b)** the BA (scale-free) model and **(c)** the hierarchical model. Panel **(c)** demonstrates the iterative construction of a hierarchical network by starting from a fully connected cluster of four nodes (blue). This cluster is then copied three times (green) while connecting the peripheral nodes of the replicas to the central node of the starting cluster. We end up with a 64-node scale-free hierarchical network by once more repeating this replication and connection process (red nodes). In panels **(a)** and **(b)** we emphasize the difference between the ER and the BA networks by coloring the five nodes with the highest number of links red and their first neighbors green. For the scale-free network we reach more than 60% of the nodes using the five largest hubs, while for the random network only 27% of the nodes are directly accessible from the five most connected nodes, demonstrating the heterogeneous nature of scale-free networks. Note that the networks in **(a)** and **(b)** consist of the same number of nodes and links.

Consequently, the typical node is well described by the average degree $\langle k \rangle$ of the network. Furthermore, for this "democratic" network model, the clustering is independent of the node degree k (Fig. 5d). This simple and beautiful model is, however, not able to accurately capture the topological properties of most real networks, as is clearly seen on comparison with e.g. Figs. 2 and 3.

3.2. SCALE-FREE NETWORK MODEL

In the network model of Barabási and Albert (BA) (Fig. 4b), the emergence of a power-law degree distribution is attributed to two crucial mechanisms, both absent from the classical random network model (Barabási, 1999). First, networks grow through the addition of new nodes linking to nodes that already are present in the system. Second, in most real networks there is a high probability that a new node link to an existing node with a large number of connections, a mechanism often referred to as preferential attachment. These two principles are implemented as follows: starting from a small core graph consisting of m_0 nodes, a new node with m links (with $m \leq m_0$) is added at each time step and connected to the already existing nodes. Each of the m new links are then preferentially attached to a node i (with k_i neighbors) chosen according to the probability

$$\Pi_i = k_i / \sum_j k_j \, . \tag{6}$$

The simultaneous combination of these two network growth rules gives rise to the observed power-law degree distribution (Fig. 5b). In contrast to a random network, the probability that a node is highly connected is statistically significant in a scale-free network; hence many network properties are determined by a relatively small number of highly connected nodes, frequently called “hubs”. In Fig. 4a and b we show an example of the effect of the hubs on the network structure by coloring the five nodes with largest degrees are red and their nearest neighbors green. While in the ER network only 27% of the nodes are reached by the five most connected ones, more than 60% of the nodes in the scale-free network are covered, demonstrating the key role played by the hubs. Additionally, the hub’s dominance of the network topology cause the scale-free networks to be highly tolerant to random failures (perturbations) while being extremely sensitive to targeted attacks (Albert, 2000). Comparing the properties of the BA network model with those of the ER model, we note that while the clustering of the BA network is larger, $C(k)$ however is approximately constant (Fig. 5e), suggesting that a hierarchical structure is absent.

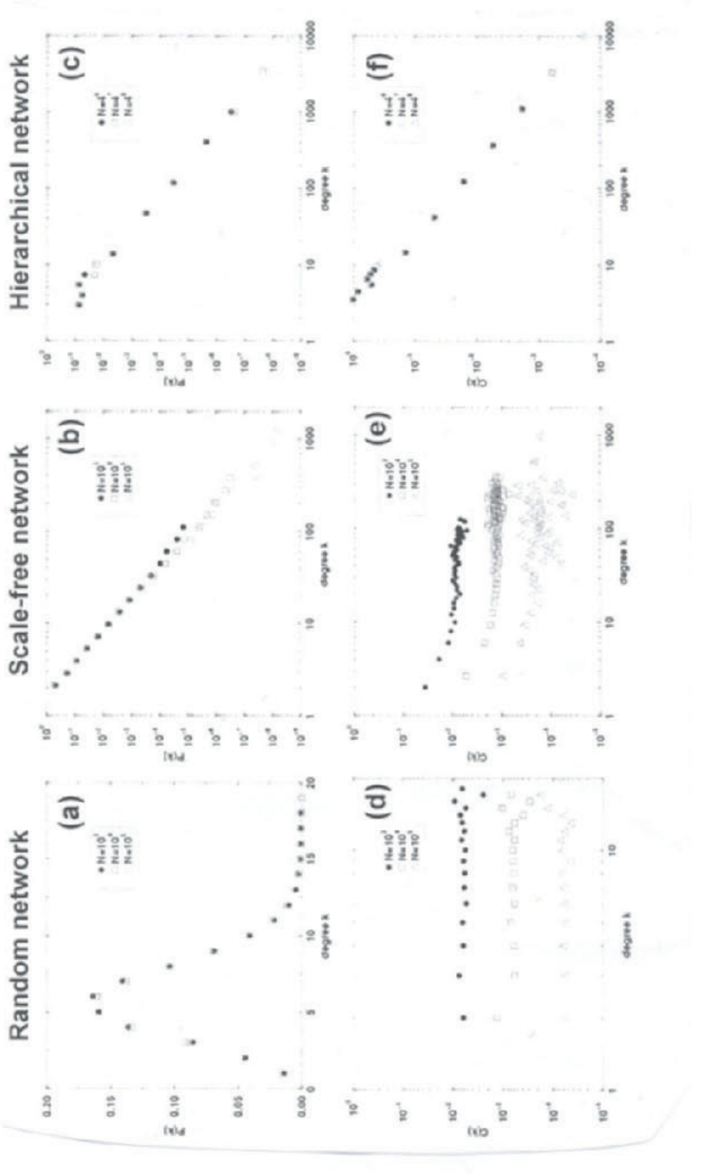


Figure 5. Properties of the three network models. **(a)** The ER model gives a Poisson degree distribution *P(k)* (the probability that a randomly selected node has exactly *k* links), being strongly peaked around the average degree $\langle k \rangle$ and decaying exponentially for large *k*. For the scale-free **(b)** and the hierarchical **(c)** network models the degree distributions instead decay according to the power-law $P(k) \sim k^{-\gamma}$. The average clustering coefficient for nodes with exactly *k* neighbors, *C(k)*, is independent of *k* for both the ER **(d)** and the scale-free **(e)** network model, while in contrast **(f)** $C(k) \sim k^{-1}$ for the hierarchical model (cf. Fig. 3).

3.3. HIERARCHICAL NETWORK MODEL

Many real networks are expected to be fundamentally modular, meaning that the network can be partitioned into a collection of modules where each module performs an identifiable task, separable from the function(s) of other modules (Hartwell, 1999; Lauffenburger, 2000; Rao, 2001). Thus, we expect a seamless combination of the scale-free property with such potential modularity. In order to account for the modularity as reflected in the power-law behavior of $C(k)$ (Fig. 3) and a scale-free degree distribution (Fig. 2), we can assume that clusters combine in an iterative manner, generating a hierarchical network (Ravasz, 2002; Vázquez, 2002) (Fig. 4c). Such a network emerges from a repeated duplication and integration process of clustered nodes (Ravasz, 2002), which in principle can be repeated indefinitely.

This process is depicted in Fig. 4c, where by starting from a small cluster of four densely linked nodes (blue), one next generates three replicas (green) of this hypothetical initial module and connect the three external nodes of the replicated clusters to the central node of the old cluster. The centers of the replicas are also connected to each other, thus obtaining a large 16-node module. Subsequently, we again generate three replicas (red) of this 16-node module, and connect the replicas as described above, obtaining a new module now consisting of 64 nodes. This (deterministic) hierarchical network model seamlessly integrates a scale-free topology with an inherent modular structure by generating a network that has a power law degree distribution (Fig. 5c) with degree exponent $\gamma = 1 + \ln 4 / \ln 3 \approx 2.26$ and a clustering coefficient $C(k)$ which proves to be dependent on k^{-1} (Fig. 5f). However, it is of importance to note that modularity does not imply clear-cut sub-networks linked in well-defined ways (Ravasz, 2002; Holme, 2003). Indeed, the boundaries of modules are often considerably blurred and bridged by highly connected nodes (hubs) which interconnect modules.

3.4. BOSE-EINSTEIN CONDENSATION AND NETWORKS

In most complex systems the nodes have differing abilities of attracting new links, which is independent of their number of nearest neighbors. For instance, some Web pages quickly acquire a large number of links through a mixture of good content and marketing, although they are just recently published on the World-Wide Web. This competition for links can be incorporated into the scale-free model by adding a "fitness" parameter, η_i, to each node, i, describing its ability to compete for links at the expense of other nodes. For example, a Web page with good up-to-date content and a friendly interface would be expected to display a greater fitness than a low-quality page that is only updated occasionally. The probability Π_i that a new node connects to one with k_i links is then modified from Eq. (4) such that $\Pi_i = \eta_i\, k_i/\Sigma_j\, \eta_j\, k_j$ (Bianconi, 2001).

The competition generated by the various fitness levels means that each node evolves differently in time compared with others. Indeed, the connectivity of each node is now given by $k_i(t) \sim t^{\beta(\eta)}$, where the exponent $\beta(\eta)$ increases with η, and t is the time since the node was added to the network (Bianconi, 2001). Consequently, fit nodes (ones with large η) can join the network at some later time and connect to many more links than less-fit nodes that have been around for longer.

Amazingly, such competitive-fitness models appear to have close ties with Bose-Einstein condensation, currently one of the most investigated problems in atomic physics (Anderson, 1995; Bradley, 1995; Legget, 2001; Greiner, 2003). In a normal atomic gas, the atoms are distributed among many different energy levels. However in a Bose-Einstein condensate, all the atoms accumulate in the lowest energy state of the system and are described by the same quantum wave function. By replacing each node in the network with an energy level having energy $\varepsilon_i = \exp(-\beta\, \eta_i)$, the fitness model maps exactly onto a Bose gas (Bianconi, 2001). According to this mapping, the nodes correspond to energy levels while the links are represented by atoms in these levels. Additionally, the behavior of a Bose gas is uniquely determined by the distribution $g(\varepsilon)$ from which the random energy levels (or fitnesses) are selected. One expects that the functional form of $g(\varepsilon)$ depends on the system. For example, the attractiveness of a router to a network engineer comes from a rather different distribution than the fitness of a dot-com company competing for customers.

For a wide class of $g(\varepsilon)$ distributions, a "fit-get-richer" phenomena emerges (Bianconi, 2001). Although the fittest node acquires more links than its less-fit counterparts, there is no clear winner. On the other hand, certain $g(\varepsilon)$ distributions can result in a Bose-Einstein condensation, where the fittest node does emerge as a clear winner. For these distributions, a condensate develops by acquiring a significant fraction of the links which is independent of the size of the system. In network language this corresponds to a "winner-takes-all" phenomenon. While the precise form of the fitness distribution for the Web or the Internet is not known yet, it is likely that $g(\varepsilon)$ could be measured in the near future.

4. Metabolic Network Utilization

It is important to realize that despite their successes, purely topological approaches have intrinsic limitations. Since the activity of the various metabolic reactions or regulatory interactions differs widely, some being highly active under most growth conditions while others are switched on only for rare environmental circumstances, it is necessary to include this information in a network description. Therefore, a biologically relevant understanding of metabolic- and other biochemical reaction networks requires us to consider the intensity (i.e., strength), the direction (when applicable), and the temporal aspects of the interactions. While so far we know little about the temporal aspects of the various metabolic reactions, recent results have added insights on how the strength of the interactions (i.e., fluxes) of the metabolic reactions are organized (Almaas, 2004).

A natural measurement of interaction strength for a metabolic network is given by the flux of the metabolic reactions, representing the amount of substrate being converted to a product within unit time. Recent metabolic flux-balance approaches (FBA) (Edwards, 2000; Edwards, 2001; Ibarra, 2002; Edwards, 2002; Segre, 2002) make it feasible to calculate the flux for each reaction. This has markedly improved our ability to generate quantitative predictions on the relative importance of the various reactions, leading to experimentally testable hypotheses. The much utilized FBA approach can be stated as follows: Starting from a stoichiometric matrix representation of the *E. coli* K12 MG1655 metabolic network, which contains 537 metabolites and 739 reactions (Edwards, 2000; Edwards, 2001; Ibarra, 2002; Edwards, 2002), the steady state concentrations of all metabolites satisfy the relation

$$\frac{d}{dt}[A_i] = \sum_j S_{ij}\, \nu_j \; = 0, \tag{7}$$

where S_{ij} is the stoichiometric coefficient of metabolite A_i in reaction *j* and ν_j is the flux of reaction *j*. We adhere to the convention of $S_{ij} < 0$ ($S_{ij} > 0$) if metabolite A_i is a substrate (product) in reaction *j*, and we constrain all fluxes to be positive by dividing each reversible reaction into two “forward” reactions with positive fluxes. Any vector of positive fluxes $\{\nu_j\}$ which satisfies Eq. (7) corresponds to a stoichiometrically allowed state of the metabolic network, and hence, a potential state of operation of the cell.

Assuming that cellular metabolism is in a steady state and optimized for the maximal growth rate (Edwards, 2001; Ibarra, 2002), FBA allows us to calculate the flux for each reaction using linear optimization. This provides a measure of each reaction’s relative activity (Almaas, 2004). In a manner similar to that of the degree distribution, the flux (or weight) distribution of *E. coli* displays a strong overall inhomogeneity: reactions with fluxes spanning several orders of magnitude coexist under the same conditions (Fig. 6a). This is captured by the flux distribution for *E. coli* which follows a power law, where the probability that a reaction has flux ν is given by $P(\nu) \sim (\nu + \nu_0)^{-\alpha}$. The flux exponent is predicted to be $\alpha = 1.5$ by FBA methods (Almaas, 2004).

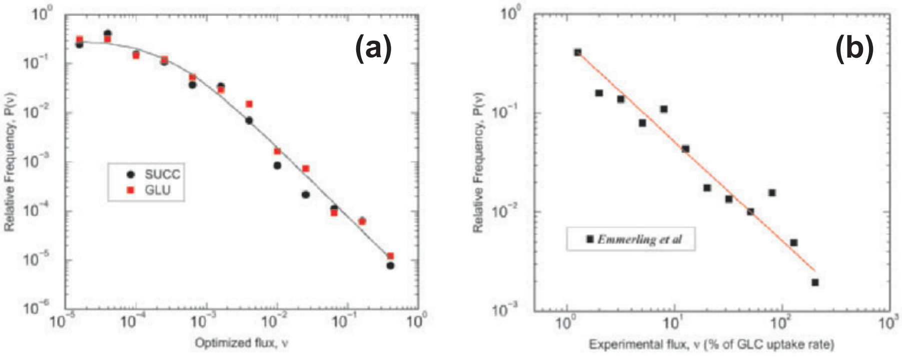


Figure 6. Flux distribution for the metabolism of *E. coli*. **(a)** Flux distribution when maximizing the biomass production on succinate (black) and glutamate (red) rich uptake substrates. The solid line corresponds to the power law fit $P(\nu) \sim (\nu + \nu_0)^{-\alpha}$ with $\nu_0 = 0.0003$ and $\alpha = 1.5$. **(b)** The distribution of experimentally determined fluxes (see Emmerling, 2002) from the central metabolism of *E. coli* also displays power-law behavior which is best fit to $P(\nu) \sim \nu^{-\alpha}$ with $\alpha = 1$.

In a recent experiment (Emmerling, 2002) the strength of the various fluxes of the central metabolism were measured in several *E. coli* mutants in addition to the wildtype, revealing the power-law flux dependence $P(\nu) \sim \nu^{-\alpha}$ with $\alpha \cong 1$ (Fig. 6b) (Almaas, 2004). This power law behavior indicates that the vast majority of the metabolic reactions have quite small fluxes, while coexisting with a few reactions with very large flux values. Repeating these simulations for the bacterial organism *H. pylori* and the eukaryotic single-celled baker's yeast *S. cerevisiae*, we find similar results.

The observed flux distribution is compatible with two quite different potential *local* flux structures. A homogeneous local organization would imply that all reactions producing (consuming) a given metabolite have comparable flux values. On the other hand, a more delocalized "hot backbone" is expected if the local flux organization is heterogeneous, such that each metabolite has a dominant source (consuming) reaction. To distinguish between these two scenarios for each metabolite i produced (consumed) by k reactions, we define the measure (Barthelemy, 2003; Derrida, 1987)

$$Y(k,i)=\sum_{j=1}^{k}\left(\frac{\hat{v}_{ij}}{\sum_{l=1}^{k}\hat{v}_{il}},\right)^2 \tag{8}$$

where $\hat{v}_{ij}$ is the mass carried by reaction *j* which produces (consumes) metabolite *i*. If all reactions producing (consuming) metabolite *i* have comparable $\hat{v}_{ij}$ values, $Y(k,i)$ scales as $1/k$. If, however, a single reaction's activity dominates Eq. (8), we expect $Y(k,i)\sim 1$, i.e., $Y(k,i)$ is independent of *k*. For the *E. coli* metabolism optimized for succinate and glutamate uptake (Fig. 7) we find that both the *in* and *out* degrees follow the power law $Y(k,i)\sim k^{-0.27}$, representing an intermediate behavior between the two extreme cases (Almaas, 2004). This indicates that the large-scale inhomogeneity observed in the overall flux distribution is increasingly valid at the level of the individual metabolites as well: the more reactions consume (produce) a given metabolite, the more likely it is that a single reaction carries the majority of the flux. This statement is valid also for the metabolism in *H. pylori* and *S. cerevisiae*.

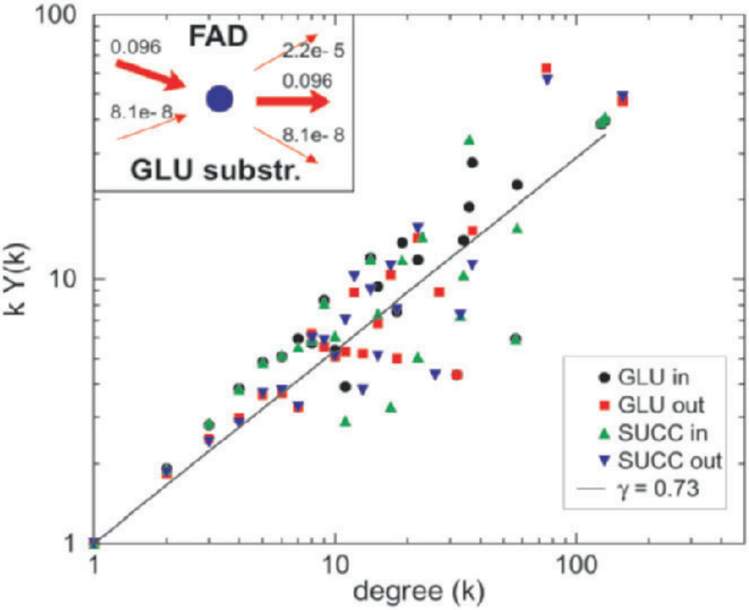


Figure 7. haracterizing the local inhomogeneity of the metabolic flux distribution. The measured *kY(k)* (see Eq. (7)) shown as function of *k* for incoming and outgoing reactions for fluxes calculated on both succinate and glutamate rich substrates, averaged over all metabolites, indicating $Y(k)\sim k^{-0.27}$, as the straight line in the figure has slope $\gamma = 0.73$. Inset: The non-zero mass flows $\hat{v}_{ij}$ producing (consuming) flavin adenine dinucleotide (FAD) on a glutamate rich substrate.

The local flux inhomogeneity described above suggests that we can identify a single reaction dominating the production or consumption of most metabolites. Henceforth, we can construct a simple algorithm which systematically removes, for each metabolite, all reactions but the one providing the largest incoming and outgoing flux contribution. When the largest outgoing flux of metabolite A is identical to the largest incoming flux of metabolite B the *high flux backbone* (HFB) of the metabolism can be uncovered, whose identity is specific to the given growth condition. The HFB mostly consists of reactions linked together, forming a giant component with a star-like topology which includes almost all metabolites produced under the given growth condition. Only a few pathways are disconnected: while these pathways are parts of the HFB, their end product serves only as the second most important source for some other HFB metabolite. It is interesting to note that groups of individual HFB reactions for the most part overlap with the traditional, biochemistry-based partitioning of cellular metabolism: e.g. all metabolites of the citric-acid cycle of *E. coli* are recovered, and so are a considerable fraction of other important pathways, such as those being involved in histidine-, murein- and purine biosynthesis, to mention a few. However, while the detailed nature of the HFB depends on the particular growth conditions, the HFB in general captures the subset of reactions that dominate the activity of the metabolism for this condition. As such, it offers a complementary approach to elementary flux mode analyses (Dandekar, 1999; Schuster, 2000; Stelling, 2002), which successfully determine the available modes of operation for smaller metabolic sub-networks, but whose application to the full *E. coli* metabolism has not yet been possible.

5. Conclusions

During the last few years, it has become evident that power laws are abundant in complex systems, affecting both the evolution and the utilization of real networks. The power-law degree distribution has become the trademark of scale-free networks and can generally be explained by invoking the principles of network growth and preferential attachment. These general principles can be realized by a wide variety of models which differ in their local growth rules. In understanding the utilization of complex networks, it is important to realize that most links represent disparate connection strengths or transportation thresholds. For the metabolic networks of *E. coli*, *H. pylori* and *S. cerevisiae*, we have implemented a flux-balance approach and calculated the distribution of link weights (fluxes), which (reflecting the scale-free network topology) displays a robust power-law independent of any environmental perturbations. Furthermore, this global inhomogeneity in the link strengths is also present at the level of the individual metabolites, allowing us to automatically uncover the high flux backbone of the metabolism. The HFB offers novel insights into the metabolic network's response to changes in the external environment. Developing universal tools for analyzing a network's response to changing conditions, and defining the nature and the degree of changes in a network under these conditions, could provide significant insights into the organization and principal forces shaping the evolution of networks.

References

Albert, R. & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, p47-97.

Albert, R., Jeong, H. & Barabási, A.-L (1999). Diameter of the World-Wide Web. *Nature*, 401, p130-1.

Albert, R., Jeong, H. & Barabási, A.-L. (2000). Attack and error tolerance of complex networks. *Nature*, 406, p378-82.

Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z.N. & Barabási, A.-L. (2004). Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, 427, p839-843.

Anderson, M.H., Ensher, J.R., Matthews, M.R. Wieman, C.E. & Cornell, E.A. (1995). Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science*, 269, p198-201.

Anderson, P. W. (1972). More Is Different. *Science*, 177, p393-6.

Barabási, A.-L. & Albert, R., (1999). Emergence of scaling in random networks. *Science*, 286, p509-12.

Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. (2004). The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA,* 101, p3747.

Barthelemy, M., Gondran, B. & Guichard, E. (2003). Spatial structure of the Internet traffic. *Physica A*, 319, p633-42.

Bianconi, G. & Barabási. A.-L. (2001). Bose-Einstein condensation in complex networks. *Phys. Rev. Lett.,* 86, 5632.

Bradley, C.C., Sackett, C.A., Tollett, J.J. & Hulet, R.G. (1995). Evidence of Bose-Einstein condensation in an atomic gas with attractive interactions. *Phys. Rev. Lett.*, 75, p1687-90.

Bollobas, B. (1985). *Random Graphs*. Academic Press, London.

Bornholdt, S. & Schuster, H. G. (2003). *Handbook of graphs and networks: From the genome to the Internet*. Wiley-VCH, Berlin, Germany.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P, Rajalopagan, S., Stata, R., Tomkins, A. & Wiener, J. (2000). Graph structure in the web. *Comput. Netw.*, 33, p309-20.

Burge, C.B. (2001). Chipping away at the transcriptome. *Nature Genet.*, 27, p232-4.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A., Heisterkamp, S., van Kampen, A. & Versteeg, R. (2001). The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. *Science*, 291, p1289-92.

Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, P. (1999). Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem. J.* 343, p115–124.

Derrida, B. & Flyvbjerg, H. (1987). Statistical properties of randomly broken objects and of multivalley structures in disordered-systems. *J. Phys. A: Math. Gen.*, 20, p5273-88 (1987).

Dorogovtsev, S.N., Goltsev, A.V. & Mendes, J.F.F. (2002). Pseudofractal scale-free web. *Phys. Rev. E*, 65, 066122.

Dorogovtsev, S.N. & Mendes, J.F.F. (2003) *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, Oxford.

Edwards, J. S., Ibarra, R. U. & Palsson, B. O. (2001). In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol* ***19***, p125-30.

Edwards, J. S. & Palsson, B. O. (2000). The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A* 97, p5528-33.

Edwards, J. S., Ramakrishna, R. & Palsson, B. O. (2002). Characterizing the metabolic phenotype: A phenotype phase plane analysis. *Biotechn. Bioeng.* 77, 27-36.

Emmerling, M., Dauner, M., Ponti, A., Fiaux, J., Hochuli, M., Szyperski, T., Wuthrich, K., Bailey, J.E. & Sauer, U. (2002). Metabolic flux responses to pyruvate kinase knockout in Escherichia coli. *J Bacteriol.*, 184, p152-64.

Erdos, P. & Renyi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5, p17-61.

Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Comput. Commun. Rev.*, 29, p251-62.

Flajolet, M., Rotondo, G., Daviet, L., Bergametti, F., Inchauspe, G., Tiollais, P., Transy, C. & Legrain, P. (2000). A genomic approach to the hepatitis C virus. *Gene*, 242, p369-79.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. & Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415, p141-7.

Greiner, M., Regal, C.A. & Jin, D.S. (2003). Emergence of a molecular Bose-Einstein condensate from a Fermi gas. *Nature*, 426, p537-40.

Hartwell, L.H., Hopfield, J.J., Leibler, S. & Murray, A.W. (1999). From molecular to modular cell biology. *Nature*, 402, C47-52.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L.Y., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W.V., Figeys, D. & Tyers, M. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*, 415, p180-3.

Holme, P., Huss, M. & Jeong, H. (2003). Subnetwork hierarchies of biochemical pathways. *Bioinformatics*. 19, p532-9.

Ibarra, R. U., Edwards, J. S. & Palsson, B. O. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420, p186-9.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.*, 98, p4569-74.

Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S. & Sakaki, Y. (2000). Towards a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci.*, 97, p1143-47.

Jeong, H., Mason, S.P., Barabási, A.-L. & Oltvai, Z.N. (2001). Lethality and centrality in protein networks. *Nature*, 411, p41-2.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature*, 407, p651-4.

Kochen, M. (ed.) (1989). *The small-world*. Ablex, Norwood, N.J.

Lauffenburger, D. (2000). Cell signaling pathways as control modules: Complexity for simplicity. *Proc. Natl. Acad. Sci.*, 97, p5031-33.

Lawrence, S. & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, 400, p107-9.

Legget, A.J. (2001) Bose-Einstein consdensation in the alkali gases: Some fundamental concepts. *Rev. Mod. Phys*., 73, p307-56.

Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E. Aberg, Y. (2001). The web of human sexual contacts. *Nature*, 411, p907-8.

Milgram, S. (1967). The small-world problem. *Psychology Today*, 2, p60-7.

Montoya, J.M. & Sole, R.V. (2002). Small-world patterns in food webs. *J. Theor. Biol.*, 214, p405-12.

Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci.*, 98, p404-9.

Newman, M.E.J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89, 208701.

Newman, M.E.J. (2003). Mixing patterns in networks. *Phys. Rev. E.*, 67, 026126.

Pandey, A. & Mann, M. (2000). Proteomics to study genes and genomes. *Nature*, 405, p837-46.

Pastor-Satorras, R., Vazquez, A. & Vespignani, A. (2001). *Phys. Rev. Lett*., 87, 258701.

Pastor-Satorras, R. & Vespignani, A. (2004). *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, Cambridge.

