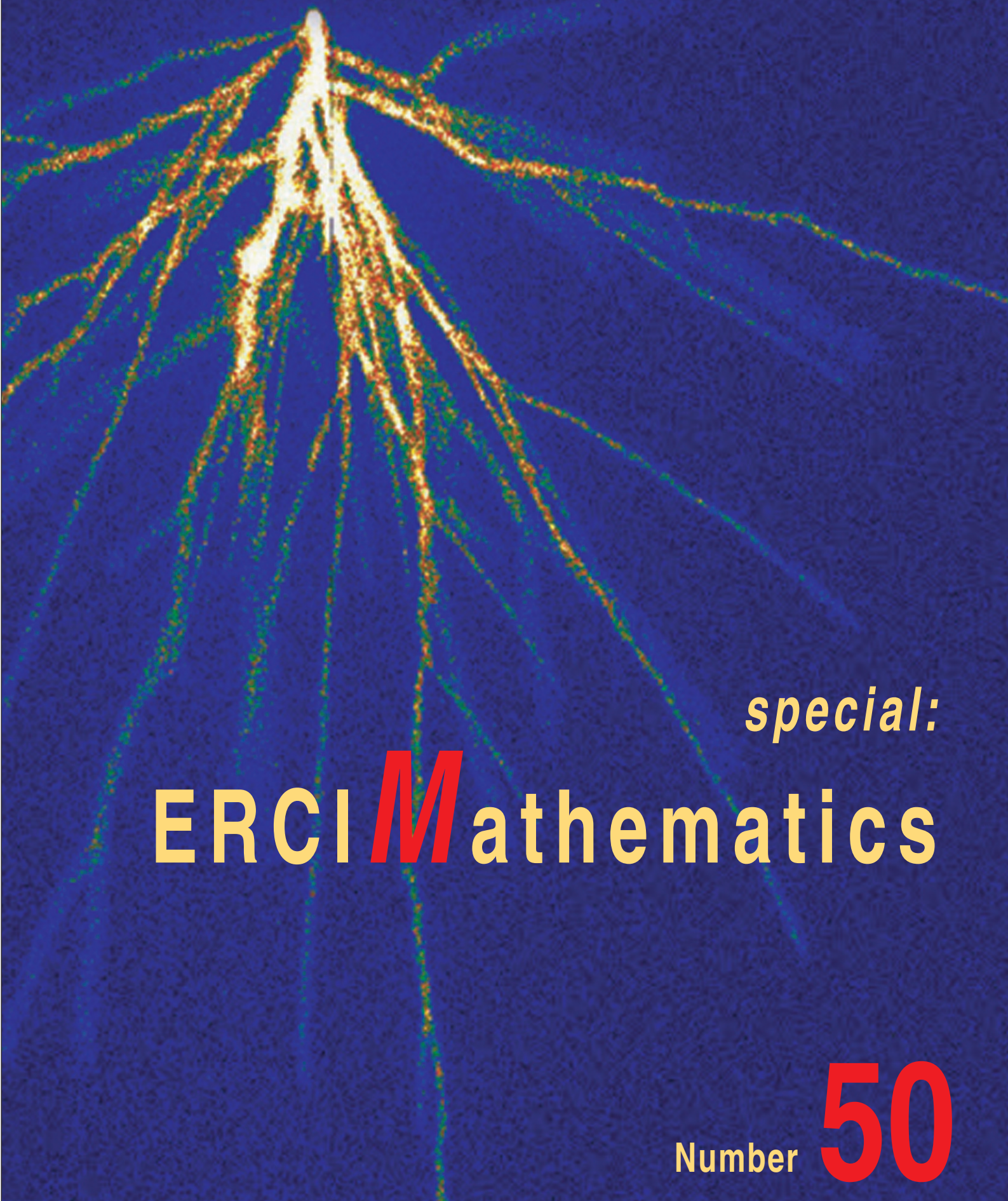
ERCIM NEWS

European Research Consortium for Informatics and Mathematics

www.ercim.org

July 2002



special:

ERCIMathematics

Number 50

CONTENTS

KEYNOTE

3 **by Rolf Jeltsch, President European Mathematical Society**

JOINT ERCIM ACTIONS

4 **50 Issues ERCIM News**
by Peter Kunz, ERCIM office

5 **Towards a Strategic Cooperation with the European Science Foundation**
by Jean-Eric Pin, ERCIM office

6 **Euro-Legal**
by Heather Weaver, CLRC

SPECIAL THEME

7 **Introduction to the Special Theme: ERCIMathematics**
by Tom Koornwinder, University of Amsterdam and CWI

9 **Nonlinear Modelling of Internet Packet Traffic**
by David K. Arrowsmith, University of London

10 **Mathematics on the Web with MathML**
by Max Froumentin, W3C

12 **Internet Protocol Table Lookup as a Geometric Problem**
by Marco Pellegrini, IIT-CNR

12 **On Large Random Graphs of the 'Internet Type'**
by Hannu Reittu and Ilkka Norros, VTT

14 **MUMPS: A Multifrontal Massively Parallel Solver**
by Patrick Amestoy, ENSEEIHT, Toulouse, France; Iain Duff, CLRC; Jacko Koster, Parallab, Bergen, Norway; and Jean-Yves L'Excellent, INRIA

16 **Simulation of Underground Water Pollution**
by Mario Arioli, CLRC

17 **Parallel Direct Solvers for Large Sparse Linear Systems**
by Iain Duff and Jennifer A. Scott, CLRC

19 **Spurious Solutions in Finite Element Simulation of Electromagnetic Cavity Resonators**
by Paolo Fernandes and Mirco Raffetto, IMATI-CNR

20 **Scalable Solvers for Sparse Linear Systems**
by Klaus Stüben, Fraunhofer SCAI

22 **A new Algorithm for Discrete Tomography**
by Rob Tijdeman, University of Leiden, The Netherlands and Herman te Riele, CWI

23 **Computational Topology Techniques for Shape Modelling**
by Silvia Biasotti, Bianca Falcidieno, Michela Mortara, Giuseppe Patané and Michela Spagnuolo, IMATI-CNR

24 **Marked Point Processes in Image Analysis**
by Xavier Descombes and Josiane Zerubia, INRIA

26 **Iterative Methods in Image Reconstruction**
by Paola Favati, Grazia Lotti and Ornella Menchi, IIT-CNR

27 **Shape Geometry and Aesthetics**
by Franca Giannini and, Marina Monti, IMATI-CNR

28 **Angry * (-1) = Surprised!**
by Aldo Paradiso , Fraunhofer IPSI

30 **Experimental Mathematics and Integer Relations**
by Jonathan M. Borwein, Simon Fraser University, Canada

31 **Algebra and Computation at SZTAKI**
by András Benczúr, Gábor Ivanyos and Lajos Rónyai, SZTAKI

33 **Renaissance of MacMahon's Partition Analysis: Symbolic Summation at RISC-Linz**
by Peter Paule, RISC-Linz, Austria

35 **ALIAS, a System Solving Library Based on Interval Analysis**
by Jean-Pierre Merlet, INRIA

36 **Theorema: A System for Computer-Supported Formal Mathematics**
by Bruno Buchberger, RISC-Linz, Austria

38 **A Colourful Theory on Graphs**
by Zsolt Tuza, SZTAKI

39 **Quantitative Domain Theory**
by Michel Schellekens, National University of Ireland, Cork

40 **Finding Dependencies in Industrial Process Data**
by Daniel Gillblad and Anders Holst, SICS

42 **Interaction of Stochastic Systems and Financial Mathematics**
by László Gerencsér, SZTAKI

43 **Algebraic Convergence of Pulled Fronts**
by Ute Ebert, CWI

R&D AND TECHNOLOGY TRANSFER

44 **Random Indexing of Linguistic Units for Vector-Based Semantic Analysis**
by Magnus Sahlgren, SICS

45 **Extending Thematic Lexical Resources by Term Categorization**
by Alberto Lavelli, Bernardo Magnini, and Fabrizio Sebastiani, IEI-CNR

46 **An End-to-End Software Suite for 3D Scanning**
by Marco Callieri, Paolo Cignoni and Paolo Pingi, IEI-CNR

48 **Exploring the Frontiers of Computability**
by Jiri Wiedermann, CRCIM and Jan van Leeuwen, Utrecht University, The Netherlands

49 **Towards the Design of Functional Materials and Drugs 'from Scratch'**
by Wanda Andreoni and Alessandro Curioni, IBM Zurich Research Lab, Switzerland

51 **Flower – Framework for Local Wireless Services**
by Tero Hakkarainen, Ali Lattunen and Vespe Savikko, VTT

53 **A Simple Model for Spark Branching**
by Ute Ebert, CWI

54 **Dyade, a Successful Cooperation between Bull and INRIA**
by Patrick Valduriez, INRIA

55 **Digital Library Competence Center**
by Maria Bruna Baldacci and Stefania Biagioni, IEI-CNR

EVENTS

56 **EU/NSF Digital Library All Projects Workshop**
by Costantino Thanos, IEI-CNR

57 **Fourth DELOS Workshop on Evaluation of Digital Libraries: Testbeds, Measurements and Metrics**
by László Kovács, SZTAKI

57 **Announcements**

55 **IN BRIEF**

Cover image: spark branching (time integrated picture, exposure time 10^{-6} seconds). Courtesy E. van Veldhuizen, Eindhoven University of Technology. Researchers at CWI have shown that branching of sparks is a direct consequence of the simplest mathematical model describing such electric discharges. See article on page 53.

Next issue:
October 2002

Special theme:
Semantic Web

Rolf Jeltsch,
President European Mathematical Society

jeltsch@math.ethz.ch
http://www.emis.de/

The dramatic increase in computing power and storage capacity since the 1940s have changed the way Applied Mathematics and Simulation can be done and have had an even greater impact on Mathematics itself. While in the early days the numerical aspect was the leading one, now more and more symbolic manipulations are done on computers. Mathematicians have started to prove theorems using computers — the Four Colour Problem is the most famous example! In the process that led to the celebrated classification of finite simple groups computers have played a crucial role at several places. Visualization of results is also beginning to make a big impact.

Constructing proofs often requires rigorous bounds; to utilise computers one had to replace the standard rounding of floating point arithmetic by a different arithmetic and rounding ['interval arithmetic'] that preserves bounds in a mathematically rigorous way. For example, Kepler conjectured in 1611 that the tightest method of packing spheres was the 'natural' pile for storing cannon balls; in 1998 T.C. Hales proved this and his proof relies on using an interval arithmetic package. Another example is the computer assisted proof by O. Lanford III in 1982 of the Feigenbaum Conjecture, a universal route for functions approaching chaos via period doubling.

Nowadays there are powerful computer algebra systems, like Magma, GAP, Macaulay2, Axiom, to name a few, which are specially designed for handling large and complicated objects in such abstract fields as group theory, geometry, and algebraic geometry.

We can also use computers to find and investigate mathematical conjectures. A visualization of the Costa minimal surface on the computer lead D.A. Hoffman to the proof that it had no self intersections, and this sparked a whole zoo of minimal surfaces. A longstanding challenge for mathematicians is the Riemann Hypothesis, that all 'interesting' zeros of the Riemann Zeta function lie on a straight line — this has been confirmed for the first 1,500,000,000 zeros using computers.

All the above mentioned approaches have been used in applications of mathematics. Formula manipulation software (eg, Maple and Matlab) can compute eigenvalues and eigenvectors of complicated matrices with symbolic entries.

Even recently students were trained to differentiate functions and learn which functions could be integrated or which differential equations solved explicitly. Today software does these things more reliably than humans.

In simulation, computing power has made a dramatic impact; nowadays we can handle much more complicated cases and more application areas than one could fifty years ago. Many of the new areas give rise to new theories in mathematics and new algorithms — most prominently genomics, finance and nanoscale technology. The European Mathematical Society (EMS) will organize a conference on application areas jointly with the French Société de Mathématiques Appliquées & Industrielles (SMAI) and the French Mathematical Society (SMF) in Nice, 10-13 February 2003.

Computers have also changed the way we store, retrieve and exchange our information. Years ago we lifted heavy volumes of the mathematical review journals while nowadays we access reviews instantly through the Internet. EMS participates in a EU research project to improve the databank 'Zentralblatt MATH'. In a world-wide effort, the mathematical community plans to digitise most mathematical literature so that it will be easily accessible and searchable electronically; EMS has sent an expression of interest for the 6th Framework Programme to cover the European part of this huge project.

The computer age and excellent modern communications systems have had an effect on teaching too. Slide rules were replaced by pocket calculators long ago, and soon every student will have a 'notebook' and Internet access. Software can be down-loaded; and the Internet provides a wealth of geometric models (which used to be done in clay or by cardboard or wires) and of video sequences demonstrating mathematical theorems (eg, how to turn a sphere inside out in three dimensional space). Software also helps high school students study interactively Euclidean geometry.

New curricula are being developed, the newest one being in Computational Science and Engineering. Use of the computer has boosted the applications of Mathematics in the last fifty years in a dramatic way. I feel it is unfortunate that this has also given rise to the fragmentation of the 'greater mathematical community' with many new organisations that do not always have strong contact with each other. Mathematics would be much stronger politically and intellectually if mathematicians would stand together, independent whether they 'care for applications' or prefer to 'stick to fundamental mathematics'.

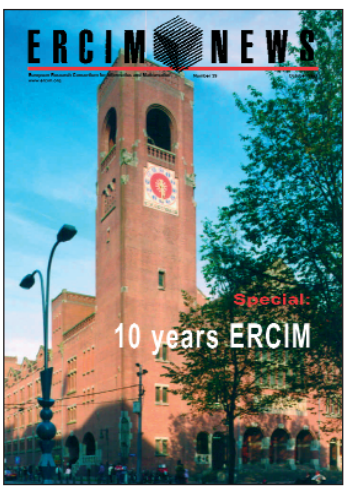
50 Issues ERCIM News

by Peter Kunz

With this number ERCIM is celebrating the 50th issue of its magazine, ERCIM News. Established in 1989, ERCIM News has witnessed significant advances in European R&D in Information Technology and Applied Mathematics in the thirteen years since it started.

The first issue of the 'INRIA, CWI, GMD newslettter' in 1989.

Number 5: the newsletter is called 'ERCIM News'.



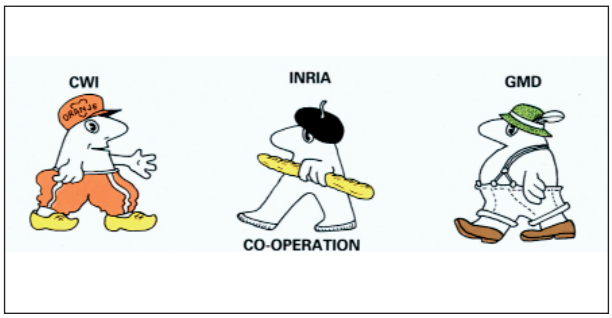
ERCIM News goes colour at the occasion of ERCIM's tenth anniversary.

When INRIA, CWI and GMD founded ERCIM in 1989, the establishment of an 'in-house magazine' with the aim of reporting on joint activities was one of the first 'joint actions'. The magazine has evolved from an 'in-house magazine' to a publication covering reports and news about scientific projects from all over Europe and even beyond, reflecting ERCIM's growth over the years.

From the early issues on, ERCIM News has been published regularly four times a year. With each issue focussing on a special theme, the ERCIM News series has become a unique collection providing an overview on different topics of Information Technology. Examples of successful topics were the two 'web' issues (no.25, April 1996 and no.41, April 2000) which coincided with the World-Wide Web conferences in Paris and Amsterdam, or the 'Grids' issue (no.45, April 2001). For each issue, ERCIM News invites a personality to write a keynote statement relevant to the European scientific community. Franco Malerba and Elly Plooij-van Gorsel, both Members of the European Parliament, the European Commissioners Erkki Liikanen and Philippe Busquin, UK Minister Lord Sainsbury and the Irish Prime Minister Bertie Ahern were among the keynote authors. Their keynotes have very much contributed to ERCIM's visibility.

The Web entered the research institutes in the early nineties and soon raised the question about an on-line version of ERCIM News. From October 1994 on, the newsletter has been published both in printed and electronic format. The electronic edition did not push back the printed edition (and the related costs for printing and distribution) but a lot of readers on the web subscribe to the printed edition, thus increasing its circulation. Beside the free subscription to the printed edition, the ERCIM News website gives access to all previous issues online back to number 19 (October 1994), offers full text search of these issues and has a form to order paper copies of back issues available on stock.

ERCIM News today has a circulation of over 9000 copies and is distributed in over 70 countries. Articles on the web have been accessed by over 60,000 different readers in 2001. Fifty issues of ERCIM News means about 1000 published articles, mainly written by scientists reporting on their research activities. If ERCIM News has achieved popularity to some extent, this is first of all the merit of the authors to whom the ERCIM Editorial Board wants to express their warmest thanks on this occasion.



"May this newsletter be one of our major channels through which we communicate with each other on our efforts, strategies, approaches and achievements, but also on our problems and questions". That was the way the directors of CWI, INRIA and GMD, Cor Baayen, Alain Bensoussan and Gerhard Seegmüller started the first issue of the 'Newsletter' in April 1989.

Please contact:
Peter Kunz, ERCIM office
E-mail: peter.kunz@ercim.org

Towards a Strategic Cooperation with the European Science Foundation

by Jean-Eric Pin

Discussions between representatives from the European Science Foundation (ESF) and ERCIM led to a proposal for cooperation between the two organisations. Both organisations signed a Memorandum of Understanding and organized an exploratory workshop to develop a joint vision for the future of e-Science. This workshop took place at CWI in Amsterdam on 30-31 May 2002.

The main thrust of the workshop was to investigate jointly a foresight for Information Technology (IT) in Europe. Twelve experts from ERCIM and twelve from ESF were invited to discuss important spearhead areas for the future, analyse strengths, weaknesses, opportunities and threats for European research and make recommendations to ERCIM, ESF and the European Union. Four discussion groups were constituted to take into account the interdisciplinary and the applied aspects of computer science.

Interdisciplinary between Mathematics and Computer Science

This group discussed the numerous interactions between the two disciplines. Logic plays a key role in semantics, in databases and in the design of modern programming languages. Probability and statistics occur in the modelling of networks, including the web, in data compression and in default analysis. Computing a realistic 3-D image or moving a robot requires a large amount of geometry. Modern cryptology relies on sophisticated methods provided by number theory. Discrete mathematics are at the heart of many algorithms. Numerical analysis, differential equations, optimisation and dynamical systems occur almost everywhere in scientific computing and their numerous applications. Perhaps more surprisingly, some concepts issued from computer science have deeply influenced modern mathematics. For instance, the famous P versus NP problem is one of the seven Millenium problems, and the automatic group concept is directly issued from the notion of automaton, the simplest model for a machine.

The discussions also included most of the topics presented in this issue of ERCIM News. Three key areas emerged from the discussion:

- Algorithms and Optimisation
- Computational Science (Modelling, Analysis, Simulation)
- Next generation computing.

The first two topics were selected for their strong interdisciplinary aspect and their wide field of applications. Europe is very strong in both areas, and should build on this competence. The last selected topic is much more prospective in nature and includes recent fields like Quantum Computing, Optical networks, DNA computing, and computation models beyond Turing machines.

The group recommended further action on Optimization, on Mathematics and next generation computing and on Games. Suitable mechanisms could be an ESF Scientific Programme, Scientific Forward Look and Exploratory Workshop respectively.

The group recommended to ESF to establish a scientific programme on Optimization, a forward looking on Mathematics and next generation computing, and an exploratory workshop on Games.

Emerging Applications in Sciences

This group explored topics covering medical informatics, bio-informatics, applications in astronomy, physics, etc. Three key topics emerged from the discussion:

- User easy access to facilities (data, software, computers, detectors / instruments with intelligent assists).
- Homogeneous view of heterogeneous data with analytical / visualisation /modelling tools
- ICT supported cooperative working among scientists in different disciplines and geographical areas.

Selected examples of scientific applications include post genomic R&D (with applications to health care), development of realistic models for the simulation of tumour growth and therapy (in order to replace experiments in vitro or with animals with in silico experiments), material characterisation and non-destructive evaluation (for improving eg, air safety and economy), intelligent transportation systems (with economic and environmental impact), environment: global warming (effects on agriculture, tourism).

ICT Services for User Communities

Information and Communication Technologies had already a huge impact on many domains of our daily life, where large databases can be used for the retrieval of information or the execution of appropriate actions. Obvious applications can be found in digital libraries. On a smaller scale, various forms of knowledge management are often used in order to control and preserve the knowledge acquired by a given group of people.

Official organisations are using their specific form of knowledge management for informing the public of all kind of useful information about the rules, laws and rights that govern the society. This domain of e-government has been extended with several electronic services and actions, including e-voting and web-supported decision making procedures (e-democracy). Internet technology for

all kinds of commercial transactions is leading to e-business and e-commerce. Two spearhead areas were selected:

- Complex interacting systems (digital libraries, knowledge management, e-business, e-government, etc.)
- Intelligent retrieval (Context awareness, multilingual queries, semantic description of data, multimedia queries and retrieval by content and context, ontologies, intelligent search engines, etc.).

The group suggested:

- to create a roadmap for intensive research in complex interacting systems and to create a research programme at the European level on high-quality automatic translation
- to organize through networking an inhanced collaboration between research groups working on intelligent retrieval of data from large and complex databases.

Advanced Communication Technologies and Scientific Computation

The demand for communication and computation solutions has been fuelled by the proliferation of wireless communication and information devices, the proliferation of the Internet, the proliferation and acceptance of e-tools as indispensable communication and computation tools, the competitive necessity of 'when-ever, where-ever' access to any community, in particular the scientific.

Of key importance in delivering solutions that can accommodate all the above demands are the computation and communication systems that are employed. Industry is leading the research for the near future, but the scientific research has a key role for investigating and discovering relevant new areas. Therefore, it is essential to promote the scientific research and to stress co-operation among scientists.
In line with this vision, the group identified the following key topics of ICT research:

- The quest for e-collaboration among scientists, whenever, wherever you are
- Meeting the challenge of distributing and integrating simulation, model and visualisation
- The need of co-operation and decision making in complex /critical situations.

Further Actions

A common recommendation by the four discussion groups was to assist ESF in IT by using ERCIM's pool of experts. It was also agreed that ESF and ERCIM jointly offer their expertise and advice to the European Commission. Scientists from ERCIM and ESF domains should also cooperate on projects in the sixth EC Framework Programme.

This meeting was a first step in the ESF – ERCIM cooperation in the fields of computer science and applied mathematics. A detailed report on the workshop will be presented to ESF's PESC Committee and Executive Board and to ERCIM's Board, as well as to Philippe Busquin, the European Commissioner for Research. The report will be accessible on ERCIM's website.

Please contact:
Jean-Eric Pin, ERCIM office
E-mail: office@ercim.org

Euro-Legal

News about legal information related to Information Technology from European directives, and pan-European legal requirements and regulations.

Dangers of Deep Linking

What is a deep link?
A deep link is a hot link to a subsidiary page of another Web site that is not its home page.

Why is this a problem?
Linking in this way may bypass registration procedures or advertising messages, depriving its publisher of value or infringe copyright.

Businesses are investing a lot of time and money into Web site development. A link to a business-critical deep page which bypasses revenue gathering pages or makes use of a page from another site which has a particular value, or was costly to produce, can have serious commercial consequences.

A deep link can create a commercially sensitive intrusion, for example by way of infringement of copyright, infringement of database rights, trade mark infringement or "passing off". Some of the best cases of deep linking come from Europe, for example Stepstone v OFIR. In this case the courts in Germany granted an injunction to prohibit a link to the claimant's web site by a rival recruitment agency which enabled it to display specific job vacancies from the claimant's site to its own site. The court granted the injunction, but the ruling was not solely based on the issue of deep linking, but rather that the process of creating the link infringed database rights which the claimant was entitled to protect. A recent case in the UK on similar lines was that of William Hill v British Racing Board. William Hill, the betting firm, had taken data derived from the Board's database to use on its own Web site. The judgment in this case stated specifically that database rights would protect virtually all collections of data in searchable form. That decision is of major importance in its relation to deep linking, as most links tend to be to pages containing useful data.

However, it is often assumed that there is an implied consent to link to a home page, but contradictory evidence exists across Europe. In Germany, case law supports the general principles of implied consent, but takes the view that deep linking is not permissible on the basis that it infringes some other right, for instance copyright. But in Denmark, music copyright owners have obtained compensation from individuals who created links from their own home pages to unlawful music files published on unrelated Web sites. In that case the infringement was regarded as an unlicensed public performance. Using a disclaimer, or linking policy statement, will not provide wholesale immunity. Whilst the Web does provide us with a free exchange of information and ideas, it is protected from being a veritable feast for digital pirates.

by Heather Weaver, CLRC
Tel: +44 1 235 446151
E-mail: H.Weaver@rl.ac.uk

Introduction to the Special Theme: ERCIMathematics

by Tom Koornwinder

Compared to special themes in earlier issues of ERCIM News the present theme, mathematics, is unusually wide and unusually old: the profession of mathematician is believed to be older than all but two other professions. So what ambition should we have to cover this theme in a few dozens of pages? Why not just refer to the marvellous 1200 pages book 'Mathematics Unlimited - 2001 and Beyond', B. Engquist & W. Schmidt (eds.), Springer? Well, the theme is not just mathematics, but ERCIMathematics. This may be a restriction which is not very drastic. In http://www.ercim.org/activity/expertise.html the scientific fields of competence of the ERCIM institutes are listed. Six out of sixteen areas listed in the rough subdivision of fields belong to mathematics:

- algebra, analysis and geometry
- combinatorics
- number theory
- control and system theory
- stochastics
- numerical mathematics and differential equations,

while four other areas (theoretical computer science, fluid dynamics, electromagnetism, and operations research) heavily use mathematics. In reality, as most readers will know, ERCIMathematics is not so comprehensive as suggested by the above list. There is a heavy bias to applied mathematics and to mathematics in interdisciplinary research, most prominently in connection with ICT.

Furthermore, there has been a dynamics in ERCIM institutes of putting less emphasis and focus on mathematics. Even members with a traditionally strong mathematics component, like INRIA or CWI, have experienced that the expansion in ICT was not matched by a proportional boost in mathematics research. Nevertheless there is still a considerable volume of mathematics research going on in ERCIM institutes. Part of it is not so explicitly visible because it is embedded in projects under a more general name. To be sure, several developments in ICT pose problems that require a mathematical approach, and for which no ready mathematical solution is at hand. In fact, these mathematical problems can be very challenging and may, when eventually solved, also open up new fruitful directions in pure mathematics.

In preparing the Call for this issue, the mathematical landscape in ERCIM has been taken into account. In particular the following possible topics for papers were mentioned:

- Computer algebra
- Computational geometry (eg, in connection with robotics)
- Automatic theorem proving (including software correctness proofs)
- Computer-assisted mathematics
- Interactive books (eg, Abramowitz & Stegun)
- Mathematics on the Web
- e-Math (electronic education, eg, online courses)
- Dynamical systems (non-linear equations)
- Image understanding
- Signal processing
- Complexity theory
- Optimisation
- Queuing theory
- Computational mathematics (numerical methods).

This is reflected in the following selection of articles, mainly coming from ERCIM institutes. The selection is neither exhaustive nor representative in the sense that the institutes with the strongest mathematics component also should have the highest weight. It just represents the spontaneous reaction on the call for contributions to this special issue. It turned out that four fields were represented with several articles: Internet/Web, Numerical Mathematics, Image Processing, and Symbolic Computation. Some other fields were represented with one article.

Some features in this harvest can be observed. First, not surprisingly, almost all papers have something to do with computers: math describing computers or computer networks, math developed by heavy use of computers, math aimed at delivering software. Second, many papers are about projects of larger size and time horizon. Third, many papers describe computer packages in an earlier or later state of completion.

Of course, with the topics mentioned and with the necessity that papers are short and do not become too technical, one cannot expect the traditional Definition-Lemma-Theorem style of mathematical papers. Also, one will not find in this volume proofs taken from the Book of Heaven. What one can pick up by browsing through and reading in this journal is a flavour of important recent ideas and concepts in applied and computer oriented mathematics, and also some trends and some surprising common concepts and keywords in different papers. Sometimes the math is mainly figuring as a useful and productive language, elsewhere deep math is used or developed for new applications. Jonathan Borwein, guest contributor from Canada, describes how experimental mathematics can lead to the discovery of mathematical gems.

The World Wide Web has had an enormous impact on society and on professional and daily life. It is also a fascinating source of mathematical questions how the Web develops as an organism and how the Web can be managed. Several papers here deal with these issues.

Stochastics has permeated the whole society and much of science. Also in this issue the importance of stochastic ideas, techniques and models becomes clear.

One editorial problem, which did not play a role so much with many earlier themes in ERCIM News, is the special way in which mathematics is expressed. The usual way of expression in ERCIM News is ordinary language combined with pictures. Mathematicians additionally want to (and have to) use formulas for making clear their ideas. In fact, this Special Theme Editor loves formulas, believes that they can have great beauty, and that they are much more efficient and precise than words for making clear what is meant. To paraphrase an old proverb, popular in image processing circles: one formula is worth a thousand words (the length of this Introduction, which I did not try to condense into one formula). Still the permanent rules of ERCIM News cannot be easily put away, and they were introduced because of the wide readership of ERCIM News, including researchers far off from mathematics, and managers. This trade-off between mathematical precision, as expressed in formalisation, and readability for a public not particularly trained in reading formulas, has led to the compromise for this Special Theme that formulas are presented as much as possible in separate boxes.

In preparing this special issue, invaluable help was provided by Henk Nieland from CWI, ERCIM Local Editor for The Netherlands.

