CWI is the National Research Institute for Mathematics and Computer Science. CWI is administered by the Stichting Mathematisch Centrum (SMC), the Dutch foundation for promotion of mathematics and computer science and their applications. SMC is sponsored by the Netherlands Organization for Scientific Research (NWO).

CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics. CWI participates in the Telematics Institute. CWI is a Member of the World Wide Web Consortium (W3C) and runs the W3C Office in the Netherlands.

General Director
G. van Oortmerssen

Colophon
Issued by the Stichting Mathematisch Centrum, June 2001 ©
Production Communication and Publication Department CWI
Design Tobias Baanders
Printing Grafisch Bedrijf Ponsen & Looijen bv

Stichting Mathematisch Centrum / Centrum voor Wiskunde en Informatica
Visiting address Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
Postal address P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Telephone +31 20 592 9333
Telefax +31 20 592 4199
Website www.cwi.nl

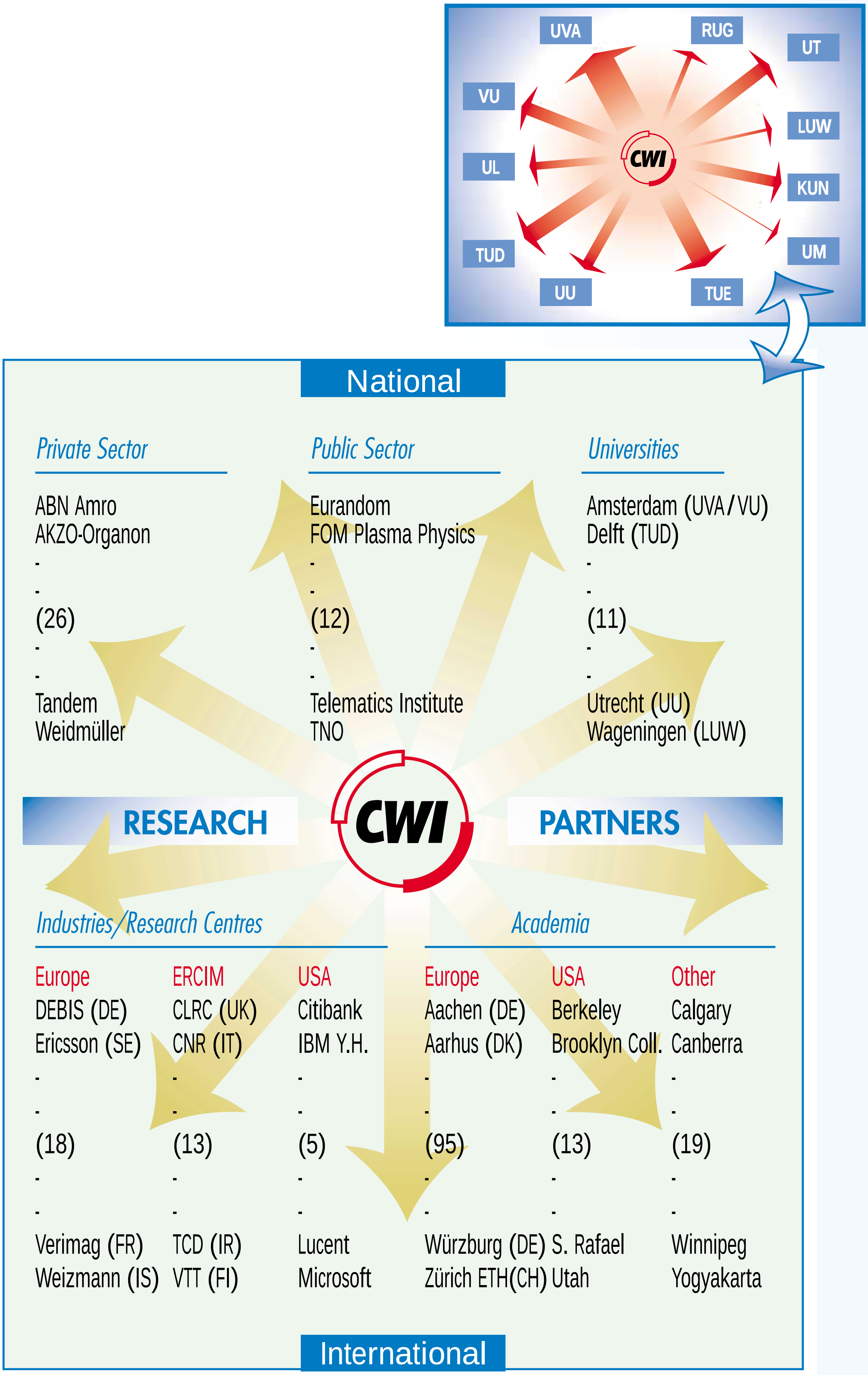
CWI-AR 2000

OVERVIEW

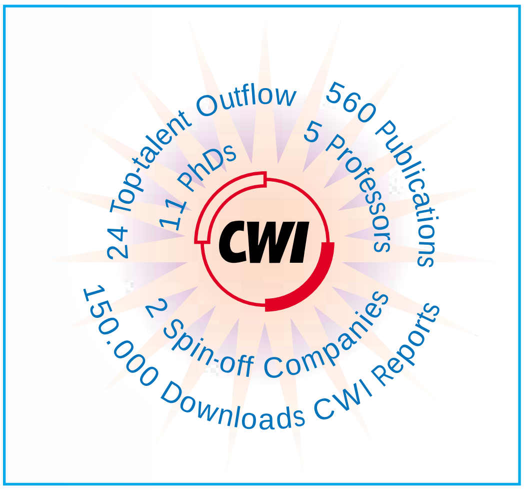
The frontiers of research in mathematics and computer science – CWI's natural field of operation – move rapidly nowadays. Directions are increasingly determined by external developments. Two examples, in which CWI is particularly active, are the Internet and the Life Sciences. CWI was prominent in the development of the Internet from the outset, and is now, e.g., disclosing new ways to store, send, process and retrieve images across the Web. Worldwide the interest in biological questions has increased enormously, possibly triggered by the recent mapping of the human genome. CWI initiated biomathematical research in The Netherlands three decades ago, focusing soon on epidemiological issues (population dynamics). The report year saw a vigorous restart of biology-related research in several fields, on which will be reported below. Remarkable results were booked in several other fields. Here we only mention the exact determination of the buffer size of heavy-tailed queues in Internet traffic, the research into destabilization