Rain, J.-C., Selig, L., DeReuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schächter, V., Chemama, Y., Labigne, A. & Legrain, P. (2001). The protein-protein interaction map of *Helicobacter pylori*. *Nature*, 409, p211-15.

Rao, C.V. & Arkin, A.P. (2001). Control motifs for intracellular regulatory networks. *Annu. Rev. Biomed. Eng.*, 3, p391.

Ravasz, E. & Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E*, 67, 026112.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. & Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, 297, p1551-5.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* 4, p131-134.

Schuster, S., Fell, D. A. & Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechn*. 18, p326–332.

Schwikowski, B., Uetz, P., & Fields, S. (2000). A network of protein-protein interactions in yeast. *Nature Biotechn.*, 18, p1257-61.

Segre, D., Vitkup, D. & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci.*, 99, p15112-7.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, p190–193.

Strogatz, S.H. (2001). Exploring complex networks. *Nature*, 410, p268-76.

Uetz, P., Giot, L., Cagney, G., Mansfield, T., Judson, R., Knight, J., Lockshorn, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M.J., Johnston, M., Fields, S. & Rothberg, J.M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403, p623-27.

Vázquez, A., Pastor-Satorras, R. & Vespignani, A. (2002). Large-scale topological and dynamical properties of the Internet. *Phys. Rev. E*, 65, 066130.

Walhout, A., Sordella, R., Lu, X., Hartley, J., Temple, G., Brasch, M., Thierry-Mieg, N., & Vidal, M. (2000). Protein interaction mapping in *C. elegans* using proteins involved in vulva development. *Science*, 287, p116-22.

Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press, Cambridge.

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of small-world networks. *Nature*, 393, p440-2.

Yook, S.-H., Jeong, H., Barabási, A.-L. & Tu, Y. (2001). Weighted evolving networks. *Phys. Rev. Lett.*, 86, p5835-38.

MATHEMATICAL MODELING OF NEURAL ACTIVITY

GAUTE T. EINEVOLL
Department of Mathematical Sciences and Technology
Norwegian University of Life Sciences, 1432 Ås, Norway
(`gaute.einevoll@umb.no`*)*

Abstract. The fantastic properties of the brain are due to an intricate interplay between billions of neurons (nerve cells) connected in a complex network. A central challenge is to understand this network behavior and establish connections between properties at the microscopic level (single neurons) and observed brain activity at the macroscopic systems level. After a brief introduction to the brain, cortex and neurons, various mathematical models describing single neurons are outlined: biophysically realistic compartmental models, simplified spiking neuron models and firing-rate models. Then examples of network modeling of the early visual system are described with particular emphasis on mechanistic ("physics-type") modeling of the response of relay cells in the dorsal lateral geniculate nucleus to visual spot stimuli. Finally an example of cortical population modeling related to the question of the neural mechanism behind short-term memory, is given.

Keywords: Computational neuroscience, neural networks, vision, cortical populations

1. Introduction

An important development in today's science is the increased use of methods from mathematics, computer science and theoretical physics in the exploration of biological systems. This is due to great advances in the understanding of living systems, establishment of new experimental techniques, methodological advances in mathematical modeling, and the continuing growth in available computer power for numerical calculations and simulations.

Neuroscience is among the biological sub-disciplines where the use of mathematical techniques are most established and recognized. An important reason for this is the success of Hodgkin and Huxley [1] 50 years ago of describing signal transport in a single neuron (nerve cell) as a modified electrical circuit where the charge carriers are Na^+, K^+, Ca^{++}, Cl^- and other ions flowing through the neuron cell membrane. This mathematical formulation, known as Hodgkin-Huxleytheory, could not only account for the results from

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 127–145.

experiments used to construct the model and fit the model parameters. From their model they could also predict the shape and velocity of the so called *action potential* which is a pulse-like electrical disturbance travelling down thin outgrowths, called *axons*, of neurons. From their model they calculated the propagation velocity of the action potential in their experimental system, the squid giant axon, to be 18.8 m/s which was roughly 10% off the experimental value of 21.2 m/s. Such quantitatively accurate model predictions are rare in theoretical biology. (Thorough introductions to Hodgkin-Huxley theory can be found in Refs. [2, 3, 4]).

Due to its obvious success in describing action potentials, the Hodgkin-Huxley approach has later been generalized to include modeling of the signal processing properties of entire neurons [5, 6]. Thus modelers now have a relatively firm starting point for mathematical explorations of neural activity. Hodgkin and Huxley's model has also been modified to account for membrane phenomena outside the nervous system, e.g., in the heart [7].

A single neuron is not particularly smart. The fantastic properties of the brain are due to an intricate interplay between billions of neurons connected in a complex network. A central challenge is to understand such network behavior, and establish connections between fairly well-understood properties at the microscopic level (single neurons) and measurements of brain activity at the macroscopic systems level using, for example, magnetoencephalography (MEG), electroencephalography (EEG), positron emission tomography (PET), or functional magnetic resonance imaging (fMRI).

Mathematical models in neuroscience can be distinguished by their purpose [4]: *Mechanistic* models aim to account for the properties of neurons or neural circuits on the basis of the underlying biophysical properties of neurons and neural networks. This corresponds to the traditional physics approach to modeling nature. *Descriptive* (or statistical) models try to account mathematically for experimental data without the aim to explain what aspects of the neurons or neuronal circuitry gives rise to the mathematical structure. *Interpretive* models aims to elucidate the functional roles of neural systems, i.e., relating neural responses to the task of processing useful information for the animal. Information theory is typically used in such modeling [4]. Interpretive modeling is unique to biological systems which have developed under evolutionary pressure. While it makes sense to ask *why* a neuron has particular properties, the question of *why* it is useful for a rock to fall to the ground is not fruitful.

The outline of this chapter is as follows: After a brief introduction to the structure of the brain and its constituent neurons, the various approaches for modeling signal processing in single neurons is outlined in section 3. This

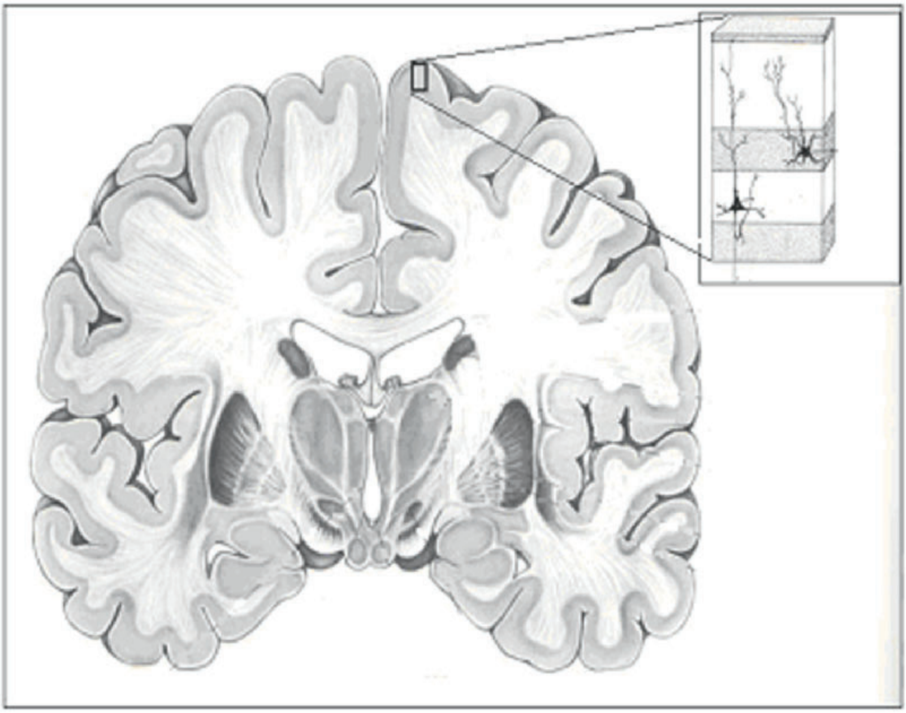


Figure 1. Cross-section of the human brain illustrating the structure of cortex. The cortical cells are located in a 2-4 mm thick layer at the brain surface (grey band). The inset illustrates the layered structure of cortex where different types of neurons are preferentially located in different layers. The inset also depicts two neurons with their cell bodies located in layers 4 and 5, respectively. (Traditionally, cortex is divided into six layers, but in the inset layers 2 and 3 have been merged.)

is followed by two examples of network modeling from our group [8] in sections 4 and 5.

2. Brief introduction to brain, cortex and nerve cells

The human brain consists of more than 100 billion neurons (nerve cells). In mammals the cortex is the most prominent structure and (as illustrated in Figure 1) consists of a convoluted thin (2-4 mm) layer of neurons located directly below the brain surface. A piece of human cortex is qualitatively similar to pieces of cortex in simpler mammals such as cats or rats. What distinguishes the human cortex from the cortices of smaller mammals is mainly the size; our cortex is much larger.

Cortex has a characteristic layered structure, where different types of neurons are preferentially located at different cortical depths (cf. Figure 1). Different parts of cortex perform different functions. For example, the part

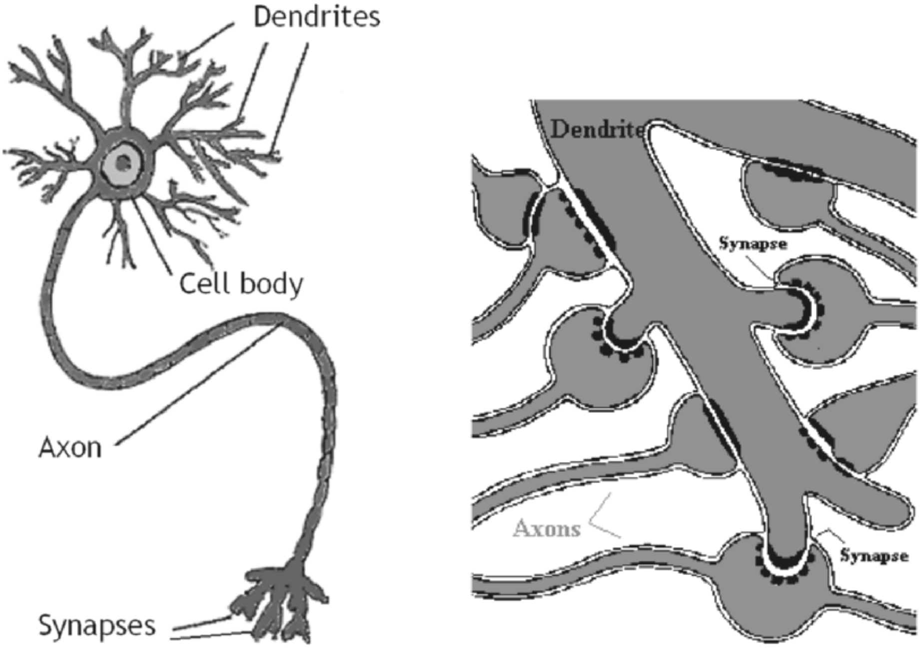


Figure 2. Left: Schematic illustration of neuron (nerve cell). Right: Schematic illustration of synapses functionally connecting two neurons.

dealing with processing of visual input is located in the back, while cortical areas associated with higher social skills are in the front. Nevertheless, the physical structure of cortex is quite similar all over. This suggests that evolution has found an efficient and flexible basic cortical circuit, which can be used in most of the various information processing and computing tasks the brain have to deal with.

In addition to cortex the brain also contains a number of smaller structures. An example is thalamus whose main task is to relay information from the sensory systems (for example visual information from the eye) to the appropriate cortical area.

A schematic illustration of a neuron is given in Figure 2 (left). Several branch-like structures protrude from the cell body. These *dendrites* receive signals from other neurons, and the signals are spread electrically to the *cell body*. If a sufficient amount of electrical charge arrives at the cell body within a certain time window, a pulse-like disturbance of the local *membrane potential*, i.e., the jump in voltage across the cell membrane, is spontaneously generated. This pulse-like disturbance is called an action potential. After generation in the cell body, the action potential will propagate without loss down the *axon*. At the other end, the axons typically branch out and form *synapses* with dendrites of other neurons. When an action potential reaches a synapse, designated molecules called neurotransmitters are released into the narrow cleft between two neurons (see Figure 2 (right)). These neurotransmitters rapidly diffuse across the cleft, and the arrival of these at designated receptors at the receiving neuron results in the injection of an electrical ion current. This electrical signal will then be spread down to the cell body of the receiving neuron, and the whole cycle can start again.

Action potentials appear to be the most important carrier of information between neurons. The shape and duration of these action potentials generated in a particular neuron vary little, and the information must thus be encoded in the sequence of time points when they are generated. But the question of what aspects of these temporal sequences carry information, i.e., the nature of the neural code, is still hotly debated. An important question is, for example, whether a single action potential encodes information independently from the other action potentials in the same neuron or whether correlations between action potentials carry significant amounts of information. It should be noted that there is no *a priori* reason that there should be a unique answer to the neural code question, i.e., different codes may be employed in different parts of the nervous system.

Neurons can be grouped in two main categories: *excitatory* and *inhibitory*. Synaptic inputs from excitatory neurons increase the probability for the receiving neurons to fire action potentials, while inputs from inhibitory neurons reduce the probability.

3. Modeling nerve cells

Various types of mathematical models are used to describe the signal processing properties of neurons. These model types are distinguished by their scope and the amount of biophysical details incorporated in the description.

3.1. COMPARTMENTAL MODELING

Compartmental modeling, illustrated in Figure 3, represents the highest level of detail. Here the neuron is divided into compartments, so small that the electrical potential can be assumed to be the same throughout the compartment. Every compartment is described as a small electrical circuit where the current is carried by ions (not electrons as in computers). The most important dynamical variable is the potential difference across the cell membrane, the membrane potential. The mathematical equation describing the dynamics of this potential follows from Kirchhoff's current law stating that current cannot vanish.

Figure 3 illustrates the construction of a compartmental model based on anatomically reconstructed dendritic trees. The mathematical equation

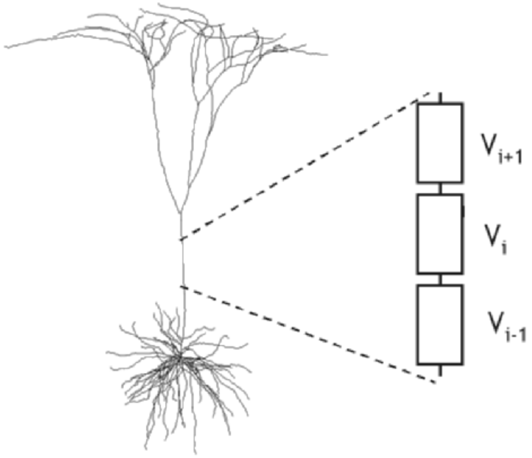


Figure 3. Illustration of principle for construction of a compartmental neuron model from an anatomically reconstructed neuron. The example neuron has been taken from Ref. [9].

describing the temporal development of the membrane potential V_i of compartment i in the figure is given by

$$g_{i,i+1}(V_{i+1} - V_i) - g_{i-1,i}(V_i - V_{i-1}) = c_i \frac{dV_i}{dt} + \sum_s I_i^s + \sum_j I_i^j \quad . \tag{1}$$

The two terms on the left hand side of the equation represent ohmic currents between compartment i and the neighboring compartments $i + 1$ and $i - 1$. The first term on the right hand side represents currents due to capacitive properties of the cell membrane, the second term represents currents due to synaptic inputs from other neurons, while the third term represents currents due to various other ion channels. Ion channels, which consists of particular proteins embedded in the cell membrane, are often selective and let only one type of ion through. The currents through the ion channels often depend in an intricate way on, for example, the membrane potential or ion concentrations. Synaptic input from other neurons are mediated through dedicated ion channels which can be opened by the arrival of particular signal molecules (neurotransmitters).

In Hodgkin and Huxley's model for action potentials in squid axons three ion channel currents were included: a sodium channel, a potassium channel and a "leak" channel incorporating other non-specified currents. They found that the generation and propagation of action potentials were due to sodium and potassium currents with a particular dependency on time and membrane potential [1].

In the full compartmental model for a neuron there will be an equation of the type shown in eq. (1) for each compartment, and the equation set must be solved numerically to determine the membrane potential over the entire neuron. The mathematical solution is in principle straightforward, and dedicated (and free) computer simulation tools such as NEURON [10] and GENESIS [6] are available. More information on compartmental modeling can be found in Refs. [4, 5, 6, 11].

3.2. SIMPLIFIED SPIKING MODELS

Compartmental modeling can be computationally expensive. Moreover, determination of the numerous parameters specifying the model is generally difficult. Various simplified neuron models are thus also used. For example, often the dendritic tree and cell body are collapsed into a single point, essentially saying that the membrane potential is the same throughout.

Further, the modeling of the dynamics of the membrane potential does often not include the generation of the action potential (*spike*) itself. Action potentials have a standardized all-or-none behavior, and it is a waste of computer resources to calculate the detailed time course every time. Instead the action potentials are generated by a separate rule: when the membrane potential reaches a preset threshold for action potential firing, an action potential is recorded, and the membrane voltage is reset to another predefined value. The *integrate-and-fire model* is the most commonly used model of this type [3, 4]. In this model the subthreshold dynamics of the membrane potential (in the single compartment) is given by

$$c\frac{dV}{dt} = -g_L(V - E_L) - \sum_s I^s \quad . \tag{2}$$

This equation follows directly from eq. (1) by (i) omitting the terms due to currents between compartments and (ii) including only an ohmic leak current (in addition to the synaptic currents). E_L corresponds to the resting potential, i.e., the membrane potential of the point neuron in the absence of synaptic inputs.

Even with this drastic simplification, it is still difficult to find analytical mathematical solutions. One generally has to resort to numerical solution also here, but they require less computer resources than full compartmental simulations.

3.3. FIRING RATE MODELS

At the coarsest level of detail, we have firing-rate models where only the *probability* for action-potential firing is modelled. Then the “activity level”

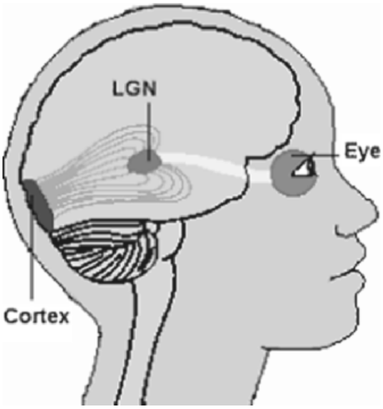


Figure 4. Illustration of early visual pathway.

(for example, membrane potential in the cell body) is typically converted to a firing rate R via a non-linear sigmoidal function, i.e., $R = P(V)$. This mimics the observed conversion of membrane potential in the cell body to firing of action potentials; the membrane potential has to be above a certain threshold before action potentials are generated.

Firing-rate models are generally described by differential or integro-differential equations, and the extensive experience in analyzing and solving such equations can be utilized. Sometimes each neuron is modelled as a single discrete unit (see example in section 4), but there is also a long tradition for considering neural field models where, for example, cortical tissue is modelled as continuous lines or sheets of neurons [4, 12, 13] (see example in section 5) .

4. Network modeling of early visual pathway

4.1. EARLY VISUAL PATHWAY

The study of vision has been a central research area in neuroscience. One reason might be that humans are visual animals in the sense that a large portion of our cortex is devoted to the processing of visual information, and that the understanding of this sense thus is of particular interest to us. Another reason is that the study of the visual system is experimentally and conceptually easier than studies of other sensory systems (hearing, smelling, etc.).

When light hits the eye, neurons in the retina on the back side of the eye get excited. Several types of neurons are involved in the signal processing in the retina, but the output action potentials, which are transmitted towards cortex, leave from *retinal ganglion cells*. The ganglion cell axons constitute the optical nerve. This nerve transmits visual information to a part of the brain called *dorsal lateral geniculate nucleus (dLGN)*, which is a part of thalamus. The relay cells in dLGN receive visual signals from retinal ganglion cells and transmit processed information to the *primary visual cortex*. From primary visual cortex the signals are then fed to other parts of cortex (as well as other brain structures), and this eventually results in a visual perception of the surrounding world.

An important notion in studies of the visual system is the *receptive field* of a neuron. This term refers to the limited area of the visual field where stimulation with light (or darkness) influences the firing of action potentials in the neuron. Visual stimulation of an area inside the receptive field of a neuron can both increase and reduce the firing activity. For retinal ganglion cells and cells in the dLGN the receptive fields are small, roughly circular, areas, and they exhibit center-surround antagonism. This means that the cells have highest response when stimulated by a circular spot of light (on a dark background) exactly covering the *receptive-field center*. Illumination of an area outside this receptive-field center, on the other hand, will reduce the firing activity. Therefore, the receptive field can be described as a circular excitatory area surrounded by a ring-shaped inhibitory area. Neurons responding to light spots in this way are called *on*-cells. *Off*-cells have opposite response, i.e., they have largest activity when a dark spot covers the receptive-field center. The antagonistic center-surround organization makes the system more suited to detect changes in the light intensity than the absolute magnitude of the intensity. In primary visual cortex the receptive fields are more complex, and one finds cells that respond best to elongated (e.g., edges, bars) and/or moving stimuli.

At the levels of retina and dLGN the on- and off-cells form two roughly independent networks of cells both covering the entire visual field. In addition, these cells also group into three distinct classes (labeled, for example, X,Y and W in cat) distinguished by their physiological properties such as receptive-field size and signal propagation velocity [14]. These cell populations seem to be essentially uncoupled before reaching the visual cortex so that there exist six independent pathways (in cat: on/off for each of X,Y, and W). These pathways carry different types of information regarding the visual input; information which somehow are merged at a higher processing level to build our mental representation of the visual world.

4.2. DESCRIPTIVE MODELING OF RECEPTIVE FIELDS

A number of descriptive models have been suggested to describe receptive fields of neurons in retina, dLGN and primary visual cortex [4]. An example is the “difference-of-Gaussians”-model (DOG-model) describing the circular

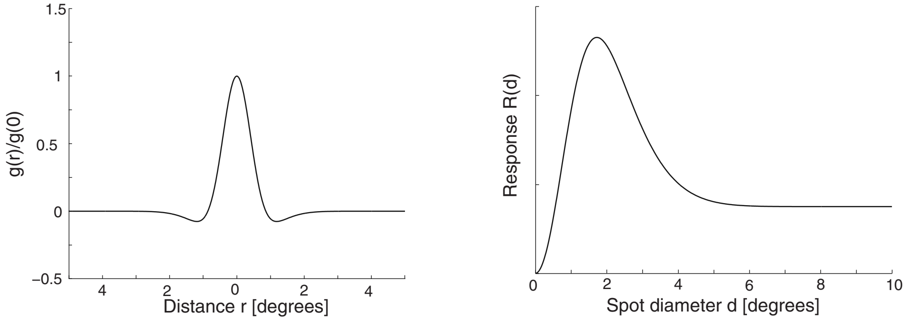


Figure 5. Illustration of DOG model for description of spatial receptive fields for neurons in retina and dLGN. Left: Example of receptive-field function $g(r)$ in eq. (3) with parameters $A_1 = 1$, $A_2 = 0.85$, $a_1 = 0.6$ deg, and $a_2 = 1.2$ deg. Right: Response for centered circular spot as a function of spot diameter for the same DOG model, eq. (4).

center- surround organization of spatial receptive fields for neurons in the retina and dLGN,

$$g(r) = \frac{A_1}{\pi a_1^2} e^{-r^2/a_1^2} - \frac{A_2}{\pi a_2^2} e^{-r^2/a_2^2} \quad . \tag{3}$$

Here A_1 and A_2 are weight parameters, and a_1 and a_2 ($a_2 > a_1$) describe the spatial extension of the receptive field. This *receptive-field function* $g(r)$ describes the change in activity (f.ex., number of action potentials fired during a particular time interval) in a particular neuron when a tiny test spot is shown in the position $\mathbf{r}$ in the visual field. The position $\mathbf{r} = \mathbf{0}$ corresponds to the center point in the receptive field of the neuron. The visual field has two spatial dimensions so the position vector $\mathbf{r}$ is two-dimensional, $\mathbf{r} = [x, y]$. The DOG-model has circular symmetry, however, so the model response to the test spot depends only on the distance $r = |\mathbf{r}|$ from the center point.

The parameters in the DOG-model for a particular neuron can be determined by fitting the formula in eq. (3) to results from a test-spot experiment. But the model would have limited interest if it only could account for a single experiment. However, many neurons in the early visual systems exhibit linear response properties, i.e., the response to other visual stimuli can be found by linear summation over contribution from many "imagined" test spots. For example, the response $R(d)$ to a circular spot of light with diameter d (not necessarily small) centered at the receptive field, can be found from a simple integral,

$$\begin{aligned} R(d) &= \Delta L^* \int_0^{d/2} \int_0^{2\pi} g(r)\, d\theta\, r dr \\ &= \Delta L^* \Big(A_1 (1 - e^{-d^2/4a_1^2}) - A_2 (1 - e^{-d^2/4a_2^2}) \Big) \quad , \end{aligned} \tag{4}$$

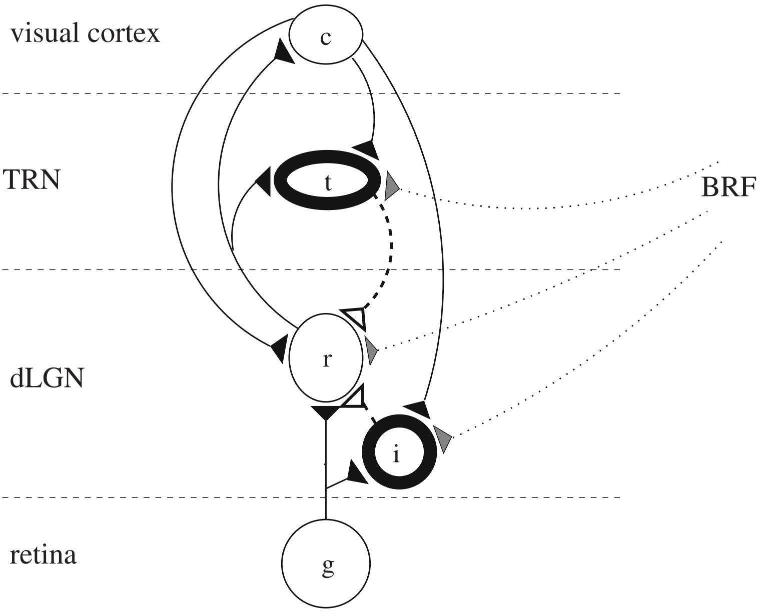


Figure 6. Schematic overview of coupling pattern in dLGN circuit. The relay cells (*r*) and interneurons (*i*) in dLGN receive their primary input from retinal ganglion cells (*g*) as well as feedback from cells (*t*) in the thalamic reticular nucleus (TRN) cells and cells in the visual cortex (*c*). The relay cells, interneurons and TRN cells also receive modulatory inputs from the brainstem reticular formation (BRF). The excitatory connections are shown as solid lines, the inhibitory as dashed lines, while dotted lines symbolize the modulatory connections.

where the constant ΔL^* depends on the luminance of the spot. Correspondingly, the response to any visual stimuli can be calculated when the parameters A_1, A_2, a_1, and a_2 are known.

An example of a DOG receptive-field function (eq. 3) with the corresponding spot-response (eq. 4) is shown in Figure 5. Note that the spot-response curve has a maximum for a diameter d_{max} around 2 degrees. This maximum corresponds to the transition from the excitatory center area to the inhibitory surround, and d_{max} is thus the receptive-field center diameter.