Please contact:
Tom H. Koornwinder,
Korteweg-de Vries Institute for Mathematics,
University of Amsterdam
(temporarily also CWI), The Netherlands.
E-mail: thk@science.uva.nl
http://www.science.uva.nl/~thk/

CONTENTS

7 **Introduction to the Special Theme: ERCIMathematics**
by Tom Koornwinder, University of Amsterdam and CWI

9 **Nonlinear Modelling of Internet Packet Traffic**
by David K. Arrowsmith, University of London

10 **Mathematics on the Web with MathML**
by Max Froumentin, W3C

12 **Internet Protocol Table Lookup as a Geometric Problem**
by Marco Pellegrini, IIT-CNR

12 **On Large Random Graphs of the 'Internet Type'**
by Hannu Reittu and Ilkka Norros, VTT

14 **MUMPS: A Multifrontal Massively Parallel Solver**
by Patrick Amestoy, ENSEEIHT, Toulouse, France; Iain Duff, CLRC; Jacko Koster, Parallab, Bergen, Norway; and Jean-Yves L'Excellent, INRIA

16 **Simulation of Underground Water Pollution**
by Mario Arioli, CLRC

17 **Parallel Direct Solvers for Large Sparse Linear Systems**
by Iain Duff and Jennifer A. Scott, CLRC

19 **Spurious Solutions in Finite Element Simulation of Electromagnetic Cavity Resonators**
by Paolo Fernandes and Mirco Raffetto, IMATI-CNR

20 **Scalable Solvers for Sparse Linear Systems**
by Klaus Stüben, Fraunhofer SCAI

22 **A new Algorithm for Discrete Tomography**
by Rob Tijdeman, University of Leiden, The Netherlands and Herman te Riele, CWI

23 **Computational Topology Techniques for Shape Modelling**
by Silvia Biasotti, Bianca Falcidieno, Michela Mortara, Giuseppe Patané and Michela Spagnuolo, IMATI-CNR

24 **Marked Point Processes in Image Analysis**
by Xavier Descombes and Josiane Zerubia, INRIA

26 **Iterative Methods in Image Reconstruction**
by Paola Favati, Grazia Lotti and Ornella Menchi, IIT-CNR

27 **Shape Geometry and Aesthetics**
by Franca Giannini and, Marina Monti, IMATI-CNR

28 **Angry * (-1) = Surprised!**
by Aldo Paradiso , Fraunhofer IPSI

30 **Experimental Mathematics and Integer Relations**
by Jonathan M. Borwein, Simon Fraser University, Canada

31 **Algebra and Computation at SZTAKI**
by András Benczúr, Gábor Ivanyos and Lajos Rónyai, SZTAKI

33 **Renaissance of MacMahon's Partition Analysis: Symbolic Summation at RISC-Linz**
by Peter Paule, RISC-Linz, Austria

35 **ALIAS, a System Solving Library Based on Interval Analysis**
by Jean-Pierre Merlet, INRIA

36 **Theorema: A System for Computer-Supported Formal Mathematics**
by Bruno Buchberger, RISC-Linz, Austria

38 **A Colourful Theory on Graphs**
by Zsolt Tuza, SZTAKI

39 **Quantitative Domain Theory**
by Michel Schellekens, National University of Ireland, Cork

40 **Finding Dependencies in Industrial Process Data**
by Daniel Gillblad and Anders Holst, SICS

42 **Interaction of Stochastic Systems and Financial Mathematics**
by László Gerencsér, SZTAKI

43 **Algebraic Convergence of Pulled Fronts**
by Ute Ebert, CWI

Nonlinear Modelling of Internet Packet Traffic

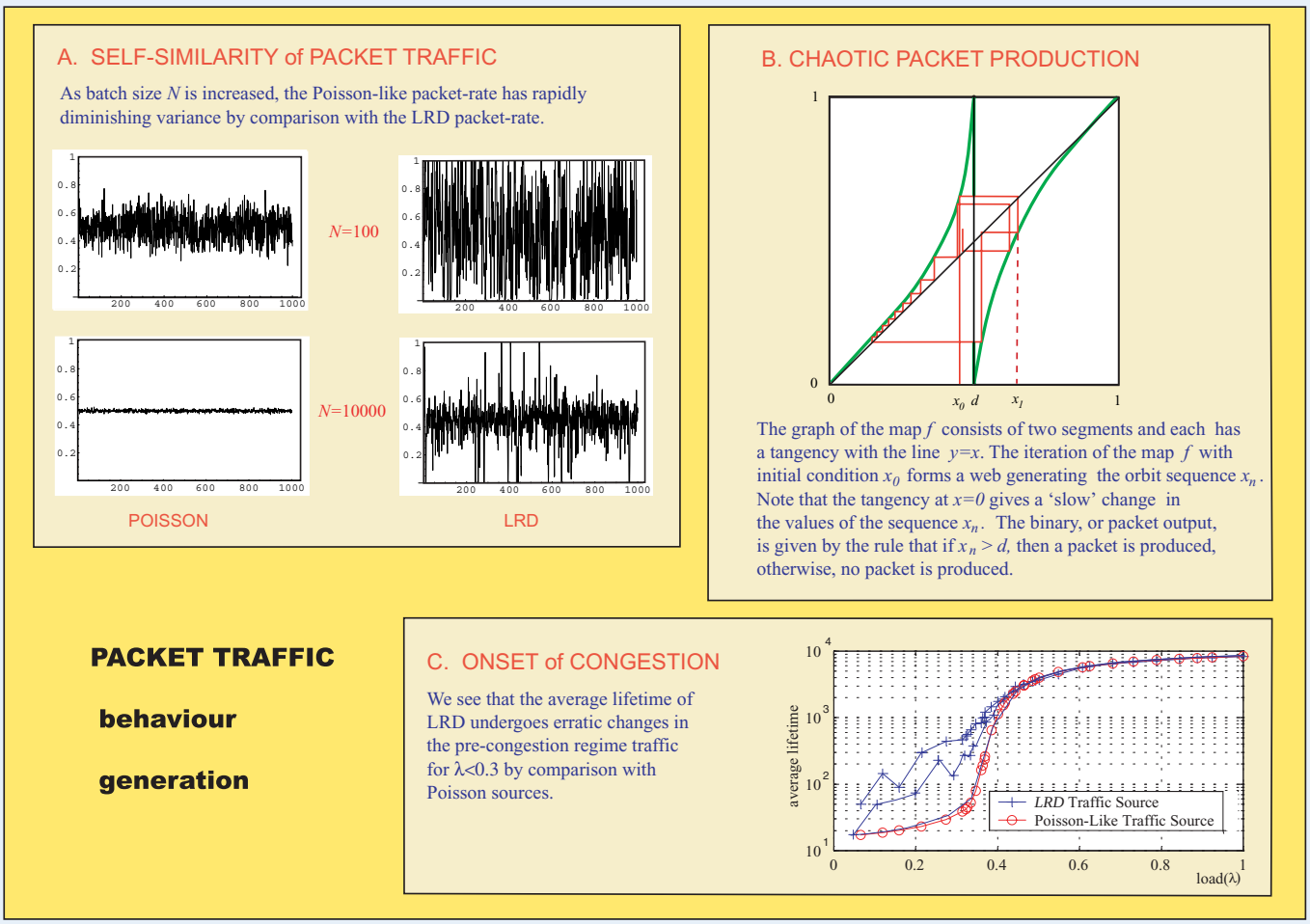
by David K. Arrowsmith

Nonlinear mathematics has recently been useful in creating packet traffic models for the Internet that can simulate the volatility of packet rates in a network on a range of time scales.

The challenge of understanding the acute problems of congestion of packet traffic in Internet networks has been helped by the use of special types of chaotic map in the modelling process. Internet traffic was shown to exhibit self-similarity or long range dependence (LRD) by Leland and others in 1993. This behaviour manifests itself along a communication channel with bursty activity in the packet rate which persists on all relevant time scales. The phenomenon makes it difficult to implement effective traffic congestion protocols (TCP).

The bursty nature of Internet traffic contrasts strongly with more traditional voice traffic, which is Poisson-like, and is known as short-range-dependent (SRD) traffic. In Figure A, we see examples which mimic the scaling of (a) short-range and (b) long-range dependent traffic for a time series of a random variable X_n, $n = 0,1,2,...$ which takes binary values, ie X_n in 0,1, when averaged in batches sizes of $N = 100$ and $N = 10000$ in each case. The standard deviation in the SRD traffic varies as the square root of the batch size, or magnification, and we see a 'smoothing' of the traffic. This means that on aggregation, the mean is an increasingly effective indicator of the instantaneous load, ie expected packet rate, in the traffic. By comparison, in the LRD traffic, we see that the variation remains relatively high and even when averaged over longer time intervals, we still see the extreme rates 0 and 1 arising. The difference between these two types of traffic shows up in the nature of the auto-correlation decay. For memoryless traffic, the decay is exponential, whereas for LRD traffic, the decay is given by a power law.

The connection with nonlinear dynamics comes in the form of a packet traffic simulator. Chaotic maps are used to provide a binary output. Specifically, maps f of the unit interval I of real numbers between 0 and 1, illustrated in Figure B, are used to provide the appropriate packet production mechanism at a host. The iteration of the map on an initial value x_0 produces a sequence of real numbers, or an orbit $\{x_n\}$ in I, for $n = 0,1,2,....$ The binary output required



to denote the production of a packet is given by an output map y which acts on the orbit $\{x_n\}$ with the rules $y_n = 1$ if $x_n \geq d$ and $y_n = 0$ otherwise, for some fixed $d \in (0,1)$. The graph of f can be easily modified to give the 'memory' effects needed for long-range dependence. Tangencies are introduced to the graph of f at both the end points of I. This ensures that any orbit which passes close to $x = 0$ or 1 spends a long time 'escaping' from $x = 0$. Such a modification can result in long strings of ones and zeros in the binary output and thus the averaging over increasingly large batch sizes can still give rise to the extreme values '0' and '1' noted earlier.

The functional form used for these packet generators was developed by Erramilli and others in the mid-nineties. The degree of the tangencies in the graph of f can be continuously changed to give varying types of traffic from Poisson (or memory-less traffic) to all levels of autocorrelation for LRD traffic.

Simple network models can be created around this efficient packet production technique. The basic model has a lattice network of interconnected nodes which are either hosts (which can both receive or transmit packets) and routers which can only transmit packets. Packets when produced have another host destination.

However, each of the network nodes has a buffer to store packets which cannot be moved immediately. When a packet arrives at the head of the queue, it then transfers to the queue at an adjacent node which is closer to its destination. One can consider in such models various performance factors such as average delivery time of packets, throughput of packets (ie packets received as their intended destination as function of packets sent).

The most common control on the network is the adjustment of the packet rate of the hosts which feed the network with packets. Performance indicators, such as packet liftetime, can change so dramatically as load is increased that they can sometimes be best described by a phase transition. This behaviour is not necessarily the bad news that it sounds. If a critical load exists at which the delivery time of packets increases dramatically, then a clear control strategy to keep the load below criticality can be implemented. However, when the transition to congestion is less clear-cut, as for LRD traffic, then the control strategy becomes much more problematical.

Our current investigations show up the latter problem very clearly. Direct comparisons of Poisson and LRD traffic in the same network and with the same load profiles show in Figure C similar network characteristics away from the congestion transition but wildly different behaviour around the onset of congestion region.

Immediate projects are to check the robustness of our results on other regular networks and extend the investigations to more realistic Internet networks. A joint EPSRC funded project began in April 2002, involving Queen Mary and the University of York, and it will investigate the different sources of LRD in a packet traffic network and assess their relative importance.

Please contact:
David K. Arrowsmith
Mathematics Research Centre,
School of Mathematical Sciences,
Queen Mary, University of London, UK
Tel: +44 1 20 7882 5464
E-mail: d.k.arrowsmith@qmul.ac.uk
URL: http://maths.qmul.ac.uk/~arrow/

Mathematics on the Web with MathML

by Max Froumentin

MathML 2.0 has become a W3C Recommendation in February 2001 and is now widely supported, making it possible at last to distribute scientific material on the Web.

Although mathematical notation has made its way into computer-generated documents with quality publishing software now correctly handling mathematics, putting mathematics on the Web has been a problem since the beginning of HTML since support for displaying mathematics was minimal. Shortly after the birth of XML, the W3C started designing MathML: an XML language for expressing mathematics, in order to make it possible to display formulas in Web browsers as well as provide a interchange format for mathematical software.

Mathematics and the Web

Mathematical notation inherits from centuries of refinement, resulting in strict rules regarding the layout of mathematical equations and formulas. These rules make it a challenge to design satisfactory mathematics typesetting software. From the beginning of the Internet the common practice for scientists was to exchange mathematics in some encoded form based on the ASCII character set. Later graphical displays became popular while computer-aided publishing software grew more and more efficient and allowed better-looking presentations of mathematics. This culminated with the advent of TeX, currently the de-facto standard for exchanging scientific documents.

As the use of the World Wide Web for distributing scientific documents increased, it became important to include mathematical representation in Web pages as one of the main uses of the Web is science education. However, the HyperText Markup Language (HTML) was not designed to describe a complete set of tags for mathematical notation, as it only defined subscripts and super-

scripts. Content authors then went back to rendering mathematics using ASCII characters or pictures from TeX renderings inserted in HTML pages. This solution is not satisfactory though, as it does not follow the principles of Web usability and accessibility: the resulting renderings of mathematics only fit one media type (screen), cannot be processed by software such as search engine indexing tools, and are not customisable by the user who might prefer the formula displayed in bigger fonts, or using different colours.

MathML

Shortly after the Extensible Markup Language (XML) was created, a W3C Working Group was chartered to create an XML language to describe mathematical notation: MathML. MathML version 1.0 was published in 1998, while version 2.0 was released in February 2001.

MathML defines markup to describe mathematical equations and formulas. The specification defines two sets of tags and attributes: Presentation MathML describes an equation almost as one would read it, specifying elements such as subscripts, fractions or operators. This dialect is somewhat similar to TeX, but it adds a few additional elements to mark identifiers, numbers and operators. For instance the TeX equation: '*$1+\sqrt{b}$*' corresponds to the following Presentation MathML markup:

```
<math>
 <mrow>
  <mn>1</mn>
  <mo>+</mo>
  <msqrt><mi>b</mi></msqrt>
 </mrow>
</math>
```

where mn, mo and mi stand for number, operator and identifier. The other type of MathML markup, called Content MathML, is meant to carry more information on the equation, in particular for exposing the semantics of functions. Our example could be written as:

```
<math>
 <apply>
  <plus/>
  <cn>1<cn>
  <root>
   <degree><cn>2</cn></degree>
   <ci>b</ci>
  </root>
 </apply>
</math>
```

Content MathML adds more mathematical meaning in its description of formulas, consequently allowing its use as an interchange format between mathematical software. However the range of mathematics covered by this markup is necessarily limited, and it was chosen to include the basic set of most standard areas of mathematics, such as arithmetic, algebra, logic, set theory, calculus, sequences and series, linear algebra, and statistics. Extension mechanisms are defined to complete this list with additional mathematical constructs.

Of course most complex equations from any area of mathematics can still be described (although in a less meaningful way) using Presentation MathML.

Because MathML is based on XML and other W3C specifications such as Cascading Style Sheets, it is fully integrated to standard Web technologies and solves the problems encountered hitherto: the rendering of MathML can be adapted to either the device used, such as a desktop computer, a calculator or a Braille device, or to the user's preferences and abilities (font size, colour) through the use of Cascading Style Sheets (CSS). Moreover, a piece of MathML markup can be annotated using additional markup in order to add more information and facilitate actions such as Web searches for a particular piece of mathematics. Finally MathML supports standard hyperlinking mechanisms, as well as interactive mathematics.

Once the definition of the language itself was finished it was up to implementors to integrate MathML in their products. Most of that work was done by the members of the Working Group, adding MathML to browsers — either natively (Amaya) or through 'plug-ins' (techexplorer, Mathplayer) — or extensions of existing mathematical software (Mathematica, Maple, etc.). As of May 2002, MathML is included in the latest versions of both Netscape Navigator and Microsoft Internet Explorer: after native MathML rendering was implemented in Mozilla, Netscape included the code in the first preview release of Navigator 7.0 (available on Windows and Unix). Similarly, all recent versions of Internet Explorer (Windows, Mac) can render MathML through plug-ins such as IBM's techexplorer or Design Science's MathPlayer, both freely available.

Although MathML is already considered a successful standard, more work needs to be done and the Working Group is not at rest. Conversion of other mathematical formats to MathML is an important task to enable publishing legacy documents on the Web. As most such documents are in TeX format, conversion tools have to be developed in order to convert it to MathML, with the extra complexity that semantic information lacking in the TeX format must be inferred. Another area of ongoing work is integrating MathML with other Web standards such as XML Schema for adding data type information to documents, Unicode for character sets, or OpenMath for higher-level mathematics markup.

Links:
The W3C Math Home Page:
http://www.w3.org/Math
The MathML 2.0 Recommendation:
http://www.w3.org/TR/MathML2

Please contact:
Max Froumentin, W3C
E-mail: mf@w3.org

Internet Protocol Table Look-up as a Geometric Problem

by Marco Pellegrini

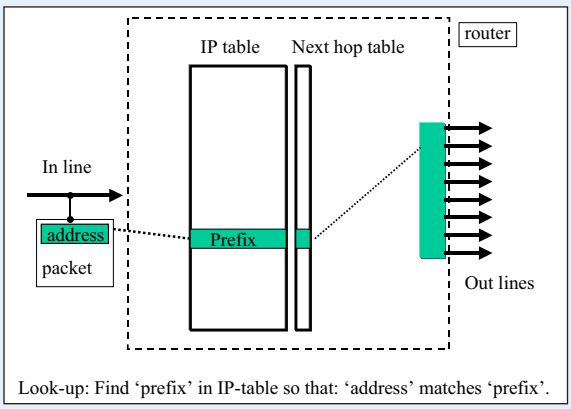
Ongoing research at the CNR Institute for Informatics and Telematics in Pisa deals with the IP table look-up problem, a critical bottleneck for high-bandwidth Internet routers. The novelty lies in mapping an essentially discrete (digital) matching problem into a continuous search over points on the real line; this permits us to invoke powerful tools from continuous mathematics.

New sectors of the economy and the public administration, ranging from e-commerce and e-entertainment to e-government, offer (or promise) consumers and citizens of the developed world a growing range of high-quality low-cost services. The hidden danger is that the underlying physical infrastructure (Internet) might not be able to cope with the increased demand for its basic delivery utilities upon which these user-oriented services are built. Already, the ever increasing number of computers and devices connected to Internet is pushing the inadequate IPv4 32-bit length standard for addresses towards the new IPv6b 128-bit length standard.

At a high level, Internet is a large network of interconnected computers and devices, where some elements, called routers, are specialized in managing the traffic of fixed-size chunks of data called packets. The raw power of an internet router is measured in packets-per-second that the router is able to receive and redirect towards neighbouring routers. The IP table look-up operation is at the core of this process and consists of finding the entry of the routing table, stored in the router, that matches the address of an incoming packet and deciding on the appropriate exit route.

Since the early 90's, the IP table look-up problem has been recognized as a bottleneck and a variety of hardware and software solutions have been proposed. A general criticism of the state-of-the-art is that most if not all schemes proposed achieve efficiency through exploitation of the particular distribution of the entries of existing routing tables, based on the present configuration of Internet. The evolution of Internet and the much longer address format of IPv6 will radically change such assumptions and will force a reconsideration of all existing table look-up algorithms. At the present, it is a matter of conjecture how the interconnetivity of Internet will evolve on the medium time scale.

In our approach to the IP table look-up problem we map the distribution of entries in a routing table to a distribution of segments on the real line. Thus the problem of matching a table entry and an incoming address maps to the purely geometric problem of finding the shortest segment that contains a point. When we see the IP table look-up problem embedded in the continuum of the real line we can call upon sophisticated probabilistic techniques, in particular urn models, that make it possible to analyze various properties of the distribution of balls into bins. Statistical tests can be used to learn approximately the input distribution and to decide on an optimal decomposition of the line into buckets to speed up the on-line search. The objective is to endow the table look-up mechanism with the capability to adapt and fine tune with long term shifts in the distribution of the entries of routing tables.



IP address look-up as a matching problem.

It has been known for a long time that (non-uniform) grids of buckets on a line can be used to test the hypothesis that a given set of points follows a certain distribution. The new twist is to add a second phase in which we also determine the uniform bin structure that is most suitable for the construction of a small depth and small memory search trees.

In addition to long term shifts, the IP table must also respond to frequent local updates that are needed to keep the overall network load balanced. The need for frequent, efficient updates of the routing tables discourages the use of powerful, but essentially off-line optimization strategies, such as dynamic programming or other heavy optimization tools. Instead, we rely on simpler local rules well grounded in mathematical theory to decide the structure of the search tree to be used in the look-up.

The three main quantities to balance are: speed of the search (worst case and average), memory consumption, and speed of updating the routing table. The optimal solution will thus consist of a profile, or workload curve that the router operator can exploit when choosing a working profile. In a typical scenario, depending on the local time, certain parts of the world wide web are more active than others. Thus, it might be convenient over a 24-hour period to make certain parts of the routing table faster by using more storage locally, and other parts slower by using less storage, in order to cope with peak demands while keeping the total storage consumption unchanged.

Please contact:
Marco Pellegrini, IIT-CNR
Tel: +39 050 315 2410
E-mail: marco.pellegrini@iit.cnr.it
http://www.imc.pi.cnr.it/~pellegrini

On Large Random Graphs of the 'Internet Type'

by Hannu Reittu and Ilkka Norros

Internet and its infrastructures are growing fast, both in size and in importance. We model the large scale structure of the Internet by a simple model, with a random graph with power-law degree distribution. The nodes represent routers or autonomous systems. We found that in such graphs a spontaneous hierarchy appears with respect to shortest path routing. The length of the paths between typical nodes grows very slowly, as the iterated logarithm (log log) of the number of nodes.

Random networks with power-law distribution of degrees of the nodes have been studied quite extensively in the past few years, at least on a heuristic level (the degree of a node is the number of links connected to that node). By power-law we mean that the probability of having some degree is proportional to that degree raised to some negative constant power, whose absolute value we call the index of the power-law. Such models seem to have some interesting practical applications. In particular, the main characteristics of the Internet graphs follow power-laws on both router and domain levels. These observations may turn out to be significant for topology and routing dependent features of the Internet.

We adopt a random graph model, introduced by Newman, Strogatz and Watts (for all references: see our link at the end). Their original work is, however, restricted to power-laws with indices greater than 3, whereas for the Internet graphs this index lies between 2 and 3, according to the famous empirical study of the Faloutsos brothers by the end of 1990s. Power tailed distributions with index between 2 and 3 have a finite mean but an infinite variance, which entails that nodes with very high degrees appear with a non-negligible frequency. In the regime of Newman et al. it has been shown that the expectation of the distance between two nodes (measured by the number of hops, the usual distance measure on a graph) scales logarithmically with respect to the number of nodes. Such graphs are sometimes called 'small worlds'. These graphs with index larger than 3 are homogeneous in the sense that in average and asymptotically all nodes have, so to say, the same kind of environment around. This is, however, not the case one would expect for the Internet graph, which has an obvious hierarchy with some powerful nodes having a key role in the functioning of the network. This appears in their large degrees in the graph. On the other hand, a 'typical' node does not possess these properties and has a small degree. The method of generating functions used by Newman et al. in their analysis does not work for power-law indices less than 3. In this case they suggested an exponential cutoff in degree distribution and found the same logarithmic scaling for the distance. The exponential cutoff removes the very large nodes and thus changes the character of the graph. We investigated the same model without cutoff and proved that the average distance is at most of the order of the iterated logarithm of the number of nodes.

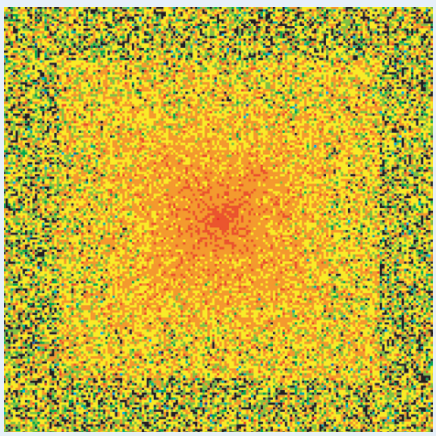


Figure 1:
31624 nodes in quadratic spiral, ordered by descending degree (largest node in the centre, outmost nodes degree 1). The colour code corresponds to closeness to the largest node, and is the same as in Figure 2.

The model is defined as follows. First we fix the number of nodes. The degrees of the nodes are then genereated as independent and identically distributed random variables following a power-law distribution. The second stage is the random formation of links between nodes. We can think this visually so that 'link stubs' sticking out from each node are joined pairwise in a random fashion (if the sum of the degrees is odd, we add one extra node with degree one). As a result, artefacts like multiple links between nodes are possible, but their role is not significant. The third stage is the investigation of properties of such graphs when the number of nodes goes to infinity, a thermodynamical limit. Several phase transitions can be identified when the power law index passes certain critical values. For instance, at value 3 the diameter of the graph changes from logarithmic to iterated logarithm between values 2 and

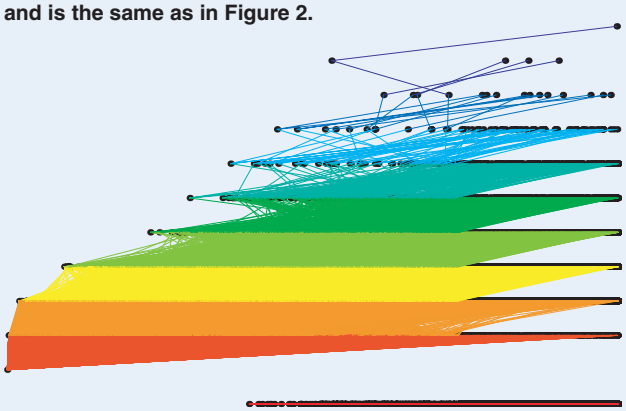


Figure 2:
Distance from the largest node. The number of nodes is 31624. The horizontal position of a node indicates its degree while the vertical position shows its distance from the largest one. The links of the nodes are coloured according to their distance from the largest node. The nodes not connected to the largest node are placed on a separate line.

3, and finally below 2 the diameter is essentially bounded. From the point of view of this physical analogy, the Internet lives in one particular phase.

Our task was to find the distance between two randomly chosen nodes in such a random graph with power-law index between 2 and 3. In the limit of large number of nodes, the largest degree in the graph is rather accurately defined. It is proportional to the number of nodes raised to a power that lies between 1/2 and 1. The existence of so huge nodes was indeed apparent from the first study of Internet topology by Faloutsos et al. Such a large node has of course many neighbours. Among them there are almost surely all other large nodes exceeding some minimum degree value. The union of the first neighbours can now be considered as a new, still larger 'supernode' having necessarily again a certain number of neighbours, and so on.Within a number of steps that is only an iterated logarithm of the number of nodes, this set is mighty enough to play the role of a 'core network'. Although the relative size of the core goes to zero as the network grows, the core can be reached from a randomly selected node with less than the same iterated logarithm number of steps. As a result any two such nodes can be connected with no more than of the order of this same number of hops. Formal proofs were done to verify this picture. In the first part, the described reasoning becomes waterproof by elementary arguments when the critical sizes of the 'layers' within the core are properly chosen. In the second part, we used a matching with a branching process.

In summary, our work reveals that in a random network where all nodes are in equal position there still exists hierarchy, if the node degrees have infinite variance. According to latest studies, this kind of hierarchy corresponds better to that found in measurements than some traditional deterministic hierarchical network models. The diameter of the graph, in number of hops between nodes, remains almost constant while the graph grows over many orders of magnitude. Both these factors are significant, at least if such a graph can model a communication network, but probably in several other contexts as well.

Link:
http://www.vtt.fi/tte/projects/cost279/

Please contact:
Ilkka Norros, VTT
Tel: +358 9 4565627
E-mail: Ilkka.Norros@vtt.fi
Hannu Reittu, VTT
Tel: +358 9 4565605
E-mail: Hannu.Reittu@vtt.fi

MUMPS: A Multifrontal Massively Parallel Solver

by Patrick Amestoy, Iain Duff, Jacko Koster, and Jean-Yves L'Excellent

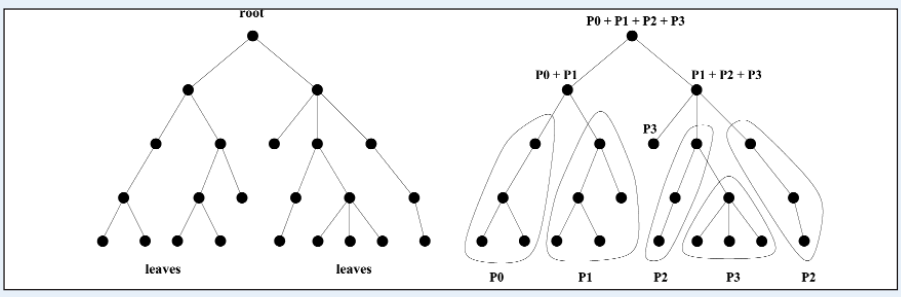
The solution of large sparse linear systems lies at the heart of most calculations in computational science and engineering and is of increasing importance in computations in the financial and business sectors. Today, systems of equations with more than one million unknowns need to be solved. To solve such large systems in a reasonable time requires the use of powerful parallel computers. To date, only limited software for such systems has been generally available. The MUMPS software addresses this issue.

MUMPS was originally developed in the context of the PARASOL Project, which was an ESPRIT IV Long Term Research Project for developing 'An Integrated Environment for Parallel Sparse Matrix Solvers'. The Project started in January 1996 and finished in June 1999 and a version of MUMPS was publicly released in March 2000. An important aspect of the PARASOL project was the strong link between the developers of the sparse solvers and the industrial end users, who provided a range of test problems and evaluated the solvers.

Since then MUMPS has continued as an ongoing research project and a new release of the package is imminent. It is the purpose of this article to describe recent work that has led to the new facilities available in this release.

The MUMPS package uses a multifrontal approach to factorize the matrix (Duff and Reid, The multifrontal solution of indefinite sparse symmetric linear systems, ACM TOMS, 9, 302-325). The principal feature of a multifrontal method is that the overall factorization is described (or driven) by an assembly tree (see Figure 1, left).

At each node in the tree, one or more variables are eliminated using steps of



Example assembly tree (left) and a possible distribution of the computation over four processors P0, P1, P2, and P3 (right).

Gaussian elimination on a dense matrix, the frontal matrix.Each edge in the tree represents the movement of data of a child node to its parent (which is the adjacent node in the direction of the root). An important aspect of the assembly tree is that it only defines a partial order for the factorization. That is, arithmetic operations at a pair of nodes, where neither lies on a path from the other to a root node, are independent. For example, work can commence in parallel on all the leaf nodes of the tree. Operations at the other nodes in the tree can proceed as soon as the data is available from the children of the node. There is thus good scope for exploiting parallelism, especially since assembly trees for practical problems contain many thousands of nodes.

The efficiency of a direct solver is very dependent on the order in which the variables of the problem are eliminated. This largely determines the amount of work that is to be done during the factorization and hence the overall time for solution. Furthermore, the elimination order of the variables also determines the number of entries in the computed factors and the amount of storage required. In the earlier versions of MUMPS, the only in-house ordering supplied was the Approximate Minimum Degree (AMD) algorithm of Amestoy, Davis, and Duff (SIAM J Matrix Anal and Applics, 17, 886-905), but recent experiments have shown that while this ordering is usually best on discretizations from 2D problems, other orderings, particularly spectral based orderings and minimum fill-in orderings, can do better on larger problems from 3D discretizations. Thus, in the new version of MUMPS, we have included a wide range of ordering packages including PORD (Jürgen Schulze), SCOTCH (LaBRI, Bordeaux), and approximate minimum fill-in orderings.

In a somewhat more loosely coupled way, orderings from the MeTiS package from Minneapolis can be used.

For nodes far from the root, to keep communication to a minimum while maintaining a high level of parallelism, MUMPS maps a complete subtree onto a single processor of the target machine (see Figure 1, right). The nodes near the root of the assembly tree normally involve more computation than nodes further away from the root. In practical examples, we observed that more than 75% of the computations are performed in the top three levels of the tree. Unfortunately, the number of independent nodes near the root is small, and so there is less parallelism to exploit. For example, the root in Figure 1 has only two neighbouring nodes and hence only two processors would perform the corresponding work while the other processors remain idle.

It is thus necessary to obtain further parallelism within the nodes near the root by partitioning these nodes among several processors. In addition, because MUMPS is designed to solve a wide range of problems including symmetric and unsymmetric problems, numerical pivoting is performed within the numerical factorization. This means that only a static analysis of the sparsity pattern and a static allocation of tasks to processors could be very inefficient if significant numerical pivoting is required. We have chosen to deal with this through using dynamic allocation of tasks during the numerical factorization. This also has the benefit of enabling the code to perform well when we are not in a single-user environment. We have done much recent research on the best way of performing this allocation. In the earlier versions of MUMPS, the processor to which a computationally expensive node is assigned, partitions the corresponding frontal matrix, and distributes the parts dynamically over processors that have a relatively low load. The elimination operations on this frontal matrix are subsequently performed in parallel. In the new version, the number of candidates which can be chosen is restricted during the prior static analysis phase resulting in a dynamic allocation that is much more more efficient in both memory estimation and usage and execution time.

The original MUMPS package was only designed for real matrices but, in the new version, complex symmetric and complex unsymmetric systems are permitted. If there is sufficient demand, a version for complex Hermitian systems might be developed in the future. The MUMPS software is written in Fortran 90. It requires MPI for message passing and makes use of BLAS, LAPACK, BLACS, and ScaLAPACK subroutines. However, in recognition that some users prefer the C programming environment, a C interface has been developed for the new release, and a version has been written that avoids the use of MPI, BLACS, and ScaLAPACK. This would be suitable for running in a single processor environment, perhaps for testing and development purposes.

MUMPS has been ported to a wide range of computers including the top-line supercomputers from Compaq, Cray, IBM, and SGI and we are currently working on the efficient exploitation of multilevel parallelism and memory management, such as is present on the new IBM SP range of computers. The MUMPS package has a good performance relative to other parallel sparse solvers. For example, we see in Table 1 comparisons with the SuperLU code of Demmel and Li. These results are taken from 'Analysis and comparison of two general sparse solvers for distributed memory computers', ACM TOMS, 27, 388-421.

Matrix	Ordering	Solver	Number of processors						
			1	4	8	16	32	64	128
bbmat	AMD	MUMPS		44.8	23.6	15.7	12.6	10.1	9.5
		SuperLU		64.7	36.6	21.3	12.8	9.2	7.2
	ND	MUMPS	—	32.1	10.8	12.3	10.4	9.1	7.8
		SuperLU	—	132.9	72.5	39.8	23.5	15.6	11.1
ecl32	AMD	MUMPS	—	53.1	31.3	20.7	14.7	13.5	12.9
		SuperLU	—	106.8	56.7	31.2	18.3	12.3	8.2
	ND	MUMPS	—	23.9	13.4	9.7	6.6	5.6	5.4
		SuperLU	—	48.5	26.6	15.7	9.6	7.6	5.6

Table: Factorization time (in seconds) of large test matrices on the CRAY T3E. '—' indicates not enough memory.

Link:
http://www.enseeiht.fr/apo/MUMPS.

Please contact:
Iain S Duff, CLRC
Tel: +44 1235 445803
E-mail: i.s.duff@rl.ac.uk

Simulation of Underground Water Pollution

by Mario Arioli

The Numerical Analysis Group at Rutherford Appleton Laboratory develops new algorithms and mathematical software that are used in several key sectors of applications. In collaboration with other teams of ERCIM (the Italian IAN-CNR and the Computer Science Department of the Czech Academy of Science), the group has started a new research project related to the study of the mathematical and numerical aspects of the underground water pollution problem.

The main objective of this research is to provide efficient and reliable numerical algorithms for the simulation of the underground flux of a liquid. Underground 3-D flow modelling plays a key role in several physical phenomena and engineering processes such as oil reservoir exploitation and underground water remediation. In particular, it plays a relevant role in the simulation of the decontamination process of the area around the Stráz pod Ralskem uranium mine in the Czech Republic, and we compare several of our algorithms on a set of 3-D test problems coming from this problem (Arioli, Maryvska, Rozlovzník, and Tuma, 2001, RAL-TR-2001-023).

Darcy's law describes the relationship between the pressure ***p(x)*** (the total head) and the velocity field ***u(x)*** (the visible effect) in ground-water flow. In Figure 1, we see how Darcy's law relates the vector field ***u*** to the scalar field p via the permeability tensor ***K(x)*** which accounts for the soil characteristics, and the divergence of ***u*** to the source-sink term ***f(x)***. In Figure 2, we give an example of the domain.