of wave fronts, the verification of the control software for a distributed truck-lifting system, and the demonstration of a lower bound for the complexity of the Shellsort sorting algorithm (an open problem for forty years). CWI research is closely coupled with knowledge transfer. Two forms are particularly successful: the flow of top talent to universities and trade & industry, and the creation of spin-off companies. During the report year eleven young researchers completed their PhD thesis, and three staff members accepted a part-time professorship, whereas another three became full-time professor. CWI created, alongside two new spin-off companies, the CWI Incubator BV, in order to reinforce its policy in this field. Finally, also the knowledge transfer in written form – here printing on demand and electronic distribution become predominant – proved to be a success: almost 900 CWI Reports were downloaded, the most popular one over two thousand times. The total number of downloads for CWI Reports was 150.000, well over a doubling compared to 1999.

Positioning and funding

All these successes are booked against a background of national and international developments, which strongly influence CWI's positioning. On the European level, politicians have expressed their worries about the unsatisfactory development of the ICT sector for which the gap with the USA is still widening. The European Council stated at its Lisbon meeting in March that the EU should create the world's most competitive and dynamic knowledge economy, and summoned its member states to meet this challenge. The Dutch ministers of Economic Affairs, and of Education, Culture & Science, thereupon produced a joint document *Competing with ICT Competencies*, in which concrete actions were announced in the field of ICT-related knowledge and innovation. It was stated that companies still lack innovative spirit, that strategic positioning of the Dutch knowledge infrastructure in ICT research and its interaction with the Trade & Industry sector is inadequate, and that the number of ICT experts does not meet the demand. In order to tackle these problems, the Dutch government