4.3. MECHANISTIC MODELING OF RECEPTIVE-FIELD PROPERTIES IN DLGN

In our group at the Norwegian University of Life Sciences we have for some time worked on mathematical modeling of the early visual pathway with a particular emphasis on dLGN (see, for example, Refs. [15, 16, 17]). This work has been done in close collaboration with the experimental group of Paul Heggelund at the Department of Physiology at the University of Oslo. Our aim has primarily been to develop mechanistic ("physics-type") models

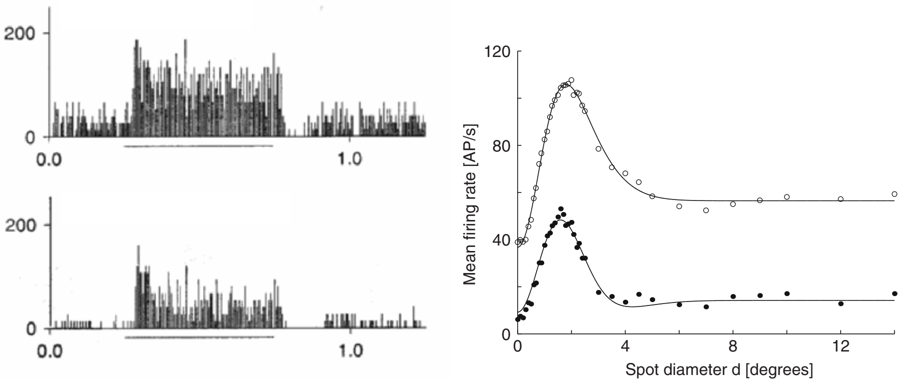


Figure 7. Example results for X relay cell in dLGN from Ruksenas *et al.* [18]. Left: Post-stimulus time histograms (PSTHs) for spot diameter $d = 1.6$ degrees for retinal input (S-potentials, top) and relay-cell action potentials (bottom). The PSTHs, which have been found by averaging data from many trials, give a measure for the time-dependent firing rate (action potentials/seconds). The horizontal lines show the time the stimulus is on. Right: Mean firing rate, found by temporal averaging of PSTHs of the type shown left, for a set of different spot diameters. Open dots correspond to retinal input, filled dots to relay-cell action potentials. The curves show the best fit to the mathematical models described in the text [15].

of the formation of receptive fields of neurons in dLGN, i.e., understand how these receptive fields are formed on the basis of the properties of the relevant neurons and their connections. This endeavor requires quite a bit of knowledge of biological details, but fortunately dLGN is (compared to other brain parts) modestly complex. Following significant experimental efforts during the last decades, the dLGN circuit is now mapped out relatively well [14]. A schematic overview is given in Figure 6, where both feedforward (from the eye) and feedback (from, e.g., visual cortex) connections are present.

The relay cells in dLGN are excited by retinal ganglion cells. In addition they receive input from inhibitory *interneurons*, which in turn are excited by retinal ganglion cells. The relay cells further receive feedback input from inhibitory neurons in a brain part called *thalamic reticular nucleus* (TRN). In addition there is a massive feedback from cortical neurons to relay cells, interneuron and TRN neurons. This feedback shows that the early visual system is more than a passive camera; cortex participates in deciding how visual information is collected.

As an example of network modeling in the early visual system I will here describe a project [15] where the response of dLGN relay cells to circular light or dark spots was modelled based on experiments done in Paul Heggelund's group. Ruksenas, Fjeld, Heggelund measured action potentials in X-type relay cells in cat dLGN [18]. For such neurons the excitatory input from retina appears to stem from a single ganglion cell. With a single

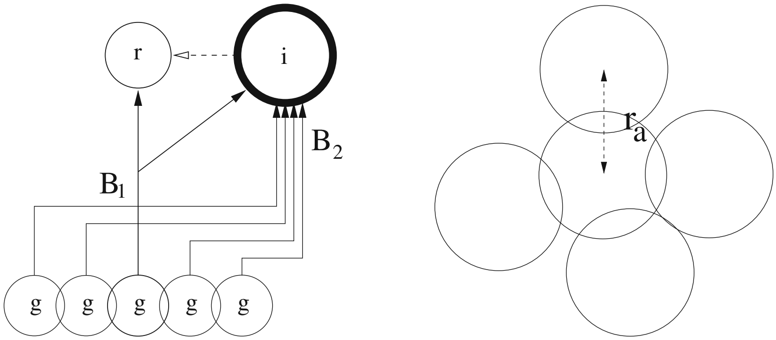


Figure 8. Illustration of synaptic coupling pattern assumed in mathematical model. Left: The model assumes a single excitatory input to the relay cell (*r*) and (in this case) five inputs to a single interneuron (*i*) which in turn gives inhibitory input to the relay cell. Right: Illustration of spatial distribution of receptive-field centers for the five ganglion cells providing input to the interneuron. The center ganglion cell gives in addition direct excitatory input to the relay cell.

electrode placed outside the relay cell, one can in addition to recording relay-cell action potentials also record so called S-potentials. These S-potentials have been shown to correspond to an action potential in the retinal ganglion cells providing the input. One thus has experimental access to both the input to and the output from the relay cell in question. The stimulus spots were displayed for half a second, and by repeating the experiment many times so called *post-stimulus time histograms* (PSTHs) for action potentials and S-potentials were measured (see Figure 7 (left)). From these measurements the mean firing rate (essentially the mean number of action potentials) while the stimulus is on, was calculated. The experiment was done for a set of spot diameters, and an example of the diameter dependence of the response is shown in Figure 7 (right). Here we see that both the retinal input and the output from the relay cell have a maximum for a particular spot diameter, in qualitative agreement with the prediction from a DOG model (cf. Figure 5 (right)). We also observe that this maximum occurs for a slightly smaller diameter for the output curve (relay-cell response) than for the input curve (retinal ganglion-cell response). Further we see that the response for very large spot diameters is more reduced for the relay cell, i.e., larger center-surround antagonism. The same observations were generally made for the 22 X-cells recorded from.

The goal of the modeling project was to investigate whether this change between input and output could be accounted for based on knowledge about the dLGN circuit.

The response of TRN-neurons and neurons in primary visual cortex to small circular spots is limited. Based also on other knowledge about the system, we thus focused on a simple feedforward circuit where a relay cell receives (i) excitatory input from a single retinal ganglion cell and (ii) inhibitory input from an interneuron which in turn is driven by a group (~ 5) of retinal ganglion cells. This group includes the ganglion cell which provides the excitatory input, and its surrounding nearest neighbors. In the model these neighbors are assumed to be located the same distance r_a from the center neuron (see Figure 8).

In our simple model we assume that the mean firing rate for the relay cell $R_r(d)$ essentially is a direct function of the weighted difference between the excitatory and inhibitory inputs, i.e.,

$$R_r(d) = P\Big(B_1 R_g(d;0) - B_2 R_g(d;r_a)\Big)\,, \tag{5}$$

where B_1 is the net weight (excitation minus inhibition) from the central ganglion cell, while B_2 is the total inhibitory weight due to the surrounding ganglion cells. $R_g(d;0)$ is the response of the central ganglion cell to a spot with diameter d. This is assumed to be given by the expression for the descriptive DOG-model in eq. (4). $R_g(d;r_a)$ is the response of the neighboring ganglion cells assumed to have the same receptive-field function given in eq. (3) (with the same model parameters). Since the circular spot is not centered on their receptive field, the diameter dependence of the response will be different. The integral in eq. (4) must be redone, and $R_g(d;r_a)$ is found to be given by a series of incomplete gamma functions [15]. To avoid unphysical negative firing rates in the model the rectifying function $P(x)$ is used in the conversion of net input to firing rate for the relay cell, i.e., $P(x) = x\,\theta(x)$ where $\theta(x)$ is the Heaviside step function.

In Ref. [15] we first fitted a DOG-model expression (analogous to eq. (3)) to the experimentally observed response curve for the retinal input in Figure 7 (right). With the retinal DOG parameters fixed, eq. (5) could then be used to determine the three remaining parameters B_1, B_2, and r_a by fitting to the relay-cell response curve shown in the same figure.

As observed in Figure 7 the mathematical models could be fitted to the example data for both the retinal input and the relay-cell response. Correspondingly, we found that the models could be fitted to data for the 21 other X-cells recorded from. But with so many model parameters, successful fits do not necessarily imply that the model reflects reality. However, the modeling also produced specific predictions which could be tested against other experiments. We found, for example, that the model predictions for the distance between neighboring ganglion cells r_a and the size of the interneuron receptive-field center were compatible with experimental results from the literature [15].

In the model above only spatial properties of the receptive field of relay cells are considered, but the temporal properties are just as important and interesting. Mathematical description for spatiotemporal receptive fields have thus also been developed [4, 16].

Firing-rate models have its limitations, however. A relay cell can, for example, in certain situations generate a so called burst of action potentials, i.e., a rapid sequence of action potentials. A more detailed neuron model is necessary to include this effect in network studies. In our group we have therefore also started computationally more demanding studies of the early visual pathway where each neuron is described using the "integrate-and-fire-or-burst" model [17] (a generalization of the integrate-and-fire model in eq. (2)). The model results here are explicit sequences of time points when action potentials are fired, and not firing rates (or mean firing rates as above). A simulation of the whole dLGN circuit for a part of the visual field will include tens of thousands of neurons, and the general network simulation program NEST [19] is used for an efficient implementation of the model.

5. Cortical population modeling

The simple network model described above relates the response of neurons to external stimuli, in this case visual stimuli. Most of the network modeling that has been done so far has been on stimulus-driven responses, but in cortex most of the neural activity is probably generated internally. A well-known example is the observation of persistent firing activity (over many seconds) in neurons in prefrontal cortex related to short-term memory [20].

As an example of modeling of internally generated activity we have recently studied a two-population model where an excitatory and an inhibitory population of neurons (for example, localized in the same cortical layer) interact, and investigated the possibility for stable spatially localized pulses ("bumps") of activity [21]. Such stable structures could possibly account for the abovementioned observed persistent firing activity.

In the model a one-dimensional continuum of neurons is assumed,

$$\begin{aligned}\frac{\partial u_e}{\partial t} &= -u_e + \int_{-\infty}^{\infty} \omega_{ee}(x-x')P(u_e(x',t)-\theta_e)dx' \\ &\quad - \int_{-\infty}^{\infty} \omega_{ie}(x-x')P(u_i(x',t)-\theta_i)dx' \\ \tau\frac{\partial u_i}{\partial t} &= -u_i + \int_{-\infty}^{\infty} \omega_{ei}(x-x')P(u_e(x',t)-\theta_e)dx' \\ &\quad - \int_{-\infty}^{\infty} \omega_{ii}(x-x')P(u_i(x',t)-\theta_i)dx' . \qquad (6)\end{aligned}$$

Here $u_e(x,t)$ and $u_i(x,t)$ represent the membrane potential in the cell body of an excitatory and an inhibitory neuron, respectively, at the position x. Further, $\omega_{mn}(x)$ describes the spatial synaptic coupling functions and is here given by

$$\omega_{mn}(x) = \frac{1}{\sigma_{mn}\sqrt{\pi}}\, e^{-x^2/\sigma_{mn}^2} , \quad m,n = e \text{ or } i \ . \qquad (7)$$

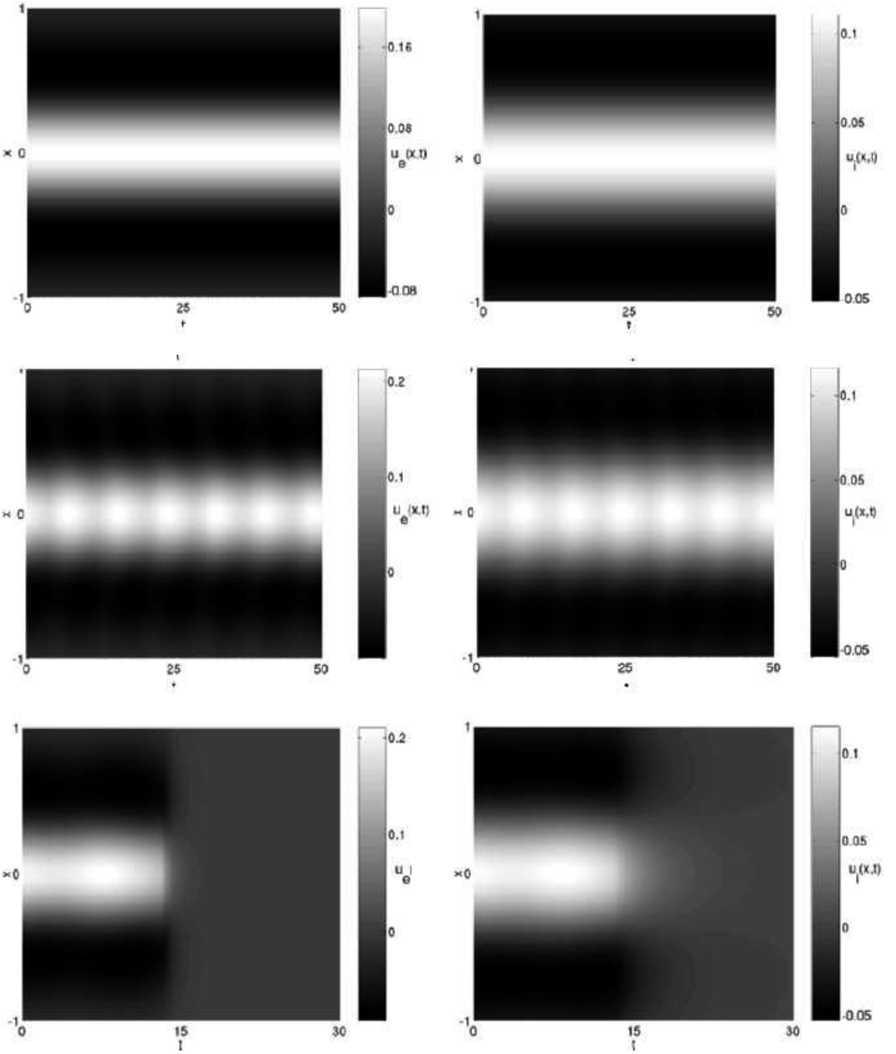


Figure 9. Temporal evolution of pulse pairs for model given by eq. (6) with ω_{mn} described by eq. (7). Excitatory activities $u_e(x, t)$ are shown in left column, inhibitory activities $u_i(x, t)$ in right column. Top row: $\tau = 3$, stable pulses. Middle row: $\tau = 3.03$, oscillatory pulses. Bottom row: $\tau = 3.5$, pulses collapse. Other parameter values: $\sigma_{ee} = 0.35$, $\sigma_{ei} = 0.48$, $\sigma_{ie} = 0.60$, $\sigma_{ii} = 0.69$, $\theta_e = 0.12$, $\theta_i = 0.08$. Adapted from Ref. [21].

$P(x)$ (assumed to be Heaviside's step function) gives the firing rate from the membrane potential, θ_n gives the threshold values for firing, and τ is the time constant for inhibition (measured relative to the time constant for excitation).

Note that there is no external input in the model; the populations receive all input from themselves and each other.

Since $P(x)$ is modelled as a Heaviside step function, the model is highly nonlinear, and the mathematical analysis more involved than for the relay-cell model in section 4. We will not go into details here, but in Ref. [21] we found that the model could support stable localized pulses (i.e., suitable pairs of pulses of excitatory and inhibitory activity) for a wide range of parameter values. As illustrated in Figure 9 we further found that the inhibitory time constant τ could control the stability of the pulses. Pulses found to be stable for small or moderate τ could start to oscillate when τ was increased beyond a critical value (by a so called Hopf bifurcation). A further increase of τ eventually resulted in collapse of the pulses.

A conclusion from our study was that persistent (stable) activity, for example related to short-term memory, *could* be explained by such a model, but this is of course not proof that this mechanism is employed by the brain.

6. Final comments

Model studies of biological neural networks have just started. While one, particularly for the early visual system, has had some success in describing stimulus-driven responses, the modeling of cortical population activity has generally been limited to qualitative studies of the type of phenomena simple models can exhibit (stable pulses, traveling waves and fronts, oscillatory activity, etc.). However, more details of the brain structure are steadily revealed due to improved experimental techniques, and the foundation for building biologically more realistic cortical models gradually improves. With access to even more powerful computers, a significant research effort on large, complex neural networks can be expected in the years to come.

Another challenge is to relate results from such network models to experiments. The traditional neurophysiological method of measuring action potentials with a sharp electrode placed inside or in the immediate vicinity of the cell body, has given a wealth of insight into the properties of single neurons. In cortex, however, information appears to be encoded in population activity and not in the firing pattern of single neurons (which typically is highly variable and “noisy”). A variety of experimental methods based on different physical effects (optical, electrical, magnetic) have been developed to study brain activity at the population level (~ 0.1-1 mm). These methods probe different aspects of neural activity. For example, the electrical potentials measured at the scalp in EEG appear to be related to ion currents across the cell membrane in dendrites following synaptic input, and not to the direct firing of action potentials. To model the contribution to EEG from a population of neurons one thus has to employ full compartmental models (even if a simpler model might be sufficient to model the firing pattern of action potentials). For electrical and magnetic measurements (for example, EEG and MEG) the link from the underlying neural activity is in principle clear, while a similar link to hemodynamic (for example, fMRI) or metabolic (for example, PET) methods presently is more obscure.

Acknowledgments

The author wish to thank Patrick Blomquist for making Figure 9.

References

1. Hodgkin, A.L. and Huxley, A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve, *Journal of Physiology* 117, 500-544.
2. Johnston, D. and Wu, S.M-S. (1994) *Foundations of Cellular Neurophysiology*, MIT Press, Cambridge.
3. Koch, C. (1999) *Biophysics of Computation*, Oxford, New York.
4. Dayan, P. and Abbott, L.F. (2001) *Theoretical Neuroscience*, MIT Press, Cambridge.
5. Koch, C. and Segev, I. (eds.) (1998) *Methods in Neuronal Modeling (2. ed.)*, MIT Press, Cambridge.
6. Bower, J.M. and Beeman, D (eds.) (1998) *The Book of Genesis: Exploring Realistic Neural Models with the General Neural Simulation System (2. ed.)*, Springer, New York.
7. Noble, D. (1979) *The Initiation of the Heartbeat*, Clarendon Press, Oxford.
8. Information about the computational neuroscience group at the Norwegian University of Life Sciences can be found at http://arken.umb.no/~compneuro.
9. Mainen, Z.F. and Sejnowski, T.J. (1996) Influence of dendritic structure on firing pattern in model neocortical neurons, *Nature* 382, 363-366. The example neuron has been downloaded from http://www.cnl.salk.edu/ zach/patdemo.html .
10. See http://www.neuron.yale.edu/.
11. De Schutter, E. (ed.) (2000) *Computational Neuroscience: Realistic Modelling for Experimentalists*, CRC Press, Boca Raton, Florida.
12. Wilson, H.R. and Cowan, J.D. (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, *Kybernetik* 13, 55-80.
13. Coombes, S. (2005) Waves, bumps, and patterns in neural field theories, *Biological Cybernetics*.
14. Sherman, S.M and Guillery, R.W. (2001) *Exploring the thalamus*, Academic, New York.
15. Einevoll, G.T. and Heggelund, P. (2000) Mathematical models for the spatial receptive-field organization of nonlagged X cells in dorsal lateral geniculate nucleus of cat, *Visual Neuroscience* 17, 871-886.
16. Einevoll, G.T. and Plesser, H.E. (2002) Linear mechanistic models for the dorsal lateral geniculate nucleus of cat probed using drifting-grating stimuli, *Network: Comput. Neural Systems* 13, 503-530.
17. Plesser, H.E., Einevoll, G.T. and Heggelund, P. (2002) Mechanistic modeling of the retinogeniculate circuit in cat, *Neurocomputing* 44-46, 973-978.

18. Ruksenas, O., Fjeld, I.T., and Heggelund, P. (2000) Spatial summation and center-surround antagonism in the receptive field of single units in the dorsal lateral geniculate nucleus of cat. Comparison with retinal input, *Visual Neuroscience* 17, 855-870.
19. See http://www.synod.uni-freiburg.de/ .
20. Wang, X.J. (2001) Synaptic reverbation underlying mnemonic persistent activity, *Trends in Neuroscience* 24, 455-463 (2001).
21. Blomquist, P., Wyller, J. and Einevoll, G.T. (2005) Localized activity patterns in two-population neuronal networks, *Physica D* 206, 180-212.

BRAIDED SPACE-TIME PARTICLE NETWORKS

K. de LANGE KRISTIANSEN[1,2], G. HELGESEN[1] AND
A.T. SKJELTORP[1,2]
[1] Institute for Energy Technology
POB 40, NO-2027 Kjeller, Norway
[2] Department of Physics, University of Oslo
POB 1048, Blindern, NO-0316 Oslo, Norway

1. Introduction

We study the complex dynamics of microspheres dispersed in ferrofluids subjected to external oscillating magnetic fields, see Fig. 1.

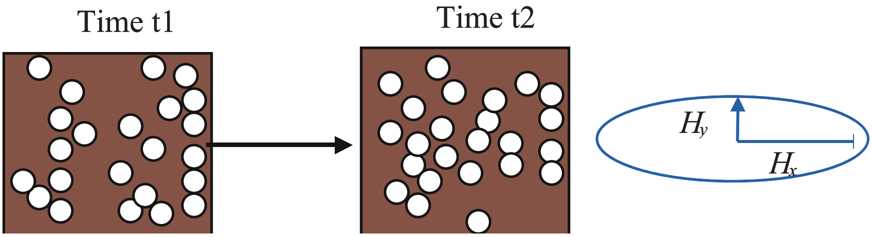


Figure 1. Typical snapshots of microspheres in a ferrofluid confined to a plane and subjected to ac magnetic fields.

The paths traced out by the microspheres may generally be very intricate and it is not trivial to give a compact description of the dynamic behavior. We have to rely on approximations and computer simulations that can predict the trajectories of the particles in the short term, but will eventually be so uncertain that we have to give up in the long term.

If one moves into the world of many-body systems of atoms and molecules, diverse cooperative phenomena occur and can be described using various well-established mathematical and computational tools such as statistical mechanics, Monte Carlo- and molecular dynamics simulations etc. It would be of interest to find out whether some mathematical tools might be found to describe also few-particle phenomena like the random motion of particles in a plane described above, and in the limit of an increasing number of particles reach the many-body situation.

What we do is to introduce the concept of “world lines”, a space-time description of the particle trajectories. One can compare this with an air show where the smoke trails left behind the planes represent the “frozen” dynamics of the planes. One can easily picture that the microspheres also trace out complicated space-time trajectories, which are interwoven in intricate patterns and networks, as shown in Fig. 2.

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 147–162.

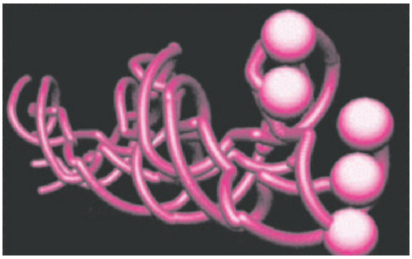


Figure 2. Space-time trajectories of microspheres in a ferrofluid subjected to ac magnetic fields.

A mathematical tool which can be used to analyze this is braid theory. It proves to be useful for a compact description of space-time particle network trajectories and statistical analysis. The complex dynamical modes describing the physical behavior of interacting particles in the present system may be distributed over a large energy range. As will be seen, it is an intermittent systems and the the modes have a broad distribution of frequencies. One way of making a systematic study of these modes is by ordering them after their occurrence frequencies. We let the mode which occur most often be ranked as number 1, the second most occurring mode be ranked number 2, and so forth to the least occurring mode which have the highest rank number. This rank-ordering statistics takes care of both common and rare events and is useful for displaying the whole range of the mode distribution.

A remarkable feature found by rank-ordering statistics in many different systems is the Zipf relation. This relation has been demonstrated for a variety of applications, such as linguistics [1], energy distribution of earthquakes [2] and in analyzing the coding and non-coding regions of DNA sequences [3]. The original Zipf relation came into being in an empirical manner in linguistics. By analyzing the occurrence of words in large written texts G.K. Zipf proposed a simple power law $\phi(r) \sim r^{-\gamma}$ with $\gamma = 1$, where $\phi(r)$ is the frequency of occurrence of a word with rank r [1]. An attempt to understand the origin of this relation was connected to the hierarchical structures of languages [4], and gives the corrected Zipf-Mandelbrot relation [5]:

$$\phi(r) = \frac{A}{(r+\zeta)^{\gamma}}, \tag{1}$$

where A is a normalization constant and ζ is a parameter. In this work we will report experiments and numerical simulations demonstrating this Zipf-Mandelbrot relation.

2. Experimental system

The experimental system is essentially made up of two advanced materials (i) ferrofluid and (ii) monodisperse polystyrene microspheres [6]. The basic features of a ferrofluid [7] are shown below in Fig. 3 with basically three length scales: (i) macroscopic length scale, it resembles an ordinary liquid; (ii) colloidal length scale, solid nanoparticle dispersed in a liquid; and (iii) nanoscale, each particle consists of a single domain magnetic iron oxide core, and a surface grafted with polymer chains (particle size ~ 10 nm).

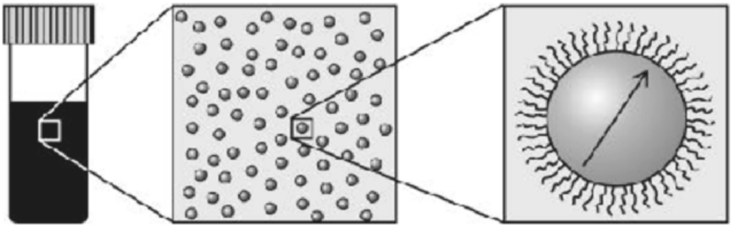


Figure 3. Characteristic features of a ferrofluid from nano- to macroscale as discussed in the text.

In order to model a many-body system, it is important to have particles of the same shape and size. The unique Ugelstad microspheres [8] offer this possibility. This is demonstrated in the picture shown in Fig. 4. Here, 10 micrometer polystyrene spheres are packed to a monolayer on a glass plate. The average uniformity is better than 2 %.

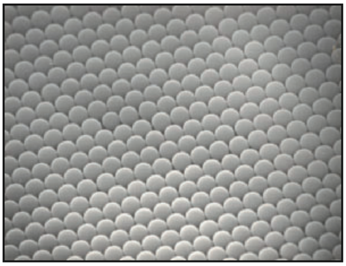


Figure 4. SEM picture of a monolayer of closed packed 10 micrometer Ugelstad polystyrene spheres.

By dispersing the microspheres into ferrofluids one obtains a so-called "magnetic hole" system [9], and a workable experimental realization is shown below in Fig. 5.

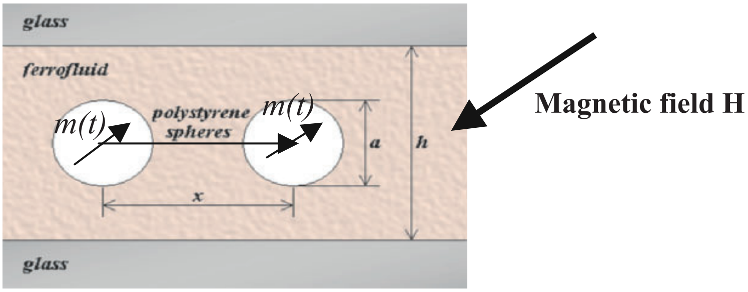


Figure 5. Cross section of an experimental cell with polystyrene microspheres dispersed in ferrofluid confined between two plane parallel glass plates. The effective magnetic moment of the "hole"created by the sphere, *m(t)*, is equal to the magnetic moment of the displaced ferrofluid, and *m(t)* and **H** are co-linear, but with opposite directions. The sphere diameter *a* is typically 10-100 μm and the separation between the glass plates typically $h \approx 2a$.