In order to solve Darcy's law, we use mixed finite-element approximation techniques. This leads to the solution of an augmented, nonsingular, and sparse

Let Ω be a simply connected, bounded, polygonal domain in $\mathbb{R}^2$, defined by a closed piecewise linear curve $\Gamma = \Gamma_D \cup \Gamma_N$, and $\mathbf{n}$ be the external normal to Γ. Darcy's equations are given by:

$$\begin{cases} \mathbf{u}(\mathbf{x}) = -K(\mathbf{x})\text{grad}\, p(\mathbf{x}), & \mathbf{x} \in \Omega \\ \text{div}\,\mathbf{u}(\mathbf{x}) = f(\mathbf{x}), & \mathbf{x} \in \Omega \end{cases}$$

with boundary conditions

$$\begin{cases} p(\mathbf{x}) = g_D(\mathbf{x}), & \mathbf{x} \in \Gamma_D \\ \mathbf{u}\cdot\mathbf{n} = g_N(\mathbf{x}), & \mathbf{x} \in \Gamma_N \end{cases}$$

The vector field $\mathbf{u}$ and the scalar field p are unknown in the interior part of Ω.

Figure 1: Darcy's law.

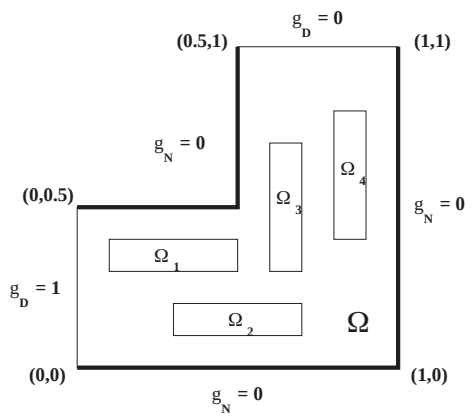


Figure 2:
Lshape domain Ω, K (x) = 10^{-8},
when x $\in \Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \Omega_4$ and
K (x) = 1 when x $\in \Omega \setminus \{\Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \Omega_4\}$.

$$\begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} - \begin{bmatrix} \mathbf{q} \\ \mathbf{b} \end{bmatrix},$$

where, $\mathbf{M} \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix and $\mathbf{A} \in \mathbb{R}^{n\times m}$ is a full rank matrix with entries equal to either 1, −1, or 0.

Figure 3: Augmented System.

Let $\mathbf{Y} \in \mathbb{R}^{n\times m}$ and $\mathbf{Z} \in \mathbb{R}^{n\times(n-m)}$ be two matrices such that $\mathbf{Y}^T\mathbf{A} = \mathbf{I}_m$ and $\mathbf{Z}^T\mathbf{A} = \mathbf{0}_{n-m,m}$. Then the steps in the null space algorithm are:

1. $\mathbf{u}_0 = \mathbf{Y}\mathbf{b}$,
2. $\mathbf{Z}^T\mathbf{M}\mathbf{Z}\,\mathbf{w} = \mathbf{Z}^T\mathbf{q} - \mathbf{Z}^T\mathbf{M}\mathbf{u}_0$,
3. $\mathbf{u} = \mathbf{u}_0 + \mathbf{Z}\mathbf{w}$,
4. $\mathbf{p} = \mathbf{Y}^T\mathbf{q} - \mathbf{Y}^T\mathbf{M}\mathbf{u}$.

The matrices $\mathbf{Y}$ and $\mathbf{Z}$ can be computed either by an orthogonal factorization method or using a shortest path tree algorithm. In the second case we do not need to store either of them.
Let us denote by $\mathbf{E}_1$ and $\mathbf{E}_2$ the matrices

$$\mathbf{F}_1 = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{n-m,m} \end{bmatrix} \quad \text{and} \quad \mathbf{F}_2 = \begin{bmatrix} \mathbf{0}_{n-m,m} \\ \mathbf{I}_{n-m} \end{bmatrix}.$$

Orthogonal factorization: $\mathbf{A} = \mathbf{H}\mathbf{E}_1\mathbf{R}$,
where $\mathbf{R}$ is an $m \times m$ sparse, nonsingular, upper triangular matrix and $\mathbf{H}$ is an $n \times n$ orthonormal matrix. Therefore,

$$\mathbf{Y} = \mathbf{H}\mathbf{E}_1\mathbf{R}^{-T} \quad \text{and} \quad \mathbf{Z} = \mathbf{H}\mathbf{E}_2.$$

Network programming: $\mathbf{PAQ} = \mathbf{LE}_1$,
where $\mathbf{L}$ is an n x n lower triangular matrix, and $\mathbf{P}$ and $\mathbf{Q}$ are permutation matrices obtained by identifying a spanning tree of a graph built using the mesh structure. Then,

$$\mathbf{Y} = \mathbf{L}^{-\mathrm{T}}\mathbf{E}_1 \quad \text{and} \quad \mathbf{Z} = \mathbf{L}^{-\mathrm{T}}\mathbf{E}_2.$$

The matrix by vector products needed by the conjugate gradient method used in Step 2 are computed by solving a lower and a upper triangular system.

Figure 4: Null Space Algorithm.

system of linear equations (see Figure 3). For real 3-D problems the system can have several million unknowns. To solve this system, we apply a specialized version of the classical null space algorithm for the minimisation of linearly constrained quadratic forms. This approach has the advantage of preserving the physical meaning of the computed velocity field $\boldsymbol{u}$ because it imposes the conservation of the flux. In the null space algorithm, first we compute a basis of the null space defined by the flow conservation equation, then we solve a reduced linear system on the null space by the conjugate gradient algorithm, implicitly computing the matrix-vector products.

Indeed, MA49 from the HSL 2002 library (see http://www.cse.clrc.ac.uk/Activity/HSL) gives the possibility of using its sparse result for implicitly computing all the matrix-vector products required by the algorithm. Moreover, the number of steps performed by the conjugate gradient method is independent of the mesh size (Arioli and Manzini, 2002, Comm. Num. Meth. Eng.). However, we point out that, for 3-D problems, storage can be prohibitive (see Figure 4). For this reason, we use a novel approach based on network programming techniques. Using the Shortest Path Tree algorithms, we identify a basis of the null space by simple permutation matrices (see Figure 4). Because we require a negligible amount of additional storage, our code is competitive for large problems where the storage required by general purpose direct solvers can be prohibitive. In particular, when the test problem has a unit square domain and $\boldsymbol{K}(\boldsymbol{x})$ has a random distribution with its values ranging from 1 to 10^{-12}, our approach is competitive in terms of CPU time with MA47 of the HSL 2002 library (Arioli and Manzini, 2001, RAL-TR-2001-037).

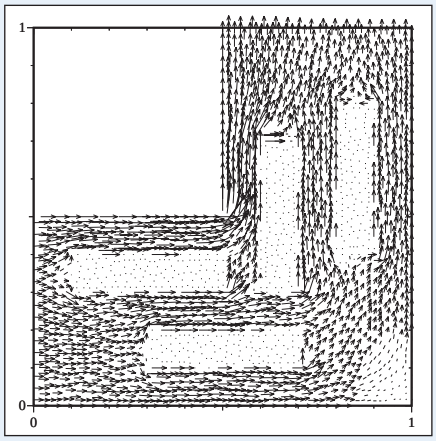


Figure 5:
Velocity flux in a domain with $K(x)$ varying between 1 and 10^{-8}.

Please contact:
Mario Arioli, CLRC
Tel: +44 1235 445332
E-mail: m.arioli@rl.ac.uk

Parallel Direct Solvers for Large Sparse Linear Systems

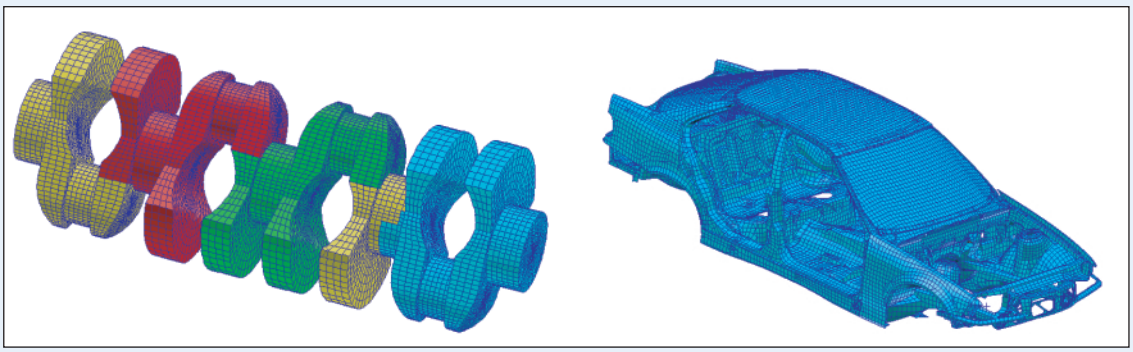
by Iain Duff and Jennifer A. Scott

In recent years, there has been an increasing demand for the efficient solution of larger and larger linear systems of equations. One way of doing this is through the development of numerical algorithms and software for parallel computers. The Numerical Analysis Group at Rutherford Appleton Laboratory is using an approach that allows to take advantage of existing expertise and experience of direct solvers to design and develop parallel solvers that offer significant speedups when used on computers with a modest number of processors.

In many industrial applications, such as the analysis of very large structures, industrial processing of complex non-Newtonian liquids, and the simulation of car bodies, the solution of large, sparse linear systems of equations is the single most computationally expensive step. Thus, any reduction in the linear system solution time will result in a significant saving in the total simulation time. As time-dependent three-dimensional simulations are now commonplace, the size of the linear systems that need to be solved is becoming ever larger, fueling the demand for algorithms and software that can be used on parallel supercomputers. Parallel computers, in addition to providing possible speedup through efficient parallel execution, are also a means for obtaining more memory for direct factorization techniques, potentially enabling the solution of problems that would otherwise be intractable.

The parallel approach that we propose is based upon first subdividing the problem into a (small) number of loosely connected subproblems by ordering the matrix to bordered block diagonal form. A direct solver is then applied in parallel to each of the subproblems. Once all possible eliminations for the subproblems have been performed, there remains an interface problem, which is much smaller than the original system. The interface solution is used to complete the solution of the subproblems.

This approach has significant advantages over attempting to design a general parallel sparse direct solver. Firstly, modern direct solvers are extremely complex codes and represent substantial programming effort by experts in sparse matrix technology. In HSL 2002, we have a number of serial sparse solvers that have been developed over many years and have proved to be very robust, reliable and efficient for a wide range of practical problems. We were therefore keen to exploit our expertise with these solvers when designing parallel software. Splitting the problem into



A crankshaft and a car body from the automative industry that are modelled by MacNeal-Schwendler as finite element problems. The problems have 1848,770 and 227,362 unknowns, respectively. By courtesy of Stefan Mayer, MSC Munich.

subproblems and applying one of our direct solvers to each of the subproblems allows us to do this. A second important advantage of this approach is that each processor can be preassigned all the matrix data required for the computations it has to perform before the factorization starts. Communications are only required to send the data that remains when all possible eliminations for the subproblems have been performed to the processor responsible for factorizing the interface problem. Interprocessor communication is thus both limited and structured. Finally, any existing sparse direct solver may be used to solve the interface problem.

Unfortunately, the factorization of the subproblems cannot be performed using an existing direct solver without some modifications. This is because standard solvers are designed to factorize the whole of the system matrix using a variant of Gaussian elimination: all the columns are eliminated in turn and the factors computed using an ordering chosen by the solver. But when applied to a subproblem that is connected to one or more other subproblems, the columns with entries in more than one subproblem cannot be eliminated. Modifications are thus needed to enable a distinction to be made between columns internal to the subproblem and those that must be passed to the interface problem.

Our work in this area began with the development of parallel frontal solvers. In recent years, we have designed and developed three codes: the first is for unsymmetric finite-element problems; the second is for symmetric positive definite finite-element problems; and the third is for highly unsymmetric linear systems such as those that arise in chemical process engineering. The codes are written in Fortran 90 and use MPI for message passing. Fortran 90 was chosen not only for its efficiency for scientific computation but also because of the features it offers. In particular, our software makes extensive use of dynamic memory allocation and this allows a much cleaner user interface. MPI is used because it is widely available and accepted by users of parallel computers. Our software does not assume that there is a single file system that can be accessed by all the processes. Thus it can be used on distributed memory parallel computers as well as on shared memory machines. Options exist for holding the matrix factors in files on disc, thus allowing the codes not only to solve very large problems but also to be used on parallel machines where each processor has access only to a limited amount of main memory.

Our parallel frontal solvers have been tested on a number of very different computing platforms, including an SGI Origin 2000, a cluster of Sun workstations, a 2-processor Compaq DS20, and a Cray T3E. For problems arising from chemical process engineering, the parallel frontal solver has been found to be significantly faster than a serial frontal code, with speedups in the range of 3 to 8 being reported on 8 processors of an Origin 2000.

Following the success of our parallel frontal solvers, we are currently working on the development of a parallel version of the well known general unsymmetric sparse direct solver MA48. This code is suitable for solving very sparse, highly unsymmetric problems and is particularly efficient when solving repeatedly for different right-hand sides. The new parallel code adopts the same design criteria as the parallel frontal solvers and, in particular, is portable, straightforward to use, efficient, and, through the range of options available to the user, flexible.

We remark that, although highly portable, our parallel solvers are only suitable for use on a modest number of processors (typically up to about 16). The reason for this limit is that, as the number of blocks in the block diagonal form increases, so too, in general, does the size of the interface problem. In our parallel solver, the interface problem is currently solved using a single processor. This presents a potential bottleneck and limits the speedups that can be achieved.

HSL Software Library

The software library HSL (formerly the Harwell Subroutine Library) is a collection of portable, fully documented Fortran packages that have been written and developed primarily by the Numerical Analysis Group at RAL. HSL has particular strengths in sparse matrix computations and large-scale optimization. HSL is marketed commercially by Hyprotech UK Ltd although there are special arrangements for licencing to academic users, and a collection of older codes from the Library are more freely available in the HSL Archive. HSL routines are regularly incorporated into advanced software applications, including chemical engineering and finite-element modelling.

Further information:
http://www.cse.rl.ac.uk/Activity/HSL/

Link:
http://www.cse.clrc.ac.uk/Activity/HSL/

Please contact:
Iain S. Duff, CLRC
E-mail: I.Duff@rl.ac.uk
Tel: +44 1235 445803
Jennifer A. Scott, CLRC
E-mail: J.A.Scott@rl.ac.uk
Tel: +44 1235 445132

Spurious Solutions in Finite Element Simulation of Electromagnetic Cavity Resonators

by Paolo Fernandes and Mirco Raffetto

The strange behaviour of some computational models of electromagnetics that produce spurious solutions in addition to physical ones is explained. The necessary and sufficient conditions to avoid these spurious solutions are given.

Wave guides and resonant cavities are basic components of microwave apparatus or particle accelerators. It is thus hardly surprising that the first numerical simulations of these devices were performed a long time ago.

One of the most powerful ways to carry out a numerical simulation of a physical device which can be modelled by a partial differential equation is to approximate that equation by a finite element method. In this kind of method the problem domain is subdivided in cells (elements) and an approximate solution is sought, which is polynomial inside each element and satisfies suitable matching conditions across the common boundaries of adjacent elements. In the case of resonant cavities, this approximate solution is determined by a matrix eigenvalue problem. The eigenvalues give the resonant frequencies while the electromagnetic field is obtained from the components of the eigenvectors (degrees of freedom), whose physical meaning depends on the specific method.

The first finite element models of resonant cavities were developed in the early seventies and it soon became apparent that, together with the resonant modes experimentally observed, many solutions having no counterpart in the real world were also obtained. The resonant frequencies of these 'spurious modes' were dramatically dependent on the element size and their field patterns did exhibit very unphysical features (eg, a large divergence, while all physical fields should have been divergence-free). A direct consequence of this difficult situation was that the outputs of numerical simulations were unreliable (as the contamination of a physical mode by a spurious one at the same frequency could not be excluded) or, at least, very difficult to use and inefficient (as, in any case, spurious modes had to be identified and discarded).

A number of attempts to overcome this problem were made, but a satisfactory solution was not proposed until the so-called "edge elements" were found. The degrees of freedom of the most obvious finite elements for vector fields are the value of the field components at specific points on the element boundary and matching these values between adjacent elements makes continuous all field components. Instead, the degrees of freedom of the edge elements are the line integrals of the field along the edges of the elements. Matching them makes continuous across any common face only the components tangential to that face. From the practical viewpoint, these elements made it possible to overcome the drawback of spurious solutions.

Unfortunately, in spite of the enormous number of papers dealing with this spurious mode issue over a period of thirty years, a theory explaining why the edge elements are spurious-free while others are not, was not correctly formulated in this vast literature. In our opinion, this state of affairs had the adverse practical consequence that new finite element models could only be developed without knowing in advance whether they would be spurious-free or not. In our study of these results, we have identified the flaws of the previous theories and we have given counterexamples to them. Quite surprisingly, we found, in particular, that even a satisfactory definition of spurious-free approximation was still lacking.

We have devoted a great deal of research effort to this issue in order to finally obtain a fully satisfactory solution of this basic theoretical problem of computational electromagnetics, including a good definition of spurious-free approximation. Furthermore, by use of results by Descloux, Nassif and Rappaz on the approximation of the spectrum of non-compact operators, we have developed a comprehensive mathematical theory of the phenomenon of spurious modes.

The above mentioned spectral approximation results played a crucial role in the development of our theory because of the following peculiar situation. Even though correct modelling of a resonant cavity leads to an eigenvalue problem for a compact operator, the resulting model includes a divergence-free constraint, which is difficult to implement in a numerical method. Hence, this constraint is usually dropped. Dropping this constraint has a twofold effect on the eigenproblem: (i) extraneous zero frequency eigensolutions are introduced; (ii) the operator is no longer compact. The idea underlying this approach is then to discard any zero (or near zero) frequency solution of the finite element approximation instead of implementing a difficult constraint. Unfortunately, this is not the whole story because conditions ensuring convergence of the spectral approximation of a compact operator are not sufficient to ensure convergence when the operator is non-compact. Therefore, if suitable additional conditions are not satisfied, the extraneous solutions are spread over the whole spectrum, thus giving rise to spurious modes.

Our theory gives some sets of conditions that are necessary and sufficient to avoid spurious modes and explains the behaviour of the spurious modes observed when not all the necessary conditions are satisfied. With the new theory, most of the edge element models previously used on empirical grounds have been rigorously proved to be

spurious-free. Both the theory and these proofs hold under very general assumptions that meet almost all the requirements of engineering practice. Simplifying assumptions — so common in mathematical work, but frequently making theoretical results almost useless to the engineer involved in real applications — are deliberately avoided. The required generality is achieved by exploiting the mathematical framework previously developed by one of the authors (P.F.) with G.Gilardi (Department of Mathematics, University of Pavia, Italy) to deal with realistic assumptions in electromagnetics. Finally, we have proved some lemmas that indicate a few modifications that can be applied to a spurious-free finite element without losing its spurious-free character. By exploiting these lemmas it is possible to develop whole families of finite elements that are spurious-free by construction. Further progress in terms of performance optimization can be obtained through the systematical investigation of these spurious-free computational models.

Links:
Detailed information about the above theory can be found in the papers listed at http://www.ima.ge.cnr.it/ima/inglese/pubbl/fr_pubbl.htm
(search for Fernandes, Raffetto and spurious).

Our paper in SIAM J. Numer. Anal. 38 (2000), 580-607: http://www.ams.org/mathscinet-getitem?mr=2001e:65172

Please contact:
Paolo Fernandes, IMATI-CNR
Tel: +39 010 6475 684
E-mail: fernandes@ima.ge.cnr.it

Scalable Solvers for Sparse Linear Systems

by Klaus Stüben

A serious bottleneck in performing large-scale numerical simulations is the speed by which the underlying sparse systems of equations can be solved. In cooperation with commercial software providers, the Fraunhofer Institute SCAI attempts to increase the efficiency of the basic numerical solvers by introducing modern solver technology.

In various application fields such as fluid dynamics and structural analysis, industry is increasingly relying on computer simulations. The physical structures to be analyzed on the computer are discretized by means of complex grids. The finer the resolution of these grids, the higher the accuracy of the corresponding numerical simulation. Unfortunately, however, increasing the grid resolution also increases the size of the corresponding (sparse) systems of equations which have to be solved numerically. Problems with many millions of degrees of freedom (unknowns) are being tackled nowadays and grid sizes will substantially grow further in the near future. The resulting large systems of equations can no longer be solved efficiently with standard numerical approaches such as conjugate gradient combined with classical preconditioners. Instead, hierarchically operating solvers are required.

The classical multigrid or multilevel approach, representing one of the major work areas of SCAI, was the first hierarchical approach to reach maturity. Rather than operating merely on the given (very fine) grid, corresponding methods combine the numerical information resulting from a (pre-defined) hierarchy of increasingly coarse grids. Compared to classical one-level solvers, the main advantage of properly designed multilevel approaches is their numerical scalability. That is, the computational work to solve the underlying systems of equations grows only proportionally with the number of unknowns. Depending on the concrete application, the computational gain may be enormous, and it increases further with increasing problem size.

Unfortunately, the integration of classical multigrid approaches into existing commercial simulation software is usually not feasible. One reason is that commercial software typically has been developed over a period of many years and the underlying data structure has not been designed with the special requirements of multigrid approaches in mind. Moreover, industrially relevant grid models are often so complex that the explicit construction of a 'natural' hierarchy of grids, as required by a classical multigrid method, is very complicated if possible at all.

Thus, a focus of the work at SCAI is on the development of 'algebraic' multigrid approaches (AMG). Corresponding solvers attempt to combine the advantages of classical multigrid approaches with those of easy-to-use plug-in solvers. To be more specific, in contrast to classical 'geometric' approaches which operate on a geometrically pre-defined hierarchy of grid levels, AMG directly operates on the linear matrix problem which corresponds to the finest-level discretization. The explicit construction of a reasonable multilevel hierarchy is part of the AMG algorithm (invisible to the user of AMG), automatically performed by exploiting algebraic properties which are easily accessible through the discretization matrix such as sign and size of entries. That is, only the discretization matrix and right hand side have to be passed to an AMG solver making it as easy to plug into an existing simulation code as any standard one-level solver. This makes AMG solvers particularly interesting for an integration into existing commercial simulation packages.

A major result of the AMG development at SCAI is an advanced solver package, SAMG, which is continuously being enhanced and extended to cover more and more applications. SAMG is most

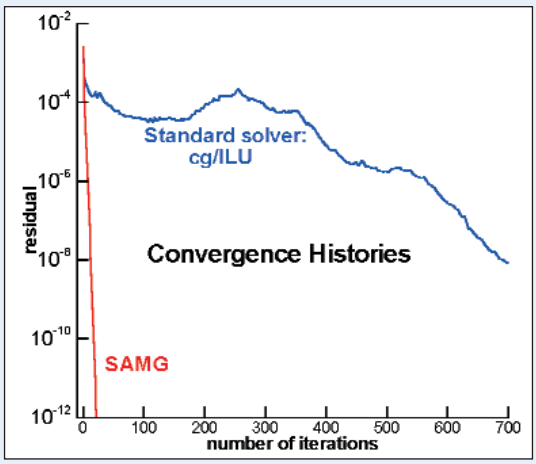


Figure 1:
Example from oil reservoir simulation: Convergence history of SAMG versus that of a standard solver.

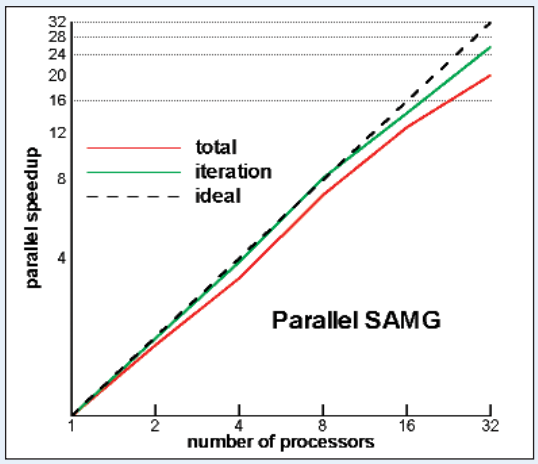


Figure 2:
Parallel performance of SAMG for the same example as in Figure 1.

mature for discretized scalar elliptic partial differential problems as arising, for instance, in computational fluid dynamics, oil reservoir simulation, ground water flow and circuit simulation. As an example, Figure 1 compares the convergence history of SAMG with that of a standard one-level solver. Here, the underlying application is the solution of a single pressure equation arising in a reservoir simulation on a grid consisting of 1.16 million cells. Compared to the standard solver, the total computational time is reduced by nearly a factor of 20. This factor will grow further with increasing grid size. For discretized systems of partial differential equations, major research still has to be performed. However, substantial progress has been achieved in various application areas such as linear elasticity and semiconductor process- and device-simulation (see Figure 3).

Further progress in numerical simulation requires the effective combination of advances in both computer and algorithm development. Ultimately, the best of all known (sequential) algorithms need to be considered for efficient parallelization. Due to their numerical scalability, hierarchical methods play an outstanding role. Consequently, a parallel version of SAMG has been developed (based on a distributed memory programming model). It can be applied to any reasonable partitioning of the given grid. Parallel SAMG scales very well as long as the number of mesh cells per processor is some 30,000 or more. Clearly, given a number of processors, the larger the grid, the higher the parallel efficiency. For the above case of 1.16 million cells, the measured speedup on 32 processors is 26 for the (iterative) solution phase, and 20 for the complete run, including SAMG's preparational phase in which the hierarchy is constructed (see Figure 2).

Link:
http://www.scai.fhg.de

Please contact:
Klaus Stüben, Fraunhofer-SCAI
Tel: +49 2241 14 2749
E-mail: klaus.stueben@scai.fhg.de

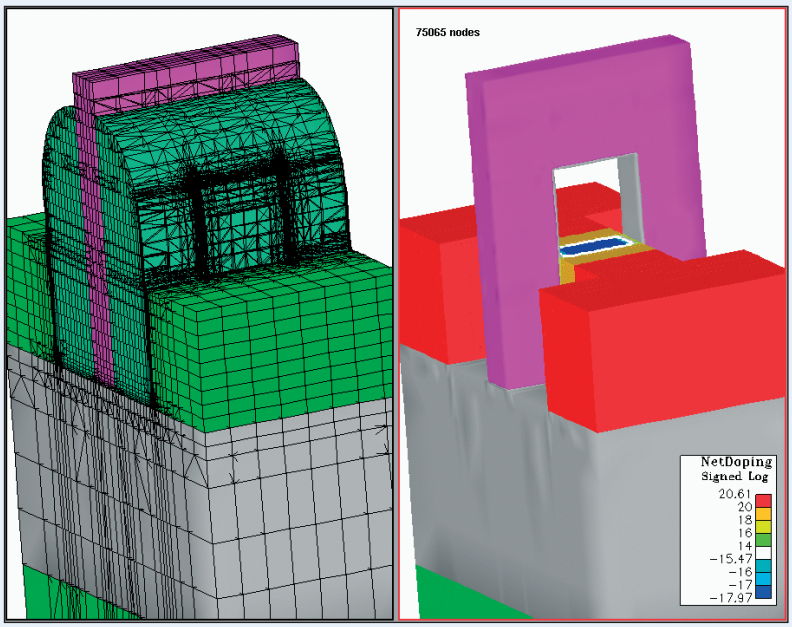


Figure 3:
FinFet transistor and underlying grid used for performing a device simulation (courtesy of Avant).

A new Algorithm for Discrete Tomography

by Rob Tijdeman and Herman te Riele

Discrete tomography concerns the problem of recovering binary images from their projections. Recently, Tijdeman (University of Leiden, The Netherlands) and Hajdu (University of Debrecen, Hungary) have made a new mathematical analysis of this problem. In a joint project starting this summer, researchers at Leiden University and CWI will implement an algorithm based on this analysis, such that it can handle large problems, and apply it to the recovery of crystalline structures and the reconstruction of the shape of heart chambers from orthogonal biplane cardiac angiograms.

Discrete tomography has its origin in the analysis of crystalline structures with the help of electron miscroscopy. In particular, one wishes to determine the presence or absence of atoms in a crystal lattice, a two-valued situation (ie, a binary image). Electron rays, transmitted through the material, yield projections of the atom structure of the crystal. The number of projections is limited because the material may get damaged by the transmitted rays.

A binary image is a rectangular array of pixels, each of which is either black (value 0) or white (value 1). A projection of a binary image is defined as a data set which for every line in some direction counts how may white pixels are intersected by that line (the transmitted ray in electron microscopy). This leads to a linear system of equations which, in general, is underdetermined and allows many solutions when only a few projections are available. It turns out, however, that discrete tomography may be quite useful in the case of convex objects such as layers and tumors. In such cases, the solution may well be unique or almost unique. In the case of tumors, the line sums are not exact in general, and so-called noise may affect the correctness of the data set.

The mathematical analysis by Tijdeman and Hajdu of the discrete tomography problem has been published in 2001 in the 'Journal für die reine und angewandte Mathematik'. Here they study the problem of reconstructing a function which maps a finite lattice of integer pairs in the plane to a subset of the integers. The sums of the function values along a finite set of lines, the projections, are given. It is shown that the possible solution functions form a grid on a linear manifold, and that the solutions with only function values equal to 0 and 1 correspond to points in the grid which have the smallest distance to the origin. It is shown further that there is a basic structure, the so-called switching component, whose translates generate the grid. A simple device is provided to derive the switching element from the set of directions. In a sequel paper, which has appeared in the journal 'Linear Algebra and its Applications', Tijdeman and Hajdu apply their theory to the special case of four directions, viz., along rows, columns, diagonals, and anti-diagonals. The resulting algorithm finds a function with exactly the prescribed line sums and with integer pixel values which are small in absolute value. From this, a 0-1 solution with approximately correct line sums is easily obtained. It turns out that for convex patterns of 1's perfect reconstruction is possible.

In the current project, an implementation of Tijdeman and Hajdu's algorithm will be improved and extended so that it can be applied to large problems coming from the two application areas mentioned above. The experience at CWI with large-scale computer calculations will be put in here. Since in many practical situations the given line sums are not exact, ie, there is noise on these data, the implementation will be adapted to handle various types of noise. If it is known that the studied objects are more or less convex, not all 0-1 solutions are equally likely to describe the original. It is then convenient to introduce a quality ranking of solutions. A device to select solutions with high ranking will be developed. Finally, the algorithm and its implementation will be extended to cope with 'grey tints' so that it will be able to find functions whose range of values is not restricted to 0 and 1, but, eg, to 0,1,2,...,15.

Example:
f_1 (original matrix),
S_1 (reconstructed matrix),
D_1 (difference matrix).

$$f_1 = \begin{pmatrix} 0&1&0&0&0&0&0&0&1&1&1&1&1&1&1&0\\ 0&1&0&0&0&0&0&1&1&1&1&1&1&1&1&0\\ 0&1&1&1&0&0&0&1&1&0&0&0&0&0&0&0\\ 0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0\\ 0&0&0&0&0&0&1&1&1&1&0&0&1&1&0&1&1\\ 1&1&1&0&0&0&1&1&1&1&1&0&0&1&1&1\\ 1&1&1&0&0&0&0&0&0&0&0&0&0&1&1&1\\ 1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&1\\ 1&1&1&1&1&1&1&0&0&1&1&0&0&0&0&1\\ 1&1&1&1&1&1&0&0&1&1&1&1&1&1&0&1\\ 0&0&0&1&1&1&0&0&1&1&1&1&1&1&0&1\\ 0&0&0&1&1&1&0&0&1&1&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1&1&1&0&0&0&0&0 \end{pmatrix} \quad S_1 = \begin{pmatrix} 0&1&0&0&0&1&0&0&1&1&1&1&1&1&0&0\\ 0&1&1&1&0&0&0&1&1&1&0&0&1&0&1&1\\ 0&0&1&1&0&0&0&0&0&1&1&0&0&1&1&0\\ 0&1&1&0&0&0&0&0&1&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&1&0&0&1&0&0&1&0\\ 0&1&0&0&0&0&0&1&1&1&0&0&0&0&0&1\\ 1&0&0&1&1&1&0&1&1&0&1&0&0&0&0&0\\ 1&1&1&1&0&0&1&1&1&0&1&0&1&0&0&1\\ 1&1&0&1&0&0&0&0&1&0&0&0&0&0&0&1\\ 1&1&1&0&0&0&0&0&1&0&0&0&0&0&0&0\\ 1&1&1&1&0&0&0&0&1&1&1&0&0&1&1&1\\ 0&1&1&1&1&1&0&0&1&1&1&1&1&1&1&1\\ 0&1&1&1&1&0&0&0&0&1&1&1&1&0&1&1\\ 0&0&0&0&1&1&0&1&1&1&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0 \end{pmatrix}$$

$$D_1 = \begin{pmatrix} \cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot\\ \cdot&\cdot&*&\cdot&\cdot&\cdot&*&\cdot&\cdot&*&\cdot&*&\cdot&*&\cdot&*\\ \cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&*&*&*&\cdot&\cdot&*&\cdot&*&\cdot\\ \cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&*&*&*&\cdot&\cdot&*&\cdot&*&\cdot&\cdot&*&\cdot&*&\cdot&\cdot\\ \cdot&*&\cdot&\cdot&\cdot&\cdot&*&*&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&*\\ *&\cdot&\cdot&*&*&*&\cdot&\cdot&*&\cdot&\cdot&\cdot&*&\cdot&\cdot&*\\ \cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&*&\cdot&*&\cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot\\ \cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&*&*&*&\cdot&\cdot&\cdot&\cdot&*\\ \cdot&\cdot&\cdot&\cdot&*&*&\cdot&\cdot&\cdot&*&\cdot&*&\cdot&*&\cdot&\cdot\\ *&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot&*\\ \cdot&*&*&\cdot&*&\cdot&*&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&*&\cdot&*&\cdot&\cdot&\cdot&\cdot&\cdot \end{pmatrix}$$

Size of f_1:	15×15
Number of differences between f_1 and S_1:	56

Please contact:
Rob Tijdeman
Leiden University, The Netherlands
Tel: +31 71 5893046
E-mail: tijdeman@math.leidenuniv.nl

Herman te Riele, CWI
Tel: +31 20 5924104
E-mail: Herman.te.Riele@cwi.nl

Computational Topology Techniques for Shape Modelling

by Silvia Biasotti, Bianca Falcidieno, Michela Mortara, Giuseppe Patané and Michela Spagnuolo

Shape understanding, as a fundamental skill for object description and representation, can traverse several application fields. In this context, mathematics provides shape descriptors, which can be successfully applied to shape analysis.