UVA
RUG
UT
VU
LUW
UL
CWI
KUN
TUD
UM
UU
TUE
National
Private Sector
ABN Amro
AKZO-Organon
.
.
(26)
.
.
Tandem
Weidmüller
Public Sector
Eurandom
FOM Plasma Physics
.
.
(12)
.
.
Telematics Institute
TNO
Universities
Amsterdam (UVA/VU)
Delft (TUD)
.
.
(11)
.
.
Utrecht (UU)
Wageningen (LUW)
RESEARCH
CWI
PARTNERS
Industries/Research Centres
Europe
DEBIS (DE)
Ericsson (SE)
.
.
(18)
.
.
Verimag (FR)
Weizmann (IS)
ERCIM
CLRC (UK)
CNR (IT)
.
.
(13)
.
.
TCD (IR)
VTT (FI)
USA
Citibank
IBM Y.H.
.
.
(5)
.
.
Lucent
Microsoft
Academia
Europe
Aachen (DE)
Aarhus (DK)
.
.
(95)
.
.
Würzburg (DE)
Zürich ETH(CH)
USA
Berkeley
Brooklyn Coll.
.
.
(13)
.
.
S. Rafael
Utah
Other
Calgary
Canberra
.
.
(19)
.
.
Winnipeg
Yogyakarta
International

has now allocated additional funds, to be spent in the ICES Programme (Interdepartmental Commission for Reinforcement of Economic Structure).



Some output figures.

CWI enjoys a strong position as a scientific ‘centre of excellence’. In addition, it possesses an extensive international network of relations with knowledge institutions and companies, and is quite successful in the field of knowledge transfer, for example by its deliberate policy of creating spin-off companies. Hence, CWI has an excellent position to make significant contributions to the realization of Dutch government policy. Of course, it should be enabled to maintain, and where possible, reinforce its prominent international role, and remain at the front of new ICT developments. The ICT sector takes an increasing share in the GNP, and develops internationally at a breathtaking pace. Substantial investments in long-term fundamental research take place in the USA, as well as in several European countries. For example, in France CWI’s sister-organization and co-operation partner INRIA will double its budget within four years. To prevent the gap from getting too wide, additional funds are necessary at short notice, for example for attracting the necessary top researchers. The needs of the rapidly expanding knowledge economy offer CWI several opportunities, but in order to realize these the institute should grow. However, there was hardly any growth over the past decade. CWI’s considerable success in acquiring externally financed projects, amounting to 30% of the budget, creates at the same time a burden due to the ‘matching’ obligation (usually on a fifty-fifty basis), thereby reducing the amount of money available for long-term research.

In order to fulfil the role outlined above, CWI needs a doubling of its basic subsidy (about nine million Euro at present) within a few years. Moreover, the resulting expansion of research staff will require new accommodation. Hence, CWI is now requesting additional funding , and negotiates with NWO about its housing problems. As an intermediate solution, CWI’s old Portakabin building was replaced by a larger one, and now houses the cluster Information Systems. The accommodation includes a mini-CAVE for research and demonstrations in the field of visualization and virtual reality.

Teaching computers to spot interesting images

Vision is an amazingly powerful source of information about the world around us. As a consequence, a considerable part of the data we collect comes in the shape of visual material, and multimedia libraries are rapidly filling up with images and video-footage. However, while progress in electronics reduces the costs of collecting and storing all these images, an efficient database search for specific material becomes increasingly difficult. As it is virtually impossible to classify images into simple unambiguous categories, retrieving them on the basis of their visual content seems to be the only viable option. Rather than asking the computer to display image *AK-C-4355.jpg*, we want to instruct it to retrieve an image showing, e.g., the *3-dimensional structure of a DNA-molecule*. Or, imagine an interior decorator looking for visual art to enliven his client's newly refurbished offices, requesting the system to find *vividly coloured abstract paintings*. These scenarios call for vision-sentient computer systems that can assist humans by sifting through vast collections of images, setting aside the potentially interesting ones for closer inspection, while discarding the large majority of irrelevant ones. Being able to do so would effectively equip the search engine with an automated pair of eyes with which it can reduce for the user the drudgery of an exhaustive search. Such systems will have to be highly adaptive. Indeed, what constitutes an 'interesting' or 'relevant' image varies widely among users, and different applications call for different criteria by which these images are to be judged. In the EU-funded FOUNDIT project, researchers at CWI are designing search engines that can be taught what images to look for on the basis of feedback supplied by the user. If the user classifies a small subset of images as similar and/or relevant, the system extracts the visual characteristics that correlate best with the user's classification, and generalizes this ranking to the rest of the database. Only images marked as highly relevant are displayed for closer inspection and further relevance feedback, thus setting up an iterative cycle of fine-tuning. The aim is to make the interaction between user and interface as natural and transparent as possible: similarities between images are indicated by dragging them into piles on the screen, and relevant images are selected by simply clicking on them. Such operational transparency is of great importance for applications, such as E-commerce. (http//www.cwi.nl/~pauwels/PNA4.1.html)