If V is the sphere volume, χ is the effective magnetic susceptibility and **H** is the external magnetic field, then the magnetic moment of the "magnetic hole" is

$$\mathbf{m}(t) = -\chi_{eff} V \mathbf{H}(t), \tag{2}$$

where $\chi_{eff} = \frac{3\chi}{3+2\chi}$ includes the demagnetization factor of a sphere, and χ is the bulk magnetic susceptibility of the ferrofluid [10]. Demagnetization and mirror effects, etc must also be considered in a more precise calculation.

Figure 6 shows the details of the experimental setup.

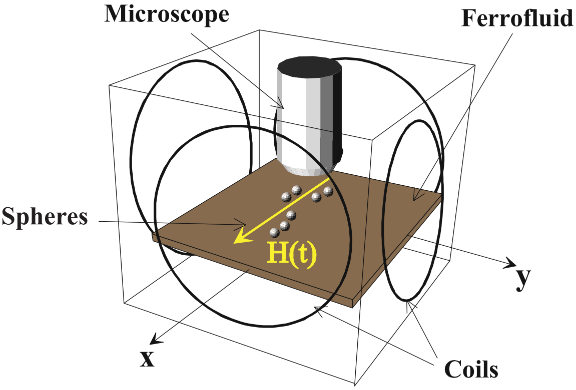


Figure 6. The experimental setup.

The external planar, elliptically polarized field $\mathbf{H}(t)$ is produced by two pairs of coils mounted perpendicular to each other carrying AC current with a phase difference of $\pi/2$. The amplitude components H_x and H_y are varied by adjusting the maximum current through the coils. The field anisotropy parameter is defined as $\varepsilon = H_y/H_x$. For the experiments presented in this study we use $H_x = 27$ Oe.

In a slowly rotating external magnetic field, the microspheres try to line up in chains parallel with the field. Due to viscous forces there will be a phase lag between the direction of the magnetic field and the chains of microspheres. The chains will furthermore break up in smaller subchains and perform a half rotation before aligning again. For example, a chain of 7 microspheres was observed to split into groups of $3+2+2$ or $3+3+1$. In an elliptically polarized magnetic field the division into subchains is so irregular that we get an apparently disordered behavior. In the frequency range studied here, the microspheres are nearly co-linear every half period of the external field. This suggests that in order to investigate the main dynamics we can focus on what happens along this axis.

3. The simulation model

The magnetic holes interact via dipole interactions, and the interaction energy U for a collection of n magnetic holes is given by:

$$U(\mathbf{r}_1,\ldots,\mathbf{r}_n,t) = \begin{cases} \sum_{i>j}^{n} \frac{m^2(t)}{r_{ij}^3} - \frac{3\cdot[\mathbf{m}(t)\cdot\mathbf{r}_{ij}]^2}{r_{ij}^5}, & r_{ij} > d \\ \infty, & r_{ij} < d \end{cases}, \tag{3}$$

where $r_{ij} = r_i - r_j$ is the vector joining the centers of the magnetic holes and d is the diameter of a microsphere. The magnetic force on particle i is then:

$$\mathbf{F}_i^M = \frac{\partial U(\mathbf{r}_1,\ldots,\mathbf{r}_n,t)}{\partial \mathbf{r}_i}. \tag{4}$$

The viscous force on a microsphere is to first order given by Stokes' law:

$$\mathbf{F}^{Stokes} = 3\pi\eta d\mathbf{v}, \tag{5}$$

where η is the viscosity of the ferrofluid and $\mathbf{v}$ the velocity of the sphere. Since the diameter of the microspheres and their velocities are relatively small, the Reynolds number, $Re = \rho v d/\eta \ll 1$, where ρ is the ferrofluid density. The system is therefore overdamped and we may neglect the inertia forces. By assuming equilibrium between the magnetic and viscous forces,

$$\sum \mathbf{F} = \mathbf{F}^M + \mathbf{F}^{Stokes} = 0, \tag{6}$$

we can easily transform the equation of motion into a numerically solvable form [11].

The angular velocity ω_H of the external field is normalized by the critical angular velocity ω_c for stable rotation of two microspheres with a circularly polarized magnetic field [11]:

$$\omega = \omega_H / \omega_c \,, \tag{7}$$

where $\omega_H = 2\pi f$ and f is the frequency of the rotating magnetic field. At ω_c a chain of two microspheres starts to show phase-slips relative to the magnetic field, and in our experiments this upper angular velocity for stable rotation is $\omega_c = 2\pi \cdot 0.62\ \text{s}^{-1}$.

4. Data analysis

The motion of the *n* interacting magnetic holes is observed with a light microscope, Fig. 7(a), and the images are acquired with an attached video camera and recorded and digitized on a workstation. A computer program is used to map the positions of the microspheres in the (*x,y*)-plane in real time, Fig. 7(b). Our next step in the analysis is to include the time, in order to obtain a space-time diagram (*x,y,t*), thus, creating the world lines of the microspheres. In this way we essentially "freeze" the dynamics of the microspheres at all times, Fig. 7(c). The geometrical braid, Fig. 7(d), will appear when projecting the space-time trajectories onto one of the spatial axis [12].

The theory of braids can be seen as a part of a larger piece of mathematic known as the theory of knots. Braids are purely abstract objects satisfying a number of rules. Emil Artin introduced the classification of braids in 1925 [13]. A thorough introduction to braid theory may be found in the book by Birman [14]. Here we will only give a brief description of the notion used in our analysis, see Fig. 8(a). The strings are labeled 1,2,3 ... from top to bottom, and the crossings are denoted as follow: If string 1 crosses over string 2, label the crossing σ_1, if it crosses under label the crossing σ_1^{-1}. It is then possible to code any braid as a sequence of these symbols, as illustrated in Fig. 8(b). The Greek letter σ is used by convention and something to attach the subscripts and superscripts to.

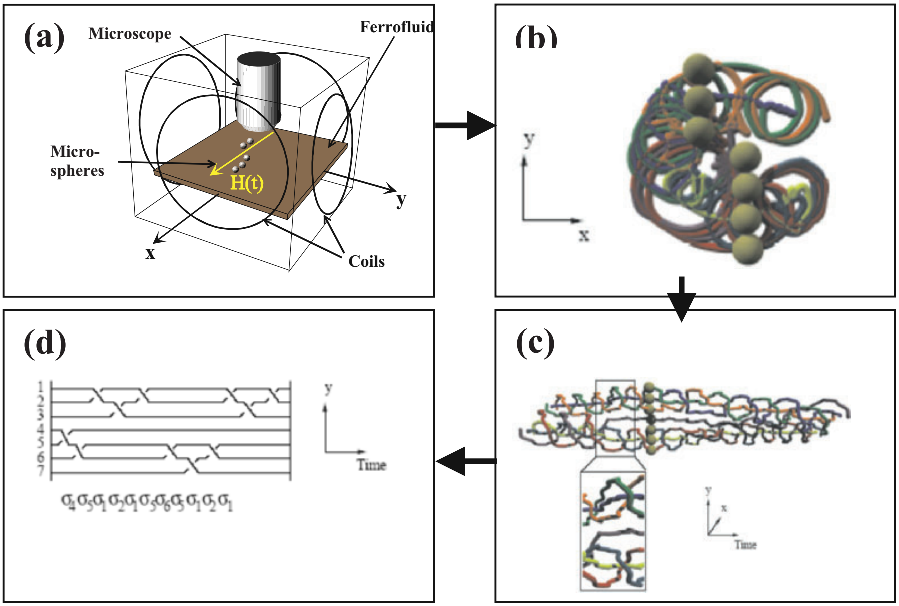


Figure 7. (a) The experimental setup shown schematically. The thin layer of ferrofluid containing the uniformly sized microspheres is confined between two plane-parallel glass plates. A pair of Helmholtz coils produces a magnetic field in the plane; (b) the traced paths of the two-dimensional motion of seven microspheres; (c) the three-dimensional space-time trajectories of the spheres; (d) the irreducible braid representation of the space-time trajectories of the spheres. The strands are numbered from top to bottom relative to the y-axis and the notation is explained in the text. The motions of the seven microspheres are taken from an experiment with $\varepsilon = 0.67$ and $\omega = 0.23$.

The braid generators, σ_i, are read out from the geometrical braid and the resulting sequence of braid generators gives the braid word. The set of all braids with n-strands is the Artin braid group B_n, containing every possible rotation of n magnetic holes. The braidwords found in experiments may not be topologically unique due to continuous deformations [15]. The challenge is to find a scheme to determine whether two braidwords are equivalent and thereby describe the same dynamics. This so-called *word problem* has been solved for B_n [16]. The *Garside algorithm* [17] is used in this analysis for a refinement of this solution [18]. After running through the Garside algorithm the braidword are represented by an ordered set of *positive permutation braids*. The positive permutation braids are small parts of the braidword and belong to a subset of B_n with two additional criteria: 1) the space-time diagram's strands have only overcrossings, and 2) two strands can cross each other only once.

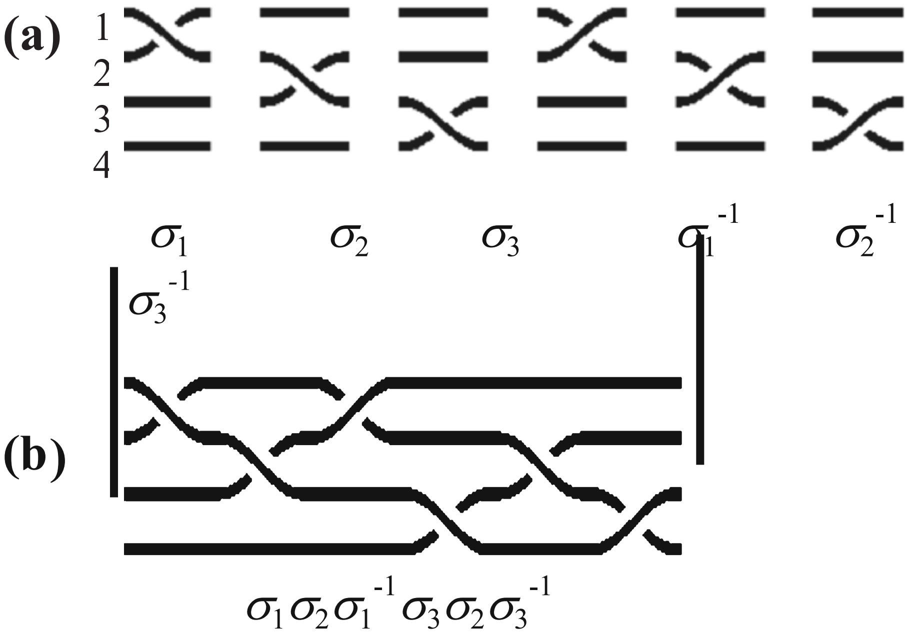


Figure 8. (a) The generators of the 4-strand Artin braid group and (b) an example of a braid B composed of these with the corresponding braid word written below.

A convenient and unique way of labeling the different positive permutation braids is by using the so-called *factorial coordinate* method [17]. The algorithm for this method is given by the expression:

$$g(\tau) = 1+g_1(\tau)1!+\ldots+g_{n-1}(\tau)(n-1)! \quad \text{with } 0 \le g_i(\tau) \le i. \tag{8}$$

The factorial coordinate $g_i(\tau)$ counts the number of crossings of string $i + 1$ with lower number strings within the positive permutation braid $g(\tau)$ after a time τ. One permutation braid typically corresponds to one half period of the rotating magnetic field. Figure 9 shows a simple example of this factorial coordinate method.

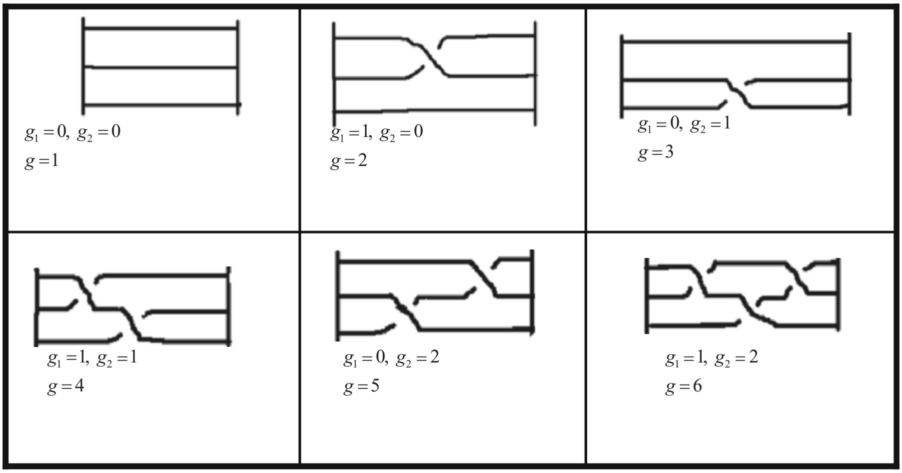


Figure 9. A simple example of labeling of braids with g-values for three spheres resulting in 6 possible modes.

5. Results and discussion

In this study we focus mainly on experiments with seven microspheres that have the anisotropy parameter ε between 0.58 and 0.85, and the angular frequency ω in the range $0.23 < \omega < 0.45$. In these parameter ranges intermittent behavior is observed. This behavior generates many different modes with different frequency of occurrence, and is easily studied with the rank-ordering statistics. Outside these parameter ranges more regular behavior is observed. For experiments with a higher number of spheres n, we have studied what happens around $\varepsilon = 0.64$. The experiments were run from 4 to 13 hours. In this work we present four different experiments with different n, ε, and ω values.

The motion is grabbed and analyzed by a computer in real time. In this analyzing process the coordinates to each sphere is found, Figs. 7(b) and 7(c), and also the braid generators, Fig.7(d). When observing the motion of the spheres two factors are striking:

1. The spheres start out in a line and get almost aligned after one half period of the external rotating field. During this half period the line of spheres break up into smaller units before realigning.
2. Some typical and stable modes occur frequently over some time. Then the motion is changing via a cascade of different and more rare modes for some time before reaching a stable mode again. This resembles intermittent behavior.

The first factor reflects the fact that the time span for a dynamical mode is around one half period of the external rotating magnetic field. The second factor suggests that the system behaves like a scaling tree, similar to the lexicographic tree described by Mandelbrot for word frequencies in linguistics [19]. He used this scaling tree to derive the Zipf-Mandelbrot relation. As will be discussed below, a similar approach can be used to derive this relation for the system under study here.

A small part of an experiment with $\varepsilon = 0.67$ and $\omega = 0.23$ is displayed in Fig. 7. Seven spheres trace out their world lines and are projected into a braidword. The different dynamical modes of this experiment are measured and displayed in terms of their corresponding g-values in Fig. 10. We obtain 365 different dynamical modes in the experiment, while for the numerical simulations with same duration we obtain 392 different modes. A significant feature is that a small set of modes appears frequently and is interrupted by more rare modes. Small variations in the system parameters might lead to large changes in the braidword and the g-values. This can be seen in the differences between the experiment, Fig. 10, and the simulation, Fig. 11, which have nominally equal parameter values. However, these minor changes do not change the overall statistical properties of the dynamics. Examples of the braids formed by the most frequent mode and one of the least frequent modes in Fig. 11 are shown in Fig. 12.

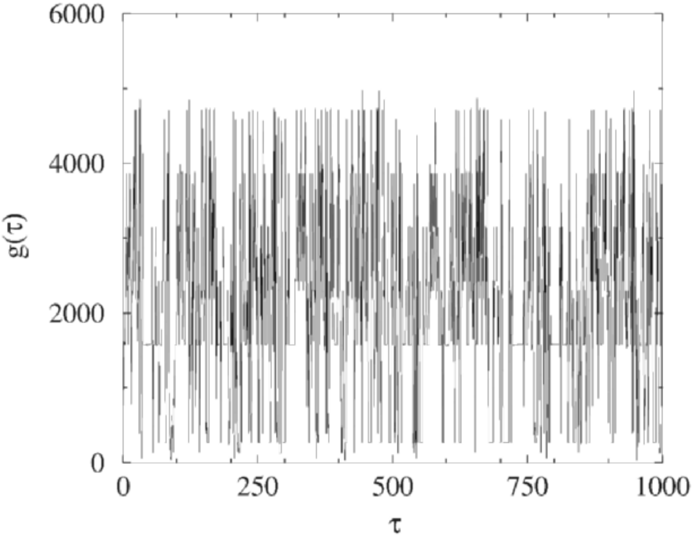


Figure 10. The dynamics of seven microspheres in an external rotating magnetic field is displayed in terms of the factorial coordinates of the positive permutation braids *g(τ)* as a function of the time τ for an experiment with $\varepsilon = 0.67$ and $\omega = 0.23$. The most frequent modes (e.g., groups of 1+3+3 spheres, $g = 1575$) reflect the symmetry in the system as opposed to rarely occurring modes (e.g., 1+1+1+3+1 spheres, $g = 265$).

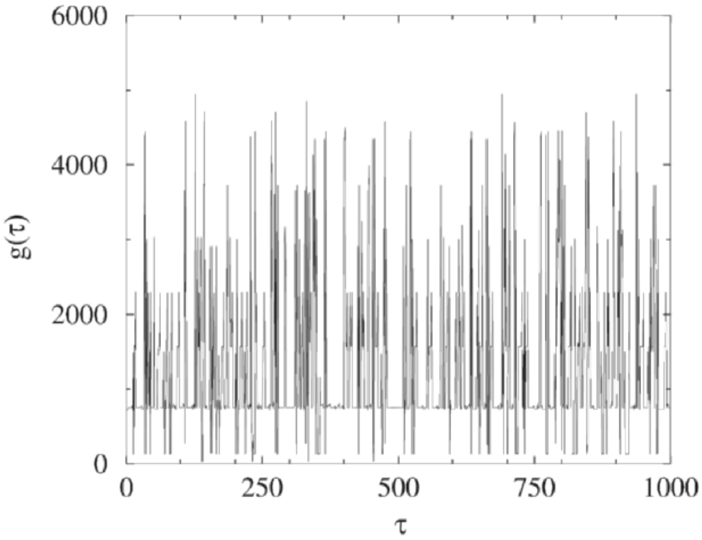


Figure 11. The dynamics of seven microspheres in an external rotating magnetic field is displayed in terms of the factorial coordinates of the positive permutation braids *g(τ)* as a function of the time *τ* for the simulation corresponding to the experiment presented in Fig. 10. The most frequent mode found here, *g* = 750, corresponds to groups of 3+2+2 spheres.

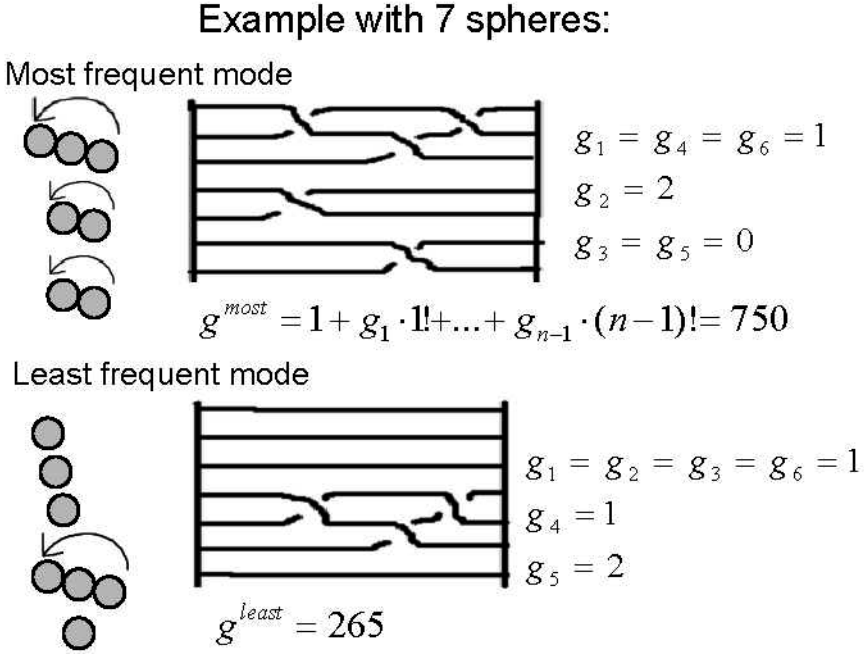


Figure 12. The most frequent mode and one of the least frequent modes in Fig. 10 with their corresponding braids and factorial *g*-values.

We then apply rank-ordering statistics to these braid permutations. In rank-ordering statistics we count all the different modes and rank them afterwards. The rank r of a mode is defined such that the most frequently occurring mode gets the rank $r=1$, the next most frequently used mode gets rank 2, etc. The frequency of occurrence $\phi(r)$ is normalized: $\sum_{r=1}^{N}\phi(r)=1$, where N is the total number of different modes. Including all types of positive permutation braids for n strands, $N=n!$. For the case of seven microspheres we have $N=5040$ possible modes.

The rank-ordering analysis of the positive permutation braids is shown in a double logarithmic plot in Fig. 13. The data may be fitted to a straight line with a slope -1.6 ± 0.1 for $r>5$. However, a non-linear fit on the whole r-range to the Zipf-Mandelbrot relation gives a better fit with parameters $A=3.2\pm0.2$, $\zeta=5.3\pm1.0$ and $\gamma=1.8\pm0.1$ for both experiment and simulation. A simulation with four times as long duration gives 448 different types of dynamical modes, however the fit gives the same parameters.

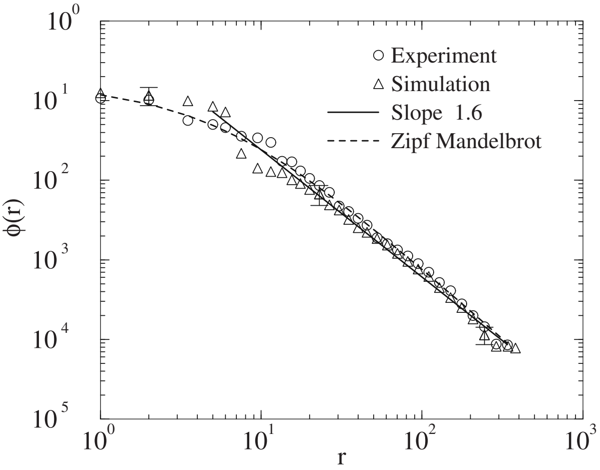


Figure 13. The frequency of occurrence of braid permutations $\phi(r)$ versus its rank r found in the dynamics of seven microspheres for an experiment with $\varepsilon=0.67$ and $\omega=0.23$. Both data sets fit well to the Zipf-Mandelbrot relation, shown as a dashed line.

Similar behavior to that discussed above is also observed for experiments with other values of ε and ω, and the exponent γ is found to be between 1.1 and 1.9, as first reported in Ref. [20]. A second example is an experiment with $\varepsilon=0.83$ and $\omega=0.44$. The rank-ordering analysis of this experiment gives a Zipf-Mandelbrot relation between $\phi(r)$ and r with exponent $\gamma=1.4\pm0.1$ as shown in Fig. 14.

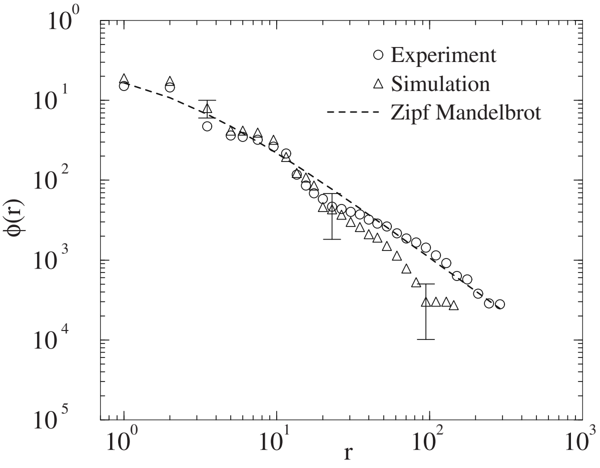


Figure 14. The frequency of occurrence of braid permutations $\phi(r)$ versus rank r found in the dynamics of seven microspheres with $\varepsilon = 0.83$ and $\omega = 0.44$.

Some minor differences between the results found in the experiments and in the simulations may be due to higher order effects neglected in the simulations, such as hydrodynamic interaction between microspheres and hydrodynamic and magnetic interactions due to confinement of the ferrofluid by the walls. Complicated flow behaviors may occur even at low Reynolds numbers in colloidal suspensions. Despite its long-recognized ubiquity, hydrodynamic coupling in these suspensions is not completely understood [21].

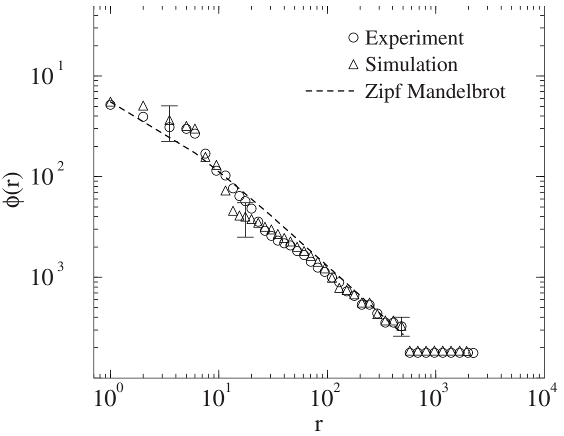


Figure 15. The frequency of occurrence of braid permutations $\phi(r)$ versus the rank r found for 11 microspheres with $\varepsilon = 0.64$ and $\omega = 0.22$. The value of the exponent is found to be $\gamma = 0.98 \pm 0.10$.

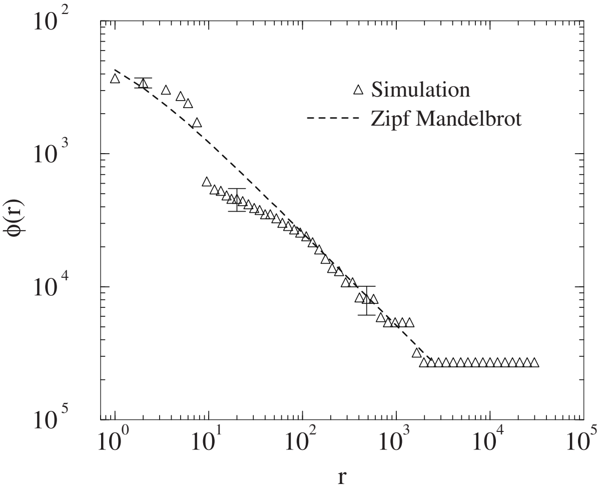


Figure16 . The frequency of occurrence of braid permutations $\phi(r)$ versus rank r found in the computer simulation of the dynamics of 20 microspheres with $\varepsilon = 0.64$ and $\omega = 0.15$. Here, $\gamma = 0.80 \pm 0.08$.

The Zipf-Mandelbrot relation is observed in systems with higher number of spheres as well. Figurs 15 and 16 are examples of 11 and 20 spheres, respectively. For both cases we used $\varepsilon = 0.64$. When increasing the number of spheres the number of possible modes increases rapidly. For eleven spheres there exist 11! ≈ $4.0 \cdot 10^7$ possible modes, while we obtain only 2200 modes in the experiment. Twenty spheres have 20! ≈ $4.0 \cdot 10^{18}$ possible modes and the simulation gave only 30,000 modes. In order to observe a behavior obeying the Zipf-Mandelbrot relation for systems with $n \geq 10$, we thus need experiments over a very long time span in order to get good statistics. The exponents found for the results presented in Figs. 15 and 16 are clearly lower than for the case with seven spheres.