Research activity at the Istituto per la Matematica Applicata e Tecnologie Informatiche (IMATI-CNR) relates to the mathematical background and application understanding that lies behind the representation and generation of shapes in computer graphics and vision. In particular, we focus on issues related to the definition of abstraction tools for high-level descriptions of complex shape models.

For many years researchers at IMATI-CNR have been working on different aspects of shape-based modelling. Characterising a shape means constructing a computational description of the most representative features of the shape, usually a few basic types, along with their relationships (structural decomposition). This approach generally produces a graph-based representation of a shape which is decomposed into basic shape elements, or features, which should have certain properties: elements should not change under translation, rotation or scaling (invariance); the description should be locally insensitive to modification of the shape occurring far from the current focus (rich local support), and a small perturbation of the shape should produce only a small perturbation of the description (tolerance to noise). Unfortunately, it is rather difficult to define shape descriptors which fulfil all these requirements, but interesting possibilities can be explored to find a reasonable trade-off among the various possibilities. Recently, computational topology has been proposed as a branch of computational mathematics with the aim of solving problems related to topological issues, without neglecting the feasibility or the computational complexity of the problem.

Shape interpretation is especially relevant for the perception of complex forms, in which the ability to vary the level of descriptive abstraction is the key to recognizing and classifying highly complex shapes. From a mathematical point of view, classical tools, such as Morse theory, homotopy and homology, would appear to be appropriate for dealing with topological questions in computer applications. In fact, homotopy and homology are branches of algebraic topology which deal with topological invariants such as the number of connected components, holes, etc. of a given manifold (in our case a surface). Our approach to shape description is based on the classical Morse theory and would seem to be suitable for analysing any data that can be modelled as a surface. More precisely, given an object surface S and a smooth real valued function f defined on it, Morse theory provides the relationship between the critical points of f (which informally can be thought as points where the tangent plane is horizontal) and the global topology of the surface S. In this context, Morse functions, ie the set of real functions whose critical points are non-degenerate (that is, the gradient of f calculated in those points is zero and the 'Hessian' matrix of second derivatives is non-singular), have been extensively studied. Under such hypotheses on f, Morse theory states that the shape of the pair (S, f) is represented by the sequence of the homology groups of the level sets, that is the set of points P on S such that $f(P)$ is smaller than a given real

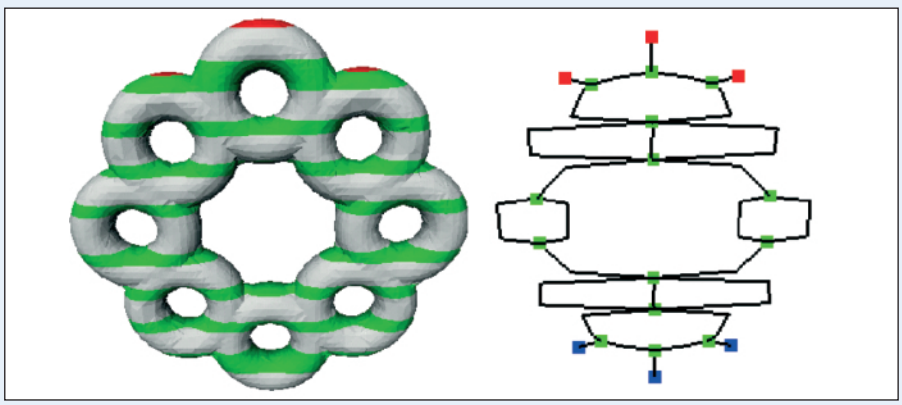


Figure 1: A topologically complex object and its Reeb graph.

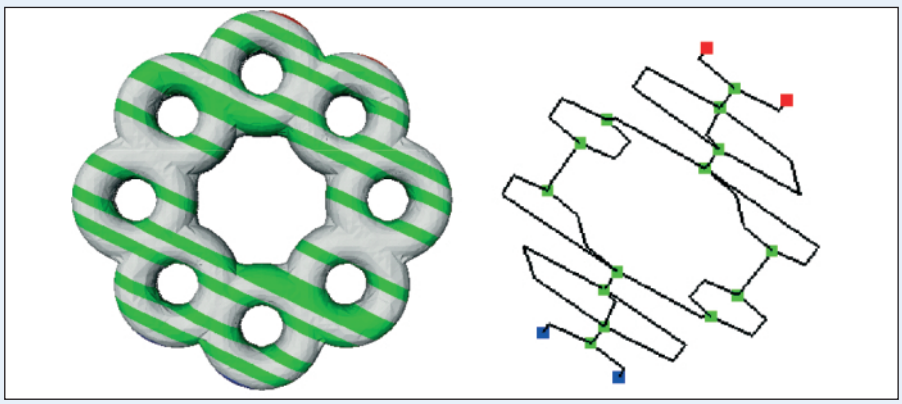


Figure 2:
The Reeb graph of the object in Figure 1 extracted with respect to a different direction.

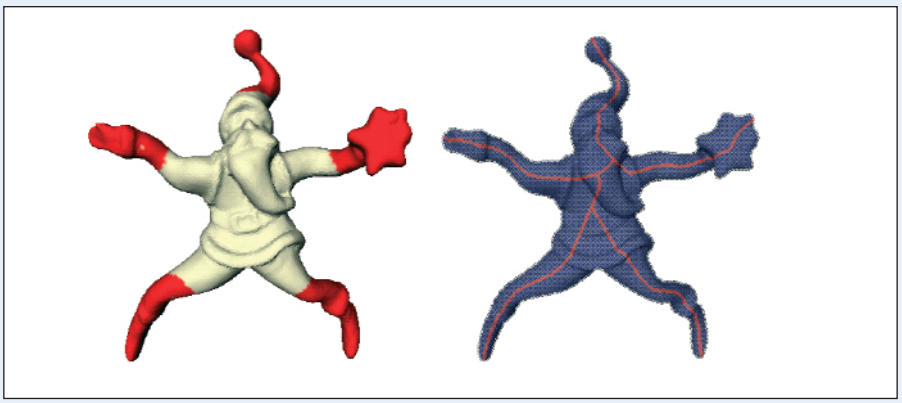


Figure 3: A skeleton of an object extracted as a Reeb graph using a non-Morse function.

number x. Moreover, the critical points of f determine the homology groups of S. Since the homology and homotopy groups codify shape properties as the number of connected components, holes and cavities of an object, it follows that a finite collection of level sets is sufficient to fully describe the surface shape, and is more complete than simply knowing the global homology. Focusing on the topological changes of level sets by varying x, we obtain a discrete description which effectively represents the shape of S and can be encoded into a topological graph, called a Reeb graph. This graph is a subset of S where points having the same value through f are identified if their pre-images are in the same connected component. Therefore, the topological changes of the surface contour levels are coded making it possible to define how the cells corresponding to several critical points glue together (see Figure 1).

Although the global homology of the surface S does not change by varying the function f, the topology of level sets depends on f. In this way the choice of the representing function f determines a collection of shape descriptors whose properties depend on the function itself, thus characterizing the object surface by taking into account the application context. Possible choices of f are the directions in the three-dimensional space (see Figure 2).

Furthermore, in cooperation with other research groups in computational mathematics, we are also trying to extend the theory to non-Morse functions and we are investigating the application of affine-independent functions. In particular, our research efforts are moving onto functions which do not depend on the orientation. For example, we have proposed a mixed strategy for extracting the skeleton of a surface represented through a simplicial complex, which combines homology with differential and computational topology techniques, (see Figure 3).

Links:
http://www.ima.ge.cnr.it/ima/inglese/predef.htm

Please Contact:
Silvia Biasotti, Bianca Falcidieno,
Michela Spagnuolo, IMATI-CNR
Tel: +39 0106 4751
E-mail: {silvia,bianca,michi}@ima.ge.cnr.it

Marked Point Processes in Image Analysis

by Xavier Descombes and Josiane Zerubia

The ARIANA research group at INRIA is developping new stochastic models based on the point process framework to tackle image analysis and feature extraction problems.

Bayesian approaches play a leading role in image analysis. They have been widely applied, among other things to quite numerous possible industrial applications, including image segmentation, image restoration and cartographic feature extraction. Basically, they consist of embedding the problem into a probabilistic framework by modelling each pixel as a random variable. Consider for instance the image segmentation problem. The goal is to obtain a partition of the image into regions which are homogeneous with respect to the image radiometry or some texture properties.

A prior, usually a Markov Random Field, is defined to model the knowledge about the expected result. In the case of segmentation, this prior can depend on the number of neighboring pixels belonging to two different regions. The prior will tend to minimize this number, penalizing partitions with numerous contours.

The data model is given by the likelihood. Without considering the neighborhood, the probability of one pixel to belong to a specific region depends on its radiometry. The result is then obtained by maximizing a Bayesian criterion such as Maximum A Posteriori which is the configuration which maximizes the product of the prior and the likelihood. This stochastic framework provides a powerful tool for the inclusion of constraints on the solution, such as homogeneity or smoothness, in the prior, leading to algorithms that are robust with respect to noise.

However, the pixelwise approach is not adequate in the case of high-resolution sensors, for which the object geometry is well defined. It seems useful to consider object level models to deal with data such as aerial images or that provided by

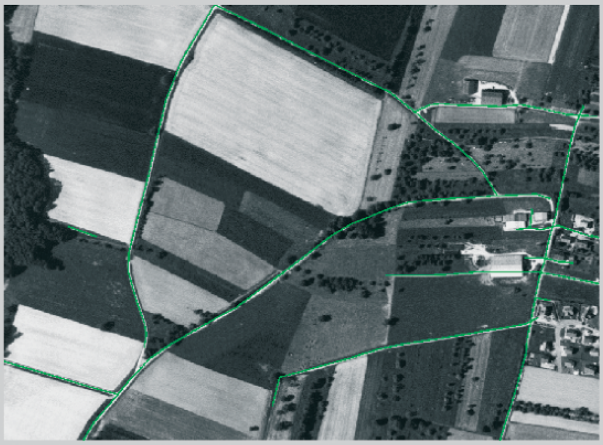


Figure 1:
Road network extraction from an aerial image provided by the French Mapping Institute IGN (result obtained by C. Lacoste).

the new generation of satellites, for example SPOT 5. In the segmentation case, these objects can be rectangles, triangles or more generally polygons, defining the geometry of objects present in the image.

To deal with this high resolution data, we study the Bayesian approach using marked point processes as priors. The main idea behind a marked point process is to model objects within a stochastic framework. These objects are random in number, are randomly localized, and their geometry is defined by random parameters. Interactions between objects are defined, allowing global properties to be embedded in the model. In addition, object shape is modelled by priors on the point parameters. In the ARIANA research group, we have developed models for image segmentation, extraction of road networks and extraction of buildings in dense urban areas. The aim of these applications is to provide tools for automatic cartography. Figures 1 and 2 show some preliminary results in road and building extraction respectively.

These models are defined on huge configuration spaces. Indeed, we do not know the number of objects. The configuration space is then the union of the spaces with a fixed number of objects. The usual Metropolis-Hasting algorithm cannot be directly applied to optimize such models. However, suitable extensions have been proposed along the same lines. The principle is to define a Markov Chain which converges to a target distribution. By including a simulated annealing scheme, the chain converges to the Maximum A Posteriori, which is the configuration which maximizes the target distribution. To define the Markov chain, an iterative process is computed. At each step, a new configuration is proposed following the so-called proposal density. This new configuration is accepted with a probability depending on an acceptance ratio which involves the target and the proposal distributions. Although convergence theorems have been proven, the efficiency of these algorithms needs to be improved in practice. A low acceptance ratio leads to a slow convergence speed. Therefore, the results obtained require huge CPU time. To fully appreciate the impact of marked point processes on image processing in term of real applications, further work on optimization techniques is necessary.

Marked point processes appear to be a promising tool for extracting objects from high-resolution remote sensing images. They define an interesting general framework for modelling cartographic features. To exploit fully their modelling power, new mathematical tools for optimizing the derived models are needed. We are currently working on this topic in ARIANA team. The ARIANA team is collaborating with Marie-Colette van Lieshout, CWI on this research topic, partially supported by ERCIM.

Link:
http://www-sop.inria.fr/ariana/

Please contact:
Xavier Descombes and Josiane Zerubia, INRIA
Tel: +33 4 92 38 78 57
E-mail: Xavier.Descombes@inria.fr, Josiane.Zerubia@inria.fr

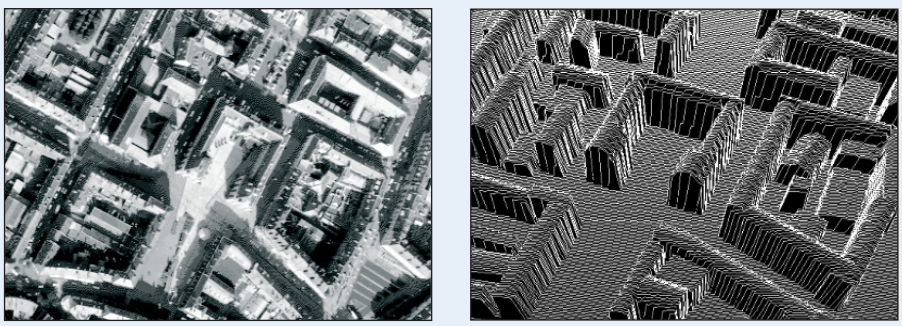


Figure 2:
Left: Aerial image provided by the French Mapping Institute IGN.
Right: extracted buildings (result obtained by M. Ortner).

Iterative Methods in Image Reconstruction

by Paola Favati, Grazia Lotti and Ornella Menchi

A research group of scientists from the Departments of Informatics and Mathematics of Pisa University and from the Institute IIT-CNR, is working in image reconstruction problems in the areas of medicine and astronomy. It was found that practical iterative solutions are possible by using suitable preconditioners which increase the rate of convergence without mixing the signal (ie the original image) and the noise (ie small errors on the data).

Image reconstruction is an inverse problem consisting of the identification of the input of a given instrument from the knowledge of its output. Inverse problems are generally ill-posed: they might not have a solution in the strict sense, solutions might not be unique and/or might not depend continuously on the data. The discrete version of the image reconstruction problem is a linear algebraic system, that is well-posed in the sense that it has a unique solution. However, this solution is completely corrupted by the noise, and is physically unacceptable because it generally does not reproduce the data within the experimental errors. This difficulty can be overcome by 'regularization', ie, by incorporating some expected physical properties of the object in terms, for example, of constraints on the size or on the smoothness of the solution. Starting from this idea, Tikhonov regularization methods solve least-squared problems depending on a regularizing parameter.

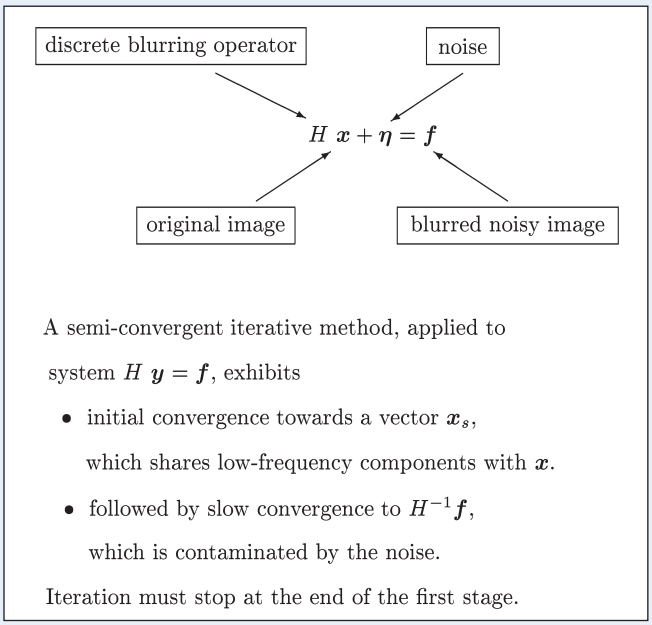


Figure 1: Discrete image reconstruction problem.

As an alternative some iterative methods, which enjoy an interesting regularization property known as semi-convergence, can be used. A semi-convergent method starts reconstructing the low-frequency components of the solution; then, as the iteration progresses, the high-frequency components are reconstructed together with the noise components. Hence the method must be stopped before it starts to reconstruct the noise. The classical conjugate gradient (CG) method, which applies to symmetric positive definite systems, has the regularizing property.

Due to ill-conditioning of the problem, the number of iterations required by CG to obtain a satisfactory result can be large and a preconditioning technique is required to increase the rate of convergence. This rate depends on the number of distinct eigenvalues of the matrix associated with the problem, and general purpose preconditioners are designed to cluster all eigenvalues around 1.

In the present context, this type of preconditioner would be harmful, mixing the signal and the noise, whereas only the eigenvalues corresponding to the low-frequency components of the signal should be clustered. In many applications, the function which describes how the imaging system affects the points of the original image (the so-called point spread function) is space invariant with respect to translations (in this case it is determined by the image of a single point source) and is

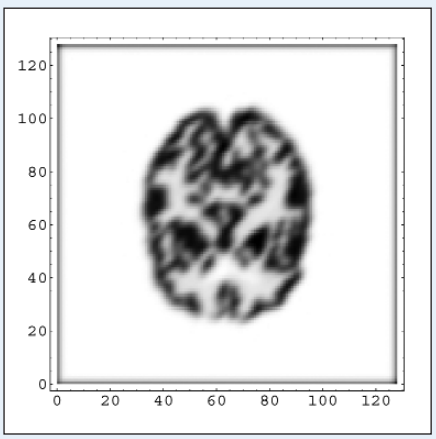


Figure 2: Blurred image of a Hoffman phantom.

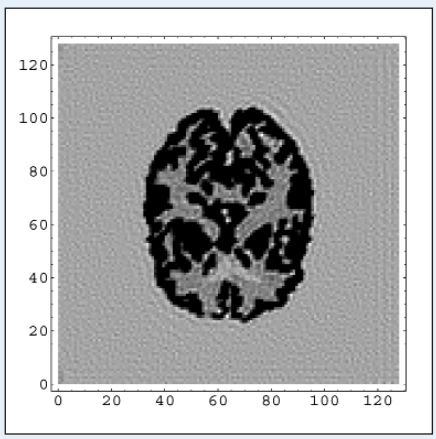


Figure 3: Reconstructed image after 5 iterations of a preconditioned conjugate gradient method.

band-limited (ie it has a local action, a single source of light is blurred into a small spot, wherever the source is located).

These properties result in a strong structure of the matrix of the linear system, which turns out to be a two-level band Toeplitz matrix. A band matrix has non-zero elements only on a few diagonals around the principal one, and a Toeplitz matrix has equal elements on each diagonal. A two-level band Toeplitz matrix is a block matrix which presents these two structures (band and Toeplitz form) both at the block level and inside each block. Circulant preconditioners, frequently applied to Toeplitz matrices, can be easily modified in order to cope with the noise. However, in the case of band Toeplitz matrices, a band preconditioner would be preferable, since it could be inverted with the same cost of a CG iteration. We have recently proposed a two-level band preconditioner, which is effective for image reconstruction problems with the above properties and has a computational cost per iteration linear with respect to the number of pixels of the image. Figure 2 shows an example of a synthetic medical image (the 2D Hoffman phantom) blurred by the instrument used for acquisition and corrupted by noise. Figure 3 shows the image reconstructed by applying a few iterations of a preconditioned CG method.

Please contact:
Paola Favati, IIT-CNR
Tel: +39 050 315 2412
E-mail: favati@iit.cnr.it

Shape Geometry and Aesthetics

by Franca Giannini and Marina Monti

FIORES-II is a research project of the European Commission that aims at investigating and identifying the links between emotional shape perception and geometry, in order to facilitate communication between designers and CAD operators and to create more user friendly tools for aesthetic design.

The market success of industrial products strongly depends on their aesthetic character, ie the emotional reaction that the product is able to evoke. To achieve their aim designers have to act on specific shape properties, but at present they are not directly supported in this by existing digital tools for model definition and manipulation, mainly because of the still missing mathematical formalisation of the properties themselves. The European project FIORES-II (GRD1-1999-10785-Character Preservation and Modelling in Aesthetic and Engineering Design), started in April 2000, aims at investigating and identifying the links between emotional shape perception and geometry and to create, through their mathematical formalization, more user friendly tools for aesthetic design

Relationships Between a Physical Form and its Emotional Message

In order to develop modelling tools to allow designers to quickly attain the desired emotional message, it is necessary to understand the procedures they follow to achieve their objectives. Within the FIORES-II project, design activities in different industrial fields have been analysed in depth and the language used in different phases of the design cycle has been studied. It emerged that the terms strictly related to emotional values (eg dynamic, aggressive, etc.) that express the objectives to be achieved by the end product are mainly used when designers talk with marketing people. On the other hand, during the creation and modification of the digital model, designers communicate their aesthetic intent using a more detailed and restricted set of terms corresponding to shape properties. In this phase they provide instructions on which elements and properties have to be changed to realise their objective (eg making a curve a bit more accelerated, or decreasing the tension of …) and to fulfil marketing directives. This second set of terms represents the first link between low-level geometric properties and the high-level features of a product. Therefore, in order to identify links between message and geometric shape, we envisage a two-level mapping: the first level links geometric properties to design terms; the second links these latter to the emotional message.

Starting from the above considerations, major attention has been given to mathematically formalising the most used terms of the Language of the Trade (ie Acceleration, Tension, Convexity, Concavity, Lead in, Crispness, Sharpness, Softness, Crown), with the objective to develop modelling tools which are fundamental for:

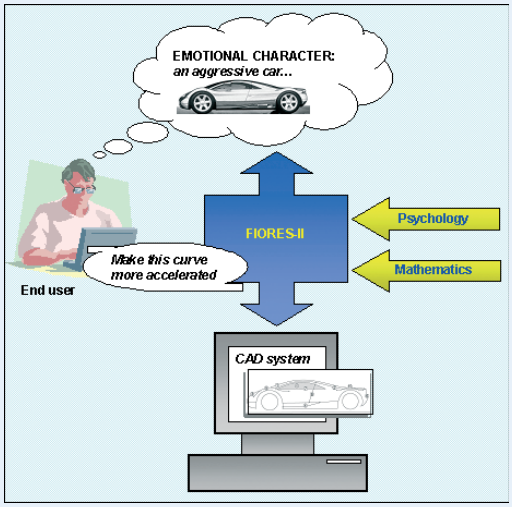
- direct shape modification with a stronger semantic control than that offered by classical methods
- specification of the aesthetic properties in objective terms
- aesthetic feature modification.

The development of these tools implies the solution of the following problems for each term:

- definition of its meaning from the design viewpoint: what shape does the designer expect when the modifier value changes for the entity considered? Which geometric properties are affected by the modifier?
- specification of the mathematical function producing the expected shape modification and the related application domain
- identification of the parameters to be provided by the user or specified automatically for character preservation, plus the specification of which parameters can be used within an optimisation process, and how

- evaluation metrics for the modifier for testing purposes.

As an example, to illustrate how the above points have been mathematically solved, the convexity property is briefly described. Traditionally a curve is convex/concave, if the curvature (ie the second derivative) along the curve has the same sign. In our case, it has a more specific meaning. Interviews with the end-users reveal that judging a curve as more or less convex depends on several factors: symmetry, roundness, curvature variation…. Many of these factors depend in turn on mathematical properties that can be calculated on the curve and have to be combined to define a suitable measure criterion, which has to be continuous and differentiable. We took into account the aspects that are implicitly judged by the users, as well as mathematical properties such as curve length, area enclosed by the curve, coordinates of the gravity centre, momentum of inertia of the lamina with respect to the axes of the local coordinate system along the curve, etc. The combination of these properties (by means of the Minkowsky measure and with the adoption of weights for better calibration) provided a measuring criterion corresponding to the user feedback in a quite satisfying way.



The FIORES-II framework.

The theoretical specification of the tools is almost complete and the implementation of a software prototype is currently under development.

It has not been easy to acquire a full understanding of how designers perceive shape and then to translate this into mathematical formalism. Even if some of the terms used have a direct mathematical counterpart, the meaning is not always the same. For example, not all curves in which the second order derivative increases are necessarily perceived as accelerating curves. Moreover, different shapes may be perceived as having the same property value. This means that several variables contribute to a single property, thus requiring a further level of interpretation to give a formal description of their interdependencies.

The preliminary results confirm the validity of the approach not only from the point of view of user interest but also from a scientific perspective, linking different disciplines such as mathematics and perceptual psychology.

Link:
Project home page:
http://www.fiores.com

Please contact:
Franca Giannini, IMATI-CNR
Tel: +39 010 6475666
E-mail: giannini@ima.ge.cnr.it

Angry * (-1) = Surprised!

by Aldo Paradiso

Facial animation systems allow reproduction of facial expressions on synthetic faces where most of the time expressions are designed by hand. In order to avoid building manually the very large set of expressions humans are able to recognize, why not generate new expressions by combining existing, even very different ones?

The Algebra of Expressions consists of a set of operators and related properties defined over a set that is a generalization of the MPEG-4 Facial Animation Parameters. It may be used to describe, manipulate, and generate in a compact way facial expressions, and adopted as a tool to further study and better understand the role of emotions conveyed by facial expressions and their relationships.

Existing animation tools allow defining facial expressions for human-like or cartoon-like synthetic faces, which are then employed for several purposes in Human-Computer interaction. Most of the tools are designed to build expressions manually, acting on low-level parameters. An example is given by MPEG-4 based tools, where each facial expression is coded as an ordered sequence of 68 integer numbers, called Facial Animation Parameters (FAPs). Parameters 3-68 (that is, 66 of them) act on points defined on a synthetic face (feature points). Each FAP represents a displacement of the associated point with respect to its original position. By displacing a point a deformation of its surrounding area is produced, resembling muscle deformations, and acting on several points facial expressions are created. A neutral expression has all values equal to zero; all other expressions have some FAP different from zero. However, the human face can generate around 50.000 distinct facial expressions, which correspond to about 30 semantic categories. Clearly the human face is extremely expressive. Is it possible to cover this amount by defining

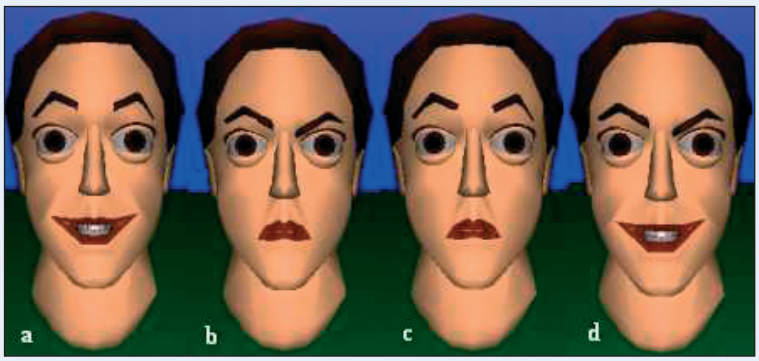


Figure 1:
Overlapping expressions. In order we have: smile (a), angry (b), one overlap (c), and a second different overlap (d).

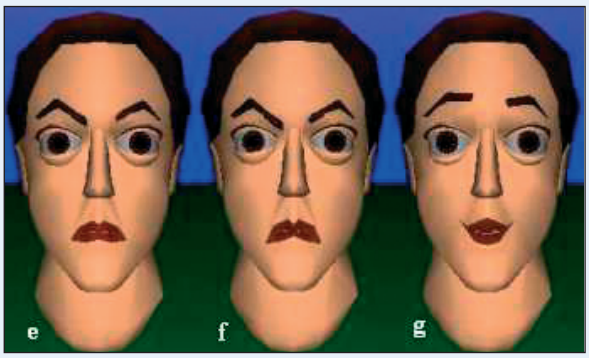


Figure 2:
Emphasizing and negating an angry expression.

a minimal set of expressions as a whole and transforming them, without manipulating directly the set of facial animation parameters? Our suggestion is to define an algebraic structure consisting of three operators and related properties defined over a set that is a generalization of the MPEG-4 FAPs. A Facial State is a set S of 66 integer numbers $(s_1, ..., s_{66})$ where each element s_i $(i = 1, ..., 66)$ satisfies the constraints defined in the MPEG-4 Facial Animation Parameter definitions table. These constraints require some of the elements to be non-negative. The first operator is called Sum: if we have two facial states S and T, and a weight v, their sum is their weighted mean (see Box 1). The set of facial states is closed under summation. In addition, the Sum operator is neither commutative nor associative in the strict sense, but in a slightly generalized sense both properties remain valid. Also, it does not have an identity. The second operator called Amplifier, is a scalar operator w (a real number) acting on a single facial state $S=(s_1, ..., s_{66})$: $|S \cdot w| = |w \cdot S| = (|w \cdot s_1|, |w \cdot s_2|, ..., |w \cdot s_{66}|)$. The symbols $||$ denote the approximation to the closest integer. The set of facial states is closed under amplification. The identity value for the amplifier is 1. The third operator is called Overlapper: we can combine two expressions S and T by choosing some regions of the face where only S applies and others where only T applies. The overlapper operator does exactly this, provided that two masks (see Box 2) are defined in order to choose the areas involved in the 'patchwork'. The set of facial states is also closed under the overlap operation. There is an identity element, but no inverse. The overlapper is both commutative and associative.

The operators introduced above have been implemented as methods included in an Application Programming Interface developed in Java. The API has been incorporated as a part of the Tinky system, a facial animation authoring tool developed by the author. Figure 1 shows the use of the overlapper. In (a) a smiling face and in (b) an angry face is shown. The mask associated with (a) covers the mouth area, while the mask associated with (b) covers the eyes and eyebrows areas. Their overlap, according to these masks, is shown in (c), where a surprised expression is resembled. This shows how to derive new expressions from old semantically different ones. By using the same expression switching the role of the masks a different overlap is produced, as shown in (d): the mouth is smiling and the eyes are angry. Again, this expression is semantically different from the previous ones, signalling smartness, arrogance or sense of superiority. In Figure 2 an angry face (e) is emphasized using the amplifier $w = 2$ (f) and has been inhibited by $w = -1$ (g). Indeed, we have found that negative values produce accordingly a negation of the expression. The negation of a smiling face seems to be a sad face, the negation of an angry face results in a surprised face, although further research needs to be done in this area to justify more general statements. Possibilities to combine expressions are countless. For example, any expression can be made asymmetric, by joining its left part with its enhanced right part (ie using the overlapper and the amplifier). Currently we are engaged in the development of ad-hoc formulas with which we may define and represent transformations of facial expressions, like asymmetric expressions, negations of expressions, derivative expressions, combinations of emotions and visemes, etc. We are also engaged in developing an animation algorithm able to exploit this algebra and use its constructs as keyframes of animations. In such a sense an effort of defining an algebra of animations, which has the task of formalizing the structure of such an algorithm, is currently under development as well.

Sum - If we have two facial states $S=(s_1,..., s_{66})$ and $T=(t_1,..., t_{66})$, with $s_i, t_i \in \mathbb{Z}$ (integers), the sum operator $+_v$ is defined as the following facial state:

$$+_v(S, T) = |s_i \cdot v| + |t_i \cdot (1-v)| \qquad i=1, ..., 66 \text{ and } v \in [0,1]$$

The symbols $|\ |$ denote the approximation to the closest integer.

Definition of Sum.

Overlapper - Let us introduce a mask $M=(\alpha_1, ..., \alpha_{66})$ where $\alpha_i=0$ or 1. If we have a facial state $S=(s_1, ..., s_{66})$, M identifies a partial region $S_M = S \cdot m_{ij}$ where m_{ij} is the diagonal matrix of M (ie, $m_{ij}=0$ for $i \neq j$ and $m_{ij}=\alpha_i$ for $i=j$; $i, j=1,..., 66$). Then, if we consider two facial states S, T with masks respectively M_1 and M_2 and priorities p_1, $p_2 \in \mathbb{Z}$, $p_1 \neq p_2$, the overlapper Ω is so defined:

$$\Omega (S_{M1,p1}, T_{M2,p2}) = (g_i) =$$

$$a)\ g_i = s_i + t_i \quad \text{if } s_i \text{ or } t_i = 0$$

$$b)\ g_i = s_i \quad \text{if } p_1 > p_2;\ g_i = t_i \quad \text{otherwise}$$

where $i=1,..., 66$

Definition of an Overlapper.