It is also possible to extend this analysis to calculate the mode-mode correlations between modes of different rank [18] and estimate the information we get from the system in terms of the so-called Shannon entropy [22].

6. Conclusions

In this review we have shown that braid theory can be used to characterize the dynamics of magnetically interacting particles. An *n*- stranded space-time braid may thus represent the motions of *n* non-magnetic microspheres in a thin ferrofuid layer subjected to oscillating external magnetic fields. It is possible to extract the positive permutation braids and use these as a measure of the collective modes in the system. The distribution of these dynamical modes was analyzed using rank-ordering statistics, and it was found to show a hierarchical structure of the Zipf-Mandelbrot type.

Acknowledgements

The work was supported in part by the Research Council of Norway through the Nanomat program and one of us, KdLK, acknowledges a PhD fellowship, Project No. 142902/432 from the Research Council of Norway.

References

1. G. K. Zipf, *Human Behavior and The Principle of Least Effort* (Addison-Wesley Press, Massachusetts, 1949).
2. D. Sornette, L. Knopoff, Y. Y. Kagan, C. Vanneste, J. Geophys. Res. **101**, 13883 (1996).
3. R. N. Mantegna, S. V. Buldyrev, A. L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, H. E. Stanley, Phys. Rev. Lett. **73**, 3169 (1994).
4. B. Mandelbrot, Word **10**, 1 (1954).
5. B.B. Mandelbrot, *The Fractal Geometry of Nature* (W. H. Freeman, New York, 1982).
6. P. Meakin and A.T. Skjeltorp, Adv. Phys. **42**, 1 (1993).
7. R.E. Rosensweig, *Ferrohydrodynamics* (Cambridge University Press, 1985).
8. J. Ugelstad, P. C. Mørk, K. Herder Kaggerud, T. Ellingsen, A. Berge, Adv. Colloid Int. Sci. **13**, 101 (1980).
9. A.T. Skjeltorp, Phys. Rev. Lett. **51**, 2306 (1983).
10. Type EMG 909 from Ferrofluidics GmbH, Hohes Gestade 14, 72622 Nürtingen, Germany, with susceptibility χ = 0.8, saturation magnetization M_s = 20 mT, viscosity $\eta = 6 \cdot 10^{-3}$ Nsm $^{-2}$, and density ρ = 1020 kgm $^{-3}$.
11. G. Helgesen, P. Pieranski, and A. T. Skjeltorp, Phys. Rev. A **42**, 7271 (1990).
12. S. Clausen, G. Helgesen, and A. T. Skjeltorp, Int. J. Bifurcation and Chaos **8**, 1383 (1998).
13. E. Artin, *Theory der Zöpfe*, Abh. Mat. Sem. Univ. Hamburg, **4**, 47-72 (1925).

14. J.S. Birman, *Braids, Links and Mapping Class Group*, Annals of Math. Study, **82** (Princeton Univ. Press, 1974).
15. C.C. Adams, *The Knot Book* (W. H. Freeman and Company, 1994).
16. F.A. Garside, Quart. J. Math. **20**, 235 (1967).
17. E.A. Elrifai and H. R. Morton, Quart. J. Math. **45**, 479 (1994).
18. K.d.L. Kristiansen, G. Helgesen, and A.T. Skjeltorp (unpublished).
19. B.B. Mandelbrot, in *Proceedings of symposia in applied mathematics vol. XII*, R. Jakobson (ed.) (New York: American Mathematical Society, 1961).
20. K.d.L. Kristiansen, G. Helgesen, and A.T. Skjeltorp, Physica A **335**, 413 (2004).
21. E. R. Dufresne, T. M. Squires, M. P. Brenner, and D. G. Grier, Phys. Rev. Lett. **85**, 3317 (2000).
22. C. E. Shannon, Bell Syst. Tech. **30**, 50 (1950).

COMBINING OPTICAL TWEEZERS AND MICROPIPETTES FOR DNA STRETCHING: ELASTICITY OF MICROPIPETTE CRUCIAL

THOMAS MØLLER HANSEN[1,2], NADER REIHANI[1,3],
AND LENE ODDERSHEDE[1*]
[1] *Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*
[2] *Institute of Molecular Biology and Fysiology, Øster Farimagsgade 2A, 1353 Copenhagen, Denmark*
[3] *Institute for Advanced Studies in Basic Sciences, PO Box 45195-159, Zanjan, Iran*

1. Introduction

Optical tweezers are often used in connection with other techniques to study physical properties of biological systems. In particular, this combination has often been used to study elastic properties of individual strands of nucleic acids. The DNA used in this study is the shortest so far reported, only 1.1 μm, 20 times its persistence length. We use two different experimental geometries, one in which the axis of the micropipette is orthogonal to that of the stretched polymer and one where the axis of the micropipette is parallel to the stretched polymer. By comparing the force-extension data to the predictions of the celebrated worm-like-chain model (Marko and Siggia, 1995), we find that the results obtained using the orthogonal geometry have severe problems, the force increases slower than expected with extension of the polymer. Also, the expected plateau at the transition away from the B-form of dsDNA is not horizontal. However, if instead one uses the parallel geometry the data obtained are fit well by the worm-like-chain model. This difference can be explained by the elasticity of the micropipette, which can be crucial to take into account when using micropipettes in connection with optical tweezers.

The elastic properties of DNA are important as it is constantly bend, pushed and pulled inside the living cell by macromolecules working on the DNA. E.g. the DNA is bend when the TATA-box binding protein attaches to the promoter during transcription initiation.

---

[*] Corresponding author, email: oddershede@nbi.dk

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 163–173.

Also,the RNA-polymerase is capable of exerting consideableforces (up to 15 pN) while transcribing the DNA (Wang et al., 1998). One of the pioneering experimental investigations of the elastical properties of DNA are presented in (Wang et al., 1997). Another impressive *in vitro* study of nucleic acids is that by (Liphardt et al., 2001), who performed a mechanical unfolding of RNA loops using an optical trap in connection with a micropipette. This setup was also used for a test of Jarzynski's equation (Jarzynski, 1997; Liphardt et al., 2002), an important equation dealing with the difficult task of relating the non-equilibrium behavior of a system with its equilibrium behavior.

Such *in vitro* experiments are good because single action-reaction couples can be isolated and understood, but on the other hand, these experiments are not per se biologically relevant as the conditions inside a living cell are totally different from that within a test tube. Recently, effort is also being put into investigating the more complex problem of motion of single molecules *in vivo*, such as e.g. the motion of proteins in the outer membrane of living E. coli bacteria (Oddershede et al., 2002).

2. Methods

2.1. OPTICAL TRAPPING

The optical trap is based on a NdYVO$_4$ laser with wavelength 1064 nm and is implemented in an inverted Leica microscope with a quadrant photodiode back focal plane detection scheme, for a full discription of the equipment see (Oddershede et al., 2001). The water immersion objective (Leica, NA=1.2) allows for optical trapping at any height within the sample and prevents effects of spherical abberations. A laser power of 0.8 W, measured at the output of the laser, was used in all experiments presented here. Data are acquired using a National Instruments card (PCI-MIO-16E-4) and the sample is mounted on a three axis translational piezoelectric stage (Physik Instrumente) with capacitative feedback control and nanometer position resolution. Data aquisition was performed using costum made Labview programs. Simultaneous control over piezo stage and output from the quadrant photodiode allows for accurate measurements of corresponding values of force and distance.

2.2. DNA CONSTRUCT

The DNA is 3256 base pairs long, corresponding to 1.1 μm. It is synthesized by PCR (Polymerase Chain Reaction) using 5' modified primers. In one end of the PCR fragment there is a biotin and the other a digoxygenin. The biotin is bound specifically to a streptavidin coated bead with diameter 2.1 μm(Bangs Laboratories). The digoxygenin end is specically bound to an anti-digoxygenin coated bead with diameter 2.88μm(Spherotech), to ensure a tight binding between the anti-digoxygenin and the bead this binding is cross-linked with protein C. The bufferused throughout the experiment contained 10 mM Tris-HCl pH 7.9,250 mM NaCl, 15 mM $MgCl_2$, 0.05 weight pct. BSA(Bovine Serum Albumin),and 10 ng/microliter carrier DNA. First, the smaller streptavidin coated beads were incubated with the PCR fragment in buffer C. The number of tethers between beads can be controlled by adjusting the ratio ofPCR fragmentto beads. Just before the experiment, the smaller beads with DNA tethers were mixed with the larger beads and further diluted in the buffer.

2.3. MICROPIPETTES, CHAMBERS

The buffer containing the DNA and beads was flushed into a microfluidic perfusion chamber. Micropipettes (outer diameter approximately 1.5 μm, inner diameter approximately 1 μm) were pointing into the chamber and immobilized with respect to the chamber. Suction could be applied to the pipettes to firmly attach the beads in the pipette. The geometry of the micropipettes with respect to the axis of the propagating laserlight is of extreme importance for this study. Figure 1 shows a schematic drawing of the geometry during a mechanical stretching of a DNA polymer. The left and right parts of Figure 2.3 show scenaria where the micropipette is respectively orthogonal and parallel to the direction in which the polymer is stretched.

2.4. EXPERIMENTAL PROCEDURE

After flushing buffer containing DNA and beads into the microfluidic chamber, one of the larger beads is optically trapped and immobilized on a micropipette. Thereafter, one of the smaller beads with DNA attached is optically trapped and brought into the same height as the larger bead on the micropipette. A timeseries of the thermal fluctuation of the bead inside the trap is monitored to calibrate the optical trap. Then, the smaller bead is brought close to the larger bead and allowed to fluctuate (and rotate) within the trap. When a DNA tether is formed between the two beads, this can be seen on the output from the quadrant photodiode if one tries to move the micropipette away. In the experiment, the optical trap is held fixed and the micropipette attached

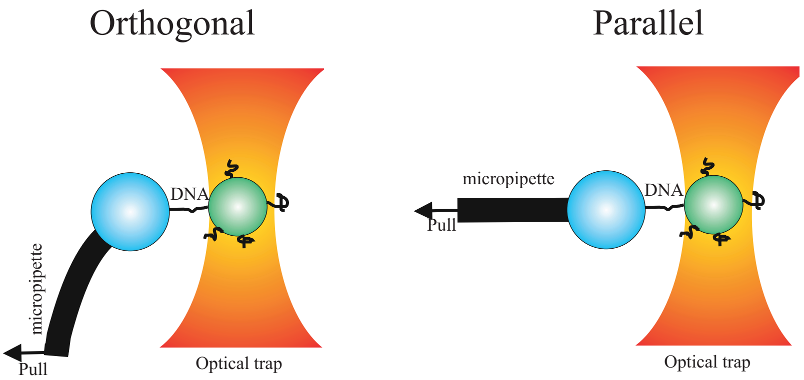


Figure 1. Two different possible geometries for combining optical tweezers with micropipettes for single molecule mechanical studies. The orthogonal geometry, shown to the left, is the most commonly used. The parallel geometry, shown to the right, is the one we suggest to use in the future to avoid bending of the micropipette. The arrow shows the direction of pull. The flexing of the micropipette in the orthogonal geometry is exaggerated.

to the piezostage is moved at a velocity of 1000 nm/s a distance of 2000 nm away from the trap. For the calibration procedure data is sampled at a 20-50 kHz rate using the quadrant photodiode. In the stretching measurements, data is sampled at 5 kHz.

3. Calibration

When the micropipette is moved that will, in turn, force the bead in the optical trap to be displaced from its equilibrium position. In order to find the forces acting on this bead as a function of its displacement from the center of the optical trap and hence to find the forces acting on the DNA polymer, we need to perform a calibration.

The three coordinates describing the thermal fluctuations of the bead are uncorrelated, and the equation of motion for each of them is the same, but with different parameter values. Therefore we may look at the equation of motion in one dimension, say, for the x-coordinate, only. The optical tweezers exert a harmonic force $-\kappa x$ on the bead. The surrounding liquid exerts a frictional force on the bead, $-\gamma \dot{x}$, with γ the friction coefficient known from Stokes' law $\gamma = 3\pi\eta d$. Here, d is the diameter of the bead and η is the viscosity of the liquid. For beads close to the surface of the coverslip, Faxén's correction to Stokes' law should be invoked too. However, in the results presented here, the bead is far away from any surface. The bead is also subject to a stochastic force $\mathcal{F}(T)$ due to the Brownian motion of the surrounding liquid at temperature T. Since the Reynolds number is very small, the inertial term can be neglected, yielding the equation of motion

$$0 = -\kappa x - \gamma \dot{x} + \mathcal{F}(t). \tag{1}$$

Upon Fourier transformation of Eq. (1), one obtaines an equation for the Fourier transformed of the position, $\tilde{x}(f)$, and the power spectrum $S_x \equiv \langle |\tilde{x}(f)|^2 \rangle$ for the position x is calculated as

$$S_x = \frac{k_B T}{\pi^2 \gamma (f_c^2 + f^2)}. \tag{2}$$

The characteristic parameter appearing in $S_x(f)$, the corner frequency is given by $f_c = \frac{\kappa}{2\pi\gamma}$, where a typical value of $f(c)$ in our experiments is ~1 kHz. Knowing γ, κ can be determined.

The recorded signal, stemming from a quadrant photo diode, gives the position x and y of the bead measured in Volts. We want to determine the factor A, translating the recorded signal, in Volts to a position in meters.

For a bead moving in a harmonic potential, the distribution of positions visited is Gaussian. Thus, from a histogram of the positions measured, we can determine the variance σ_V^2 (subscript V to indicate measured in Volts)

$$\sigma_V^2 = \left\langle x_V^2 \right\rangle - \langle x_V \rangle^2 . \tag{3}$$

In a harmonic potential, the equipartition theorem gives

$$\sigma^2 = \frac{k_B T}{\kappa} \quad . \tag{4}$$

With κ known from the relation $f_c = \frac{\kappa}{2\pi\gamma}$, σ (measured in meters) can be found. Now, A can be determined from σ_V and σ found from Eqs. (3) and (4)

$$A = \frac{\sigma}{\sigma_V}. \tag{5}$$

For a very precise analysis of the fluctuations of a bead in an optical trap, additional terms must be taken into consideration, of which the most severe is the filtering effect of the quadrant photodiode (Berg-Sørensen et al., 2003), but also aliasing must be taken into account and an analytical fit can be performed(Berg-Sørensen and Flyvbjerg, 2004). In our data analyses, we use a program (Hansen et al., 2005) which does all this.

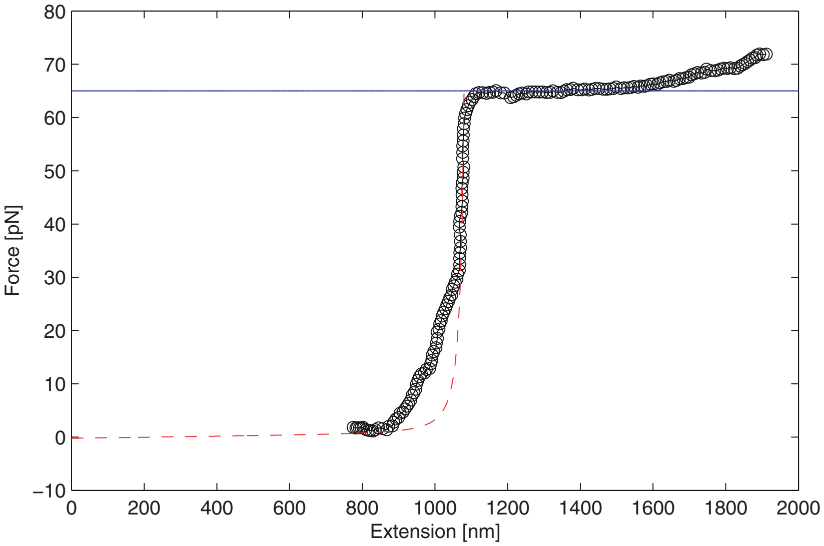


Figure 2. Force versus extention of DNA using the **parallel** geometry. Black circles: Data. Dashed red line: worm-like-chain theory prediction. Full blue line: Expected location of the plateau where dsDNA changes conformation.

4. Results and discussion

Using the two different geometries shown in Figure 1 yield quite different force-extention curves while stretching 1.1 μ DNA. Figure 2 shows the force extension curve resulting from stretching DNA using a geometry where the axis of the micropipette is parallel with the stretching of DNA (right scenario in Figure 2).

Figure 3 shows the force-extension curve resulting from stretching DNA with a setup where the micropipette is orthogonal to the direction of the stretching.

Numerous models have predicted the force-extension behavior of biopolymers such as DNA. One of the most famous models is the so-called 'Worm-Like-Chain' (WLC) (Marko and Siggia, 1995), predicting that the force F nescessary to extend the polymer to a distance x is given by:

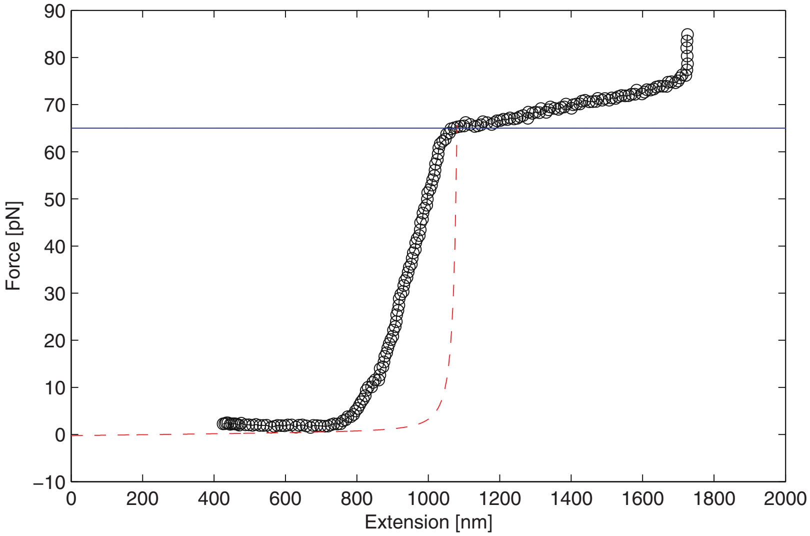


Figure 3. Force versus extention of DNA using the **orthogonal** geometry. Black circles: Data. Dashed red line: worm-like-chain theory prediction. Full blue line: Expected location of the plateau where dsDNA changes conformation.

$$F = \frac{k_B T}{\xi_p}\left(\frac{1}{4(1-x/L_0)^2} - \frac{1}{4} + \frac{x}{L_0}\right), \tag{6}$$

where ξ_p is the persistence length and L_0 the contour length. In our experiment $L_0 = 1.1\mu$m and we use $\xi_p = 48$ nm as the DNA persistence length under the given electrolyte conditions (Wang et al., 1997). The first quantitative experimental study of the elastical properties of DNA (Wang et al., 1997) showed that the force-extension curve is well fitted by a modified version of the Marko-Siggia WLC. However, for simplicity, and as our goal was not to find the exact elastic properties of the DNA, but rather to make some statement about the equipment, we chose to use the WLC of equation 6, which actually is a good approximation both at high and low forces. At intermediate forces WLC is expected to deviate up to 10 pct.

On both Figures 2 and 3 the prediction of WLC (equation 6) is shown as a dashed red line. It is clear that the increase of the force accordingly to WLC is significantly steeper than that observed with the orthogonal geometry. On the other hand, using the parallel geometry, a much nicer resemblence is found. It should be noted, that in our experiments only the relative, not the absolute, extension x is known. Therefore, the data has been artificially shiftet horisontally to reach the plateau simultaneously with the predictions of WLC.

It has been shown experimentally (Smith et al., 1996) that exerting a sufficiently large force on DNA will cause the DNA to change conformation and undergo a highly cooperative transition into a stable form which is 70 pct. longer than the usual B-form dsDNA. This transformation happens around F=65 pN. In both Figures 2 and 3 we observe this transformation happening around 65 pN (solid blue line in Figures). Using the orthogonal setup geometry this transition does not appear to happen at a constant force, whereas it does appear to happen at constant force using the parallel setup. In both Figure 2 and 3 we also observe the onset to melting transition which happens at an extension of about 1.8 μm.

If the micropipette is not sufficiently stiff with respect to bending, then in the orthogonal setup, the elasticity of the micropipette might cause the tip of the micropipette to bend when acted upon by a force applied orthogonal to the axis of the micropette (the end of the micropipette is fixed in the perfusion chamber, which is moved by the piezo electrical stage). This would give rise to a behavior as observed in Figure 3, namely that the increase in force is slower than predicted by WLC and that the plateau does not appear horisontal. In the parallel geometry, on the other hand, these deviations are not observed, and largely, the force-extension curve follows WLC and the plateau is horizontal. Of course, in the parallel configuration, the micropipette is stretched lengthwise, but it appears that the micropipette is stiffer in the direction along the pipette than orthogonal to it.

The stiffness of the micropipette used in a setup with orthogonal geometry (Wuite et al., 2000) was estimated on the basis of the thermal fluctuations of a bead attached to the micropipette to be $\kappa_{\mu pipette}$ $\sim$4nN/μm. Our measurements give a different number: Judging from Figure 2 and 3, it appears that at a force of F=65 pN, the pipette has been stretched by an additional $\sim$200 nm, corresponding to a stiffness of $\kappa_{\mu pipette}$ $\sim$0.3 nN/μm. Of course, individual pipettes have varying inner and outer diameters and have differing suspension lengths from their fixed point to the tip, so their flexibility in the orthogonal direction can be expected to vary substantially.

It is important to be aware of this artefact of the micropipette when one uses a micropipette e.g. in connection with optical traps to study mechanical behavior of biological molecules. For longer molecules as

λ-DNA it is not so important, but for smaller polymers it becomes increasingly important. Especially when the precision claimed significantly exceeds the effect brough up here. It should be noted, that a number of experiments reported in literature, e.g. in (Smith et al., 1996) and (Liphardt et al., 2001),(Liphardt et al., 2002) use the orthogonal geometry. A close inspection of the force extension data presented in these papers show that stretching DNA (Smith et al., 1996) does not follow WLC and the plateau is not totally horisontal. Stretching DNA-RNA hybrid handles (Liphardt et al., 2001; Liphardt et al., 2002) does not follow WLC either. One reason for this could be the geometry of the setup as examined in the present paper.

By no means do we claim to have solved the problem, there might be other experimental artefacts that are even more severe. Another issue which we did not consider, but which might be important is that of the pulling speed; the amount of irrevserible work put into the system might change the force-extention relations. However, as our two types of experiments were done with the same pulling speed, they are directly comparable, and our conclusion about the flexibility of the micropipette as crucial, holds.

Our DNA had a contour length of 1.1 μm, and hence it is the shortest so far reported in stretching experiments. The DNA used in (Wang et al., 1997) was 1328 nm and that e.g. in (Smith et al., 1996) was 16 μm λ-DNA. WLC is derived with the assumption that the contour lenght is significantly longer than the persistence length. In our case, this ratio is 23, and one obvious challange would be to go to even shorter DNA's and see the gradual increase in deviation from WLC. Another subject that deserves investigation is whether DNA-RNA hybrids have different force-extension properties than dsDNA.

5. Conclusion

By using optical tweezers we have stretched 1.1μm DNA to its transition point. The DNA was tethered between two beads, one held by a micropipette and the other by optical tweezers. The optical tweezers were equipped such as to enable accurate measurements of corresponding values of force and extension of the DNA tether. We used two different setups: One in which the micropipette was orthogonal to the direction in which the polymer was being stretched, in the other the micropipette was parallel to the stretching direction. Using the parallel setup yielded results where the force-extension data resembled worm-like-chain predictions and a horisontal transion plateau was observed as expected at 65 pN. However, using the orthogonal geometry, which is widely reported in literature, we observed the force to increase slower than expected and the plateau was not horizontal. This can be explained by the fact that the micropipette bends and hence, its elasticity must be taken into account when pulling orthogonal to it at its tip. Using a setup geometry where the pulling direction is parallel to the micropipette decreases this effect, and is the route we devise out of this problem.

Acknowledgements

We acknowledge discussing the biological construct with Michael A. Sørensen, Institute of Molecular Biology and Fysiology, University of Copenhagen, as well as discussions with the other members of the Optical Tweezers Group, Niels Bohr Institute. We acknowledge financial support from the Carlsberg Foundation, the BIOP Graduate School and from the Villum Kann Rasmussen Foundation.

References

Berg-Sørensen, K. and H. Flyvbjerg: 2004, 'Power spectrum analysis for optical tweezers'. *Rev. Sci. Ins.* **75**, 594–612.

Berg-Sørensen, K., L. Oddershede, E. Florin, and H. Flyvbjerg: 2003, 'Unintended Filtering in a Typical Photodiode Detection System for Optical Tweezers'. *Journal of Applied Physics* **93**, 3176–3177.

Hansen, P. M., I. Tolic-Nørrelykke, , H. Flyvbjerg, and K. Berg-Sørensen: 2005, 'Faster version of MatLab package for precise calibration of optical tweezers'. *submitted to Comp. Phys. Comm.*

Jarzynski, C.: 1997, 'Nonequilibrium Equality for Free Energy Differences'. *Physical Review Letters* **78**, 2690–2693.

Liphardt, J., S. Dumont, S. Smith, I. T. Jr., and C. Bustamante: 2002, 'Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality'. *Science* **296**, 1832–1835.

Liphardt, J., B. Onoa, S. Smith, I. T. Jr., and C. Bustamante: 2001, 'Reversible Unfolding of Single RNA Molecules by Mechanical Force'. *Science* **292**, 733–737.

Marko, J. and E. Siggia: 1995, 'Stretching DNA'. *Macromolecules* **28**, 8759–8770.

Oddershede, L., J. Dreyer, S. Grego, S. Brown, and K. Berg-Sørensen: 2002, 'Motion of a single molecule, the lambda-receptor, in the outer bacterial membrane'. *Biophysical Journal* **83**, 3152–3161.

Oddershede, L., S. Grego, S. Nørrelykke, and K. Berg-Sørensen: 2001, 'Optical Tweezers: Probing Biological Surfaces'. *Probe Microscopy* **2**, 129–137.

Smith, S. B., Y. Cui, and C. Bustamante: 1996, 'Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules'. *Science* **271**, 795–798.

Wang, M., M. Schnitzer, H. Yin, R. Landick, J. Gelles, and S. Block: 1998, ‘Force and Velocity Measured for Single Molecules of RNA Polymerase’. *Science* **282**, 902–907.

Wang, M., H. Yin, R. Landick, J. Gelles, and S. Block: 1997, ‘Stretcing DNA with optical tweezers’. *Biophysical Journal* **72**, 1335–1346.

Wuite, G., R. Davenport, A. Rappaport, and C. Bustamante: 2000, ‘An Integrated Laser Trap/Flow Control Video Microscope for the Study of Single Biomolecules’. *Biophysical Journal* **79**, 1155–1167.

UNIVERSAL NETWORKS AND PROCESSES IN SOFT AND COMPLEX MATTER: FROM NANO TO MACRO

Interconnected physical phenomena in clays

JON OTTO FOSSUM
Norwegian University of Science and Technology (Ntnu),
Department of Physics, Hoegskoleringen 5,N-7491 Trondheim, Norway, Jon.Fossum@Ntnu.No

Abstract. Here we give examples of different physical phenomena observed in physical networks (networked aggregates and/or pore networks) realized in clays. We study the SAME material, from different points of view simultaneously, and in parallel. Knowledge gained from one type of study is utilized in another, for example: We demonstrate that studies of colloidal network structures of clay gels teaches us things about collapsed surface networks and surface roughness, and vice versa, we show that diffusion processes in clay samples teaches us things that are relevant for the polarizibility of clay particles exposed to external electric fields and vice versa, and we will show that water suspensions of nano-layered clay particles teaches us lessons to be used for explaining adaptive structures of clay particles suspended in oil. The phenomena under study are independent, in their own right, each with their own level of description, but still interconnected. We use examples from our activities on nano-layered silicates (i.e. clays), in order to illustrate complexity and universality in materials physics. This research may be said to link nano physics with macro physics, and the systems we study, namely clays, enable us to interact with ideas both towards nano-technological applications (smart materials, diffusion in membranes, nanocomposites etc), and towards geology and environmental soil science. This is the strength of physics, namely the universal approach to problems.

Keywords: Nanostructured materials, soft matter, complex matter, networks in materials, macroscopic stability and dynamics, complex physical phenomena

1. Introduction

Physical phenomena are often not independent of each other, i.e. interconnected, and this may be said to be the basic idea behind a project in our group [1] concerned with studies of many complex physical phenomena

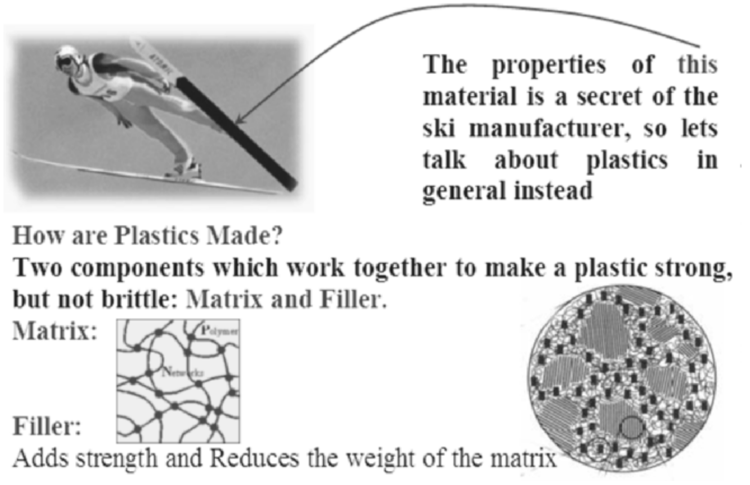
A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 175–190.

in parallel, in one single material, namely clays. Clays were chosen for this project because so many different complex physical phenomena can be realized and studied experimentally in this one physical model system [2], and in particular clays are good representatives of macroscopically soft materials built from structures and networks of nano-particles [3]. Clays represent one of the traditional natural materials that humans have used since the beginning, and as such, there are both traditional and very modern industrial uses of clays: For example, clays are these days some of the "hottest" nano-particles for making new nano-composites [4]. Clays have important uses as rheology modifiers in for example paints, cosmetics, toothpaste and household cleansers [5]. Clays in the form of so-called "pillared layered structures" are used as molecular sieves and porous catalytic materials [6]. Clay land formations are important geologically, and thus both land surface stability, and stability of sub surface oil-wells possess problems that require basic understanding of all such relevant physical phenomena. It has also been suggested that clays may have played an essential and instrumental role for the origin of life, as we know it [7], and this evidently provides additional inspirational motivations for our research work on complex physical phenomena in clays.

In this review, we present work that we have undertaken in our group [1], and results we have achieved [8]-[17] during recent years, on studies of interconnected complex and universal physical phenomena in our selected physical model system: Clays.

Figure 1. The oldest sample known of baked clay: A figurine called "Venus of Vestonice" found in 1920 in the Czech Republic. Approximate age: 23000 B.C. (Dated from mammoth bone ash). The picture is scanned from [18].

The following two figures describe examples of networks in materials: Polymer nanocomposits.



The properties of this material is a secret of the ski manufacturer, so lets talk about plastics in general instead

How are Plastics Made?

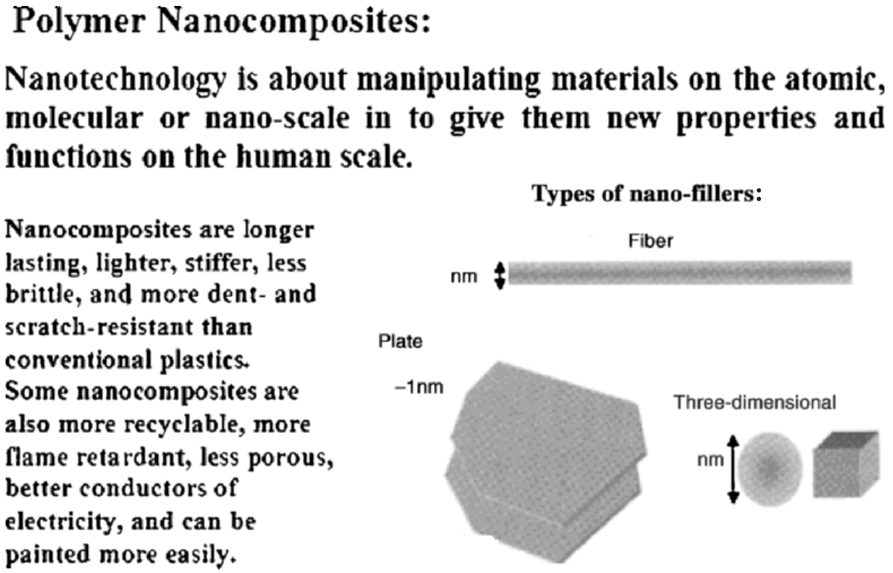
Two components which work together to make a plastic strong, but not brittle: Matrix and Filler.

Matrix:

Filler:

Adds strength and Reduces the weight of the matrix

Figure 2. Polymer nano-composites.



Polymer Nanocomposites:

Nanotechnology is about manipulating materials on the atomic, molecular or nano-scale in to give them new properties and functions on the human scale.

Nanocomposites are longer lasting, lighter, stiffer, less brittle, and more dent- and scratch-resistant than conventional plastics. Some nanocomposites are also more recyclable, more flame retardant, less porous, better conductors of electricity, and can be painted more easily.

Types of nano-fillers:

Figure 3. Nano fillers in polymer nano-composites.

2. Selected examples of interconnected physical phenomena in clays:

2.1. CLAYS AS PHYSICAL MODEL SYSTEMS FOR STUDIES OF NETWORKS AND PROCESSES IN SOFT AND COMPLEX MATTER

2.1.1. *Synthetic fluorohectorite*

Fluorohectorite belongs to the 2:1 type clays (generic terms: Swelling clays, smectite clays). The fundamental level of description for the 2:1 clays is at the mesoscopic scale. i.e. the "building blocks" are 1 nanometer thick sheets with effective sheet "radius" in the 20 nm to 10 μm range, These sheets have charged surfaces (negative) and edges (small positive), and there are counter ions present in order to compensate for these charges.

The properties of fluorohectorite are summarized in the following figure:

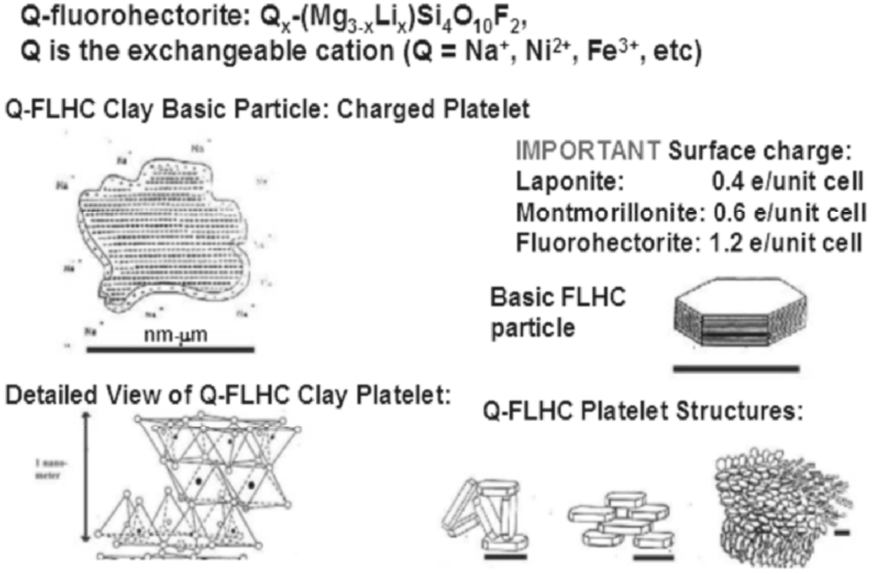


Figure 4. Description of 2:1 clays.

Because of the high surface charge, the florohectorite 1 nm thick sheets will not disperse independently in water, but instead the "building blocks" are stacked particles consisting of about 100 sheets [9], resulting in platelets of about 100 nm thick. This is unlike laponite which is discussed below, for which the lower surface charge allows individual sheet dispersion in water.

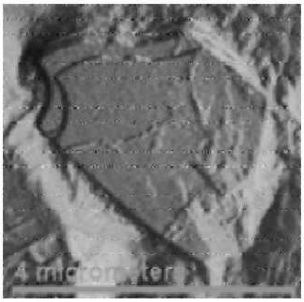


Figure 5. AFM image of synthetic fluorohectorite particle.

2.1.2. *Laponite*

Laponite is (like fluorohectorite) a synthetic 2:1 clay, but in this case lower surface charge allows for dispersion of individual 1 nm thick platelet sheets in water. This is at the basis of the laponite phase diagram explored investigated in [19] and explored by simulations in [20]. Laponite is (unlike fluorohectorite) monodisperse, i.e. the 1 nm thick sheets all have roughly the same radius (about 25 nm). The following figure illustrating possible networked structures based on nano-particles in laponite (adopted from [20]), showing from left to right, a liquid suspension of colloidal aggregates, an isotropic percolating gel structure, and to the right a percolating nematic structure. This networking is simply put stabilized by competition between van der Waals attractive forces and electrostatic repulsive forces screened by ion contents in the suspending liquid (water in this case), i.e DLVO theory, see [21].

Figure 6. Organization of laponite platelets depending on density into (from left to right) sol, isotropic and nematic ordering. The figure is taken from [20].

2.2. ANISOTROPIC NETWORKS IN MATERIALS: LIQUID CRYSTALLINE STRUCTURES AND STUDIES OF ORDERED PHASES IN CLAY COLLOID PLATELET SYSTEMS [15]

This project within our group [1] deals with Na-fluorohectorite synthetic clay suspended in salt water (NaCl), in which the following phase separation in gravity has been observed by us [2], [9], [15] (the figure is adopted from [2] and [15]):

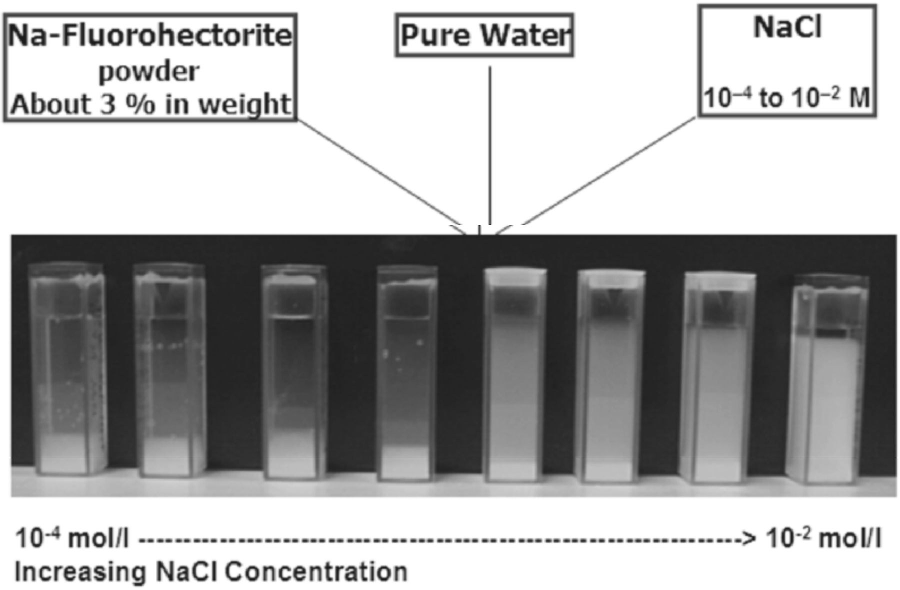


Figure 7. Phase separation in aqueous dispersion of Na-fluorohectorite [15].

Based on synchrotron x-ray scattering experiments [9] as well as on visual observations in crossed polarizers [15], we have found that the phase just above a lowest sedimented phase has nematic like ordering of fluorohectorite platelet particles, and the following phase diagram has been established (adopted from [15]):

Some of he first ideas and observations of nematic like ordering [22] in clays date back to Langmuir [23], and it has been a topic of several investigations both in clays and in other systems during recent years. Nematic networking in materials may have practical relevance for proerties of nanopatterns, nanocomposites, and macroscopically anisotropic gels.

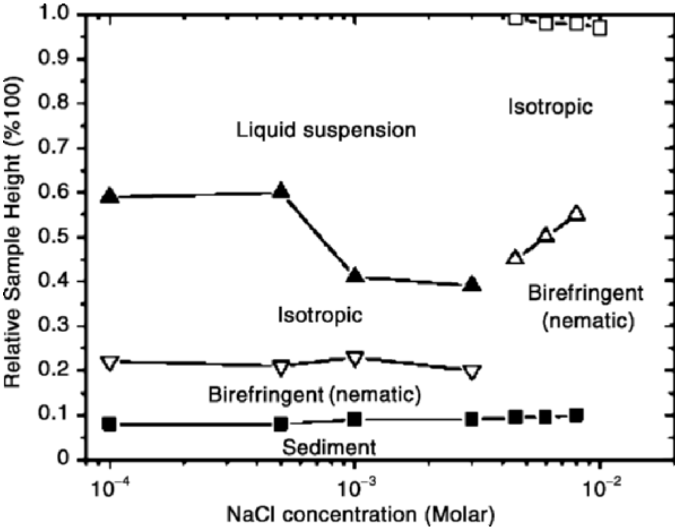


Figure 8. Phase diagram of aqueous dispersion of Na-fluorohectorite [9].

2.3. NETWORKS IN SMART MATERIALS: ELECTRO-RHEOLOGICAL BEHAVIOR OF CLAY SYSTEMS [17]

This activity started out using *the same particles as studied in activity 2.2*, above, but instead of immersing the particles in water, we immerse them in oil, and apply electric fields to the suspension, thus polarizing the clay particles, which in turn form chains in between the electrodes due to dipolar interactions.

Later we have shown that also natural clays show the same behavior, which maybe not so strange because our synthetic fluorohectorite particles are very much like natural illite particles for example quick clays in Norway. We have so far studied these problems using optical microscopy and using both SAXSandWAXS synchtrotonscatteringtechniques.Wehavethus investigated time scales for polarisibility of clay particles, we have determined the "directions of easy polarisibility" of clay particles, we have determined distributions of relative orientation of clay particles inside the dipolar chains, and we have determined field strengths needed for chain formation. The following collection of figures summarizes our own optical microscope observations, as well as our recent wide-angle-x-ray-scattering (WAXS) experiments performed at the Swiss-Norwegian Beam Lines (SNBL) at ESRF in Grenoble. In the upper right hand panel: Observations in an optical microscope of chaining of polarized copper electrodes. In the bottom left-hand corner panel: Bragg-scattering from oil suspension of Na-fluorohectorite in between Cu-electrodes in zero and in finite electric field showing change from pure random powder Bragg-scattering to ansitropic

Bragg-scattering due to orientation of clay particles in the electric field (one dimensional representation is shown un the lower right panle). The experiments show that the particles polarize perpendicular to the particle stacking direction (100 layers stack where each layer is 1 nm thick), and the data also measures the degree of misalignment of particles trapped in the chain-structure, for more details see [17].

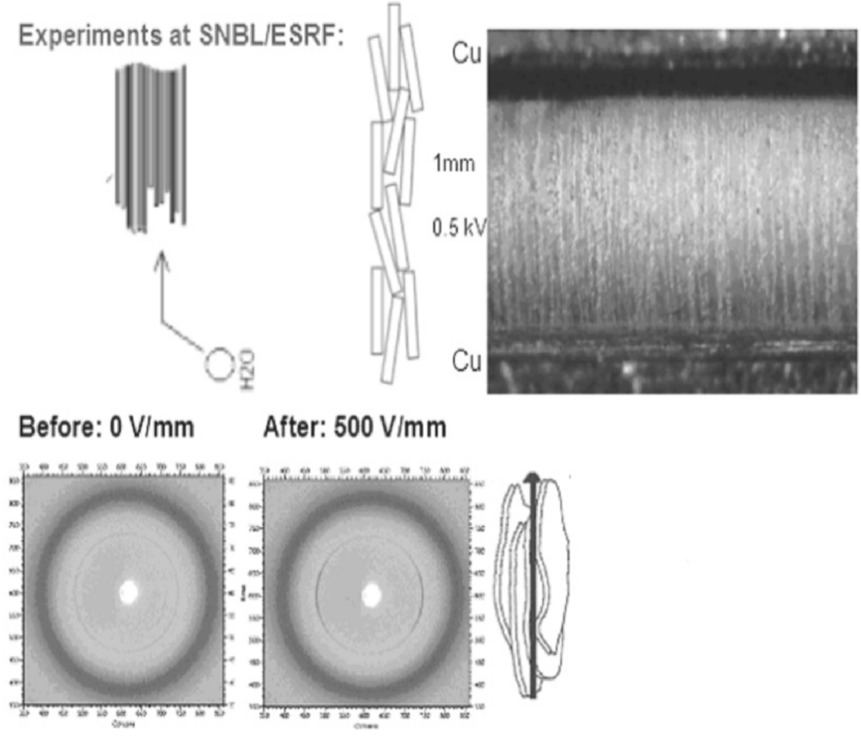


Figure 9. Upper panels: Left: Scetch of fluorohectorite particle. Middle: Scetch of dipolar chain with disorder. Right: Microscope image of dipolar chain formation between copper electrodes. Lower panels: Left: Narrow ring is isotropic powder scattering from oil-suspended fluorohectorite particles (E=0). Broader ring scattering from silicone oil. Middle: Anisotropic scattering from dipolar chains (E = 500 V/mm). Right: Arrow indicationg polarization direction of clay particles.

Our present clay-in-oil-suspension observations exemplifies what happens in general when a colloidal suspension of electrically-polarizable particles in non-conducting fluids. When such suspensions are subjected to an external electric field, usually of the order of 1kV/mm, the particles become polarized, and subsequent dipolar interactions are responsible for aggregating a series of interlinked particles that form chains and columns parallel to the applied field. Like in our case of clay particles, this structuring occurs within seconds, and disappears almost instantly when the field is removed [30]-[34]. It coincides with a drastic change in rheological properties (viscosity, yield stress, shear modulus, etc.) of the suspensions [35], which is why they are sometimes called electro-rheological fluids (ERFs). This makes the mechanical behavior readily controllable by using an external electric field [30]-[36], like magneto-rheological fluids may be controlled by means of applied magnetic fields [37]. These structures thus are linear anisotropic networks that may be tuned on the nano-scale level by particle "design", and at the human scale by applied electric (or magnetic fields). Our own research in this field currently also includes electro-magneto-rheological behaviors in mixtures of laponite clay nano-platelets and magnetic ferro-fluid particles [38].

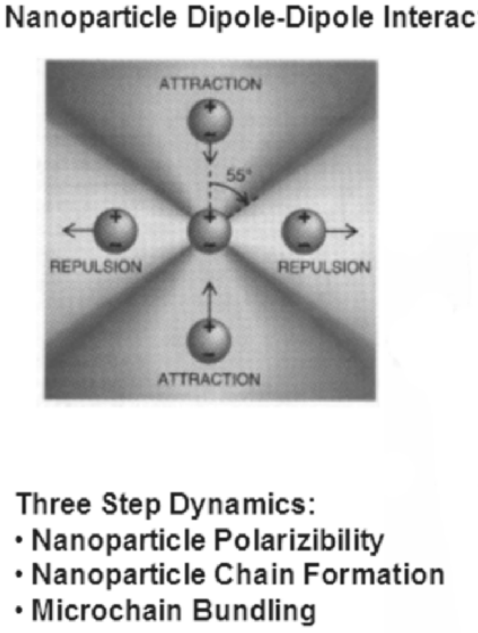


Figure 10. Physics of formation of dipolar networks, figure is scanned from [33].

This project is of relevance to applications for smart materials, smart structures, smart and adaptive materials networks etc.

We want to understand how clay platelet nanoparticles interact, what kind of structures they form and how these structures result in variuous macroscopic behaviors, like flow-, elastic- and optical- properties.

"LINEAR" NETWORKS

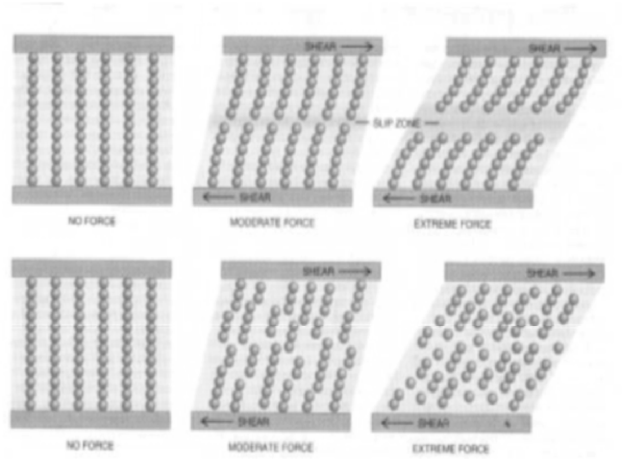


Figure 11. Possible breakdown mechanisms of dipolar chain networks, the figure is scanned from [33].

2.4. POROUS NETWORKS IN COMPLEX MATTER STRUCTURES: ANOMALOUS DIFFUSION FROM INTERPLAY OF MOLECULAR TRANSPORT NANOPORES AND MESOPORES? [16]

Global transport differing from the normal diffusion equation, or anomalous diffusion, attracts much attention within complexity physics. Recently it has been demonstrated from NMR-imaging experiments that water transport in porous networks in zeolites may be used as model systems for anomalous diffusion, or subdiffusion [39]. This indicates that global water transport in zeolite porous networks (for which the pore-network consists of both nano-pores and meso-pores) may be described in terms of a generalized diffusion equation including fractional derivatives, see [39] and references therein. It could be that this global water transport, and thus the numerical value of the fractional derivative "exponent" in this case reflects the underlying interplay between transport dynamics in nano-pores and meso-pores. Motivated partly by such questions, our group [1] has undertaken studies of water transport in systems of pressed, as well as of loosely packed powder samples of clays, in particular in fluorohectorite samples [16]. In our experiments, we have studied directly the water contents in the nano-pores by means of synchrotron wide-angle-x-ray-scattering (WAXS). As described in section 2.1 above, individual fluorohectorite particles are stacks of bout 100 x 1 nm thick sheets. We know that water may be intercalated in between the sheets and into the particles, mono-layer by mono-layer, up to 3 mono-layers of water for the case of fluorohectorite [9],[10],[12],[16]. This means that we by means of synchrotron WAXS can map out the global water-transport in such a system of a porous network defined by packed clay nano-particles by measuring the water-layer intercalation state as a function of space and time.

The following picture shows an example of a simple sample in a holder used during our recent experiments (2004-5) at the National Brazilian synchrotron source, LNLS, in Campinas Brazil, where we studied intercalation fronts supplying humid athmosphere to one end of the sample, while keeping the other end of the sample dry, and at the same time controlling the temperature by means of a Peltier element (also seen below the sample in the picture). The Bragg-scattering from the 3 cm long sample can in this case be accessed at any time, through four x-ray transparent windows seen along the sample. For details of these experiments, see [16].

The water transport (from left to right in the picture), surrounding our clay particles and also possibly their intercalation properties are essential for the polarizing properties discussed above (section 2.3), thus demonstration the interconnection of these different physical phenomena: “Nano-diffusion”and elctrorheology.

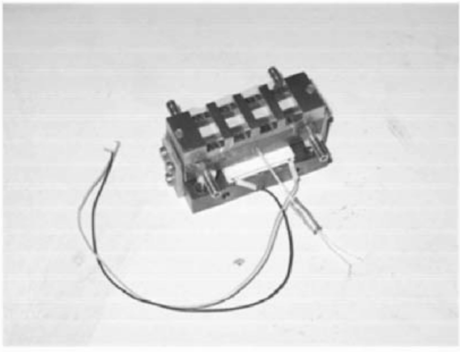


Figure 12. Temperature controlled sample cell for studies of "nano-diffusion" in clays.

There are still many open questions in this research [16], [39]-[41], and this project has relevance for universal description of molecular transport in systems with nano-pores, as well as for environmental questions connected to molecular transport in clayey soils.

2.5. SURFACE STRUCTURING FROM COLLAPSE OF 3-DIMENSIONAL NETWORKS [13]

Surfaces and profiles as characterized in terms of self-affinity are abundant in nature [42][43], and self-affine surfaces are characterized in terms of roughness Hurst, exponents [42]. One class of natural surfaces, which has attracted considerable attention during a number of years, are those associated with surface growth [42][43], where several experimental and theoretical model systems have been studied [43]. This is the background for a project in our group [1] concerned with collapsing 3-dimensional sol/gel structures into rough surfaces, and then asking the question whether there is any information left in the surface roughness (Hurst exponent) from the sol/gel network (fractal dimension).

We conducted an experiment, preparing samples as shown in the following picture, starting with droplets of 3 dimensional laponite samples taken from the laponite phase diagram discussed in section 2.1.2 above.

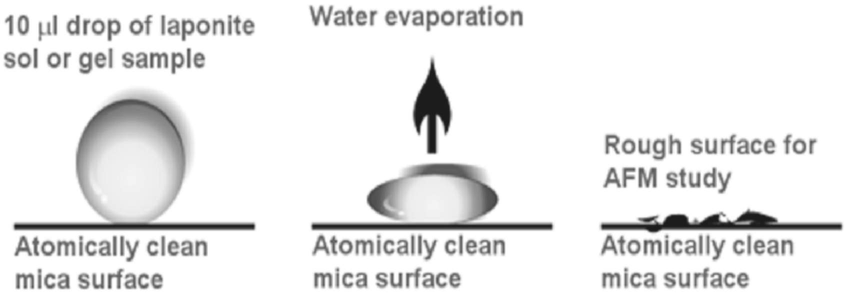


Figure 13. Scetch of sample preparation for AFM studies of rough deposited surfaces of laponite [13].

After drying, the rough surfaces are studied by means of Atomic Force Microscopy (AFM), and by analyzing the power spectra (PSD), of the 2-dimensional rough surfaces, the following was found for dryed laponite surfaces:

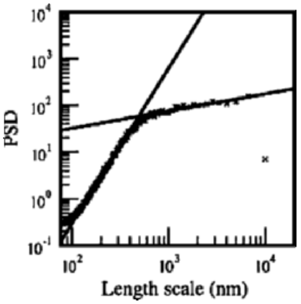


Figure 14. Power spectrum (PSD) of rough deposited clay surface [13].

At small length scales, the observed roughness presumably is a result of the collapsing of the clay aggregates due to the passing of the water meniscus.