Link:
http://www.ipsi.fhg.de/~paradiso/

Please contact:
Aldo Paradiso, Fraunhofer IPSI
Tel: +49 6151 869843
E-mail: paradiso@ipsi.fraunhofer.de

Experimental Mathematics and Integer Relations

by Jonathan M. Borwein

The emergence of powerful mathematical computing environments, growing availability of correspondingly powerful (multi-processor) computers and the pervasive presence of the Internet allow mathematicians to proceed heuristically and 'quasi-inductively'. Its exciting consequences have been studied over the past ten years at the Centre for Experimental and Constructive Mathematics of Simon Fraser University, Canada. They include several recent results connected with the Riemann Zeta Function.

Mathematicians increasingly use symbolic and numeric computation, visualisation tools, simulation and data mining. This is both problematic and challenging. For example, we mathematicians care more about the reliability of our literature than other sciences. These new developments, however, have led to the role of proof in mathematics now being under siege.

Ten years ago I founded the Centre for Experimental and Constructive Mathematics (CECM, http://www.cecm.sfu.ca) and wrote: "At CECM we are interested in developing methods for exploiting mathematical computation as a tool in the development of mathematical intuition, in hypotheses building, in the generation of symbolically assisted proofs, and in the construction of a flexible computer environment in which researchers and research students can undertake such research. That is, in doing Experimental Mathematics." At present, the 'mathematical universe is unfolding' much as anticipated.

Many of my favourite examples originate in between mathematical physics and number theory/analysis/knot theory and involve the ubiquitous Zeta Function, of Riemann hypothesis fame. They rely on the use of Integer Relations Algorithms: A vector $(x_1, x_2,..., x_n)$ of real numbers possesses an integer relation if there are integers a_i not all zero with

$$0 = a_1 x_1 + a_2 x_2 + ... + a_n x_n.$$

The goal is to find a_i if such exist, and if not found, to obtain lower bounds on the size of possible a_i.

For $n = 2$, Euclid's algorithm gives the solution. The first general algorithm was found in 1977 by Ferguson and Forcade, followed by Lenstra, Lenstra and Lovász (LLL) in 1982, implemented in Maple and Mathematica, and my favourite PSLQ (Partial Sums using matrix LQ decomposition) in 1991, with a parallel version in 1999. Recently, J. Dongarra and F. Sullivan ranked Integer Relation Detection among 'the 10 algorithms with the greatest influence on the development and practice of science and engineering in the 20th century'. (Some others: Monte Carlo, Simplex, FFT.)

A CECM interface (http://www.cecm.sfu.ca/projects/IntegerRelations/) allows one to find relations and explore the underlying algorithms. Three examples are presented in the Boxes: 1. Algebraicness, 2. Riemann Zeta Function, 3. Binomial Sums.

Results in Box 3 were obtained via Gosper's (Wilf-Zeilberger type) 'telescoping algorithm' (see also Paule's contribution in this issue). Identity (3) was recently proved by Almkvist and Granville (Experimental Math, 1999) thus finishing the proof of (2) and perhaps shedding light on the irrationality of $\zeta(7)$? — $\zeta(2N+1)$ is not proven irrational for $N > 1$, though it is now known that one of $\zeta(3,5,7,9,11)$ is!

Many other similar triumphs, and an equal number of failures are described in the references given below, in fields as diverse as quantum field theory, nuclear magnetic resonancing, probability, complexity, algorithm design, and coding theory:

- The role of proof in mathematics: J.M. Borwein and P.B. Borwein, Computing in Science & Engineering,

1. Algebraicness

An immediate very useful application is to determine if a number is algebraic. We compute α to sufficiently high precision (up to n^2 digits) and apply LLL to the vector $(1, \alpha, \alpha^2, ..., \alpha^{n-1})$. Solution integers a_i are coefficients of a polynomial likely satisfied by α. If no relation is found, firm exclusion bounds are obtained.

2. Riemann Zeta Function

The Riemann Zeta Function is defined, for $s > 1$ by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Thanks to Apéry (1976) it is known that

$$\zeta(2) = 3\sum_{k=1}^{\infty} \frac{1}{k^2 \binom{2k}{k}}$$

$$\zeta(3) = \frac{5}{2}\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^3 \binom{2k}{k}}$$

$$\zeta(4) = \frac{36}{17}\sum_{k=1}^{\infty} \frac{1}{k^4 \binom{2k}{k}}$$

These suggest that

$$Z_5 := \zeta(5) / \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^5 \binom{2k}{k}}$$

is a simple rational or algebraic number.

Sadly, or happily, PSLQ tells us that if Z_5 satisfies a polynomial of degree $\leq$ 25 the Euclidean norm of coefficients exceeds 2×10^{37}. And the order and norm can be extended at will.

3. Binomial Sums

(iii) *Binomial sums.* A less well-known formula for $\zeta(5)$ suggested generalizations for $\zeta(7), \zeta(9), \zeta(11) \ldots$. Again coefficients were found by PSLQ. *Bootstrapping* the earlier pattern kept the search space of manageable size. Thus, Bradley and I (1997), quickly found:

$$\zeta(7) = \frac{5}{2}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k^7\binom{2k}{k}} + \frac{25}{2}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k^3\binom{2k}{k}}\sum_{j=1}^{k-1}\frac{1}{j^4} \tag{1}$$

Finding integer relations for $n = 1, 2, \ldots, 10$ led to a single conjectured generating function, entirely *ex machina*: for any complex z,

$$\sum_{n=1}^{\infty}\zeta(4n+3)z^{4n} = \sum_{k=1}^{\infty}\frac{1}{k^3(1-z^4/k^4)} = \frac{5}{2}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k^3\binom{2k}{k}(1-z^4/k^4)}\prod_{m=1}^{k-1}\frac{1+4z^4/m^4}{1-z^4/m^4} \tag{2}$$

The first '=' is easy, the second quite unexpected! Thus, $z = 0$ yields Apéry's formula for $\zeta(3)$ and the coefficient of z^4 is (1).

How it was found. The first ten cases show (2) has the form

$$\frac{5}{2}\sum_{k\geq 1}\frac{(-1)^{k-1}}{k^3\binom{2k}{k}}\frac{P_k(z)}{(1-z^4/k^4)}$$

for undetermined P_k; with abundant data to compute

$$P_k(z) = \prod_{m=1}^{k-1}\frac{1+4z^4/m^4}{1-z^4/m^4}$$

We found many reformulations of (2), including a marvelous finite sum:

$$\sum_{k=1}^{n}\frac{2n^2}{k^2}\frac{\prod_{i=1}^{n-1}(4k^4+i^4)}{\prod_{i=1,\,i\neq k}^{n}(k^4-i^4)} = \binom{2n}{n} \tag{3}$$

Vol.3, 48-53 (2001): http://www.cecm. sfu.ca/ preprints, and http://www.cecm.sfu.ca/personal/jborwein/cmesg25.html

- Experimental Mathematics: D.H. Bailey and J.M. Borwein in: Mathematics Unlimited — 2001 and Beyond, B. Engquist and W. Schmid (Eds.), Springer-Verlag, Vol.1, 51-66 (2000).
- Top 10 algorithms: Computing in Science & Engineering, Vol.2, 22-23 (2000): http://www.cecm.sfu.ca/personal/jborwein/algorithms.html
- Integer Relations: http://www.cecm.sfu.ca/projects/IntegerRelations/

Please contact:
Jonathan M. Borwein,
Simon Fraser University, Canada
Tel: +604 291 3070/5617
E-mail: jborwein@cecm.sfu.ca

Algebra and Computation at SZTAKI

by András Benczúr, Gábor Ivanyos and Lajos Rónyai

There has always been a strong and productive connection between mathematics and computing. Mathematical ideas turned out to be useful in the realm of computation, and, conversely, advances of computational nature have led to mathematical discoveries. Algebra and computation play a major role in some research projects at SZTAKI's Informatics Laboratory.

Symbolic Computation with Algebras, Modules and Representations

Calculating the structure of matrix algebras is at the core of several algorithmic tasks related to algebraic problems from representation theory and its applications in physics and chemistry (such as the investigation of molecular symmetries).

The study of computational complexity of the structure of associative algebras was initiated by researchers of our Laboratory in the eighties. Since then several polynomial time algorithms have been proposed for such problems. Our experience with associative algebras turned out to be eminently useful, when, together with our colleagues at the Technische Universiteit Eindhoven, we developed efficient algorithms for computations with finite-dimensional Lie algebras. The methods have been implemented as part of ELIAS, the Eindhoven LIe Algebra System by Willem de Graaf. Now these procedures are parts of the standard libraries of the computational algebra systems GAP and MAGMA.

Interestingly, ideas that surfaced during the development of Lie algebra algorithms appeared to be fruitful in the case of associative algebras as well. In particular, such constructions can be used to generate random elements of the radical (maximal nilpotent ideal) of a finite dimensional associative algebra. This enables us to avoid solving the huge systems of linear equations used in former methods. This technique, supplemented with methods of graph theoretic flavour led to an almost optimal randomized procedure for computing the struc-

ture of a matrix algebra over a finite field. The idea of randomization appears to be new in the context of radical computation. The basic method is of Monte Carlo type. There is, however, a somewhat slower Las Vegas variant (it never gives incorrect results, but — with small probability — may report failure).

Based on very similar ideas, together with Klaus Lux, one of the developers of the C-MeatAxe package for computations with modular representations of finite groups, we found a simple extension of the constructive irreducibility test of the MeatAxe to those cases where the original method did not work efficiently. The extension has been implemented in the new release of C-MeatAxe. The interested reader is referred to the homepage of MeatAxe: http://www.math.rwth-aachen.de/LDFM/homes/MTX/.

Theoretical Applications of Gröbner Bases

Gröbner basis techniques proved to be tremendously successful tools in algebraic computation. They are penetratingly effective in computational tasks related to polynomials, such as finding solutions of systems of polynomial equations. A Gröbner basis is a transparent and computationally useful generating system of a polynomial ideal. Practical applications abound from robot motion planning to combinatorial optimization.

Perhaps less known is the fact that Gröbner bases are valuable in theoretical investigations as well. As early as in the 1920's Francis S. Macaulay used them in his treatment of ideals in polynomial rings. It was recognised more recently that they may be applied to the study of finite (combinatorial) structures.

Along this line, we considered the following setting. Let $\mathcal{F}$ be a set system over the n-element base set $[n]$ (ie, $\mathcal{F}$ is a collection of subsets of $[n]$). Represent the elements X of $\mathcal{F}$ via their characteristic vectors. These are vectors of length n; the ith component of the vector of X equals 1 if i is in X, and 0 otherwise. One can view polynomials in n variables as functions on these vectors. The ideal I of all polynomials vanishing on the characteristic vectors of $\mathcal{F}$ carries a lot of information about the set system $\mathcal{F}$. For this reason, it is of interest to know the Gröbner bases and standard monomials of such ideals. We found that standard monomials are in close connection with the combinatorial concepts of shattering and higher incidence matrices.

We have described Gröbner bases and the standard monomials for the complete uniform families (when $\mathcal{F}$ is the set of all k-subsets of $[n]$, for some fixed k). This way, we could extend results of Richard M. Wilson and Péter Frankl on inclusion matrices. In the future we hope to compute Gröbner bases and attached information for more general families $\mathcal{F}$. Prospective application areas are combinatorics and the theory of computation (circuit complexity in particular).

Singular Value Decomposition and the WWW

The World Wide Web is an invaluable source of information with over two billion web pages at our disposal. However, as the size of the Web grows, it becomes exceedingly difficult to locate the quality Web pages in a topic as there may be millions of pages to choose from.

Good search engines are essential in helping to find our way across the ocean of resources. With some simplification, there are two major approaches to aid Web navigation, one based on the similarity of word frequencies across documents and another on the hyperlink structure of the Web — the Web Graph. Apparently the major engines use these approaches more or less separately. For example Google ranks its hits by using the hyperlink structure, the Northern Light clusters the hit list by word frequencies, while ResearchIndex provides results of various available ranking methods separately.

In a recently launched project we aim to combine word frequency similarities with the hyperlink structure and tailor a unified method for documents written in Hungarian.

Surprisingly, algebraic techniques can be applied in both navigational aid methods. The Singular Value Decomposition (SVD) — sometimes also called Lánczos decomposition — has long been used by statisticians (in principal component analysis) to extract essential information stored in matrices.

It is only recently realized that, when applied to the matrix of word frequencies in documents, the truncated SVD drastically reduces the dimension of the space representing the documents, providing both compression, noise filtering and a so-called Latent Semantic Indexing of the documents. Finally it is very recently realized that SVD, when applied to the adjacency matrix of the Web Graph, is capable of ranking pages by their relevance to a given query. The first such method from 1998 is the HITS algorithm of Kleinberg that approximates the first singular vector of the Web Graph of the search hit list neighborhood.

In our experiments we test variants of the SVD computation and its interpretation in order to obtain more accurate and more efficient retrieval methods.

Please contact:
Lajos Rónyai, SZTAKI
Tel: +36 1 279 6193
E-mail: lajos@csillag.sztaki.hu

Renaissance of MacMahon's Partition Analysis: Symbolic Summation at RISC-Linz

by Peter Paule

The Combinatorics Group at the Research Institute for Symbolic Computation (RISC, Kepler University Linz, Austria) has specialised in developing computer algebra algorithms to simplify sum expressions arising in combinatorics and related fields. One of the recent achievements is the revitalisation of a hundred years old method of MacMahon.

In his famous book 'Combinatory Analysis' (1916) MacMahon introduced partition analysis as a method in additive number theory for solving problems with constraints in the form of linear diophantine (ie, with integer coefficients) inequalities or equations, respectively. For several decades this method has remained dormant. In recent research work carried out jointly with G.E. Andrews (PennState, USA) who also initiated the project, we developed a new algorithmic approach to MacMahon's method and implemented it in the form of the Mathematica package Omega. Omega can be used in various ways: for finding all non-negative integer solutions to a given (homogeneous) linear diophantine system of inequalities or equations, respectively; for deriving closed form representations of multivariate generating functions with linear diophantine constraints on the summation parameters; or as a heuristic tool for combinatorial studies.

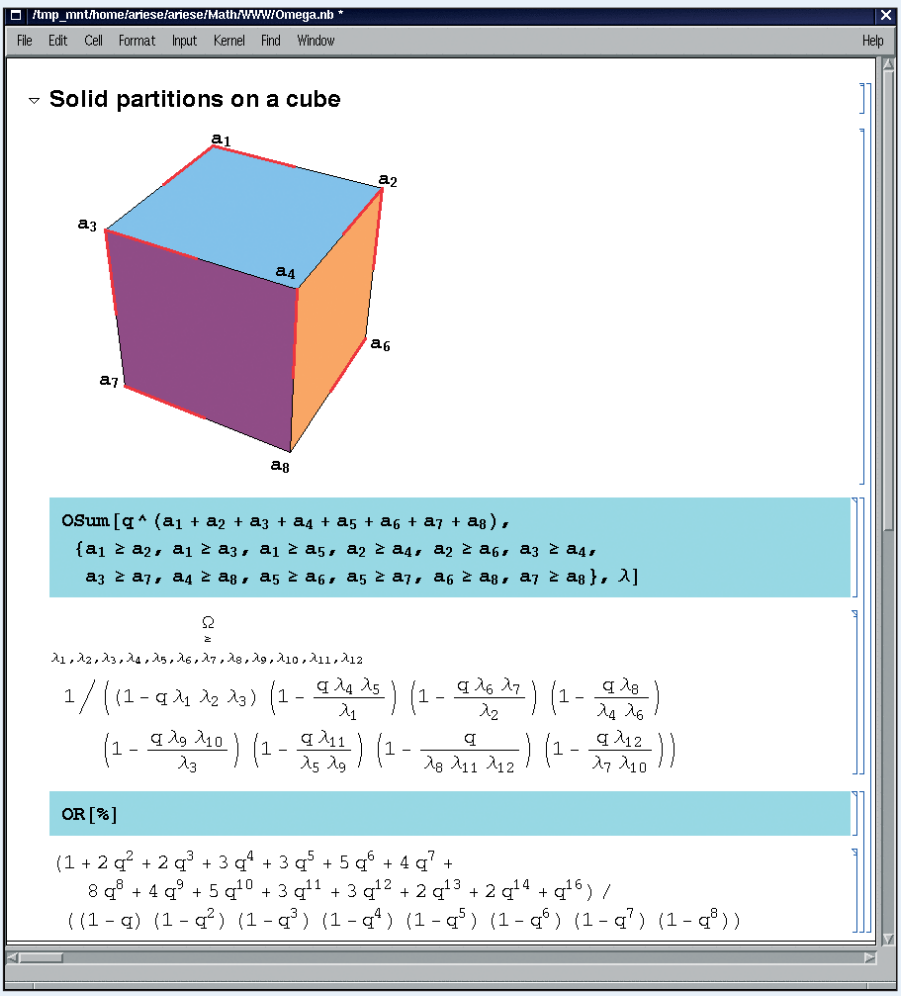


Figure 1:
MacMahon's solid partitions on the cube. The computation of the generating function.

Example (MacMahon's solid partitions on the cube): Consider all partitions of a given integer into eight non-negative integer parts a_1 up to a_8 where a_i must be greater than or equal to a_j whenever an edge is directed from a_i to a_j on the corresponding cube. How Omega is used to compute the generating function is shown in Figure 1. The computation takes only a few seconds. At MacMahon's time the situation was quite different: he cleverly split the problem into eighteen (!) subproblems which add up to the desired result. We cite from MacMahon's paper: "Mr. A.B. KEMPE, Treas. R. S., has verified this conclusion by a different and most ingenious method of summation, which also readily yields the result for any desired restriction on the part-magnitude." For Omega also this refined problem is a routine exercise!

The Omega project started some four years ago as a new branch in our longer existing and ongoing project on symbolic summation. Historically, the starting point of modern symbolic summation is Gosper's algorithm (1978) which constitutes a decision procedure for indefinite hypergeometric summation. A summand term $f(k)$ is called hypergeometric, if $f(k+1)/f(k)$ equals a fixed rational function $r(k)$ for all k. The algorithm computes a hypergeometric term $g(k)$ such that $g(k+1) - g(k) = f(k)$ for all k; if no hypergeometric solution exists, the algorithm detects this. So in the positive case the sum $f(0)+f(1)+\ldots+f(n)$ has the closed form presentation $g(n+1) - g(0)$.

Despite being a first breakthrough, for a long time its applicability has been considered as quite restricted since most hypergeometric summations arising in practice are definite. Informally, definite sums admit closed form representations only for a particular choice of the summation bounds as, for instance, in the binomial theorem.

Zeilberger's 'creative telescoping' (1990) dissolved this limitation (see Figure 3). He observed that only a slight variation of Gosper's method is needed to turn it into an algorithm which computes recurrences as most useful presentations of definite hypergeometric sums. For details we recommend the book 'A=B' by Petkovsek, Wilf and Zeilberger.

Due to the success of Zeilberger's breakthrough, symbolic summation has developed into a well-recognized discipline of computer algebra. The primary interest of the RISC Combinatorics Group is to contribute to this field by the design of new algorithms and by providing prototype implementations in the form of computer algebra packages. Obviously the applicability of these tools is not restricted to purely combinatorial investigations. The problem of simplifying complicated sum expressions arises also in other mathematical areas, for example, in number theory (see Figure 2), in special functions, in the analysis of algorithms, or in statistical mechanics.

Other applications for such procedures concern theorem proving as well as automatic validation of mathematical tables. There is a strong interaction with the THEOREMA group at RISC; see Buchberger's article in this issue. Automatic validation plays an important role in projects like the NIST Digital Library of Mathematical Functions for which Paule serves as editor; see http://dlmf.nist.gov/.

We note that the software developed by our group was already used to generate computer proofs of identities which were conjectured by (human) researchers, but for which no human proof has been found before; eg, see Figure 2 with Ahlgren's identity.

An identity first proved by the computer:

$$\sum_{j=0}^{n}(1-5\,j\,\mathrm{H}_j+5\,j\,\mathrm{H}_{n-j})\binom{n}{j}^5=(-1)^n\sum_{j=0}^{n}\binom{n}{j}^2\binom{n+j}{j}$$

This identity arose in number theoretic work of Scott Ahlgren (University of Illinois). The first proof of it has been found automatically by Schneider's package Sigma. Here H_j denotes the harmonic number $1+1/2+\cdots+1/j$.

Figure 2: Ahlgren's identity.

Creative Telescoping: Given $f(n,k)$, find $c_0(n),\ldots,c_d(n)$ free of k and $g(n,k)$ such that for all n and k,

$$g(n,k+1)-g(n,k)=c_d(n)\,f(n+d,k)+\cdots+c_0(n)\,f(n,k).$$

With mild extra conditions this implies that $S(n):=\sum_{k=0}^{n}f(n,k)$ must satisfy the recurrence

$$0=c_d(n)\,S(n+d)+\cdots+c_0(n)\,S(n).$$

In practice, one starts the procedure with $d=1$. If the algorithm succeeds, $S(n)$ is of closed form; if not, one tries to find a recurrence of order 2 with $d=2$, and so on.

Figure 3: Creative telescoping.

A significant part of the achievements of the RISC Combinatorics Group concerns various extensions and generalizations of Zeilberger's theory. For instance, with Riese's package '*q*Zeil' for *q*-hypergeometric sums one can generate automatic proofs of formulae which in the limit yield the celebrated Rogers-Ramanujan identities.

Another important object is to design algorithms for multiple (*q*-)hypergeometric sums. Based on Sister Celine's method, described also in Rainville's book 'Special Functions', Wilf and Zeilberger established a corresponding algorithmic proof theory. For transforming their setting into efficient procedures, we combined it with ideas of Verbaeten and with techniques from modular arithmetic. This enables to find closed form evaluations of two- and three-fold sums over (*q*-)hypergeometric summands which up to now have been considered as algorithmically infeasible.

An important step toward extending summation theory from hypergeometric input to summands of more general type (eg, including harmonic numbers) has been made by Schneider. In his PhD work he was able to generalize Karr's approach which can be seen as a difference field analogue to Risch integration. Schneider was the first who noticed that Karr's summation algorithm from 1980 makes implicit use of the 'creative telescoping' principle. This observation, in combination with other new ideas, enabled him to design powerful algorithms for simplifying indefinite and definite nested sums and for solving linear recurrences of arbitrary order in certain difference field extensions. His PhD thesis (2001) presents an impressive list of nontrivial applications, including nested sums originating from a vicious walker problem in statistical mechanics.

Link:
http://www.risc.uni-linz.ac.at/research/combinat/

Please contact:
Peter Paule,
RISC, Kepler University Linz, Austria
Tel: +43 732 2468 9920
E-mail: Peter.Paule@risc.uni-linz.ac.at

ALIAS, a System Solving Library Based on Interval Analysis

by Jean-Pierre Merlet

ALIAS is a C++/Maple library based on interval arithmetics, designed for system solving and optimization problems. It has been developed in the framework of the COPRIN project since 1999. Next to general purpose schemes it addresses systems with a particular structure, and intensively uses symbolic computation.

Interval arithmetics is a well known method that enables one to compute bounds for almost any analytically defined expression $F(X)$, being given ranges for the unknowns.

Such computations are required to determine, for example, the range of action of a robot, or possible molecular structures, satisfying certain constraints. The interval obtained by applying interval arithmetics on a mathematical expression is called the 'interval evaluation' of the expression and can be implemented so that numerical round-off errors are taken into account, ie the interval evaluation will always include the exact value of the mathematical expression for any instance of the unknowns within their ranges.

A direct application of this property for systems solving is that if the interval evaluation of the left hand side of at least one equation of the system $F(X) = 0$ does not include 0, then there is no solution for the system within the given ranges for the unknowns. A straightforward solving algorithm for determining the real roots of a system within some search space can easily be implemented using a branch-and-bound method. A 'box' is defined as a set of ranges, one for each unknown, and a list L of boxes B_0, B_1, B_2, ...,B_n will be maintained during the solving procedure and will be initialized with a box B_0 which describes the initial search space. An index i, initialized to 0, will indicate the number of the currently processed box in L.

The solving algorithm proceeds along the following steps:

- compute the interval evaluation of all equations of the system for the unknowns with the range indicated in box B_i
- if the interval evaluation of one equation does not include 0, then increment i and goto step 1
- if all interval evaluations include 0:
  a) if the widths of the ranges for all unknowns is lower than a given threshold, then store B_i as a solution. Increment i and goto step 1
  b) otherwise choose one unknown and bisect its range. Two new boxes are created and are put at the end of L. Increment i and goto step 1.

This algorithm stops when all boxes in L have been processed, and returns as solutions a set of boxes. It may be seen already that this basic algorithm may be extended without effort to deal with inequalities constraints and global optimization. Furthermore, distributed implementation is possible as the processing of one box is independent from the processing of all other boxes in L.

This basic algorithm may be drastically improved by using:

- 'filtering operators': these operators use the structure of the equation to reduce the width of the ranges of a given box
- 'exclusion operators': these operators determine that there is no solution to the system within a given box. Interval evaluation is one example of an exclusion operator
- 'existence and uniqueness operators': these operators may determine that there is one unique solution within some sub-box of a given box and, if this is the case, provide a numerical scheme that allows to safely compute this solution. One example of an existence operator is the Kantorovitch theorem, which uses the Jacobian and Hessian of the system, and allows to determine a box such that the Newton scheme will always converge toward the unique solution that lies within the box.

Possible operators have been proposed from different disciplines, numerical analysis and constraints programming, but no available software offers the possibility of hybrid solving based on a combination of operators. However, various experimental trials have shown that an appropriate combination of numerical analysis and constraints programming operators leads to the most efficient solver. Hence, our objectives are twofold:

- implement and combine already proposed operators
- develop new operators and improve existing ones. New operators and their mathematical analysis have already been provided by COPRIN (for example the 3B operator), but we are also working on the mathematical background of existing operators. For example we have proven that for distance equations (see the molecular biology example below) the result of the general purpose existence operator based on the Kantorovitch theorem may drastically be improved.

ALIAS is a relatively large C++ library (200,000 lines of code) that implements various general purpose solving schemes for systems solving and optimization problems. It includes a large library of filtering, exclusion and existence operators, some of them quite original in this field.

ALIAS also includes a set of specific purpose solving schemes devoted to classes of systems with a particular structure (such as distance equations). A theoretical analysis based on the structure of the equations has been used to optimize the efficiency of the general purpose operators and to develop operators that are specific to the system, thereby drastically improving the efficiency of the schemes.

Another important feature of ALIAS is its intensive use of symbolic computation. First of all, the ALIAS C++ library may be used directly from Maple: within a Maple session it is possible to call a specific solving Maple procedure that will automatically create the necessary C++ code, compile and execute it, and return the result to the Maple session. Symbolic computation is also used to improve the efficiency of the C++ code and to create the C++ code of operators that are completely specific to the system at hand.

ALIAS is the software platform of the COPRIN project and has been in development since 1999. The library can freely be downloaded (see the information at http://www-sop.inria.fr/coprin/logiciels/ALIAS/ALIAS.html). ALIAS has been used to solve various problems in very different domains. Here we give four examples.

Robotics 1
We consider a robot with 3 translation degrees of freedom x, y, z. Due to the mechanical structure of the robot only a limited workspace may be reached by the hand of the robot. Find the largest cube enclosed in this workspace such that all real roots r of a given polynomial, whose coefficients depend on x, y, z, satisfy $1/3 \leq r \leq 1$.

Robotics 2
We consider a robot with 3 successive revolving joints. The geometry of the robot is defined by a set of 30 parameters that indicate the direction of the joint axis, the lengths and respective orientations of the links connecting the joints, and the location of the base of the robot. Find the possible values of these parameters such that the hand of the robot may reach 5 pre-defined poses. This is a very demanding problem that has never been solved previously. After 5 days of computation on a cluster of 25 PCs we have been able to find 98 possible solutions within a relatively large search space.

Molecular Biology
Being given a molecule with approximately 100 atoms and constraints equations indicating that the distances between some pairs of atoms should be equal to a constant, find all possible 3D shapes of the molecule, ie find all possible locations of the atoms. This involves solving a system of about 400 distance equations. The results have been obtained in 15 minutes on a laptop.

Networking
Design of an algorithm for processor sharing policy in an integrated service network. We have here a problem in two variables a, b and it must be shown that there exists at least one solution for $b > 1$ to a system of one equation $F(a,b) = 0$ and one inequality $G(a) \leq 0.2$, where F and G contain algebraic and exponential functions.

Links:
ALIAS software library:
http://www-sop.inria.fr/coprin/logiciels/ALIAS/ALIAS.html

Please contact:
Jean-Pierre Merlet, INRIA
Tel: +33 4 92 38 77 61
E-mail: jean-pierre.merlet@inria.fr

Theorema: A System for Computer-Supported Formal Mathematics

by Bruno Buchberger

The software system Theorema provides a uniform logic and software technologic frame for proving, solving, and simplifying formulae in all areas of mathematics. Theorema is developed at the Research Institute for Symbolic Computation (RISC), Austria.

Theorema was designed by this author, and is under development since 1996. It aims at supporting all phases and aspects of mathematical exploration: defining concepts, stating theorems, formulating algorithms, proving theorems, solving constraints, simplifying formulae, executing algorithms, composing and manipulating large mathematical knowledge bases. The logic and language frame of Theorema is a version of higher order predicate logic. Theorema is programmed in Mathematica so that it can be used on all platforms on which Mathematica is available.

Theorema provides a growing library of provers, solvers, and simplifiers for general predicate logic and various special mathematical theories (set theory, elementary analysis, various inductive domains, combinatorial identities, boolean combinations of complex equalities, coordinate geometry, and others). For most of the provers, Theorema also provides human-readable output and emphasis is laid on generating proofs whose structure and presentation resembles the style mathematicians typically follow when generating and presenting proofs.

For proving theorems in certain areas of mathematics, Theorema also applies sophisticated computer algebra methods. For example, for geometric theorem proving, the Gröbner bases method, the characteristic sets method, the area method, and Collins' cylindric algebraic decomposition method are used. The application of such methods has two aspects: the methods are applied for proving theorems but they also need to be proved correct! The proof of the correctness of such methods is on a layer (object layer) which is different from the layer of applying the method (meta-layer) for proving. For example, exploring the theory of Gröbner bases needs set theory proving, inductive proving, and other proof techniques, whereas the application of the Gröbner bases method for geometric theorem proving proceeds in the frame of the

boolean and equational theory over fields. The transition from the object layer to the meta-layer seems to be an important technique that makes human theory exploration and proving efficient and feasible. For this reason, in Theorema, emphasis is put on providing a frame for doing formal mathematics in sequences of formal layers.

Another application of this principle in Theorema makes it possible to prove the correctness of an algorithm in a first phase and then, in a second phase, to execute the algorithm without leaving the logic and language frame of Theorema. In fact, in Theorema, algorithms are just special logic formulae.

Theorema also has an elaborate functors construct by which towers of mathematical domains can be generated. Following the principle of switching between object and meta-layers, functors in Theorema are not only used for generic programming and computing in towers of domains but also for constructing corresponding towers of provers.

Theorema also provides various tools for organizing and structuring the proofs automatically generated by the system so that studying such proofs should be easier than studying human generated proofs. Two of these techniques are the 'focus windows' technique by which, in a given proof step, all and only the relevant formulae are displayed and the 'logicographic symbols' technique by which pictures can be introduced for logic predicates and functions.

The Theorema Group 2002: Standing: Koji Nakagawa, Bruno Buchberger, Wolfgang Windsteiger, Christian Vogt, Adi Craciun, Temur Kutsia. Sitting: Florina Piroi, Tudor Jebelean, Gabor Kusper, Nikolaj Popov.

The objective of Theorema is necessarily open-ended. We try to incorporate more and more proving, solving, and simplifying algorithms for various parts of mathematics and to build up a network of interactions between these algorithms. Also, equally important, hierarchies of mathematical knowledge in towers of domains is being built up. Currently we are working on two benchmark application problems: first, a formal development of the theory of Gröbner bases, which — since its introduction by B. Buchberger in the late sixties — has traditionally been a research topic at RISC, and second, a formal development of the theory of Hilbert spaces, which is the frame for joint research with numerical analysts in the project 'Scientific Computing' at the Johannes Kepler University of Linz, to which RISC belongs. Also, within this project, there are particularly strong ties to the work of Peter Paule's group on combinatorial identities. The Theorema Group is also a founding member of the European Calculemus Network in which all European groups working in the interaction between computer algebra and automated theorema proving cooperate.

Links:
Scientific Computing at the Johannes Kepler University of Linz:
http://www.sfb013.uni-linz.ac.at/

Peter Paule's group on combinatorial identities: http://www.risc.unilinz.ac.at/research/combinat/risc/

European Calculemus Network:
http://www.calculemus.net/

Papers on Theorema:
http://www.Theorema.org/

Some of them can also be downloaded from
http://www.risc.uni-linz.ac.at/people/buchberg/

Please contact:
Bruno Buchberger, RISC,
Kepler University, Linz, Austria
Tel: +43 664 4211646
E-mail: buchberger@risc.uni-linz.ac.at

A Colourful Theory on Graphs

by Zsolt Tuza

The concept of List Colouring is a relatively new issue in the field of graph colouring theory. It is related to hot topics of applications, most notably the frequency assignment problem in wireless communication. In a somewhat different form, its study at SZTAKI dates back to the late 1980s when we encountered a practical scheduling problem. Since the mid-1990s we have carried out related work in active collaboration with researchers from several countries.