This is supported by the fact that our AFM images of very dilute sol samples, for which isolated aggregates can be seen, suggest considerable vertical collapse as water is removed. The observed Hurst exponent of the surface reflects the structure of the clusters before they collapse. This may be understood as follows: A fractal structure in three-dimensions is projected onto a two-dimensional plane. The height of the projected structure at a given point is proportional to the fractal sitting above this point. All our AFM images are taken with a pixel size above the individual laponite particle size, which is about 1x25 nm^2. Thus, at our smallest length scales we always measure collapsed clusters of such individual particles, and the measured Hurst exponent *H* at small length scales, thus, tells us something about the structure of clusters before collapse. It can be shown that the Hurst exponent of a collapsed structure *H* is related to the fractal dimension of an un-collapsed structure *D* by the equation $D=2+H$. Hence, when we find a Hurst exponent of one this tells us that the fractal dimension of the suspended and un-collapsed clusters is three.

At large length scales, we essentially see the results of the sedimentation process, and a Hurst exponent of essentially zero signals a logarithmically rough surface, which is consistent with the annealed-disorder Edwards-Wilkinson growth equation in 2+1 dimensions [43]. For further details of this project, see [13].

3. Summary and concluding remarks

In the present review we have given examples of different physical phenomena observed in physical networks (networked aggregates and/or pore networks) realized in 2:1 clays. It is an important aspect of our different projects that we study the SAME material, from different points of view simultaneously and in parallel. Knowledge gained from one project is utilized in another, for example: We have demonstrated that studies of colloidal network structures of clay gels teaches us things about collapsed surface networks and surface roughness, and vice versa, we have seen that diffusion processes in clay samples teaches us things that are relevant for the polarizibility of clay particles exposed to external electric fields and vice versa, and we have seen that water suspensions of nano-layered clay particles teaches us important lessons to be used for explaining adaptive structures of clay particles suspended in oil. The phenomena under study are thus independent, in their own right, each with their own level of description [44], but still interconnected.

Here we have written about our activities on nano-layered silicates, This research may be said to link nano physics with macro physics, and the systems we study, namely clays enable us to interact with ideas both towards nano-technological applications (smart materials, diffusion in membranes, nanocomposites etc), and towards geology and environmental soil science. This is the strength of physics, namely the universal approach to problems [44].

Acknowledgements

The following collaborators have been especially instrumental in forming and performing the research described here:
K. D. Knudsen, Adjunct Professor, Department of Physics, NTNU and Research Scientist, Institute for Energy Technology, Kjeller, Norway
Y. Meheust, Research Scientist, Department of Physics, NTNU, Norway
A. Gmira, PostDoc, Department of Physics, NTNU, Norway
D. Fonseca and K. Parmar, PhD students, Department of Physics, NTNU, Norway
S. Berg Lutnaes, Master student, Department of Physics, NTNU, Norway
E. Gudding, Master student Department of Physics, NTNU, Norway
H. Huru Bergene, Master student Department of Physics, NTNU, Norway
K. J. Måløy, Professor, Physics Department, University of Oslo, Norway
G. Løvoll and B. Sandnes, PostDocs, Physics Department, University of Oslo, Norway
G. Helgesen, Research Scientist, Institute for Energy Technology, Kjeller, Norway
A. Elgsaeter and A. Hansen, Professors, Department of Physics, NTNU, Norway
J. Depeyrot, G. J. da Silva and F. A. Tourinho, Professors,, Institute of Physics and Institute of Chemistry, University of Brasília, Brasília, Brazil
R. Droppa, Research Scientist, LNLS, Brazilian Synchrotron Light Laboratory, Campinas, Brazil
E. DiMasi, Brookhaven National Labroatory, USA.
And in addition: All other coauthors on relevant publications cited below: [8]-[17].
This research is supported by the Research Council of Norway (RCN), through the NANOMAT Program: RCN project numbers 152426/431, 154059/420 and 148865/432, as well as through RCN project number 138368/V30 and RCN SUP154059/420.

References

1. This research is part of a Norwegian nationally coordinated research program on Complex Systems and Soft Materials: www.complexphysics.org . For specifics about our clay activities, see also: www.ntnu.no/~fossumj .
2. J.O. Fossum, "Complex Physical Phenomena in Clays", in Soft Condensed Matter: Configurations, Dynamics and Functionality, A.T. Skjeltorp and S.F. Edwards eds. 269-279, Kluwer Academic Publishers, the Netherlands (2000)
3. H. van Olphen, "Clay Colloid Chemistry", Krieger Puplishing Company Florida, 2nd ed.(1991)

4. A. Usuki, N. Hasegawa, M- Kato, “Polymer-clay nanocomposites”, INORGANIC POLYMERIC NANOCOMPOSITES AND MEMBRANES ADVANCES IN POLYMER SCIENCE 179: 135-195 (2005)
5. “Handbook of Clay Science”, Edited By F. Bergaya, B.K.G. Theng, G. Lagaly ISBN: 0-08-044183-1, Elsevier (2006)
6. S.A. Solin, "Clays and Clay Intercalation Compounds: Properties and Physical Phenomena", Annu. Rev. Mater. Sci. 27, 89-115 (1997)
7. “Clay minerals and the origin of life”, edited by A. G. Cairns-Smith and Hyman Hartman. Cambridge University Press (1986)
8. A. Bakk, J.O. Fossum, G. J. da Silva, H. M. Adland, A. Mikkelsen and A. Elgsaeter, Viscosity and Transient Electric Birefringence Study of Clay Colloidal Aggregation, Phys.Rev.E 65, 21407-1, (2002)
9. E. DiMasi, J.O. Fossum, T. Gog and C. Venkataraman, Orientational order in gravity-dispersed clay colloids: A synchrotron x-ray scattering study of Na fluorohectorite suspensions, Phys.Rev.E 64, 61704, (2001)
10. G. J. da Silva, J.O. Fossum, E. DiMasi, K. J. Maloy and S. B. Lutnaes, Synchrotron X-Ray Scattering Studies of Water Intercalation in a Synthetic Layered Alumino-Silicate, Phys.Rev E 66, 011303, (2002)
11. K-D. Knudsen, J.O. Fossum, G. Helgesen and V. Bergaplass, Pore characteristics and water absorption in a synthetics smectite clay, J. Appl. Cryst. 36, 587–591, (2003)
12. G. J. da Silva, J. O. Fossum, E. DiMasi, and K. J. Måløy, Hydration transitions in a nanolayered synthetic silicate: A synchrotron x-ray scattering study**,** , Phys. Rev. B 67, 094114, (2003)
13. J.O. Fossum, H. Huru Bergene, A. Hansen, B. O’Rourke and G. Manificat, Self-Affine Crossover Length in a Layered Silicate Deposit, Phys.Rev.E 69, 011303 (2004)
14. K-D. Knudsen, J.O. Fossum, G. Helgesen and M.W. Haakestad, Small-angle neutron scattering from a layered synthetic silicate, Physica B 352, 247, (2004)
15. J.O. Fossum, E. Gudding, D.M. Fonseca, Y. Meheust, E. DiMasi, T. Gog and C. Venkataraman, Gravity Dispersed Clay Colloids: Synchrotron x-ray Scattering Studies and Visual Observations of Orientational Order in Na-Fluorohectorite Suspensions, ENERGY The International Journal 30, 873, (2005)
16. G. Løvoll, B. Sandnes, Y. Méheust, K.J. Måløy, J.O. Fossum, G.J. da Silva, M.S.P. Mundim, R. Droppa Jr. and D.M. Fonseca, Dynamics of water intercalation fronts in a nano-layered synthetic silicate: A synchrotron X-ray scattering study Physica B: Condensed Matter, Volume 370, Issues 1-4, 90-98, (2005)
17. J. O. Fossum, Y. Meheust, K. P. S. Parmar, K. D. Knudsen, K. J. Maloy, D. M. Fonseca, “Intercalation-enhanced electric polarization and chain formation of nano-layered particles”, Preprint submitted (2006)
18. R.E. Hummel*,* "Understanding Materials Science", Springer-Verlag NY (1998)
19. A. Mourchid, E. Lecolier and H. van Damme, P. Levitz, "On Viscoelastic, Birefringent, and Swelling Properties of Laponite Clay Suspensions: Revisited Phase Diagram", Langmuir 14, 4718-4723 (1998)
20. M.J. Dijkstra, J-P. Hansen and P.A. Madden, "Gelation of a Clay Colloid Suspension", *Phys. Rev. Lett.* 75,2236 (1995), and M.J. Dijkstra, J-P. Hansen and P.A. Madden, "Statistical model for the structure and gelation of smectite clay suspensions", *Phys. Rev. E* 55, 3044-3053 (1997)
21. J. Israelachvili, "Intermolecular and Surface Forces", Academic Press London (1992)
22. L. Onsager, "The Effects of Shape on the Interaction of Colloidal Particles", Ann. N.Y. Acad. Sci. 51,627 (1949)

23. I. Langmuir, The role of attractive and repulsive forces in thr formation of tactoids, thixotropic gels, protein crystals and concervates, J.Chem.Phys. 6, 873 (1938)
24. J.P. Gabriel, C. Sanchez and P. Davidson, "Observation of Nematic Liquid-Crystal Textures in Aqueous Gels of Smectite Clays", *J. Phys. Chem.* 100, 11139-11143 (1996)
25. V.J. Andersen, H.N.W. Lekkerkerker. Nature 416(6883):811 (2002)
26. M.A. Bates and D. Frenkel, "Nematic-isotropic transition in polydisperse systems of infinitely thin hard platelets", J. Chem. Phys, 110, 6553-6559 (1999)
27. F.M. van der Kooij, K. Kassapidou, H.N.W. Lekkerkerker, Nature 406:868 (2000)
28. J-CP. Gabriel, F. Camerel, B.J. Lemaire, H. Desvaux, P. Davidson, P. Batall, Nature 413:504 (2001)
29. M. Kawasumi, N. Hasegawa, A. Usuki and A. Okada,"Nematic liquid crystal/clay mineral composits" *Mat. Science and Engineering* C6, 135-143 (1998)
30. A. P. Gast and C. F. Zukoski. Adv. Coll. Int. Sci., 30:153-202 (1989)
31. J. C. Hill and T. H. Van Steenkiste. J. App. Phys., 70(3):1207-1211 (1991)
32. T. C. Halsey. Science, 258:761-766 (1992)
33. T. C. Halsey and J. E. Martin. Scientific American, 269(4):58, october (1993)
34. W. D. Wen, W. Zheng, and K. N. Tu. J. App. Phys., 85(1):530-533, 1999
35. W. Wen, X. Huang, S. Yang, K. Lu, and P. Sheng. Nature Materials, 2:727, (2003)
36. A. F. Sprecher, J. D. Carlson, and H. Conrad. Material Science and Engineering, 95:187-197 (1987)
37. S. Odenbach, “Magnetoviscous Effects in Ferrofluids” (Lecture Notes in Physics), Springer Verlag (2002)
38. L.O. Paula, J. Depeyrot, G.J. da Silva, R. Aquino, F. A. Tourinho, J.O. Fossum, K.D. Knudsen, “SANS and SAXS Investigations of Colloidal Dispersions of Clay Nanoplatelets and Magnetic Nanoparticles”, preprint to be submitted (2006)
39. E.N. de Azevedo, P.L. de Sousa, R.E. de Souza, M. Engelsberg, “Concentration-dependent diffusivity and anomalous diffusion. A magnetic resonance imaging study of water ingress in porous zeolite”, preprint accepted for publication in Physical Review E (2006)
40. B.Y. Chen, H. Kim, S.D. Mahanti and T.J. Pinnavaia, Z.X. Cai, "Percolation and diffusion in two-dimensional microporous media: Pillared clays", J. Chem. Phys. 100, 3872-3880 (1994)
41. F.P. Duval, P. Porion and H. Van Damme, "Microscale and Macroscale Diffusion of Water in Colloidal Gels. A Pulsed Field Gradient and NMR Imaging Investigation", J. Phys. Chem. B 103, 5730-5735 (1999)
42. J-F. Gouyet, “Physics and Fractal Structures”*,* Masson, Paris, (1996)
43. A.-L. Barabasi and H.E. Stanley, “Fractal Concepts in Surface Growth”*,* Cambridge University Press, Cambridge, UK, (1995)
44. N. Goldenfeld, L.P. Kadanoff, “Simple lessons from complexity”, SCIENCE VOL 284, 2 APRIL (1999)

WHAT ECONOMISTS SHOULD LEARN FROM ECONOPHYSICS

JOSEPH L. MCCAULEY, GEMUNU H. GUNARATNE,
AND KEVIN E. BASSLER
Physics Deptartment
University of Houston
Houston, Tx. 77204 USA
jmccauley@uh.edu

Abstract. We state the usual postulatory approach used by economists and then contrast it with our empirically based discovery of the dynamics of financial markets, where all predictions are calculated from 'the market Green function'. In particular, we predict option prices in agreement with traders' valuations, but without using any nonempirically determined parameters. Both global and local volatility are defined via the noise traders' diffusion coefficient, and a new dynamic definition of 'value' is given. Self-fulfilling prophecies are discussed in the context of complexity.

Keywords. Market dynamics, Markov processes, volatility, option prices, economics and finance

1. Introduction

Economists postulate a model, nearly always stationary (near statistical equilibrium), and try to force fit it to empirical data by a best choice of arbitrary parameters (see, e.g., Chow & Kwan, 1998). They conclude that the data are too hard to fit over long time intervals. One object of this lecture is to explain that economic data are too easy to fit. In particular, finance data are very easy to fit over all observable time scales, and other economic data are far more sparse than finance data.

Neo-classical economic theory (utility maximization) is a falsified model but is still taught in all leading textbooks as if it would describe the ideal standard to be achieved by real markets (the model is used by The World Bank, the IMF, the EU, the US Treasury, and the US Federal Reserve (Stiglitz, 2002)). Another main aim of this work is to make economists aware that they must discard all existing standard texts and start

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 191–202.

over again, using empirically-based modelling (see also Soros (2000) for a related viewpoint).

By market dynamics we mean excess demand dynamics: $dp/dt=\varepsilon(p,t)$ where $\varepsilon(p,t)$ = excess demand = $D(p.t)-S(p,t)$, and where $D(p,t)$ is demand and $S(p,t)$ is supply at price p and time t. Financial markets suggest stochastic dynamics: price changes on the smallest time scales ($\Delta\tau\approx 1$ sec.) are not predictable, whereas in deterministic dynamics, even in chaotic and complex systems, changes on the smallest time scales are easily predictable due to local integrability.

In stochastic dynamics, excess demand is modelled as drift plus noise. Ignoring for the time being the fact that price and time changes in markets are discrete, excess demand dynamics in finance markets is pretty well described by the stochastic differential equation (sde)

$$dp=prdt+\sqrt{p^2d(p,t)}dB(t) \tag{1}$$

where r is an interest rate, $p^2d(p,t)$ is the price diffusion coefficient and $dB(t)$ is the Wiener process (dB/dt is white noise, but we use Ito calculus in order to preform coordinate transformations on (1) easily and systematically). The function $d(p,t)$ characterizes the market and must be discovered from the data (McCauley, 2004). If $d(p,t)$ would be constant then the price distribution would be lognormal (and the returns $x=\ln p$ would be Gaussian). But for real markets $d(p,t)\neq$constant and cannot be merely postulated.

Laws of physics are based on the four standard space-time symmetry principles (Wigner, 1967). Are there any corresponding symmetry principles for markets? Only one is known: the ‘no arbitrage’ condition applied to a single asset in spatially separated markets is a geometric invariance principle analogous to rotational invariance. There are no other known space-time invariance principles for markets.

1.1 PROVING ‘EQUILIBRIUM’ WITHOUT DYNAMICS

Economists love to prove that equilibrium ‘exists’ mathematically in a model, but dynamics is generally ignored. We now illustrate why existence proofs without dynamics are dangerous. The lognormal pricing model is defined by the sde (1) with $d(p,t) = \sigma_p$ = constant. The corresponding price density $g(p,t)$ satisfies the Fokker-Planck partial differential equation (pde)

$$\frac{\partial g}{\partial t} = -r\frac{\partial}{\partial p}(pg) + \frac{{\sigma_p}^2}{2}\frac{\partial^2}{\partial p^2}(p^2 g) \qquad (2)$$

whose fundamental time-dependent solution (Green function) is the lognormal distribution. The condition for statistical equilibrium is solved by the nonnormalizable function $g(p)=Cp^{-2r/\sigma 2}$ but statistical equilibrium is never reached by the (normalizable) lognormal Green function, which instead vanishes as t goes to infinity. Unbounded prices mean that equilibrum can't be attained (due to the continuous spectrum of the Fokker-Planck operator). Price controls, bounds on p(t), produce statistical equilibrium asymptotically via a discrete spectrum. The lognormal model is nonstationary, describing a hypothetical unstable market where the Gibbs entropy increases without bound. In spite of this fact finance theorists still talk about 'equilibrium markets'.

2. The Myth of the Invisible Hand

How would a hypothetical equilibrium market behave empirically? Market equilibrium would require that g(p,t) is asymptotically stationary (t-independent) over an observable market time scale (a week, a month, a year), or equivalently, that all moments of the distribution g(p) must become constants, independent of time t. Statistical equilibrium demands a stationary process asymptotically. In particular, the average drift <pr(p,t)> must vanish, guaranteeing that <ε(p,t)>=0, and because all higher moments of the distribution must be constant as well, the variance

$$\sigma^2 = \left\langle \Delta p^2 \right\rangle = \left\langle p^2 \right\rangle - \left\langle p \right\rangle^2 = {}_{t}^{t+\Delta t}\left\langle p(s)^2 d(p,s) \right\rangle ds \qquad (3)$$

must be asymptotically constant for a long enough time interval (t, t+Δt). This condition on the variance is easily tested and is badly violated by real markets, as is the condition for vanishing excess demand. In other words, real markets are far from statistical equilibrium, equilibrium (stationary process) is a completely illegal zeroth order approximation to market reality.

Why should anyone care about equilibrium? Because were equilibrium to hold, then we could take the equilibrium price p* to be either the average or most probable price, and this would yield a t-invariant definition of "value", in agreement with neo-classical economic theory. In the absence of equilibrium there is no t-invariant definition of "value": we will explain in part 6 that the ideas ‘undervalued’ and ‘overvalued’ when applied to future asset prices are effectively subjective. Next, we compare the equilibrium predictions with the empirical facts about financial markets.

3. Real financial markets

It is necessary empirically to study logarithmic returns

$$x(t) = \ln(p(t)/p_o) \tag{4}$$

rather than prices p(t), price increments δp, or small returns $\delta p/p$, where p_o is a reference price, because the variable x is both additive and units-free. A Markov process (1) is a good zeroth order approximation because it agrees with the efficient market hypothesis (EMH), which simply means that markets are very hard to beat. Using Ito calculus, the sde for x is

$$dx = (r - D(x,t)/2)dt + \sqrt{D(x,t)}dB(t) \tag{5}$$

where D(x,t)=d(p,t). "Volatility" is defined by the variance of x

$$\sigma^2 = \left\langle \Delta x^2 \right\rangle \approx c\Delta t^{2H} \tag{6}$$

where H is the Hurst exponent and $\Delta x=x(t+\Delta t)-x(t)$. For stationarity H=0 is required, but real markets yield $H\approx 1/2$ There is much extensive about stationarity in the finance literature). Real markets are nonstationary/unstable, there is no Invisible Hand to produce market stability. Traders, unable to know 'value' (as we explain in part 6), are uncertain and trade often, contributing to nonstationarity and volatility. These are the noise traders, the traders who provide liquidity in normal markets (Black, 1986).

Here’s how we constructed our finance market model. Start with an empirical time series x(t) and construct the market density f(x,t), f(x,t)dx=g(p,t)dp, as unmassaged histograms. The empirical distribution is approximately exponential, is far from Gaussian for small to moderate intraday returns x. Discovering the dynamics means discovering the t-dependence of three parameters (γ,ν,δ) defined in McCauley (2004) in the distribution. We used the global volatility $\sigma^2=\Delta t$ to discover that $\gamma,\nu=\Delta t^{-1/2}$, yielding

$$f(x,t) = \frac{A_\pm}{\sqrt{\Delta t}} e^{-|x-\delta|/\sqrt{\Delta t}} \tag{7}$$

Then, we plugged f(x,t) into the Fokker-Planck equation

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}(Rf) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(Df) \tag{8}$$

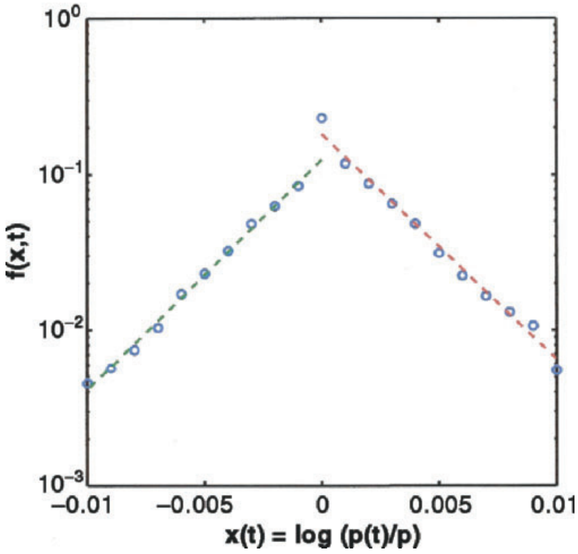


Figure 1. The empirical distribution of financial returns is exponential for small to moderate intraday returns.

where R is defined below and solved the 'inverse problem' to find D(x,t). The exponential distribution with x-dependence appearing in the form $x/\sqrt{\Delta t}$ is generated by the diffusion coefficient

$$D(x,t) = 1 + |x - \delta| / \sqrt{\Delta t} \tag{9}$$

and where

$$\delta = \langle \Delta x \rangle \approx R\Delta t \tag{10}$$

locates the peak of the distribution, the most probable return. In other words, and this is the main point, *we discovered the form of the noise the market*. We emphasize that the ‘local volatility’ D(x,t) haracterizes the so-called ‘noise traders’.

4. Volatility and option prices

Quite generally, the average/global volatility is

$$\sigma^2 = \left\langle \Delta x^2 \right\rangle = \int_t^{t+\Delta t} ds \left\langle D(x,s) \right\rangle = \int_t^{t+\Delta t} ds \int_{-\infty}^{\infty} dx' G(x,t;x',s) D(x',s) \tag{11}$$

where G(x,t;x’,t’) is the market Green function satisfying the Fokker-Planck pde (8) and D(x,t) is the ‘local volatility’, $\sigma^2 \sim D(x,t)\Delta t$ for $\Delta t \ll 1$. The empirical distribution (7) is the market Green function for x’=0, G(x,t;0,0)=f(x,t). Our option pricing prediction, based on the exact formula

$$C(p,K,T-t)e^{r(T-t)} = \left\langle (p_T - K)\theta(p_T - K) \right\rangle e^{r(T-t)} = \int_{\ln K/p}^{\infty} (pe^{x_T} - K) G(x_T,T;x,t) dx_T \tag{12}$$

agrees with traders’ prices without using any adjustable parameters (falsifiable model). In (12) C is the call price, T is the expiration time, t is the present time, p is the known price at time t, and K is the strike price at expiration. The reason that we can calculate option prices from the Market Green function is that, with the choice of R(x,t))=r-D(x,t)/2 satisfying the risk neutral hedge condition, the correct ‘Black-Scholes’ pde is, to within a time transformation, just the backward time Kolmogorov pde corresponding to the market Fokker-Planck pde (8) (McCauley, 2004).

There is nonuniqueness in deducing the t-dependence of the empirical density f(x,t) from the data, but we have the luck that for option pricing the nonuniqueness doesn't matter on a time scale small compared with 100 years. Due to uniqueness in modelling empirical data via infinite precision dynamics, this leads to the viewpoint that the data are *too easy* to fit over long times To be honest, we already know this important lesson from nonlinear dynamics (Chhabra et al., 1989).

4.1 LIQUIDITY, NOISE TRADERS, AND CRASHES

The essential unstated assumption so far is that we have an adequate 'liquidity bath'. By a normal market we mean the following: A liquidity bath is assumed, meaning that approximately reversible trades are possible via your discount broker in real time over the shortest time intervals (Δt is on the order of a few seconds) on your Imac or PC. This assumption is represented by the noise term $\sqrt{D(x,t)}dB(t)$, which describes the uncoordinated actions of the "noise traders". Noise traders provide the liquidity/entropy in the market. Mathematically seen, noise traders *are* the market ('with measure one'). Noise traders, uncertain about 'value', buy and sell often: a financial market is largely noise because most traders don't have either inside or other knowledge to trade on. Actually, it was von Neumann who suggested to Shannon to look for market entropy in liquidity. The liquidity/money bath is analogous to a thermal heat bath, but the liquidity bath cannot be described by equilibrium ideas like temperature.

Fat tails do not describe market crashes, fat tails describe large returns that occur during perfectly normal markets. In contrast, a market crash is a *liquidity drought* (the noise traders can't sell because there are no buyers) and is described qualitatively by $R<<0$ and $D(x,t)\approx 0$.

5. Three Easy Pieces

We study the pde

$$\frac{\partial f}{\partial t} = -R\frac{\partial f}{\partial x} + \frac{1}{2}\frac{\partial^2}{\partial x^2}(Df) \tag{13}$$

with R=constant (Alejandro-Quinones et al, 2004). To satisfy the replicating self-financing hedge condition in option pricing (risk neutral hedge) we need $R=r-D(x,t)/2\neq\text{constant}$, but we can take $R \approx$ constant on any time scale small relative to 100 yrs: this is part of the nonuniqueness. Therefore, we can study

$$\frac{\partial f}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}(Df) \tag{14}$$

and then replace x by x-RΔt. This partial differential equation has nice scaling properties. In order to find out, set u=x/√Dt, f(x,t)=F(u)/√Dt, and D(x,t)=D(u). The result: given the returns density f(x,t) we can calculate the local volatility D(x,t), and vice/versa, and we can do that analytically for at least three essential cases.

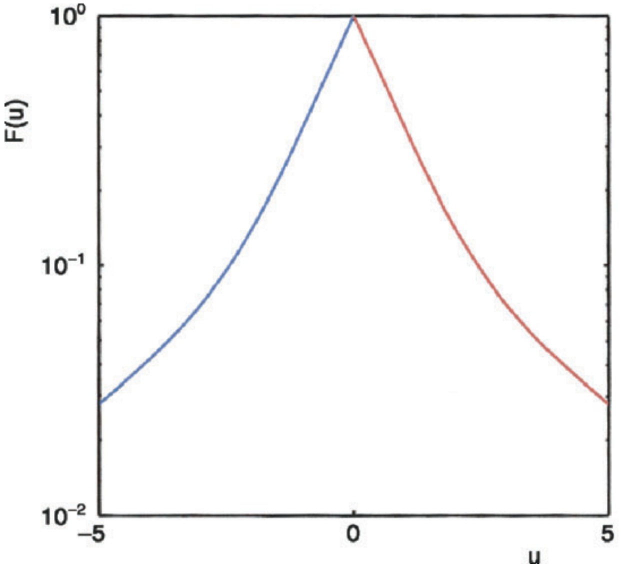


Figure 2. The market distribution is exponential for moderate returns but has fat tails for large returns. The tail exponents are nonuniversal and range from 2 to 7 (see also Dacorogna et al., 2001).

If

$$D(u) = 1 + \varepsilon|u| \tag{15}$$

then

$$F(u) = Ce^{-|u|((\varepsilon|u|+1)^{\alpha-1})/\varepsilon} \tag{16}$$

where $\alpha = \varepsilon^{-2}$. As ε increases then the tails of the distribution decay more slowly. The limit ε=0 yields the Gaussian, and ε=1 yields the exponential (7).

Next, if

$$D(u) = 1 + \varepsilon u^2 \tag{17}$$

then the result is surprising: we can generate (nonstationary!) fat tails

$$f(x,t) \approx |x|^{-\mu}, |x| >> 1 \tag{18}$$

for all tail exponents $2<\mu<\infty$: the exact solution is

$$F(u) = \frac{C}{(1+\varepsilon u^2)^{1+1/2\varepsilon}} \approx u^{-2-1/\varepsilon}, |u| >> 1 \tag{19}$$

where $\mu=2+1/\varepsilon$. Note that were even one extra higher order term required in (17) to generate fat tails, then the empirically observed tail exponent μ would not determine all the free parameters in D(x,t).

We can generate the observed financial distribution (Fig. 2) via a 'noise trader function'

$$D(u) = 1 + |u| + \varepsilon u^2 \tag{20}$$

where the tail exponent $\mu=2+1/\varepsilon$ uniquely determines ε. A decisive test of our model would be to measure D(x,t) empirically, which is very hard: Peinke (2001) tried, but his results fail for large returns x because he inadvertently made a small returns approximation. Finally, One sees immediately from (12) that option prices diverge if fat tails are included: *option traders do not and cannot insure against fat tails.*

6. Our new dynamic definition of "value"

Both f(x,t) and D(x,t) have extrema at $x=\ln p_c/p_o=\delta$ where $p_c=p_o e^{\delta}$ is the most probable price, and p_o is the initial most probable price. The price p_c defines the 'consensus price' and so represents the most widely agreed upon "value" of the asset at time t. This is our non-neo classical definition of value. However, the peak $\delta\approx R\Delta t$ of f(x,t) does not stand still, it can shift suddenly in a market crash and even in normal intraday trading (the expected return R can change suddenly, discontinuously, with sudden changes in noise traders' sentiments). In other words, "value" is very far from a time-invariant idea and depends on what the noise traders believe about an asset at any given time. In particular, 'value' is impossible to know in advance (complexity), we can at best know what value was at different times in the past. This means that notions like 'overvelued' and 'undervalued' are knowable only at the present time or historically, but cannot be predicted any degree of confidence for the future. This viewpoint is completely non-neo classical. Pricing options by using the empirical density (7) rather than the market Green function (still analytically unknown, hard to calculate) means approximating the present observed asset price p(t) by the consensus price.

Fischer Black (1986) was wrong: there is no tendency for price to 'return to value' because market dynamics are unstable, prices always diffuse away from 'value', there being no 'springs' in the market to pull prices back to value. But Soros (1998) was right, *financial markets are dynamically unstable*.

7. Market Complexity

So far, we've discussed nothing but simple stochastic dynamics that generates the historic statistics, so where's the complexity? Predictions based on past statistics hold so long as there are no basic market shifts, or 'surprises'. Surprises generated by UTM (universal Turing machine) dynamics and undecidability were discussed by Moore (1990, 1991) and at the 2001 Geilo School (Skjeltorp and Viscek, 2002). Insurance companies assume that tomorrow will be statistically like yesterday. This assumes that the noise traders never change their diffusion coefficient D(x,t), never change their noisy behavior/psychology. This assumption will fail in an unknown way at some unknown time in the future.

There are also *self-fulfilling expectations* that are not merely a repetition of past statistics, but represent the creation of something new. Examples of self-fulfilling expectations are communism via dictatorship (regulatory extremism) and globalization via deregulation (free market extremism). Feyman contrasted nonthinking nature with socio-economic phenomena and pointed out that the latter are very different from physics because wishful thinking can be made into reality by acting on it. But just as Turing said of numbers and arithmetic, we can assert about physics that mathematical laws of nature are beyond human invention, convention, and intervention, whereas all market phenomena are human-made, are invented by human will and actions. Here, we connect with Wigner (1967): a single space-time invariance principle (the assumption that arbitrage is to zeroth order impossible between spatially separated markets) is inadequate to pin down time-invariant mathematical laws of motion. Our market distribution is only a model, not a fixed market law, and will fail when the noise traders eventually change their habits enough that (21) no longer describes the market.

Neo-classical economics, the theoretical basis for globalization via deregulation that assumes that perfect knowledge of the future on the part of all traders, is a falsified model (McCauley, 2004) is *mathematized ideology*. There is neither simple uncertainty nor complexity in that ideology. Nor can the model be relaxed to make it perturbatively realistic in any sense: instead of approximately perfect knowledge (vanishing entropy), real markets reflect large and ever increasing entropy

$$S(t) = -\int_{-\infty}^{\infty} f(x,t)\ln f(x,t)dx \qquad (21)$$

due to liquidity. Far from randomness and other simplicity, a very few traders do not generate noise but also do not behave predictably. E.g., George Soros (1998) defeats self-fulfilling expectations of opponents (e.g., the Bank of England) by generating *surprises* (or psychological tricks). Soros (1998, 2000) tries to describe how traders behave and discusses self-reference and the Cretan Liar with in light of Gödel's incompletelness theorem. However, it's not clear to this writer that Soros does anything more complicated that to play winning poker with an adequate bankroll (he avoids the gamblers' ruin).

JMC is extremely grateful to Harry Thomas for a very careful reading of the preliminary manuscript and for suggesting improvements in the presentation (part 1.1 remains in spite of his objection), and likewise to Cornelia Küffner who also read the earlier ms and suggested clarifications of several key phrases. GHG and KEB (DMR # 0406323) thank the NSF for financial support.

References

Alejandro-Quinones, L., Bassler, K.E., Gunaratne, G.H., McCauley, J.L., Field, M,. Nicol, I., Timofeyev, A., and Török, A. 2005. *A Theory of Fluctuations of Stock Price Fluctuations*, submitted.

Black, F. 1986. *J. of Finance* **3:** 529.

Chhabra,A., Jensen., R.V., & Sreenivasan, K.R. 1988. Phys. Rev. **A40**: 4593.

Chow, G.C., and Kwan, Y.K. 1998. *J. Monetary Econ***.** **41**: 30.

Dacoroga, M.M., Ramazan, G., Müller, U.A., Olsen, R.B., and Picte, O.V. 2001. *An Intro. to High Frequency Finance.* Academic Pr., N.Y.

McCauley, J.L. 2004. *Dynamics of Markets: Econophysics and Finance.* Cambridge Univ. Pr., Cambridge.

Moore, C. 1990. *Phys. Rev*, Lett. **64**,23;**1991.** *Nonlinearity* **4**, 199 & 727.

Renner, C., Peinke, J., & Friedric, R. 2001. *Physica* **A298,** 49.

Skjeltorp, A.T. and Viscek, T. 2002. *Complexity from Microscopic to Macroscopic Scales: Coherence and Large Deviations*. Kluwer, Dordrecht.

Soros, G. 2000. *Open Society, Reforming Global Capitalism*. Public Affairs: N.Y.
G. Soros *The Crisis of Global Capitalism* (Little, Brown & Co.: N.Y., 1998).
J.E. Stiglitz *Globalization and its Discontents* (Norton: N.Y., 2001).
E.P. Wigner *Symmetries and Reflections* (Univ. Indiana: Bloomington, 1967).

THE MINORITY GAME: STATISTICAL PHYSICS OF COLLECTIVE BEHAVIOUR OF ADAPTIVE AGENTS IN A COMPETITIVE MARKET

DAVID SHERRINGTON*
Rudolf Peierls Centre for Theoretical Physics, University of Oxford,
1 Keble Road, Oxford OX1 3NP, United Kingdom
(`sherr@thphys.ox.ac.uk`*)*

Abstract. A brief review is given of the minority game, an idealized model stimulated by a market of speculative agents, with emphasis on its statistical physics.

The minority game is a minimalist model stimulated by considerations of a stockmarket of independent speculative agents trying to profit by buying low and selling high. It has potential interest both from the perspective of economics (or econophysics) and from that of statistical physics. In this paper I shall concentrate on the latter, in whose community there is currently much interest in the emergence of complex cooperative behaviour through the combination of competition (frustration) and inhomogeneity (disorder), even when the microscopic entities and interaction rules are simple.

The model describes a system of a large number N of agents each of whom at each step of a discrete dynamics makes a bid that can be either positive or negative (buy or sell). The objective of each agent is to make a bid of opposite sign from that of the sum of all the bids (*i.e.* a minority choice)[1]. Each agent decides his/her bid through the application of a personal strategy operator to some common information, available identically to all, but without direct knowledge of the strategies or actions of other agents[2]. We here restrict to the simplest case in which the strategy operators are allocated randomly and independently for each agent before play commences and are not modified during play. Each agent has a finite set of strategies, one of which is chosen and used at each step; for simplicity discussion below is restricted to two

---

[1] This is to emulate the way that a price rises if there are few sellers and falls if there are few buyers.

[2] This is to emulate the fact that in a stockmarket all have access to the stockmaket indices, world news etc., but not to the minds of their competitors.

A.T. Skjeltorp and A.V. Belushkin (eds.),
Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, 203–210.

strategies per agent. The choice is determined by ‘points’ allocated to the strategies and at each step increased if the strategy yields the actual minority choice and decreased if it yields the majority choice[3]. This is the only mechanism for co-operation but is sufficient to yield complex macroscopic behaviour.

In the original version of the model [2] the information on which decisions were made was the history of the actual play over a finite window (the last m time steps). However, simulations demonstrated that utilising instead a random fictitious ‘history’ (information) at each time-step produces essentially identical behaviour[3], suggesting that its relevance is just to provide a mechanism for an effective interaction among the agents. A natural non-trivial measure of the macroscopic behaviour is the volatility, the standard deviation of the total bid. It demonstrates statistical physics interest in several ways: (i) in exhibiting non-trivial scaling behaviour as a function of $d = D/N$, where D is the information dimension[4][4], (ii) in exhibiting a cusp at a critical d_c following a *tabula rasa* start , and especially (iii) in that the system is ergodic with volatility independent of starting point allocations for $d > d_c$ but non-ergodic and preparation-dependent for $d < d_c$; these features are illustrated in fig 1 which shows both simulation and calculation (discussed below), a variety of different intial conditions and both uncorrelated and anti-correlated strategies[5].

Since the information on which the agents act is the same for all, this problem is manifestly mean-field [6]. It therefore offers the potential for exact solution for its macro-behaviour in the sense of the elimination of the microscopic variables in favour of self-consistently determined macro-parameters in the limit of large N [5]. The physics appears to be robust to variations of detail, but for completeness the version discussed explicitly is detailed.

Each agent i, $i = 1, \ldots, N$, is taken to have two $D = dN$-dimensional strategies $\mathbf{R}_{ia} = (R^1_{ia}, \ldots, R^{dN}_{ia})$, $a = \pm 1$, with each component R^{μ}_{ia} chosen independently randomly ± 1 at the outset and thereafter fixed. The common random information enters in that $\mu(t)$ is chosen stochastically randomly at each time-step t from the set $\mu(t) \in \{1, \ldots, D\}$

---

[3] Note that the minority choice is determined by the cumulative actions of all the agents.

[4] $D = 2^m$ in the original case but can be generalized to any value

[5] In the case of anti-correlated strategies each agent’s two strategies yield opposite choices when they act on any given information, although there is still no correlation between the actions on different information fields, nor between agents.

[6] Indeed it is similarly clear that the real stockmarket is dominantly mean-field, albeit temporally non-local.

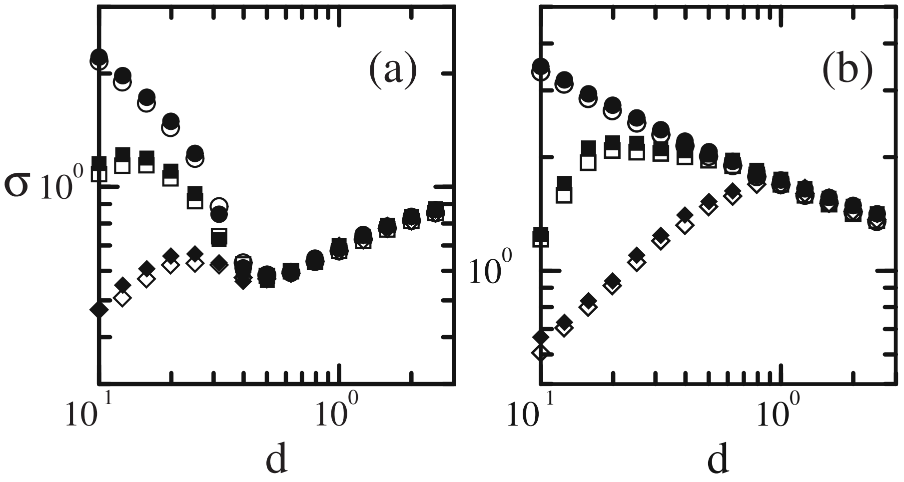


Figure 1. Volatilities in batch minority games with 2 strategies per agent; (a) with completely uncorrelated strategies, (b) with each agent's 2 strategies mutually anti-correlated but with no correlation between agents. Shown are three different bias asymmetries between the points allocated initially to each agent's 2 strategies: $p_i(0) = 0.0$ (circles), 0.5 (squares) and 1.0 (diamonds). Also exhibited is a comparison betweeen the results of simulation of the deterministic many-agent dynamics (open symbols) and the numerical evaluation of the analytically-derived stochastic single-agent ensemble dynamics. From [7].

and each agent plays one of his/her two strategies $R^{\mu(t)}_{ia_i}$, $a_i = \quad 1$. The actual choices of a_i used, hereafter called $b_i(t)$, are determined by the current values of point differences $p_i(t)$; for the deterministic case through $b_i(t) = \text{sgn}(p_i(t))$[7]. The $p_i(t)$ are updated every M time-steps according to

$$p_i(t+M) = p_i(t) - M^{-1} \sum_{\ell=t}^{t+M-1} \xi_i^{\mu(\ell)} \left\{ N^{-1/2} \sum_j (\omega_j^{\mu(\ell)} + \xi_j^{\mu(\ell)} \text{sgn}(p_j(t))) \right\} \quad (1)$$

where $\boldsymbol{\omega}_i = (\mathbf{R}_{i1} + \mathbf{R}_{i2})/2$, $\boldsymbol{\xi}_i = (\mathbf{R}_{i1} - \mathbf{R}_{i2})/2$.In the so-called 'online' game $M = 1$ but as M is increased to $M \geq O(N)$ the so-called 'batch'

---

[7] Generalization to stochastic update is straightforward but not discussed here [6].

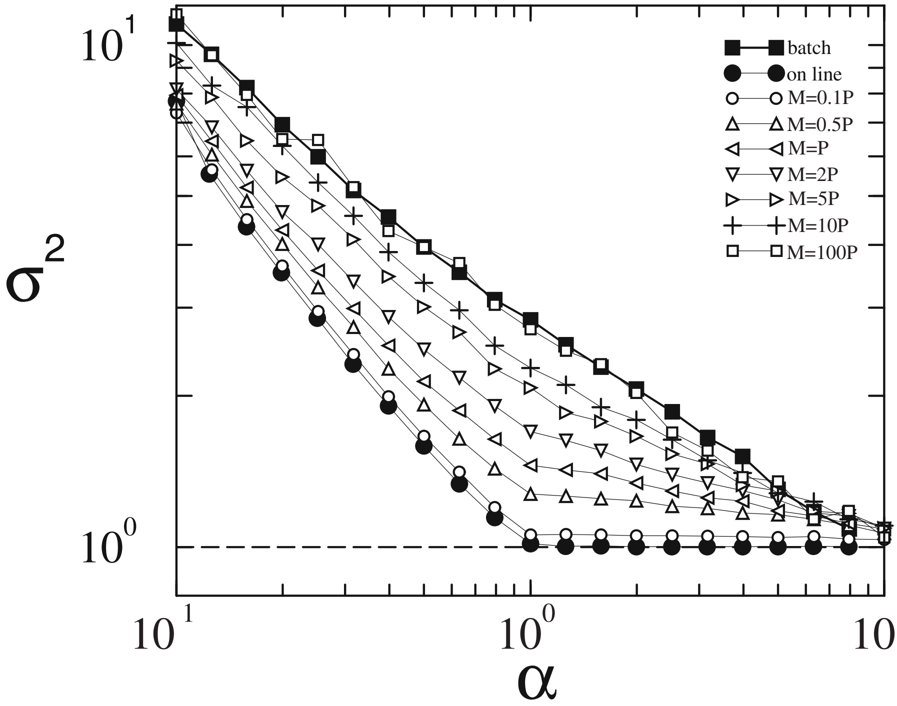


Figure 2. Volatility of batch minority games with anti-correlated strategies for a range of values of the number of time-steps M between point updates, going between $M = 1$ (on-line) on the left to $M >> N$ (equivalent to batch) on the right. ($P \equiv N$) From[8].

game results in which $\mu(\ell)$ in (1) may be replaced by an average [8] so that[8,9]

$$p_i(t+1) = p_i(t) - \sum_j J_{ij} \mathrm{sgn}(p_j(t)) - h_i \equiv p_i(t) - \partial H / \partial s_i|_{\, s_i = \mathrm{sgn}(p_i(t))}; \quad (2)$$

where $J_{ij} = N^{-1} \sum_{\mu=1}^{D} \xi_i^\mu \xi_j^\mu$, $h_i = N^{-1/2} \sum_{\mu=1}^{D} \omega_i^\mu \xi_i^\mu$, $H = \sum_{(ij)} J_{ij} s_i s_j + h_i s_i$. The results in fig 1 are for the batch limit; for uncorrelated patterns the two limits yield similar results but for anti-correlated

---

[8] The time unit is rescaled here.

[9] Thus the system behaves as if controlled by a 'Hamiltonian' of effective spins interacting through random exchanges and in random fields. The random exchange term has a form analagous to that of the Hopfield neural network model, but crucially with opposite sign; the μ are analogues of the labels of the stored memories and the ξ of their states, but they are now repellers rather than attractors. In Hopfield model convention the reduced variable d is usually called α and that notation is employed in figs 2 and 3.

strategies there are significant differences of detail, as can be seen in fig 2.

To proceed analytically a dynamical generating functional method [9] is employed, with

$$Z = \int \prod_t d\mathbf{p}(t) W(\mathbf{p}(t+1) \mid \mathbf{p}(t)) P_0(\mathbf{p}(0)), \tag{3}$$

where $\mathbf{p}(t) = (p_1(t), \ldots, p_N(t))$, $W(\mathbf{p}(t+1) \mid \mathbf{p}(\mathbf{t}))$ denotes the transformation operatation of eqn. 2 and $P_0(\mathbf{p}(0))$ denotes the probability distribution of the initial score differences from which the dynamics is started. The typical case results by averaging over the specific choices of quenched strategies. The averaged generating functional may then be transformed exactly into a form involving only macroscopic but temporally non-local variables ($\tilde{\mathbf{C}}$, $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{K}}$) relatable to the correlation and response functions of the original many-agent problem:

$$Z = \int D\tilde{C}(t,t') D\tilde{G}(t,t') D\tilde{K}(t,t') \exp\Big(N\Phi(\tilde{\mathbf{C}}, \tilde{\mathbf{G}}, \tilde{\mathbf{K}})\Big), \tag{4}$$

where Φ is independent of N and the bold-face notation denotes matrices in time. This expression is extremally dominated in the large N limit and consequently permits steepest descents analysis. This yields an effective stochastic single agent dynamics[10]

$$p(t+1) = p(t) - \sum_{t' \leq t} (\mathbf{1} + \mathbf{G})^{-1}_{tt'} \mathrm{sgn} p(t') + \sqrt{\alpha}\eta(t), \tag{5}$$

$$\text{where} \quad \langle \eta(t)\eta(t') \rangle = [(\mathbf{1} + \mathbf{G})^{-1}(\mathbf{1} + \mathbf{C})(\mathbf{1} + \mathbf{G^T})^{-1}]_{tt'} \tag{6}$$

and the $\mathbf{G}$ and $\mathbf{C}$ are two-time response and correlation functions determined self-consistently as averages over an ensemble of such single agents[11]. In the limit of large N this analysis is believed to be exact, but it is highly non-trivial. Empirical evidence for its correctness is shown in fig 1 where comparison is made between the results of simulations over many instances of the many-agent eqn. 2 and numerical evaluations of the analytically-derived single-agent dynamics of eqn. 5, including extension to anti-correlated strategies [8].

Hence, naive characterization in terms of a unique deterministic ‘representative agent’, as common in conventional economics theory, is not possible. However, a single effective-agent description is available in a much more subtle sense. This is that one can consider the system to behave as though one has a **‘representative stochastic ensemble’**

---

[10] For simplicity we restrict to uncorrelated patterns here, although extension to correlated strategies is possible and is discussed in [8] and is used in fig 1b.

[11] $\mathbf{C}$ and $\mathbf{G}$ are the extremally dominating ‘values’ of $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{G}}$. See [9] for details.

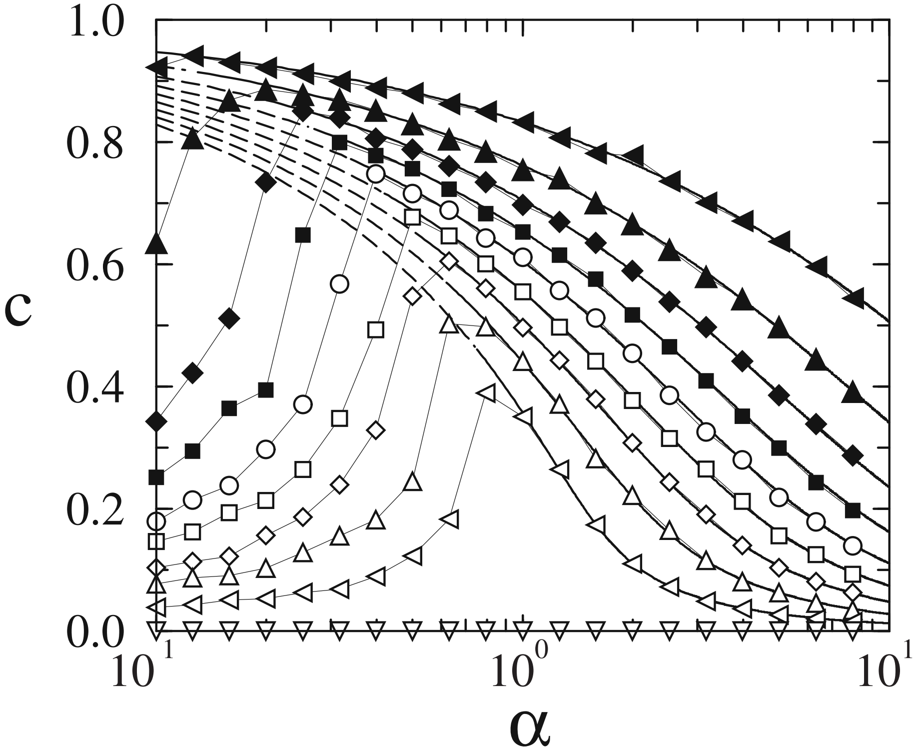


Figure 3. Persistent part Q of the correlation function for the batch MG with *tabula rasa* initial conditions. Symbols are simulation data. Solid lines are the theoretical predictions for the ergodic regime, extrapolated as dashed lines into the non-ergodic phase below d_c (where they are no longer valid), the changeover signalling the predicted breakdown of the ergodic assumption. The different curves are for different degrees of mutual correlation between agents' two stategies; from anticorrelated at the bottom to highly correlated at the top. From[8].

of non-interacting agents experiencing memory-weighting and coloured noise, both determined self-consistently over the ensemble[12].

A complete closed-form solution the effective single-agent ensemble dynamics is currently not available. However, one can solve for aymptotic long-time behaviour in the ergodic equilibrating region. where stationarity ensures that the two-time correlation and response functions become functions only of the relative times. Assuming also finite integrated response and weak long term memory leads to a formulation determining self-consistently the asymptotic order parameters $Q = \lim_{\tau\to\infty} C(\tau)$ and the integrated response $\chi = \sum_{\tau} G(\tau)$. Breakdown of the ergodic regime is signalled by diverging integrated response. Again the analytic theory works well within this ergodic regime, as

---

[12] Note that eqn. 5 is stochastic even though eqn. 2 is deterministic.

is demonstrated in fig 3. The volatility, however, requires also the non-stationary parts of C and G and remains incompletely solved in general, even in the ergodic regime [8].

It is of interest to query the origin of the large volatilities found for *tabula rasa* starts in the non-ergodic regime. These can be ascribed to oscillatory behaviour, clearly visible empirically in the temporal correlation function

$$C(\tau) = \lim_{t\to\infty} N^{-1} \sum_i \mathrm{sgn}(p_i(t+\tau))\mathrm{sgn}(p_i(t)), \tag{7}$$

which exhibits persistent oscillations (with period 2 in the rescaled time units of eqn. 2) for $d < d_c$ [8]. *Tabula rasa* starts in this region exhibit essentially no frozen agents, whereas highly biased starts result in mostly frozen agents and hence reduce the oscillations and with them the excess time-averaged volatility. The oscillations and the excess volatility are also reduced by random asynchronous point updating [8] and by adding appropriate stochasticity to the original MG dynamics [10][6].

Thus, as well as its possible relevance as an idealized economics model, the Minority Game is of interest as a novel complex many-body system with both similarities and differences compared with other problems previously studied in statistical physics. Techniques developed within the spin glass community have proven useful in its analysis and suggest extensions to other dynamical many-body systems characterised by a combination of local/personal and global/range-free parameters.

Acknowledgements

Thanks are due to A. Cavagna, T. Coolen, T. Galla, J. Garrahan, I. Giardina, A. Heimel and E. Moro for collaborations and to EPSRC, ESF (SPHINX) and EC (STIPCO) for financial support.

References

1. See e.g. H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, 2001)
2. D. Challet and Y.-C. Zhang, Physica A **246** (1997) 407
3. A. Cavagna, Phys. Rev. E **59** (1998) R3783
4. D. Challet and Y.-C. Zhang, Physica A **256** (1998) 514

5. D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35** (1975) 1792
6. A.C.C. Coolen, A. Heimel and D. Sherrington, Phys. Rev. E **65** (2001) 16126.
7. T. Galla and D. Sherrington Physica A 324, (2003) 25
8. T. Galla and D. Sherrington Eur. Phys. J. B **46** (2005) 153
9. A.C.C. Coolen, *Mathematical Theory of Minority Games* (Oxford University Press, 2005)
10. A. Cavagna, J.P. Garrahan, I. Giardina and D. Sherrington, Phys. Rev. Lett. **83** (1999) 4429

INDEX

adaptive agent, 203
alpha helix, 61

biosignalling, 71
braids, 147

cell, 41, 129
clay, 169
complex matter, 173, 182
complex systems, 107
computational neuroscience, 127
connectivity, 5
cortical populations, 127
crack propagation, 49

depinning transition, 49
diffusion, 66
DNA, 4, 41, 163
dynamical scaling, 49

econophysics, 191
electro-rheology, 179

ferrofluid, 147
flux balance, 107
fragmentation, 61

Garside algoritm, 153
gene regulation, 1
graph theory, 2

hierarchical network, 116

interfacial fracture, 49

macroscopic stability and dynamics, 173
market dynamics, 191

Markov processes, 191

metabolism, 107
Metropolis rewiring algorithm,36
Minority Game, 203
mutagenesis, 41

nanostructured materials, 173
network, 1, 31, 107, 173
neural networks, 127
nucleotide, 41

optical tweezers, 163

option prices, 191

photoreception, 71
phototransduction, 92
protein, 65

random network, 113
retina, 72
RNA, 24, 42

scale-free network, 107, 113
self-affinity, 49
Small World Network, 4
social networks, 20
soft matter, 173

topological hierarchy, 1
transcription, 6

vision, 71, 127
visual cycle, 99
volatility, 196

World Wide Web, 31, 107

Zipf law, 23, 148
Zipf-Mandelbrot relation, 148