In the standard formulation of the List Colouring problem, a graph (network) and lists of colours allowed for its vertices (nodes) are given. The goal is to select one colour for each vertex from its list in such a way that adjacent vertices never get the same colour. An alternative approach, called Precolouring Extension, starts from a feasible colouring of a part of the graph in question, and asks for the minimum total number of colours admitting a colouring of the entire graph that extends the partial colouring given. A more general variant is the Set Colouring problem, where one has to select several colours for each vertex from its list. (The colours selected for adjacent vertices must be distinct.)

In this area we carry out research in various directions. A fundamental question is whether a given collection of lists admits any feasible colouring. One goal of our research is to draw a sharp line separating the efficiently (ie, polynomial-time) solvable problem instances from the intractable classes. Another interesting question is how much larger subsets can be selected if the lists of allowed colours are larger with a given amount. Particular types of graphs, eg planar networks, are of importance, too.

In one of the applications the vertices represent transmitters, the list of colours allowed for a vertex contains the frequencies locally usable, and adjacency means that interference may occur between the corresponding two stations and hence they should not use the same frequency. In this interpretation, selecting large subsets of the colour lists ensures that the stations can operate on a fairly wide range of frequencies without the danger of interfering with each other.

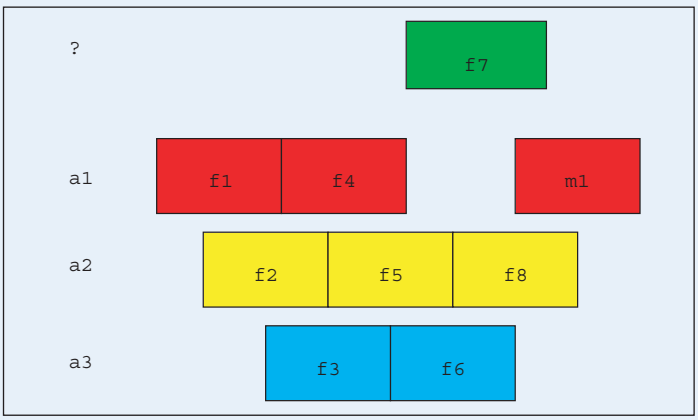


Figure 1: Greedy algorithm requires 4 colours.

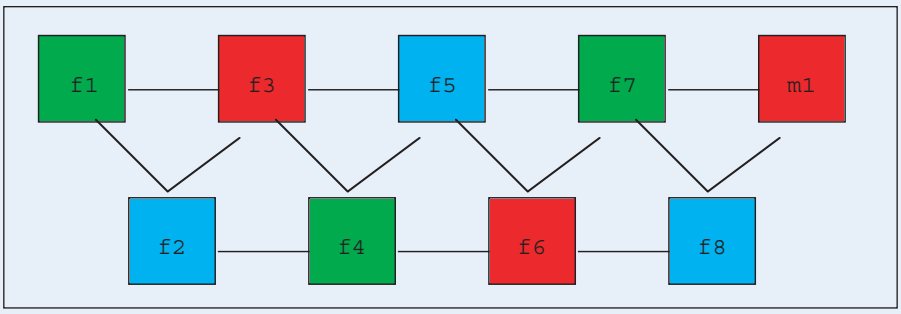


Figure 2: A better solution with 3 colours.

In another application — what was the starting point of our research on Precolouring Extension — a mixed type of scheduling has to be found. Given a maintenance plan for a fleet of airplanes, and given the flights in the timetable, the problem is to assign the flights to the airplanes in such a way that the assignment be compatible with the maintenance plan. In this model the vertices represent time intervals (maintenances and flights), the colours correspond to the airplanes, the precoloured part is the set of maintenances, and adjacency means that two represented time intervals overlap.

In the paper where we initiated the study of Precolouring Extension with Miklós Biró and Mihály Hujter (both ex-SZTAKI), we designed an efficient algorithm that finds, in terms of the mixed aircraft scheduling problem, a feasible schedule whenever it exists, provided that each airplane has just one maintenance within the time horizon considered. In later works with Mihály Hujter we investigated the algorithmic complexity of the graph-theoretic problem under various conditions, and also studied related structural questions on some interesting particular classes of graphs. We have discovered good characterisations (efficiently testable necessary and sufficient conditions) for the extendability of partial colourings with a given number of colours.

We have explored the area of Set Colouring in joint research with Margit Voigt (TU-Ilmenau). With the collaboration of Noga Alon (Tel-Aviv University) we have proved that for sufficiently large lists of allowed colours it is precisely the

graph invariant called 'fractional chromatic number' that expresses the proportion of colours that can be selected from each list. Moreover, in joint work with Jan Kratochvil (Charles University, Prague) we proved that this estimate, which is tight in general, can significantly be improved when stronger conditions hold for the pairs of adjacent lists.

The concepts described above have many interesting connections to other areas, too, with an extensive and continuously growing literature. (For an overview of results discovered until mid-1997, see Zs. Tuza: Graph colorings with local constraints — A survey, Discussiones Mathematicae Graph Theory, 17:2 (1997), 161–228.)

The area offers lots of challenging open theoretical questions as well. Among others, this includes problems on graphs where feasible colourings exists whenever the size of every list is at least as large as the 'chromatic number'; and on the largest possible number of vertices in partial colourings if the lists are not large enough to make the entire graph colourable.

It has also turned out that List Colouring can be formulated as a subproblem of the Colourability problem of 'mixed hypergraphs', a concept introduced by Vitaly Voloshin (Academy of Sciences of Moldova, Kishinau) about a decade ago. In mixed hypergraphs, two types of sets are distinguished, and vertex colourings are to be found where each set of the first type has two vertices of the same colour, while each set of the second type has two vertices of distinct colours. We carry out research in this direction with Gábor Bacsó (SZTAKI) and Lorenzo Milazzo (University of Catania).

We are open to extend our broad international cooperation to further research centres of Europe, both in theory and applications.

Please contact:
Zsolt Tuza, SZTAKI
Tel: +36 1 297 6110
E-mail: tuza@sztaki.hu

Quantitative Domain Theory

by Michel Schellekens

Domain Theory, a formal basis for the semantics of programming languages, originated in work by Dana Scott in the mid-1960s. Models for various types of programming languages, including imperative, functional, nondeterministic and probabilistic languages, have been extensively studied. Quantitative Domain Theory forms a new branch of Domain Theory and has undergone active research in the past three decades. The field involves both the semantics of programming languages and the mathematical field of Topology. Researchers at The National University of Ireland, Cork, focus on a recent development which has led to a partial solution of an open problem in Topology. Interestingly, the solution was guided by Computer Science intuitions.

Domain Theory originated in the revolutionary work by Strachey and Scott on the development of models (semantics) for programming languages. Scott's work quickly led to connections with the mathematical fields of Topology and Category Theory. In this framework, the least fixed point of a functional, defined on a domain of Scott continuous functions and with range in the same set, is obtained as a topological limit. The topology involved is now known as the Scott topology.

An alternative to this essentially order theoretic approach was proposed by the Dutch school at CWI and originated in work by De Bakker, Zucker and America. This approach advocates the use of metrics.

In order to reconcile both approaches to Domain Theory, Smyth pioneered at Imperial College the use of methods from the field of Non-Symmetric Topology. This field traditionally studies 'quasi-metrics' which are obtained from classical metrics by removing the symmetry requirement. Hence, for quasi-metrics, the distance from a given point to a second point need not be the same as the converse distance. A simple example of a quasi-metric is the 0-1 encoding of a partial order which defines the distance between two points x and y to be 0 in case x is below y in the order and 1 otherwise.

Recent developments in Domain Theory indicate that additional concepts are required in order to develop the corresponding applications. These developments include domain theoretic approaches to dataflow networks, logic programming, domain theoretic approaches to integration, models for probabilistic languages as well as models which incorporate complexity analysis. An extensive series of papers has been published in this last area, by the author in collaboration with Salvador Romaguera (Polytechnical University of Valencia). Each of these applications involve 'real number measurements' in some sense, and hence the adjective 'quantitative' is used as opposed to the adjective 'qualitative' which indicates the traditional order theoretic approach.

The question remains as to how 'quantities' can be introduced to classical Domain Theory in a simple and elegant way. Even though we favour weightable quasi-metric spaces, introduced by

Matthews at Warwick, our approach is to view Quantitative Domain Theory in first instance as an extension of traditional Domain Theory via a minimal fundamental concept: that of a semivaluation. A semivaluation is a novel mathematical notion which generalizes the fruitful notion of a valuation on a lattice to the context of semilattices.

The semivaluation approach has the advantage that it allows for a uniform presentation of the traditional quantitative domain theoretic structures and applications, as for instance the totally bounded Scott domains of Smyth and the partial metric spaces of Matthews.

A topological problem stated by the Swiss Mathematician Hans-Peter Kunzi essentially required the mathematical characterization of partial metrics. Interestingly, such a characterization has been obtained based on the notion of a semivaluation. This notion was directly motivated by computer science examples. Hence the result forms an example of recent developments where the mathematical area of Topology is influenced by Computer Science. Traditionally the influence has largely been in the opposite direction.

The benefit to Computer Science is that semivaluations allow for the introduction of a suitable notion of a 'quantitative domain' which can serve to develop models for the above mentioned applications. This work has been reported in two recent papers, both of which will appear in the Elsevier journal 'Theoretical Computer Science', and which are available from our webpage: 'The correspondence between partial metrics and semivaluations' and its sequel: 'A characterization of partial metrizability. Domains are quantifiable.'

Current investigations in this area are under way at UCC as part of a Boole Centre for Research in Informatics BCRI-project led by the author, in collaboration with Maria O'Keeffe and, via Enterprise Ireland and Royal Irish Academy projects in collaboration with researchers from the university of Siegen, the university of Birmingham and Imperial College London.

Please contact:
Michel Schellekens
National University of Ireland, Cork
University College Cork (UCC)
Department of Computer Science
Tel: +353 21 490 3083
E-mail: m.schellekens@cs.ucc.ie
http://www.cs.ucc.ie/~mpcs/

Finding Dependencies in Industrial Process Data

by Daniel Gillblad and Anders Holst

Dependency derivation and the creation of dependency graphs are critical tasks for increasing the understanding of an industrial process. However, the most commonly used correlation measures are often not appropriate to find correlations between time series. We present a measure that solves some of these problems.

Today, many companies keep close track of their systems by continuously recording the measurements of a large amount of sensors. At the same time, the hard competition in many of these industries leads to a need to increase the knowledge of the process for further optimization. Much of the recorded data is in the form of time series, ie sequences of observations indexed by the time at which they were taken. To gain thorough insight into the process, an understanding of how each measured sequence affects the others is necessary. More specific, the general task is to find how strongly the time series are correlated to each other, how long the time lag is between the correlations and, if possible, the causal direction.

For example, consider a process where you would like to find out how long it takes for material to flow from one point in the process to another. Injecting tracing substances and measuring the time directly might be expensive, difficult and interfere with the process in a negative way. Instead, it might be possible to measure two related variables, for example temperature, at the two points in the process and find out the time lag between the strongest correlations. This time lag can then give a very good indication of how long it takes for material to travel from one point to the other.

There are a number of applicable correlation measures, the perhaps most common one being the correlation coefficient. Unfortunately, many correlation measures, the correlation coefficient included, do not consider the specific properties of time series. Therefore they give a too smooth correlogram to be practically useful and are not sensitive and exact enough to determine the delay. Here, a measure based on the mutual information rate is described which solves some of the problems encountered using other correlation measures.

Mutual information is a general correlation measure that unlike the correlation coefficient can be generalized to all kinds of probability distributions. (The correlation coefficient assumes a Gaussian distribution.) Mutual information is based on the information theoretic notion of entropy, (see Eq. (1)), which is the expected amount of information given by a stochastic variable. The mutual information can be written as in Eq. (2) and be interpreted as the part of the information that is common to both variables. Given an appropriate model of the distributions, this measure can potentially detect non-linear dependencies between variables. However, it is still a general measure of correlation not specific for time series, with similar problems as the other general measures.

Entropy of the variable X:

$$H(X) = -\sum_{k \in \chi} P(x_k) \log P(x_k) \quad (1)$$

Mutual information between variables X and Y:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (2)$$

Entropy rate of the time series X:

$$H_{rate}(X) = \lim_{n \to \infty} H(X_n | X_{n-1}, X_{n-2}, \ldots, X_1) \quad (3)$$

Mutual information rate between time series X and Y:

$$I_{rate}(X;Y) = H_{rate}(X) + H_{rate}(Y) - H_{rate}(X,Y) \quad (4)$$

Markov assumption:

$$P(X_n | X_{n-1}, X_{n-2}, \ldots, X_1) = P(X_n | X_{n-1}) \quad (5)$$

Figure 1: Equations 1-5.

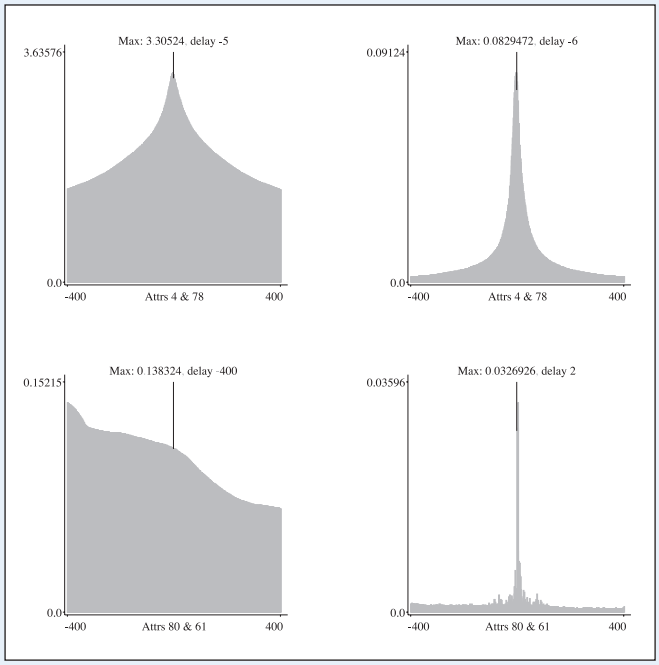


Figure 2: correlograms generated for chemical plant data.

The problem with these is that in a time series, one value is not independent of the previous values. For example, the current value of a time series is probably close to the previous value. This means that several time steps contribute partly with the same information, which causes the general correlation measures to give too high values of the correlation. To construct a more sensitive and accurate measure specific for time series, we can instead use an expression for the amount of information that is new in each time step. This is the entropy rate given in Eq. (3). Using this we can also define the mutual information rate as in Eq. (4). This way, the mutual information rate measures the complete dependence between the sequences.

Working with finite amounts of data, the entropy rate in Eq. (3) is impossible to estimate perfectly. Using a Markov assumption (Eq. (5)) the information rate can be computed. Because of this, we also have to restrict ourselves to measuring influences with specific time delays between series. This means that we construct a correlogram measuring the direct dependencies between one series and the other for different time shifts between them.

The measures presented here have successfully been used on several real data sets, mainly from a paper mill and a chemical plant. Figure 2 shows correlograms generated for chemical plant data. The correlograms to the left in the Figure show the mutual information and the correlograms to the right the mutual information rate between two different pairs of variables. In each correlogram, the *x*-axis represents the delay between the two series and the *y*-axis the degree of correlation between them.

The correlation between the first pair of variables (top) is an example of a well behaved, linear correlation with a short and reasonable time delay. The mutual information correlogram shows just one clear peak at delay -5, and the mutual information rate correlogram for the same attributes shows the same behaviour. Note that the correlation in the right Figure is much lower, which is closer to the truth than the high correlation to the left. The correlation is also much more sharply peaked at the correct time delay of the influence.

The correlogram for the linear mutual information between the second pair of variables (bottom) is very smooth, although somewhat low, but the measure is obviously fooled by some general trend in the data since it is constantly increasing with decreasing values of the delay. The mutual information rate on the other hand shows a clear peak at delay 2. That is a plausible value of the delay between the sequences, although the value of the correlation is rather low. The information rate diagram is not at all as smooth as the mutual information, showing several small spikes which are very likely effects of noise and oddities in the data. This is because the mutual information rate requires a more complicated model to estimate, and therefore is slightly more sensitive to noise.

Using mutual information or the correlation coefficient tends to give a too high value of the correlation. This happens because if the series evolve slowly enough, pure random coincidences between the series get multiplied with a factor depending on how slow the series are. The information rate, which only considers new information in every step, correctly compensates for this effect.

All in all, the mutual information rate gives more reliable indications of correlations between time series than the general correlation measures not specifically adapted for time series data.

Please contact:
Daniel Gillblad, Anders Holst, SICS
Tel: +46 8 633 1500
E-mail: dgi@sics.se, aho@sics.se

Interaction of Stochastic Systems and Financial Mathematics

by László Gerencsér

This article summarizes recent results obtained in the area of stochastic systems and financial mathematics, the focus being on the statistical theory of Hidden Markov Models, financial time series and option pricing on incomplete markets.

Hidden Markov Models (HMM) and financial mathematics have become our main technical area of research in recent years, with a strong interaction between the two areas. We have carried out research in Hidden Markov Models to derive new techniques in modelling economic time series, such as stochastic volatility processes, while research on financial mathematics has been focused on more conceptual problems, such as hedging under the presence of transaction costs. The main advance in these areas is that we have established a link between HMM-s and linear stochastic systems, enabling us to use a well-developed arsenal for the statistical analysis of HMM-s. Furthermore, in a cooperative effort we have developed new techniques for analyzing portfolio processes under transaction costs, leading to new, fundamental theorems on option pricing.

Our interest in Hidden Markov Models originates in our interaction with the group of Ian Hunter at MIT on control problems related to master-slave microrobots for eye surgery and heart surgery. To achieve high-accuracy control with low-accuracy sensors has led us to the mathematical problem of identifying a Gaussian ARMA-process under quantized observations, meaning that the actual signals are observed with finite precision. This seemingly innocent problem has turned out to be unexpectedly hard and is still unsolved: neither solid statistical theory, nor a computationally viable procedure is available today. An advance in this area is that the identifiability of quantized Gaussian ARMA processes has been established recently by a former member of our group (Ádám Szeidl et al.).

Hidden Markov Models constitute a significant extension of the classical theory of linear systems developed by Kalman. Its development was prompted by speech processing in the sixties. In a Hidden Markov Model we have a state, such as a phoneme, that follows a Markovian dynamics and we have an observation, called a read-out, which is a probabilistic function of the state, such as a digitized version of a phoneme. A fundamental problem of HMM theory is to identify the dynamics and the read-out kernel based on observed values of the process and to estimate the state. The significance of HMM-s in control theory has been recognized in the past few years. HMM-s have become a key tool for modelling not only in control but also in communication, biological systems and economics.

For the statistical analysis of HMM-s we have developed, with Gábor Molnár-Sáska, György Michaletzky and Gábor Tusnády, a fundamentally new tool establishing a link with the statistical theory of linear stochastic systems.

While the computation of the maximum-likelihood estimate of HMM parameters, such as system parameters for a quantized Gaussian ARMA-processes, are hopelessly hard, we did have significant advance in the problem of estimating a quantized linear regression. With Ildikó Kmecs we have developed a computationally feasible method for implementing the well-known EM-method, in which the unknown log-likelihood function is replaced by its conditional mean, given the observations and a tentative, a priori value for the unknown parameter. The key computational difficulty, as in all HMM estimation problems, is to compute a large number of conditional expectations. Using a Markov-Chain Monte-Carlo method we have arrived at a randomized EM-method, the theoretical justification of which has been given by a novel application of the theory of recursive estimation as developed by Benveniste, Metivier and Priouret. Extensive experimental work has proved the viability of the method.

We have cooperated in the area of HMM-s with Francois Legland, IRISA, Rennes, Lorenzo Finesso, LADSEB, Padova and Jan van Schuppen, CWI, Amsterdam, all members of the ERCIM Working Group on Systems and Control.

Our research activity in financial mathematics has been motivated by the emerging interest in option pricing within the control community. In the original theory of Black-Scholes a key condition is that of completeness, meaning the possibility of exact hedging, or synthetizing of any terminal claim starting with an initial capital which is the price of the option, and re-balancing our portfolio over time with no additional cash movement. However, real-world markets are incomplete, and thus hedging against future risk can be done at best with a prescribed probability. This so-called quantile-hedging requires tools of stochastic programming and control.

A prime example of an incomplete market is a market with friction: these are markets with positive transaction costs. A basic problem for the seller of the option is an initial endowment being sufficiently high and well-structured to superreplicate a given contingent claim at the time of maturity. This fundamental problem has been solved by a member of our group, Miklós Rásonyi, jointly with Youri Kabanov and Christophe Stricker, of the Université de Besançon.

To connect theory and practice we need to understand data. The analysis of economic time series is a fascinating area where well-established techniques of system identification can be mixed with novel ideas. With Zoltán Reppa we have analyzed the marginal distributions of the daily returns of some of the most liquid shares traded on the Budapest Stock Exchange. This study has led us to the analysis of ARMA-processes with heavy-tailed, normal inverse Gauss (NIG) innovations. We have extended basic results of system identification,

implying significant reduction of secondary sources of prediction errors, and developed a computationally feasible estimation method.

A model class particular for financial mathematics is the class of stochastic volatility models, in particular GARCH models. These have been studied in cooperation with Zsuzsanna Vágó. Detection of changes in the dynamics of physical systems, such as the vibration characteristics of an off-shore oil platform is a hot area of control theory. It is natural to ask if and how the techniques of these areas are applicable to the detection of changes of financial time series. Our earlier work with Jimmy Baikovicius on real-time change-point detection of ARMA-processes has been extended to the analysis of stochastic volatility processes.

Please contact:
László Gerencsér, SZTAKI
Tel: +36 1 279 6138
E-mail: gerencser@sztaki.hu

Algebraic Convergence of Pulled Fronts

by Ute Ebert

Propagating fronts which separate regions of different behaviour are quite common in Nature. A particular situation arises if the state ahead of the front is linearly unstable, like in the spreading of an epidemic. Recently researchers at CWI and Leiden University isolated the universal part in the expression for the speed of such fronts. Their results bear on several nonlinear phenomena including gas discharges.

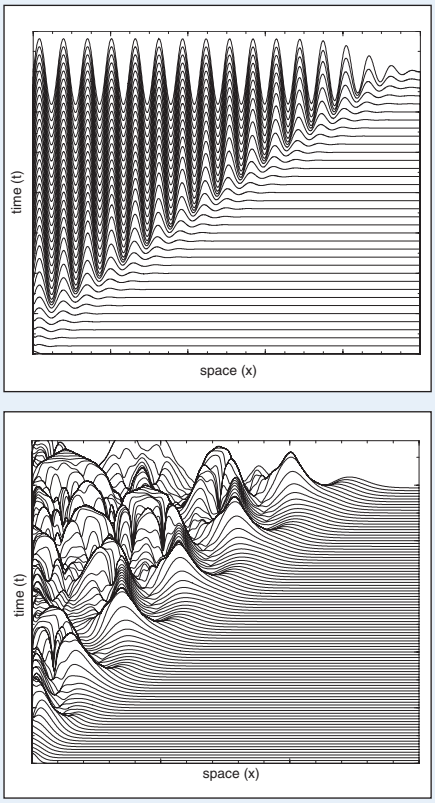
Examples of fronts penetrating into an unstable state range from the expansion of bacterial colonies to the branching of electric discharges. The mother of all equations describing such phenomena is a nonlinear reaction-diffusion partial differential equation called the Fisher-Kolmogorov equation, originally designed in the 1930s to describe epidemics:

$$\partial_t u = \partial_{x^2} u + u - u^2$$

The state $u = 0$ can be shown to be unstable, ie, a small disturbance can very quickly grow out to substantial size. The state characterized by $u = 1$ turns out to be stable. If the system was initially in the state $u = 0$, a wave front between these two states propagates with a certain, time-dependent velocity v, whose asymptotic expansion for large times is given by:

$$v(t) = a + b/t + c/t^{3/2} + O(1/t^2)$$

where a, b and c are numerical constants (equal to 2, -3/2, and $3\sqrt{\pi}/2$, respectively). Note that the leading correction terms for large, but finite times are algebraic, not exponential (as it would be for fronts propagating into a metastable rather than an unstable state). The universal term $\sim 1/t$ was derived by Bramson (1983) in a lengthy paper. Using asymptotic expansions and resummations, U. Ebert and W. van Saarloos recently determined the next order universal term $\sim 1/t^{3/2}$ (the time-dependent shape of the wave front). Moreover, they showed that the higher order terms $O(1/t^2)$ are not universal, but depend on details of the initial conditions. They also derived a simple universal expression for the convergence of the front profile. Amazingly, their method can also be applied to many other equations, including those with more complicated differential operators (eg, the Swift-Hohenberg equation), certain integro-differential equations, coupled sets of reaction-diffusion equations, and the equations describing an electric discharge. In all these cases the velocity $v(t)$ of the 'pulled' wave front is given by the above expression, only with different equation-specific numerical constants. The front is called a 'pulled' wave front because it seems that there is a leading edge pulling along the full front with all its complex system-dependent structure, with an almost universal speed. This is the result of a new 'center manifold theorem' for the expansion about the asymptotic solution: in the limit of large times, not the full equation is relevant, but only a restricted mathematical substructure that can be identified by a saddle-point expansion of the linearized equation. However, although the analysis of the linearized equation plays a central role in the calculation, the convergence law differs quantitatively from that of the linearized equation. Wave fronts with such an algebraically slow relaxation pose various problems to analytical perturbation theory and to numerical schemes with local refinement. Both are presently under investigation at CWI.



Two examples of fronts propagating into an unstable state. The dynamics are shown by plotting the fronts one above the other at successive moments. The front on the top gives rise to a periodic state, the one below creates a chaotic state. Nevertheless, the way in which these fronts 'mature' (ie approach their asymptotic speed and front shape) is given by one and the same formula.

Please contact:
Ute Ebert, CWI
Tel: +31 20 592 4206
E-mail: Ute.Ebert@cwi.nl
http://www.cwi.nl/~ebert/

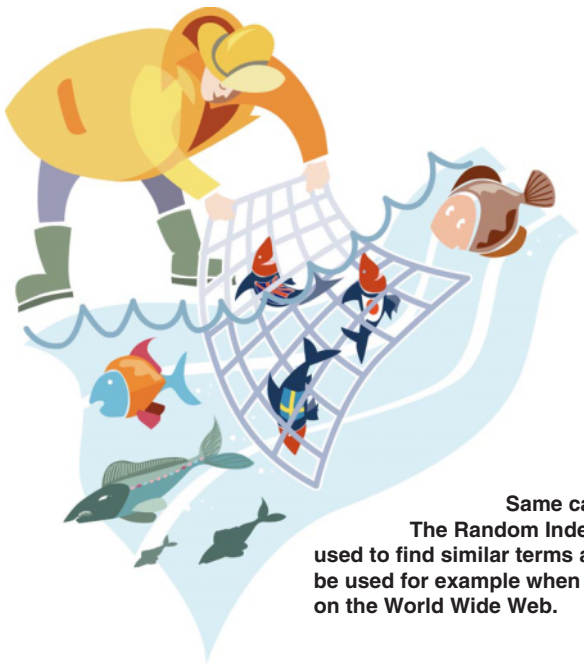
Random Indexing of Linguistic Units for Vector-Based Semantic Analysis

by Magnus Sahlgren

The Stochastic Pattern Computing project at SICS studied the mathematical foundations of human-like flexible information processing methods that compute with high-dimensional random vectors. The project ended in 2001 and led to the development of the Random Indexing technique for acquiring and representing semantic information about linguistic units.

Vector-space models have been used for over a decade to automatically acquire and represent semantic information about words, documents and other linguistic units. In short, the idea is to use co-occurrence information to construct a multi-dimensional semantic space in which the linguistic units are represented by vectors whose relative distances represent semantic similarity between the linguistic units. The space is constructed by collecting co-occurrence information in a words-by-contexts frequency matrix where each row represents a unique word and each column represents a context (typically a word or a document). The cells of the co-occurrence matrix records the (normalised) frequency of co-occurrence with the given context.

The rows (and the columns) of the frequency matrix can be interpreted as multi-dimensional context vectors where the elements are (normalised) frequency counts, and the dimensionality is the number of contexts in the text data. Thus, the representations are local. Now, the inherent problem with using local representations in natural language processing is that the size, or dimensionality, of the representations will grow with the size of the data. This means that the model will not scale very well, and that the co-occurrence matrix will soon become computationally intractable when the vocabulary and the document collection grows. To make the method practically feasible, it is necessary to reduce the dimensionality of the matrix.



Same catch, different languages. The Random Indexing technique can be used to find similar terms across languages. This can be used for example when searching for information on the World Wide Web.

However, dimension reduction techniques tend to be computationally very costly, which means that if efficiency is important, it may not be practicable to use such techniques. Furthermore, dimension reduction is a one-time operation, with a rigid result, which means that new data cannot be added to the model once a dimension reduction has been performed. As an alternative to vector-space models that use local co-occurrence matrices and some form of dimension reduction, we have studied the use of distributed representations that eliminates the need for separate dimension reduction of the co-occurrence matrix. The technique, which we call Random Indexing, accumulates a words-by-contexts co-occurrence matrix by incrementally adding together distributed representations in the form of high-dimensional (ie on the order of thousands) sparse random index vectors. The index vectors contain a small number of non-zero elements, which are either +1 or -1, with equal amounts of both. For example, if the index vectors have eight non-zero elements in, say, 1,800 dimensions, they have four +1s and four -1s.

The index vectors serve as indices or labels for words or documents, depending on which kind of co-occurrences we want to use. When using document-based co-occurrences, the documents are represented by high-dimensional sparse random index vectors, which are used to accumulate a words-by-contexts matrix by the following procedure: every time a given word occurs in a document, the document's index vector is added to the row for the word in the matrix. The procedure is similar when using word-based co-occurrences: first, we assign a high-dimensional sparse random index vector to each word type in the data. Then, every time a given word occurs in the data, the index vectors of the surrounding words are added to the row

for the focus word. Words are thus represented in the co-occurrence matrix by high-dimensional context vectors that contain traces of every context (word or document) that the word has co-occurred with (or in).

Note that the same procedure will produce a local frequency matrix if we use unary vectors of the same dimensionality as the number of documents (when using document-based co-occurrences) or the size of the vocabulary (when using word-based co-occurrences). These index vectors would have a single 1 marking the place of the context (word or document) in a list of all contexts - the nth bit of the index vector for the nth document or word would be 1. Mathematically, the unary local vectors are orthogonal, whereas the random index vectors are only nearly orthogonal. However, since there exist a much larger number of nearly orthogonal than truly orthogonal directions in a high-dimensional space, choosing random directions gets us sufficiently close to orthogonality to provide an approximation of the unary vectors. The amount of noise we introduce by choosing random directions is so small that it does not have any noticeable effect on the similarity relations between the entries, which means that the local frequency matrix and the Random Indexing matrix contain approximately the same information. By using, for example, 1,800-dimensional random index vectors with 8 non-zero elements, we may accumulate approximately the same co-occurrence information in a 50,000 by 2,000 matrix (assuming a vocabulary of 50,000 words in 30,000 documents) as we do in a 50,000 by 50,000 or 50,000 by 30,000 matrix using local representations for word-based vs. document-based co-occurrences.

By using the random index vectors to accumulate the co-occurrence matrix, we effectively perform a dimension reduction of the data, without the need for an explicit dimension reduction phase. This makes Random Indexing more efficient than techniques using computationally heavy dimension reduction techniques. It also makes the technique more flexible towards new data, since it may be immediately added to the model without the need to recompute the entire matrix - a new word only needs a new row in the co-occurrence matrix, and existing context vectors may simply be updated with the new information. Furthermore, the high-dimensional sparse random index vectors may be used to cover basically any size of the vocabulary, without the need to increase the dimensionality of the vectors, which makes the technique extremely scalable.

The Random Indexing approach thus presents an attractive alternative to vector-space models using local representations and computationally heavy dimension reduction techniques.

Please contact:
Magnus Sahlgren, SICS
Tel: +46 8 633 1604
E-mail: mange@sics.se
http://www.sics.se/~mange

Extending Thematic Lexical Resources by Term Categorization

by Alberto Lavelli, Bernardo Magnini, and Fabrizio Sebastiani

Researchers from IEI-CNR, Pisa, and ITC-irst, Trento, are currently working on the automated construction of specialized lexicons, as part of an ongoing collaboration in the fields of Machine Learning and Information Retrieval.

Increasing attention is being given to the generation of thematic lexicons (ie sets of specialized terms, pertaining to a given theme or discipline). Such lexicons are useful in a variety of tasks in the natural language processing and information access fields, including supporting information retrieval applications in the context of thematic, 'vertical' portals.

Unfortunately, the manual generation of thematic lexicons is expensive, since the intervention of lexicographers and domain experts working in collaboration is required. Furthermore, a manual approach does not allow for fast response to rapidly emerging needs.

We have developed a methodology for the automatic generation of thematic lexicons by 'term categorization', employing a combination of techniques from information retrieval (IR) and machine learning (ML). We view the generation of such lexicons as an iterative process of learning previously unknown associations between terms and themes. The process is iterative in that, for each item in a set of predefined themes, it generates a sequence of lexicons, bootstrapping from an initial lexicon given as input. Associations between terms and themes are learnt from a sequence of sets of generic documents ('corpora'). In this way, the lexicon can be enlarged as new corpora become available. At any given iteration, the process builds the lexicons for all the themes in parallel, and from the same corpus.

The method we propose is inspired by recent work in text categorization, the activity of automatically building programs capable of labelling natural language texts with zero, one, or several thematic categories from a predefined set. The construction of an automatic text classifier requires the availability of a corpus of preclassified documents. A general inductive process (called the learner) automatically builds a classifier for the categories of interest by learning their characteristics from a training set of documents.

While the purpose of text categorization is that of classifying documents represented as vectors in a space of terms, the purpose of term categorization is (dually) that of classifying terms repre-

sented as vectors in a space of documents. This means that, as input to the learning device and to the term classifiers that it will eventually build, we use 'bag of documents' representations for terms, dual to the 'bag of terms' representations commonly used in text categorization (and information retrieval). In our task, terms are thus items that may belong, and must thus be assigned, to (zero, one, or several) themes belonging to a predefined set. In other words, starting from a set of 'training' terms, a set of 'test' terms is classified, and the test terms which are deemed to belong to a theme are added to the corresponding lexicon; they can then be used as new training terms in the next iteration of the process.

The novelty of this 'supervised' approach is that there is basically no requirement on the corpora employed in the process. This differs from the classic unsupervised approach in which, in order to generate a thematic lexicon for a given topic, a corpus of documents labelled with respect to that topic is needed. This may be problematic, since labelled texts are often hard to obtain, and labelling them requires expert manpower.

As our learning device we adopt AdaBoost.MH(KR), a more efficient and effective variant of the well-known AdaBoost.MH algorithm, developed at IEI-CNR. Both algorithms are an implementation of boosting, a method for supervised learning which has proven one of the best performers in text categorization so far. Boosting is based on the idea of relying on the collective judgment of a committee of classifiers that are trained sequentially. When training the k-th classifier special emphasis is placed on the correct categorization of those training examples which have proven more difficult to classify (ie have been misclassified more frequently) for the previously trained classifiers.

We have chosen a boosting approach not only because of its state-of-the-art effectiveness, but also because it naturally allows for a form of 'data cleaning' if a lexicographer wants to inspect the classified terms for possible misclassifications. At each iteration, in addition to generating the new lexicon, the algorithm ranks the training terms according to their 'difficulty', ie how successful the classifiers generated in the process have been at correctly recognizing their label. Since the highest ranked terms are the ones with the highest probability of having been misclassified in the previous iteration, lexicographers can remove the misclassified examples by scanning the list downwards, stopping when they want. The process of generating a thematic lexicon then becomes an iteration of generate-and-test steps.

We are currently experimenting this methodology using a benchmark called WordNetDomains[42], an extension of WordNet in which each word has been labelled with one or more from a set of 42 themes (such as eg Zoology, Medicine, Sport) commonly used in dictionaries. We have randomly partitioned each thematic lexicon (ie the set of words labelled by one of the 42 labels) into a training and a test set, so that the purpose of the experiment is to check the ability of the algorithm to extract the terms of the test set from the corpus. As our evaluation measure, we are using F1, a combination of precision and recall, commonly adopted in text categorization. As our 'learning' texts, we are using a sequence of subsets of the Reuters Corpus Volume I. Our preliminary experiments have shown this approach to outperform previous approaches to thematic lexicon expansion based on unsupervised methods.

Links:
http://faure.iei.pi.cnr.it/~fabrizio/Publications/LREC02.pdf

Please contact:
Fabrizio Sebastiani, IEI-CNR
Tel: +39 050 3152 892
E-mail: fabrizio@iei.pi.cnr.it

An End-to-End Software Suite for 3D Scanning

by Marco Callieri, Paolo Cignoni and Paolo Pingi

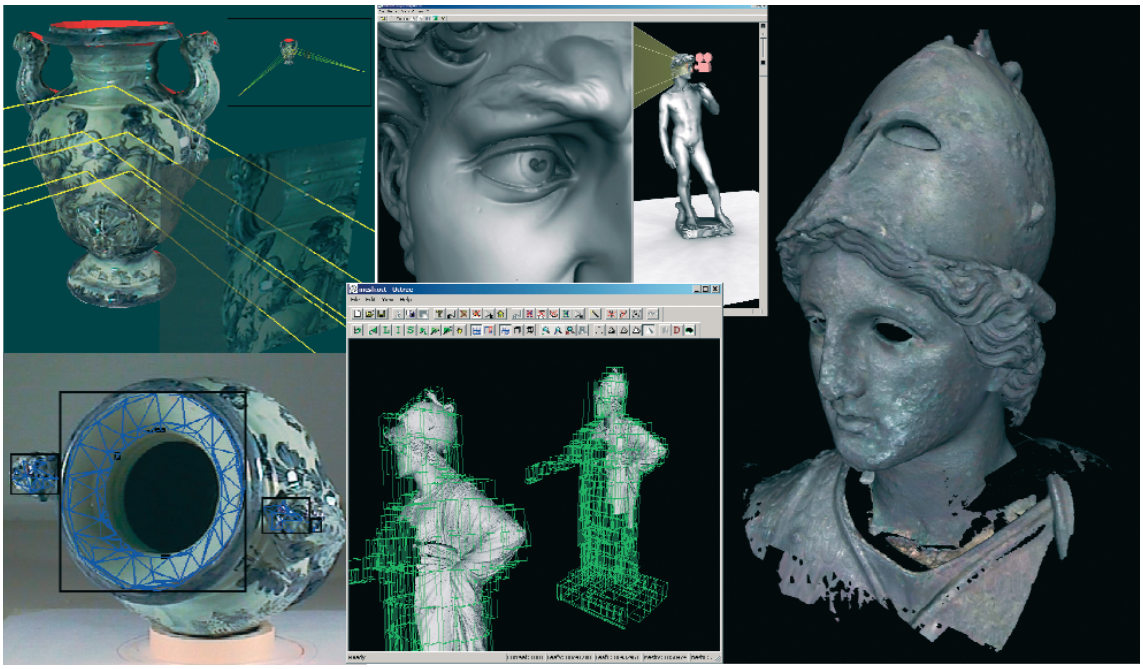
The Visual Computing Group of IEI-CNR, Pisa, has developed a suite of software procedures for the processing of 3D scanning data. The software package can be used for the automatic acquisition of 3D Cultural Heritage and for the creation of realistic 3D digital models.

3D scanning is a rapidly developing activity: every year a number of new, more efficient and cheaper scanner devices become available. Huge 3D datasets — so large that they must be processed before they can be efficiently used — are thus now easily produced. The acquisition phase is just the beginning of the whole scanning pipeline.

Our research group has been working in the field of 3D scanning for the last five years. One of our first studies resulted in the design and implementation of a cheap 3D scanner based on structured light (see our Web site); recent efforts have been focused on the development of software able to work efficiently with the huge amount of data produced by 3D scanners.

We have developed a standard software package. The scanning pipeline can be summarized in five steps:

Acquisition: the first step is to capture geometric data with the scanner hardware. Depending on the type of the scanner employed, the execution of this phase can vary considerably. Whatever the approach used, eventually we have a series of partial scans, called range maps, each of them covering a part of the object. This is because a number of scans from different positions are needed to entirely capture the geometry of an object. We wrote a program that tailored



Part of the Graphical User Interface of the scanning tools and the results.

the functioning of our current (commercial) 3D scanner to our needs: the program manages a rotary stage and a lighting system.

Alignment: range maps taken from different viewpoints have to be aligned. This process can be completely automatic if the exact position of the scanner during each acquisition is known. Otherwise, manual intervention is needed to obtain the initial placement, then refined automatically. This manual intervention is the most time-consuming phase of the whole process. In our package, a graphic program is used to obtain the first user-driven placement of the range maps. Then an automatic iterative process, based on the ICP algorithm, determines the best alignment .

Fusion: once aligned, the partial scans are merged into a single 3D model. A careful use of resources is important in this stage since the data dimensions can easily exceed the capacity of a standard PC. A variation of the Marching Cubes algorithm which uses less memory and a hierarchical approach has been adopted to compute precise and hole-filling fusions.

Decimation: 3D scanners provide a huge amount of data. For an effective use of the model we have to reduce the size of the geometric information acquired, especially with respect to the less significant parts of the object (a sample rate of 0.25 mm can be useful for very complex sections, but not on flat surfaces). Our decimation software is based on edge collapsing and error driven simplification. For very complex 3D models (more than 6 million faces), an new algorithm that works with secondary memory has been designed to overcome memory shortage. In this way, we were able to build the official simplified models of the David (56M faces) and the San Matteo (300M faces) produced by the Digital Michelangelo Project.

Texturing: for certain applications (eg reverse engineering) information on colour is not needed. On the contrary, for a wide class of objects, additional information about the real appearance of the object must be provided. This is mostly done by texturing the final model, using photos taken during acquisition. Our texturing procedure is performed in two steps: the photos are first aligned to the geometry (manually or automatically) and then mapped to the model. Alignment of the images and production of the textures correspond to two distinct procedures in our package.

In addition to the above package, we have developed a number of converters from/to the most used 3D file formats and some utilities to clean unnecessary data, fill holes, smooth data, geometry and topology correction.

It should be noted that the processing of 3D data does not depend on the 3D scanner used. Different scanning technologies obviously produce different results in terms of precision, size of captured object, types of data acquired, but the required processing pipeline remains unchanged.

We obtained good results working on data coming from different 3D devices: triangulation laser scanners, structured light scanners and time-of-flight scanners. Our software package is currently used for the automatic acquisition of 3D small to medium size objects (vases, statues, jewels) within several national and international projects for the preservation of cultural heritage. The 3D digital models obtained can be used in many ways, ranging from the cataloguing of 3D cultural objects to the construction of virtual museums, from the planning of the restoration process to the production of replicas.

Links
http://vcg.iei.pi.cnr.it/

Please Contact:
Paolo Cignoni, IEI-CNR
Tel: +39 050 315 2927
E-mail cignoni@iei.pi.cnr.it

Exploring the Frontiers of Computability

by Jiří Wiedermann and Jan van Leeuwen

Modern networked computers challenge the assumption of classical computability theory: they are always on, interact with each other and their environment, and their functionality can change with time. At the Institute of Computer Science of the Academy of Sciences of the Czech Republic (a member institute of CRCIM) in Prague and the Institute of Information and Computing Sciences at Utrecht University, extensions of computability theory are developed that capture the current ways of information processing in computer networks.

In 1936 A. M. Turing invented a universal model of computation that has served as the basis of computability theory ever since. The principles of operation of a Turing machine capture the way computational tasks are performed by humans as well as by digital computers. The presumed generality of the model is expressed in the so-called Church-Turing Thesis, stating that whatever is computable by an algorithm, in a reasonable sense of the word, is computable by the Turing machine. Note that this thesis cannot be proved as a mathematical theorem, since it does not specify what is a reasonable sense of the word 'computable' nor what an algorithm is.

Computations challenging the Church Turing thesis have been studied since Turing's times. As a rule, the respective computational models were hardly 'reasonable' ones: they made use of (physically infeasible) in finite precision real number arithmetic, used auxiliary information from non-computable 'oracles' and so on. Nevertheless, as new insights into the nature of computations from physics (cf. quantum computing), biology and contemporary information processing technologies emerge, we are witnessing an upsurge in exploring the limits of computation. More than half a century after its invention, the presumed generality of the Turing machine is at stake.

Modern networked computers challenge the assumptions of the classical computability theory in several ways. For instance, they operate practically uninterruptedly since the moment of their installation. They obtain their inputs via many different channels at unpredictable times, deliver responses continuously, accumulate information over the course of their entire existence and use it across the boundaries of separate 'runs'. Also, parts of their underlying hardware and software are upgraded whenever an 'external agent' decides to do so, without loss of vital data. The most prominent example is the Internet. It can be seen as a wide-area computing infrastructure, a kind of global computer, evolving unpredictably, with unpredictable computing characteristics.

Compared to the classical computing scenario, the changes in modern computing technology can be subsumed under three complementary issues: interactivity, non-uniform evolution (or adaptivity), and non-stop operation 'ad infinitum'. The systems performing according to these three qualities together constitute what is meant by evolving interactive computing.

Many people, especially in the software engineering community observed that these systems do not not fit the classical Turing machine paradigm. Surprisingly, evolving interactive computing not only arises in current, highly networked computer systems but also in the the information processing in (societies of) living organisms. At the Institutes of Computer Science in Prague (part of CRCIM) and Utrecht, a mathematical theory of computation by 'lineages' of machines is being developed to capture the principles of evolving interactive computing. Lineages of machines operate on unbounded (infinite) streams of inputs and have attractive computational properties. Lineages can be shown to be equivalent to 'interactive' Turing machines that are extended by an oracle-type mechanism known as advice in the context of computational complexity theory. As these machines are provably computationally more powerful than classical Turing machines, even if the latter are extended so as to work on infinite input streams, it has formally exhibited that evolving interactive computing goes beyond the realm of Church-Turing computability. Presently the theory of lineages of machines is being extended so as to cover further phenomena and complexity-theoretic properties.

The quest for computing beyond Turing machine limits has also gained interest in theoretical physics. It has been recently shown that general relativity theory permits the idea of observing the infinity of certain discrete processes in finite physical time. The idea is based on speculations about effects of strong gravitational fields on time and has lead to the concept of relativistic computing. From the viewpoint of computability theory, relativistic computers can be seen as Turing machines with a suitable oracle, but the fact that these computers are based on apparently realistic 'thought experiments' gives them a special status. The consequences for the foundations of computing have begun to occupy philosophers, physicists, and recently also mathematicians and computer scientists.

The study of the computational power of relativistic computations has given rise to an interesting instance of the use of oracle calls as a resource, and several results have been proved for it by the authors. One must bear in mind that, unlike the evolving interactive systems that do exist, the existence of relativistic computing models has not been experimentally verified.

As a result, it appears that classical Turing machine computing not only fails to determine the limits of what is computable, it even fails to capture the computational power of current algorithmic systems. The quest for a better, more efficient, more powerful, and more general realistic model of computing is a highly interesting challenge. No matter how the resulting theory would look like, it is clear that what computers can or cannot compute, and how efficiently they will do what they can, is not only determined by pure mathematics, but also by the laws of physics.

The research reported here was partially supported by GACRgrant No. 201/02/1456 and by EC Contract IST-1999-14186 (Project ALCOM-FT).

Links:
See the recent technical reports by the authors at http://www.cs.cas.cz/ and http://www.cs.uu.nl/

Please contact:
Jiří Wiedermann, CRCIM
Tel: +4202 6605 3520
E-mail: jiri.wiedermann@cs.cas.cz

Towards the Design of Functional Materials and Drugs 'from Scratch'

by Wanda Andreoni and Alessandro Curioni

High Performance Computing is a promising future technology for the design of novel materials including pharmaceuticals. Sustained innovation in computational methodologies is a key requisite for its success. Work in this direction is actively performed at the IBM Zurich Research Laboratory in Rüschlikon.

Our daily life is dominated by innovations in materials and in drugs. One need only think of how the replacement of steel with aluminum alloys has made cars and airplanes lighter and faster, how our eyeglasses have become lighter and thinner thanks to the introduction of titanium for the lenses and titanium-rich metal alloys for the frames, how much brighter and finer the screens for our laptop displays are getting day after day as well as how much progress medicine has made over the years, thanks to the continuous development of new drugs. Over time the demands for new materials and drugs have changed, becoming much more specific and detailed. Replacing one material in a multicomponent device implies the optimization of many requirements under complex boundary conditions; for example, a novel material that may one day replace silicon dioxide as gate dielectric in CMOS technology not only has to provide higher permittivity but also has to preserve sufficiently high electron mobility and be integrable in current processing technology. Drug design has always been known to be extremely complex, but lately it has become clear that real progress will only be possible through personalized medicine. Given that the need is increasingly more sophisticated, one must rely on increasingly powerful and sophisticated design methodologies and on their sustained innovation.

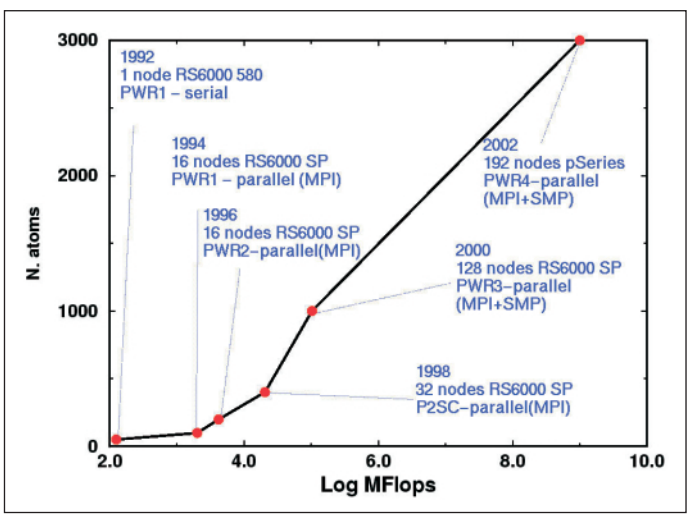


Figure 1: Size of the computationally treatable systems versus hardware speed.

Computers nowadays have the power to store several terabytes of data and to process them at the speed of 10-20 teraflops. As this power becomes more widely available, the door opens for the creation of powerful simulation techniques. Their development is being pursued by several groups in well-known research and industrial institutions worldwide. The ultimate goal of this global effort is that of creating a virtual laboratory in which new compounds are synthesized, screened and optimized through the direct observation of their evolution in time in response to the change of physical conditions (eg, temperature and pressure) and/or to the action of chemical agents.

At the IBM Zurich Research Laboratory, the Computational Biochemistry and Materials Science Group is active in the development of accurate and efficient methodologies for the simulation of diverse materials, ranging from silicon to enzymes, at the atomistic level. The specific purpose of this work is that of obtaining the above-mentioned lead optimization 'from scratch', namely, guided only by the knowledge of the fundamental parameters of the atomic components, and from the direct computation of the intra- and intermolecular interactions based on the laws of

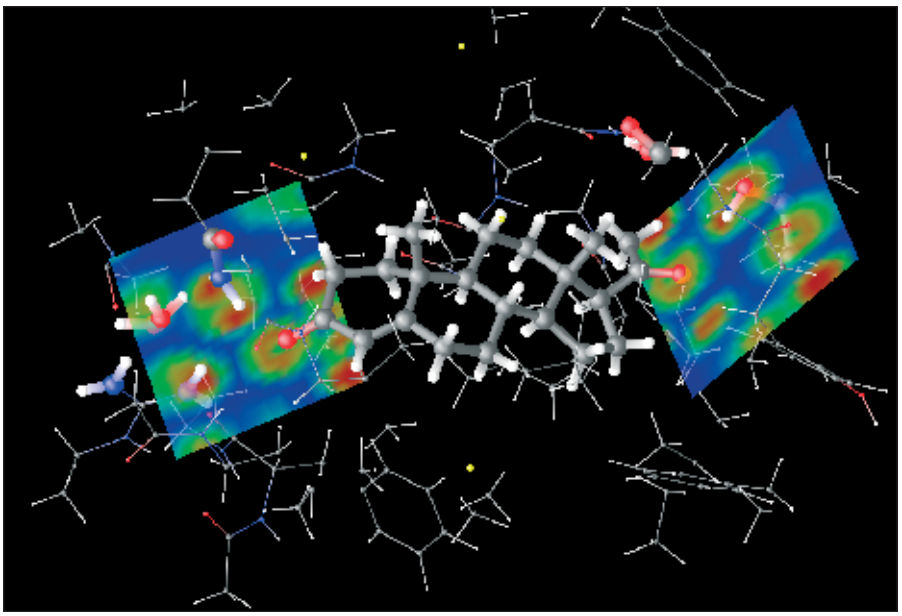


Figure 2: Computer simulations have unraveled how progesterone (the 'hormone of pregnancy') binds to its receptor, thus solving a long-standing puzzle. Clouds represent how electrons distribute in the bonding regions.

quantum mechanics. A fully quantum-mechanical approach to a given compound has long been hindered by the limited size (number of atoms) of the model. The required analysis is becoming feasible through modern computing power; the upper limit to the size of this model is bound to increase over the years with the increase of the power of the hardware and the simultaneous progress in the algorithms. As an example, Figure 1 shows how the size of the systems one can simulate from first-principles has changed in the past decade with the increase in computer power. The leap forward in 1994 was due to the advent of parallel computers and the use of a highly parallelized code.

The parallelization of such a code is a non-trivial task since it is composed of several parts that scale non-homogeneously with the system size and the dominant parts (both in terms of memory use and computational time) which are explicitly dependent on system size, ie, elements, that dominate the computation at small/intermediate sizes, become irrelevant at larger sizes. In particular, there are two different classes of algorithms in the code: 1) Fast Fourier Transform related algorithms (which scale as $V^2 \log V$, where V is the 'volume' of the system) and 2) Linear algebra related algorithms (such as orthogonalization, that scale as V^3). At small/intermediate system sizes, type 1 algorithms are dominant and these were parallelized in our laboratory in 1994 using an MPI (Message Passing Interface) implementation on distributed memory computers. Type 2 algorithms, however, only become important for molecular systems containing more than 1000 atoms and were parallelized with OpenMP (Open Message Passing) on mixed shared/distributed memory computers in 2000. These two steps were instrumental in extending the application realm of our simulation techniques. For an extended description of the method see W.Andreoni and A.Curioni, 'New Advances in Chemistry and Material Science with CPMD and Parallel Computing', Parallel Computing 26 (2000) 819.

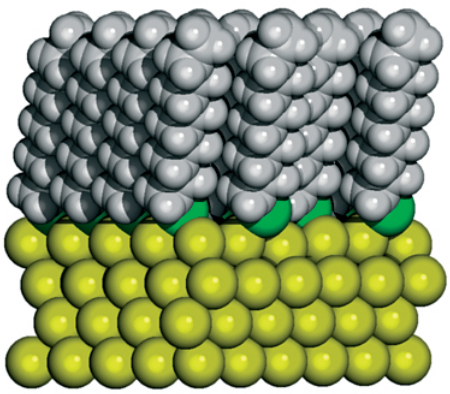


Figure 3: Computer simulations have determined the morphology of self-assembled monolayers of alkylthiolates on gold.

Currently our simulations of large systems (thousands of atoms) rely mostly on a kind of hybrid modelling in which that part of the compound where 'the action takes place' is treated at the quantum level and the rest ('the environment') at the classical level, namely with interactions modeled using semi-empirical parameters. The former can go up to sizes of ~1000 atoms, but calculations would not achieve the required degree of accuracy if the effects of the interactions with its environment were neglected. Examples of biochemical processes we have studied in this way are enzymatic reactions in which the chemical event activated by the presence of the enzyme is relatively confined, and the binding of ligands to protein (receptors) in aqueous solution (Figure 2), which is related to the design of optimal drugs. Examples of advanced materials for technological use are the organic components of light-emitting devices (LED), namely amorphous molecular compounds interacting with the electrodes, and functionalized nanostructures of self-assembled monolayers of organic chains on metal substrates (Figure 3) that are used in diverse technological applications such as the fabrication of sensors and transducers and as patternable materials.

The success of High Perfomance Computing-driven design clearly depends on the steady increase of computational capabilities. One usually refers to Moore's law, which predicts that transistor density doubles every 18 months. It is no secret, however, that IT industries are currently struggling to maintain this pattern because of the intrinsic physical limits of the technology. As history teaches us, crises are often overcome by the advent of new concepts. A most promising one is that of GRID computing.

Link:
http://www.zurich.ibm.com/st/compmat/

Please contact:
Wanda Andreoni
IBM Zurich Research Laboratory,Switzerland
E-mail: and@zurich.ibm.com

Flower – Framework for Local Wireless Services

by Tero Hakkarainen, Ali Lattunen and Vespe Savikko

Local wireless services enable the distribution of location-dependent information using wireless short-range communication methods. At VTT Information Technology, a novel framework solution for local wireless services has been developed.

Local wireless services provide content that is accessible only at the immediate vicinity of the service access point. Local services are accessed normally either with mobile phone or palmtop computer equipped with appropriate content, eg, WAP (Wireless Application Protocol), browser and suitable short-range communication method like Bluetooth or WLAN (Wireless Local Area Network). The limited service area distinguishes local services from cellular network, eg, GSM, based services that are accessible in fairly broad geographical area. Proximity of the user enables local services to provide content that is both location-dependent and meaningful for the user only at that precise location. This fact and charge-free local communication bearer services make local service concept ideal for providing services like local information kiosks, guidance, m-payment and ticketing services, games, entertainment, advertising and remote control solutions, as illustrated in Figure 1. Effort to build local wireless services is substantially decreased by the novel framework called Flower (Framework for Local Wireless Services) which provides an easy-to-use programming interface for the developers of local service applications.

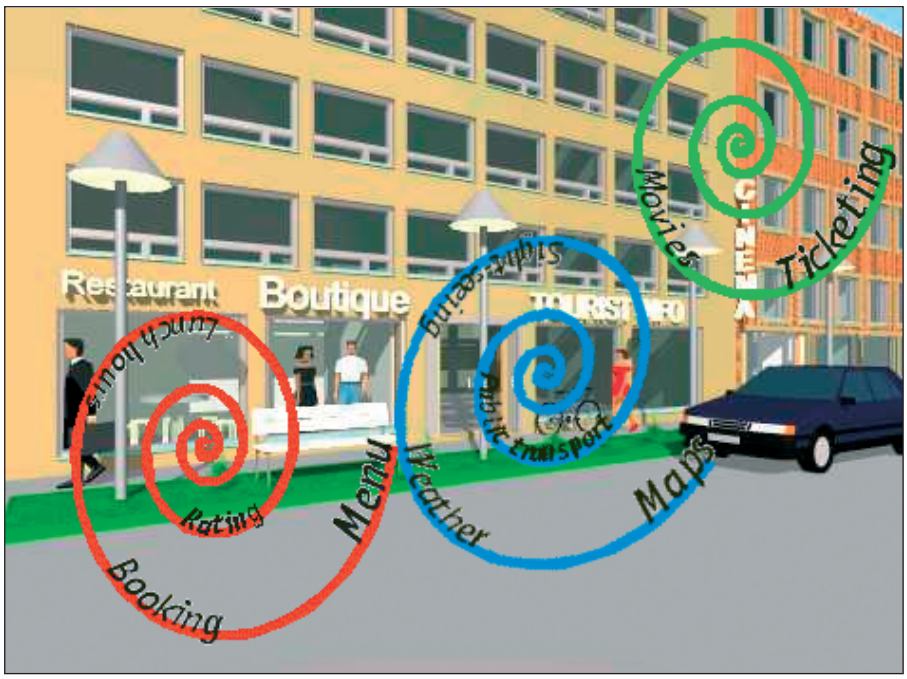


Figure 1: Local wireless services in urban environment.

Framework Architecture

The Flower framework hosts local wireless service applications capable of responding to requests and events in the service environment. As depicted in Figure 2, the content generated by an application is conveyed through a local area network to the access point that hosts the terminal that requested the service. Access points transfer content to terminals' browsers using short-range wireless communication methods.

The Flower server executes the local applications and offers them a view to the physical service environment. Applications generate content typically in either HTML (HyperText Markup Language) or WML (Wireless Markup Language) format. HTTP (Hypertext Transfer Protocol) protocol is used for transferring content data, eg, HTML, from Apache Tomcat Web Server to HTTP enabled terminals. For WAP terminals, WAP protocol must be used instead. The essential part of the latter communication route is the WAP Gateway, which transforms content data, eg, WML, to WAP protocol messages.

The Flower system may contain an arbitrary number of independent access points. Access points may autonomously choose the wireless short-range communication method they use. It would therefore be fully possible to have eg, Bluetooth and WLAN access points simultaneously in a Flower system. The primary function of access points is to relay protocol data in either WAP or HTTP format to the terminals. Each terminal belongs to only one access point at a time. Additionally, access points monitor the terminals in their service area. A new or a lost terminal is signaled to a special server in the framework, the Manager, which is a database containing the current state of the whole system. A change detected by some access point causes an event that can be caught and handled by the local applications.

Creating local Wireless Services

The use of the WAP Gateway makes it possible to use established web technologies in local wireless service development. Flower development framework is implemented with Java servlets. This solution offers the following benefits:

- servlet developers can easily migrate to local wireless applications domain
- availability of free servlet engines (eg, Tomcat)
- Java Native Interface (JNI) is used to map the Flower API (Application Programming Interface) to C++.

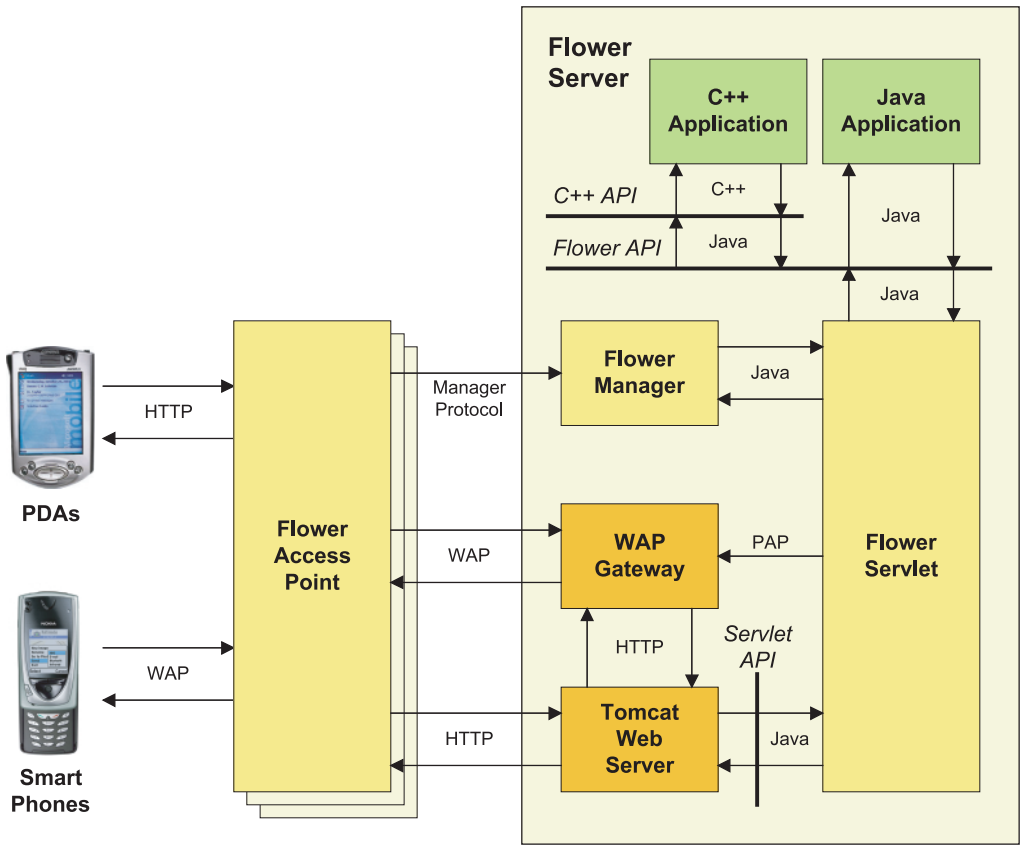


Figure 2:
Flower framework architecture.

The Flower API is actually a layer encapsulating the Servlet API, providing Flower specific alternatives to HTTP specific classes (eg, Request 'HttpRequest, Session' HttpSession) and hiding those parts of the Servlet API that do not have a Flower counterpart (eg, cookies). As the Flower API reflects the concepts of the Servlet API it feels familiar to servlet developers, thus making them productive after short orientation. Additional, Flower specific, features are seamlessly integrated into the Flower API, encouraging a developer to create genuine local wireless applications with them:

- Interface to the Manager provides the current state of the system (access points and terminals connected to them).
- Due to the dynamic nature of the system, where a terminal can leave the service area of its access point at any time, a simple acknowledgement mechanism is devised. Access point signals through the Manager to the originating Flower application whether the message could be sent to the target terminal.
- WAP push support enables a Flower application to approach terminals instead of passively waiting for requests. The actual push mechanism (Push Access Protocol) is implemented by the WAP Gateway.

Conclusions

The Flower framework supports currently WAP services over Bluetooth communication. Since actual WAP-over-Bluetooth terminals are not yet available, laptops with Bluetooth adapters have been used as substitutes for smart phones. In the future, the system development will be targeted to completing the HTTP protocol support and adding support for new types of terminals and protocols. With the experience and knowledge gained from the pilot applications already running over Flower it is likely that the forthcoming technical improvements will make the framework even more attractive for research purposes as well as for real-life applications.

Links:
http://www.vtt.fi/tte/samba/projects/flower/
http://www.vtt.fi/tte/projects/WAP/

Please contact:
Tero Hakkarainen, VTT
Tel: +358 3 316 3355
E-mail: Tero.J.Hakkarainen@vtt.fi

A Simple Model for Spark Branching

by Ute Ebert

Researchers at CWI in Amsterdam have shown that branching of sparks is a direct consequence of the simplest mathematical model describing such electric discharges. This finding is against previous expectations.

The work by Manuel Arrayás (now University of Madrid), Ute Ebert and Willem Hundsdorfer was published in April in the Physical Review Letters. It applies to various natural phenomena, from small-scale sparks on the ground up to the recently observed upward lightning: discharges from a thunder cloud up to the far higher ionosphere occurring on a scale much larger than ordinary cloud-to-earth lightning, and exhibiting a much stronger branching pattern. The study of spark branching is also relevant for combustion gas cleaning or industrial ozone production.

If a strong electric field is suddenly applied to a non-conducting medium like a gas, narrow channels of ionized matter may be formed. Through such a conducting channel the charge surplus at one end (cloud) 'discharges' to the other end (earth). The conducting regions emerge because in a sufficiently strong electric field free electrons can ionize gas molecules by impact. One single electron may thus induce a chain reaction. In the field, positive and negative charges then drift into different directions. Hence, electrically charged regions form in particular at the boundaries of the ionized region. These charges modify the externally applied field and therefore the ionization rates and drift velocities, making the problem non-linear. In general, non-linear dynamics leads to spontaneous formation of patterns in space and time. In this particular case, these patterns take the form of narrow channels: in the conducting interior the field is suppressed, while at the tip of the growing channel it is largely enhanced, leading to high local ionization rates and rapid particle drift.

The mathematical model studied at CWI describes the generation and motion of free electrons and ions in a gas under the influence of an external electric field in its simplest form (minimal discharge model). Since the degree of ionization stays low (less than 10^{-5}) we can neglect gas convection and thermal effects. The rate of change in the electron (ion) density then equals the divergence of the local electron (ion) current density, plus a source term representing the generation rate of electron-ion pairs due to the impact of field-accelerated electrons onto neutral gas molecules. This generation rate is proportional to the electron density, but also depends on the local field strength in a non-linear way. It becomes substantial if both the electron density is non-vanishing and the local electric field is sufficiently strong. Since the motion of the free electrons and ions is damped by collisions with the gas atoms, the current is composed of a drift term proportional to the field and a diffusive term.

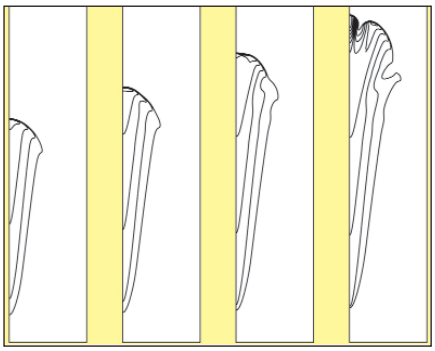


Figure 2: Four snapshots of the evolution of a streamer, resulting from a numerical simulation on the basis of the minimal discharge model studied at CWI. The streamer tip is seen to become unstable and branch. (A. Rocco, U. Ebert, W. Hundsdorfer, CWI.)

This simple but realistic model can exhibit already spontaneous spark branching, in contrast to previous expectations. This is revealed by precise numerical simulations at CWI, combined with a new analytical insight: the instability that leads to branching, is of Laplacian type. This puts spark branching into the same class of phenomena as viscous fingering, dendritic solidification or coral growth. The theoretical predictions are presently being investigated in experiments at the Technische Universiteit Eindhoven, where now discharge phenomena on time scales shorter than a nanosecond can be observed.

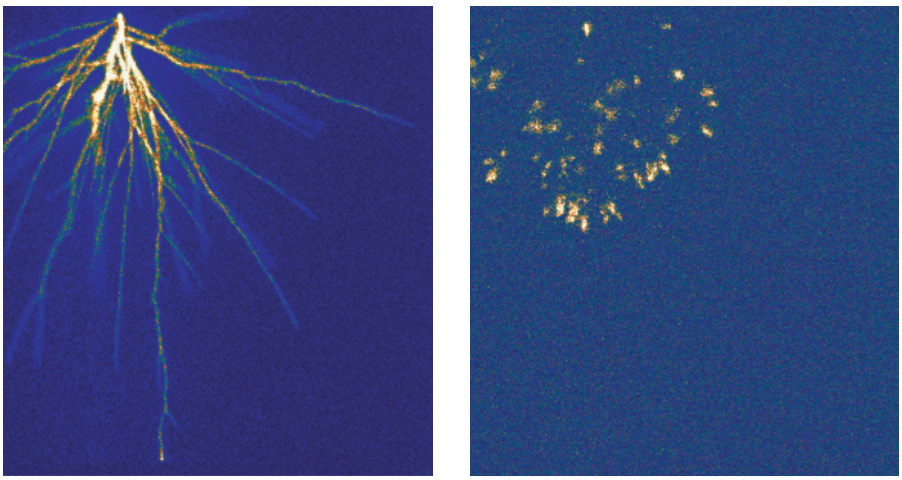


Figure 1: Two views of a streamer (the phase immediately preceding a discharge). To the left a time integrated picture (exposure time 10^{-6} seconds); to the right a snapshot of the same event (exposure time 10^{-9} seconds), showing the propagating heads of the streamer channels. (Courtesy E. van Veldhuizen, Eindhoven University of Technology.)

Please contact:
Ute Ebert, CWI
Tel: +31 20 592 4206
E-mail: Ute.Ebert@cwi.nl
http://www.cwi.nl/~ebert/

Dyade — A Successful Cooperation between Bull and INRIA

by Patrick Valduriez

Dyade (1986-2001) was a strategic R&D joint venture between Bull and INRIA in information and communication technology. Innovative technologies produced by Dyade were transferred to Bull product divisions or released as open source software. Dyade also gave birth to four start-up companies: TrustedLogic, KelKoo, Jalios and Scalagent.

Recognizing the new dimension of Information (ICT) and its major impact on economical and social growth, Bull and INRIA created Dyade in march 1996 as a five-year Economic Interest Group with about 50 people. Dyade's mission was to identify and develop innovative ICT with strong market potential, to demonstrate the industrial strength of the developed technologies with pilot users, and to transfer the results to either Bull or start-ups from Bull and INRIA. The incentive for Bull was to increase its technology portfolio and gain access to a fast-growing, high-potential market. The motive for INRIA was to transfer research results to industry and identify research directions that are strategic for industry.

Dyade engaged in a number of R&D actions, each sponsored by a Bull division and focused on an ICT domain where INRIA could bring its researchers' expertise. An action would then associate INRIA researchers and Bull engineers at a single location (an INRIA site) to work together on a specific topic with the objective of demonstrating the industrial interest within one or two years. Actions were conducted in the following domains:

- formal methods, based on either mathematical logic or model checking, applied to security (eg certification of security protocols and architectures) or the validation of complex computer architectures such as multiprocessor or smartcard
- multimedia technologies applied to graphic interface development or intelligent video-surveillance
- network technologies and services including the implementation of IPv6 and ATM protocols, video-conferencing services over TCP/IP and network management services
- distributed information management technologies including agent-based technologies for building and deploying distributed applications, mediator technologies for discovering and accessing distributed data sources, and content-management technologies for sharing information on the Web.

Transfer to Bull's Product Divisions

Formal methods were successfully used to test and validate multiprocessor and smartcard products. In particular, the Java Virtual Machine in Bull's smartcard Odyssey was implemented by Dyade. Agent-based technologies were used to improve the Netwall product. Graphic interface development technologies and network management technologies were integrated in the OpenMaster platform. Dyade's IPv6 implementation was integrated in the IBM's AIX operating system by Bull and IBM.

Open source software

The main technologies released were JORAM, an implementation of the Java Messaging Service on top of the agent middleware; RMI-JDBC, a Java gateway allowing transparent remote database access; Koala-Graphics, a graphics layer over Java AWT library; and JSMAN, a number of Java components for network management. The success of these technologies led to the creation of the ObjectWeb consortium in February 2002 (see ERCIM News no. 48) whose objective is to federate the development of an OSS object infrastructure.

Creation of start-ups by Dyade members. TrustedLogic, created in 1999, specializes in security technologies and services based on formal methods. KelKoo, created in 2000, uses Dyade's mediator technology for comparing Web merchant sites. It is now the number one Internet buying guide in Europe. Jalios, created in 2001, develops and sells Web content-management solutions tailored to companies' business needs. Scalagent, created in 2002, specializes in developing mediation solutions for B2B applications using agent-based technologies.

In parallel to technology transfer, Dyade produced excellent research results: 15 PhD theses and over 80 research papers which were published in the best specialized journals and conferences. Within Dyade, researchers could take advantage of a real experimental environment for proof-of-concept. For instance, the formal validation of cache coherency protocols for large multiprocessors helped finding new solutions for scaling up model-checking tools to very high numbers of states.

Retrospection

In retrospect, Dyade has been quite successful in terms of technology transfer and research results. The first reason is obviously the high quality and motivation of the INRIA researchers and Bull engineers who made Dyade real. But there are two other important reasons. One is the major commitment of Bull and INRIA's management which defined the original vision and allocated significant resources to the actions. The other reason was an efficient, lightweight management structure with independent actions, each sponsored by a Bull product division having strong interest in getting back results.

Links:
http://www.dyade.fr/
http://www.trustedlogic.com/
http://www.kelkoo.com/
http://www.jalios.com/
http://www.scalagent.com/
http://www.objectweb.org/

Please contact:
Patrick Valduriez, INRIA
E-mail: Patrick.Valduriez@inria.fr

Digital Library Competence Center

by Maria Bruna Baldacci and Stefania Biagioni

IEI-CNR, Pisa, Italy is operating a Digital Library Competence Centre. The Centre provides specific user communities with tutorials, demonstrations and case-studies of digital library technologies, services and expertise. It also offers hands-on experience with advanced digital library testbeds.

Digital library technologies will dominate the Internet of the 21st century. As the global information infrastructure becomes an essential part of everyday life, digital libraries will become a necessary component of this infrastructure. They will become one of the major tools that provide European citizens with access to scientific knowledge and to information on their cultural heritage. Independently of the extent to which librarians, archivists and museum curators are happy with this vision of the future, they are now demanding knowledge of digital library technologies and of the new type of services that digital libraries can deliver.

To satisfy this demand, a Digital Library Competence Center has been set up in Pisa, Italy, coordinated by IEI-CNR, under a contract from the EC IST Programme (IST-2001-32587). The Institute has a well-consolidated experience in the research and development of various types of advanced digital library systems. The objective of the Centre is to provide specific user communities, eg librarians, archivists, scholars and technicians, with access to advanced digital library services, expertise and training that will allow them to understand and use these new technologies.

The following courses have been scheduled for April 2002 – November 2003. Each course is supported by a specific DL prototype, mainly developed at the Institute:

- **Self-publishing/Scholarly Publication:** The aim is to facilitate transition from the current centralised, discrete publishing model, to a distributed, continuous, self-publishing model. Test-bed: ETRDL – The ERCIM Technical Reference Digital Library

- **Digital Library Based Collaboration:** The objective is to demonstrate new ways in which digital libraries, ie, collections of documents and associated services, can facilitate collaboration within and between user groups.
  Test-bed: SCHOLNET — A Digital Library Testbed to support Networked Scholarly Communities

- **Open Access to Digital Libraries:** The Open Archives philosophy is explained; the aim is to facilitate the implementation of infrastructures that provide open access to digital libraries. This environment is built on top of the Open Archive Initiative which develops and promotes standards for interoperability that aim to encourage the efficient dissemination of content.
  Test bed: CYCLADES — an open collaborative virtual archive service environment

- **Permanent Access to On-line Journals:** The aim is to enhance the confidence of the librarian community in the Web as a more effective medium for scientific and technical communication than material published on paper. A prototype system preserving access to scientific journals published on the Web will be demonstrated. This system has been designed and implemented at Stanford University Libraries with funding from the National Science Foundation and Sun Microsystems, Inc.
  Test-bed: LOCKSS — A Permanent Web Publishing and Access System

- **Indexing and Searching Audio-visual Material:** Archivists and other interested user communities will be trained in the (semi)-automatic indexing and searching of audiovisual material. The aim is to improve the productivity and cost effectiveness of producing and using digital audiovisual archives. An advanced metadata editing environment integrated within a film digital library will be demonstrated. Test bed: ECHO — European Chronicles Online

The courses are free of charge. Registration is requested and no more than ten participants are accepted per course.

Link:
http://dlibcenter.iei.pi.cnr.it/

Please contact:
Maria Bruna Baldacci or Stefania Biagioni, IEI-CNR
Tel: +39 050 315 2899/2900
E-mail: baldaccilbiagioni@iei.pi.cnr.it

EU/NSF Digital Library All Projects Workshop

by Costantino Thanos

A concertation meeting among the digital library projects funded by the European Union and the National Science Foundation, also including invited projects of other relevant digital library initiatives in Europe, took place on 25-26 March 2002 in Rome, Italy.

The meeting was organized by the DELOS Network of Excellence, in cooperation with the Cultural Heritage Applications Unit of the EC Information Society Technologies programme, and the US National Science Foundation. More than 130 participants attended the meeting.

The main idea behind this workshop was to bring together researchers from projects funded by the Information Society Technologies Research Programme of the EC and the Digital Library Initiative (Phase 2) of the NSF. Both are supporting substantial research concerned with digital libraries and expect to develop the enabling technologies as well as basic principles for DL design, implementation and operation, and to better understand the social and organizational contexts in which DLs will operate. By and large, scientists involved in the EC and NSF projects have not had the opportunity to meet regularly. A coordination effort in the DL field can help avoid duplication of effort, prevent the development of fragmented digital systems, and encourage productive interchange of scientific knowledge and scholarly data around the world.

Given this premise, DELOS has identified the 'DL All projects workshop' as the most appropriate event allowing representatives of these projects and other interested parties to meet regularly. Unlike a conference, researchers engaged in projects meet with their colleagues for the purpose of:

- presenting on-going research
- exchanging working experiences and information
- identifying areas of synergy
- identifying themes and topics which need further investigation
- establishing collaborative links between on-going projects.

The meeting was opened by Costantino Thanos, Director of DELOS, who welcomed the participants and briefly outlined the aim of the meeting. William Bainbridge (National Science Foundation) and Patricia Manson (European Commission) welcomed the participants on behalf of the NSF and the European Commission respectively. The meeting then broke up into various parallel sessions (18 in all), grouped into the following broad categories:

- Architecture
- Preservation
- Basic DL technologies
- DL services
- DL applications
- National Initiatives.

By grouping projects with the same topics of interest together, the meeting provided a focus for the various projects to map across, thereby highlighting the commonality of interests between the projects. Each project presentation emphasized the results thus far achieved, allowing participating researchers to benefit from the experiences gained within similar projects to their own. The parallel sessions lasted one and one half days. In all, 52 presentations were given: 22 EU projects, 10 European National Initiative projects from the UK, Germany, Denmark and Finland, and 20 US NSF projects.

On the afternoon of the second day, a plenary session was held, chaired by Erich Neuhold (FhG-IPSI), dedicated to the presentation and discussion of a DELOS report on 'Future Directions for a European Research Programme on Digital Libraries'. The session began with a presentation by Patricia Manson on the new concepts and instruments of the 6th Framework Programme, with particular emphasis on the priority thematic area of 'Information Society Technologies'. William Bainbridge then illustrated the NSF research policy in the field of digital libraries. The session continued with a presentation by Yannis Ioannidis (University of Athens) of the DELOS report, followed by a discussion by all participants. This report is the result of a DELOS brainstorming meeting, held in June 2001 and attended by 15 EU and 5 US high-level researchers, whose objective was to contribute to the definition of a new vision of digital libraries, and to translate it into concrete research actions to be undertaken by the 6FP.

Complete details about the meeting (programme, projects, presentations, list of participants) can be found at URL: http://delos-noe.iei.pi.cnr.it/activities/internationalforum/All-Projects/us.html

Please contact:
Costantino Thanos, Director of the DELOS Network of Excellence
Tel: +39 050 3152910
E-mail: Thanos@iei.pi.cnr.it

Fourth DELOS Workshop on Evaluation of Digital Libraries: Testbeds, Measurements and Metrics

by László Kovács

The Fourth DELOS Network of Excellence Workshop on Evaluation of Digital Libraries: Testbeds, Measurements, and Metrics, was held at SZTAKI's Distributed Systems department in Budapest from 6-7 June 2002.

The workshop was organized by the DELOS Working Group responsible for maintaining a Digital Library Evaluation Forum and a Digital Library Test Suite which will provide researchers with information for the selection of potential digital library testbeds and new research or evaluation targets. 28 participants attended the workshop, 20 of them came from Europe, seven from the US and one from Japan.

Digital libraries (DL) can be viewed from a number of perspectives. They can be new forms of information institutions, multimedia information retrieval systems, or information systems that support the creation, use, and search of digital content. Digital libraries are not an end in themselves; rather, they are enabling technologies for digital asset management, electronic commerce, electronic publishing, teaching and learning, and other activities. Accordingly, digital libraries need to be evaluated in the context of specific applications. The methods and metrics for evaluating digital libraries will vary by whether they are viewed as institutions, as information systems, as new technologies, or as new services.

The DL research communities need large test beds in order to be able to evaluate and demonstrate new concepts. Research results are most valuable when they are compared with other approaches and validated against many sets of data. Considering users, collections and systems as the major components of a digital library, evaluations may involve only a subset of these components. Besides the evaluation context, the underlying criteria have to be specified. Also, for comparing different evaluations, specification of the test material employed is essential. There are no standard description schemes yet and only little work has been done in this area so far.

The programme for the Workshop included ten long papers and six short papers from Europe, Japan and the US, and was presented in five sequential sessions:

- Setting the background on Digital Library Evaluation: reports from working groups
- Users and User Interfaces
- Evaluation in Context
- Metrics and Testbeds
- Evaluation of Digital Library Services and Scalability.

During the first session an overview of some evaluation initiatives was provided. Ron Larsen (US) gave us an insight view of some of the experiences from DARPA's DLib Test Suite project in the US, and gave the evaluation metrics from the Dlib Metrics Working Group. Michael Mabe (UK) presented a publisher's view on the conceptual schema developed by the DELOS Evaluation Forum Test Suite team. His message is "WHO is using WHAT and WHY" . The "HOW" is the next question to ask, also in evaluation. Nariko Kando (Japan) reported about the initiative 'Evaluation of Information Access Technologies', and the results from three previous evaluation workshops. This was a valuable starting-point for the other presentations where attendees got snapshots of ongoing projects and ideas.

The discussions in the lively 'mini-panels' following each session were stimulating. Discussions and results were also the results from the breakout groups, four altogether:

- Users and User Interfaces, chaired by Nicholas Belkin
- Evaluation in Context, chaired by Christine Borgman
- Metrics and Testbeds, chaired by Ingeborg T. Sølvberg
- Strategies for funding of future Evaluation programs, chaired by Ron Larsen.

The results of the breakout groups were reported in plenary sessions, and the written reports will be made available on the homepage of the workshop (http://www.sztaki.hu/conferences/deval). The final proceedings of the Workshop will be published as an ERCIM report. In addition to the papers the proceedings will include reports from the four breakout groups as well. The proceedings and the power point presentations are also available at the workshop homepage.

Links:
Workshop:
http://www.sztaki.hu/conferences/devalDELOS
DELOS Test Suite :
http://www.sztaki.hu/delos_wg21/
SZTAKI, Distributed Systems Department:
http://dsd.sztaki.hu

Please contact:
Laszlo Kovacs, SZTAKI
Tel: +36 1 279 6212
E-mail: laszlo.kovacs@sztaki.hu

ERCIM NEWS

ERCIM News is the magazine of ERCIM. Published quarterly, the newsletter reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of over 9000 copies.

ERCIM News online edition is available at http://www.ercim.org/publication/ERCIM_News/

ERCIM News is published by ERCIM EEIG, BP 93, F-06902 Sophia-Antipolis Cedex
Tel: +33 4 9238 5010, E-mail: office@ercim.org
ISSN 0926-4981

Director: Bernard Larrouturou
Central Editor:
Peter Kunz
peter.kunz@ercim.org
Local Editors:
AARIT: Wolfgang Hawlik
w.hawlik@ocg.at
CLRC: Martin Prime
M.J.Prime@rl.ac.uk
CRCIM: Michal Haindl
haindl@utia.cas.cz
CWI: Henk Nieland
Henk.Nieland@cwi.nl
CNR: Carol Peters
carol@iei.pi.cnr.it
FORTH: Leta Karefilaki
karef@ics.forth.gr
FhG: Michael Krapp
michael.krapp@iuk.fhg.de
INRIA: Bernard Hidoine
bernard.hidoine@inria.fr
NTNU: Truls Gjestland
truls.gjestland@ime.ntnu.no
SARIT: Harry Rudin
hrudin@smile.ch
SICS: Kersti Hedman
kersti@sics.se
SRCIM: Gabriela Andrejkova
andrejk@kosice.upjs.sk
SZTAKI: Erzsébet Csuhaj-Varjú
csuhaj@sztaki.hu
TCD: Ann McNamara
Ann.McNamara@cs.tcd.ie
VTT: Pia-Maria Linden-Linna
pia-maria.linden-linna@vtt.fi

Free subscription

You can subscribe to ERCIM News free of charge by:

- sending e-mail to your local editor
- contacting the ERCIM office (see address above)
- filling out the form at the ERCIM website at http://www.ercim.org/

CALL FOR PARTICIPATION

ESORICS 2002, Seventh European Symposium on Research in Computer Security RAID 2002, Recent Advances in Intrusion Detection

Zurich, Switzerland 14-16 and 16-18 October 2002 respectively

Organized in a series of European countries, ESORICS is confirmed as the European research event in computer security. The symposium started in 1990 and is now held every two years in different European countries and attracts audience from both the academic and industrial communities.

The RAID symposium, the fifth in an annual series, brings together leading figures from academia, government, and industry to discuss state-of-the-art intrusion detection technologies and issues from the research and commercial perspectives. The RAID International Symposium series is intended to further advances in intrusion detection by promoting the exchange of ideas over a broad range of topics.

More information:
http://www.zurich.ibm.com/~gka/Esorics/
http://www.raid-symposium.org/

SPONSORED BY ERCIM

SOFSEM 2002 - XXIX-th Seminar on Current Trends in Theory and Practice of Informatics

Milovy, Czech Republic, 24-29 November 2002

SOFSEM is an annual conference devoted to the theory and practice of software systems. Its aim is to foster co-operation among professionals from academia and industry working in various areas of computer science. The program consists of a series of Invited Talks, Contributed Talks, and accompanying actions. SOFSEM offers a unique opportunity to obtain quickly a representative overview of the areas of computer science which are selected as the topics of the year. SOFSEM'2002 will feature invited talks on the following topics:

- Computer Science Theory,
- Education in Computer Science,
- Multimedia, Databases, and Vision.

Accompanying actions:
- Third Workshop on Soft Computing (organized by the ERCIM working group on Soft Computing)
- Workshop on Multimedia
- PhD Student Research Forum.

More information:
http://www.sofsem.cz/

CALL FOR PAPERS

Euroweb 2002 Conference 'The Web and the GRID: from e-science to e-business'

Oxford, UK, 17-18 December 2002

The conference theme is intended to prompt debate on a convergence of developments pioneered for e-science on the Grid and as web services to provide business applications. How will frameworks like web services, Grid Services, and .net address the issue of Internet-aware programs using the services that are offered by other programs? What is XML's role in managing and routing data and services? What kinds of problems lend themselves to resource sharing? EuroWeb 2002 follows on from the success of the EuroWeb 2001, which was held in Pisa in December, 2001 on the topic of the web in public administration.

Conference Deadlines
- paper submission: Friday 27th September 2002.
- paper acceptance details: Friday 25th October 2002
- final camera ready copy: Friday November 8th 2002

EuroWeb 2002 invites submissions related to all aspects of GRID technologies and Web Services. Both technical and survey/overview papers are solicited, as well as descriptions of working GRIDs and Web Services systems, position statements and reports on work in progress.

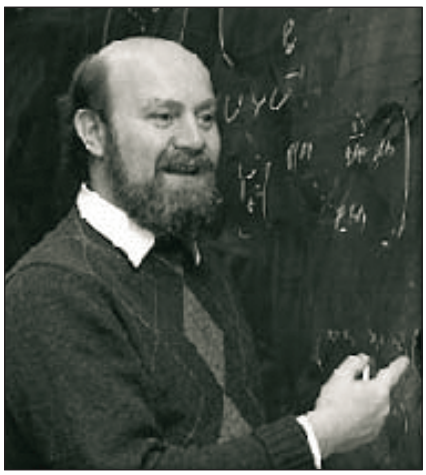
More information:
http://www.w3c.rl.ac.uk/Euroweb/

CLRC — **Five New Projects in Business and Information Technology** — The Business and Information Technology Department of CLRC has recently been awarded 5 grants for new collaborative projects under the EU Framework 5 IST programme. The projects are in strategically important areas of e-Information including, Web and Grid technologies, Knowledge and Trust management, and e-Learning and will strengthen BITD's existing R&D programme which already includes projects in Business Process Modelling, Distributed Information Systems, Knowledge Management and Data Mining, Security and Trust Management, Software Technology, and Web and Grid Technologies. The new projects are:

- Semantic Web Advanced Development for Europe (http://www.bitd.clrc.ac.uk/Activity/SWAD-Europe/)
- Grid architecture for Application Service Provision (http://www.bitd.clrc.ac.uk/Activity/GRASP/)
- An Agent Based Platform for Organisationally Mobile Public Employees (http://www.bitd.clrc.ac.uk/Activity/PELLUCID/)
- Trust Management in Dynamic Open Systems (http://www.bitd.clrc.ac.uk/Activity/itrust/)
- E-Learning within a Grid environment (http://www.bitd.clrc.ac.uk/Activity/lege-wg/)

For more information, please contact Juan Bicarregui (j.c.bicarregui@rl.ac.uk).

CWI — **The 2002 Brouwer Memorial Medal** was awarded to Michael Aizenman of Princeton University on April 4, during the annual Dutch Mathematical Congress in Eindhoven. Aizenman's research focuses on systems with many degrees of freedom, in particular critical behaviour and disorder, both classical and quantum-mechanical. Most phenomena in this area do not allow analytical solutions, and also perturbation theory breaks down. Nonetheless Aizenman could provide answers on central questions regarding such phenomena. The Brouwer Medal, named after the famous Dutch mathematician L.E.J. Brouwer (1881-1966) and awarded every three years, can be considered as the Dutch version of the Fields Medal, which is often called the Nobel Prize in mathematics. In his lecture following the award, Aizenman addressed stochastic geometry and interacting fields.



Michael Aizenman.

SZTAKI, Fraunhofer — **The Virtual Institute for Production and Business Management** was founded on 7 May 2002 as the first step of a co-operation between SZTAKI and Fraunhofer Institute for Manufacturing Engineering and Automation (IPA). The envisaged fields of activity of the Virtual Institute address product management, factory and logistics planning, organisation and quality management, order and process management, supply chain management, production networks, and E-business. As the first sign of co-operation, a workshop on production and business management was held at SZTAKI at the same day. The main lines of activities at the two co-operating institutions were introduced, and the workshop served as a forum for discussing problems and potential methods, tools and solutions for handling them in the information area. On the subsequent days, specialists of IPA and SZTAKI visited several companies interested in this initiative. Co-ordinated efforts are also foreseen within the 6th Framework of the EU.

NTNU — **The Norwegain University of Science and Technology has recently obtained 13 Marie Curie Training Sites** (MCTS) for doctoral training funded by the EU. This places NTNU among the top five universities regarding number of MCTSs and designated centers of excellence. An MCTS at NTNU support young researchers pursuing doctoral studies in the EU Member states, providing them the possibility of undertaking part of their doctoral studies in Norway, and allowing them the benefit of working with internationally recognized groups in their specialized areas of research. The MCTSs are supported by the European Commission. The MCTSs relevant for the ERCIM community are 'SMART-SoC' — System-on-Chip Design for Smart Applications, and 'CyberMar' — Marine Cybernetics —Modelling and Control of Hydroelastic Structures. SMART-SoC (http://www.ntnu.no/intersek/mcts/smartsoc/) comprises modelling, design, testing, characterisation and measurement methods and techniques for nanometer electronics, including interfacing with Micro Electro Mechanical Systems. The research fields of CyberMar (http://www.itk.ntnu.no/marinkyb/CyberMar/) are addressing the integration of mathematical modelling of marine structural dynamics and hydrodynamics, control engineering and information technology.

CWI — **Two CWI researchers have received honorary doctorates**: Lex Schrijver from the University of Waterloo, Canada, and Jan Willem Klop from the University of East Anglia, UK. One of Schrijver's major achievements is his book 'Theory of Linear and Integer Programming', described as "the most outstanding piece of scholarly work in the field of mathematical programming". Klop developed process algebra, jointly with Jan Bergstra and Jos Baeten, and is presently involved in term rewriting systems. In this field he recently completed a voluminous book, together with several co-authors.

ERCIM – The European Research Consortium for Informatics and Mathematics is an organisation dedicated to the advancement of European research and development, in information technology and applied mathematics. Its national member institutions aim to foster collaborative work within the European research community and to increase co-operation with European industry.

AARIT
Austrian Association for Research in IT
c/o Österreichische Computer Gesellschaft
Wollzeile 1-3, A-1010 Wien, Austria
Tel: +43 1 512 02 35 0, Fax: +43 1 512 02 35 9
http://www.aarit.at/

CLRC
Central Laboratory of the Research Councils
Rutherford Appleton Laboratory
Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom
Tel: +44 1235 82 1900, Fax: +44 1235 44 5385
http://www.cclrc.ac.uk/

CNR
Consiglio Nazionale delle Ricerche, IEI-CNR
Area della Ricerca CNR di Pisa,
Via G. Moruzzi 1, 56124 Pisa, Italy
Tel: +39 050 315 2878, Fax: +39 050 315 2810
http://www.iei.pi.cnr.it/

Crcim
Czech Research Consortium
for Informatics and Mathematics
FI MU, Botanicka 68a, CZ-602 00 Brno, Czech Republic
Tel: +420 2 688 4669, Fax: +420 2 688 4903
http://www.utia.cas.cz/CRCIM/home.html

CWI
Centrum voor Wiskunde en Informatica
Kruislaan 413, NL-1098 SJ Amsterdam,
The Netherlands
Tel: +31 20 592 9333, Fax: +31 20 592 4199
http://www.cwi.nl/

FORTH
Foundation for Research and Technology – Hellas
Institute of Computer Science
P.O. Box 1385, GR-71110 Heraklion, Crete, Greece
Tel: +30 81 39 16 00, Fax: +30 81 39 16 01
http://www.ics.forth.gr/

Fraunhofer Gesellschaft
Fraunhofer-Gesellschaft,
Information and Communication Technology Alliance
Schloß Birlinghoven, D-53754 Sankt Augustin, Germany
Tel: +49 2241 14 0, Fax: +49 2241 14 2889
http://www.fraunhofer.de/

INRIA
Institut National de Recherche en Informatique
et en Automatique
B.P. 105, F-78153 Le Chesnay, France
Tel: +33 1 3963 5511, Fax: +33 1 3963 5330
http://www.inria.fr/

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and
Electrical Engineering, N 7491 Trondheim, Norway
Tel: +47 73 59 80 35, Fax: +47 73 59 36 28
http://www.ntnu.no/

SICS
Swedish Institute of Computer Science
Box 1263,
SE-164 29 Kista, Sweden
Tel: +46 8 633 1500, Fax: +46 8 751 72 30
http://www.sics.se/

SARIT
Swiss Association for Research in Information Technology
Dept. Informatik, ETH-Zentrum, CH-8092 Zürich,
Switzerland
Tel: +41 1 632 72 41, Fax: +41 1 632 11 72
http://www.sarit.ch/

SRCIM
Slovak Research Consortium for Informatics and
Mathematics, Comenius University, Dept.of Computer
Science, Mlynska Dolina M, SK-84248 Bratislava,
Slovakia, Tel: +421 7 726635, Fax: +421 7 727041
http://www.srcim.sk

MTA SZTAKI
Magyar Tudományos Akadémia
Számítástechnikai és Automatizálási Kutató Intézete
P.O. Box 63, H-1518 Budapest, Hungary
Tel: +36 1 279 6000, Fax: + 36 1 466 7503
http://www.sztaki.hu/

Trinity College
Department of Computer Science,
Dublin 2, Ireland
Tel: +353 1 608 1765, Fax: 353 1 677 2204
http://www.cs.tcd.ie/ERCIM

VTT
Technical Research Centre of Finland
P.O. Box 1200
FIN-02044 VTT, Finland
Tel:+358 9 456 6041, Fax :+358 9 456 6027
http://www.vtt.fi/tte

Order Form

If you wish to subscribe to ERCIM News
free of charge
or if you know of a colleague who would like to receive regular copies of ERCIM News, please fill in this form and we will add you/them to the mailing list.

send, fax or email this form to:
ERCIM NEWS
Domaine de Voluceau
Rocquencourt
BP 105
F-78153 Le Chesnay Cedex
Fax: +33 1 3963 5052
E-mail: office@ercim.org

Name:

Organisation/Company:

Address:

Post Code:

City:

Country

E-mail:

Data from this form will be held on a computer database.
By giving your email address, you allow ERCIM to send you email

You can also subscribe to ERCIM News and order back copies by filling out the form at the ERCIM website at http://www.ercim.org/publication/Ercim_News/