Interactive definition of image similarity. Left: The interface displays five images grouped together to reflect its current model of image similarity. Right: If this grouping does not match the user's appreciation of their similarity, he can rearrange them on screen into a more satisfactory configuration by a simple drag-and-drop operation. For example, if the user considers the top left image to be more reminiscent of mountains, he will drag it over to the image group on the right. In the next iteration step, the search engine will adapt its features to reproduce this user-defined similarity as faithfully as possible.

Spin-off policy

An important step in the realization of CWI's policy to promote knowledge transfer was the creation of CWI Incubator BV in the summer of 2000. Its function is to stimulate the generation and support the initial phase of a spin-off company, in which fundamental results obtained in a research group are transformed into market-able products, at the same time maintaining working relations with the original research group. CWI Inc is owned for 75% by CWI, 25% of the shares being in the hands of CWI personnel. CWI aims at creating two such spin-off companies annually, in order to bring the developed technology quickly to the market. During the report year two spin-off companies were created: SIG (Software Improvement Group), gen-erated from the research theme Interactive Software Development & Renovation, and Epictoid, related to a system for facial animation developed at CWI. The first financial fruits of this spin-off policy were harvested when the institute partly sold in 2000 its shares in Data Distilleries, an earlier spin-off company involved in data mining. The proceeds are re-used for new research at CWI through a 'revolving fund'.

Life Sciences

CWI also started research in several new areas. In particular research connected with the Life Sciences was stimulated. CWI has excellent opportunities here because its expertise in mathematical modelling, numerical analysis, visualization and bio-informatics is well-suited to this field. A dozen projects started in 2000, and some others were prepared. In particular we mention here the creation of the Silicon Cell Consortium, in which the University of Amsterdam and the Free University in Amsterdam collaborate with CWI. The long-term goal is the computation of Life at the cellular level. CWI's annual day for Trade & Industry, *CWI in Bedrijf*, is an important instrument for knowledge transfer. The year 2000 edition focused on the role of mathematics and computer science in the biosciences. Some hundred participants coming from several companies and institutions were informed about CWI's research activities in this field. External speakers included: Harry van der Laan (astronomer and director of PROFAST Consultancy), Kees Vijlbrief (ministry of Economic Affairs, director General Technology Policy), Jacob de Vlieg (Unilever Research), and Roel van Driel (University of Amsterdam, director of the Swammerdam Institute of Life Sciences).

Other themes

Developments in the field of Scientific Visualization and Virtual Reality went at such a pace over the past few years that CWI decided to upgrade its research group in that field. A pilot research theme started from 2001, led by Robert van Liere. At the same time the pilot theme Financial Mathematics was terminated. Current research in this field was transferred to the group Evolutionary Systems, where E-commerce is emphasized, for example in the new Telematics Institute project ASTA (Autonomous Systems of Trade Agents in E-commerce). CWI research into performance analysis of networks gains more weight, in particular because of the rapid developments in broadband

CWI Research in Life Sciences

Started in 2000:

biomodels	Mathematical models of biological and physical processes with self-organized critical behaviour
agriculture	Image segmentation with applications to agriculture
living cells	Modelling of biochemical processes in living cells, and of signal transduction
nervous system	Numerical methods for modelling the development of neural connections in the nervous system
phytoplankton	Numerical methods for modelling the growth of phytoplankton in the oceans Analysis of phytoplankton bloom and diffusion of macromolecules
biofilms	Development of a dynamic model for the exchange of particles and solutes between biofilms and water
cell biology	VR techniques in cell biology
genomic data	Computational tools for the exploration of genomic data

In preparation:

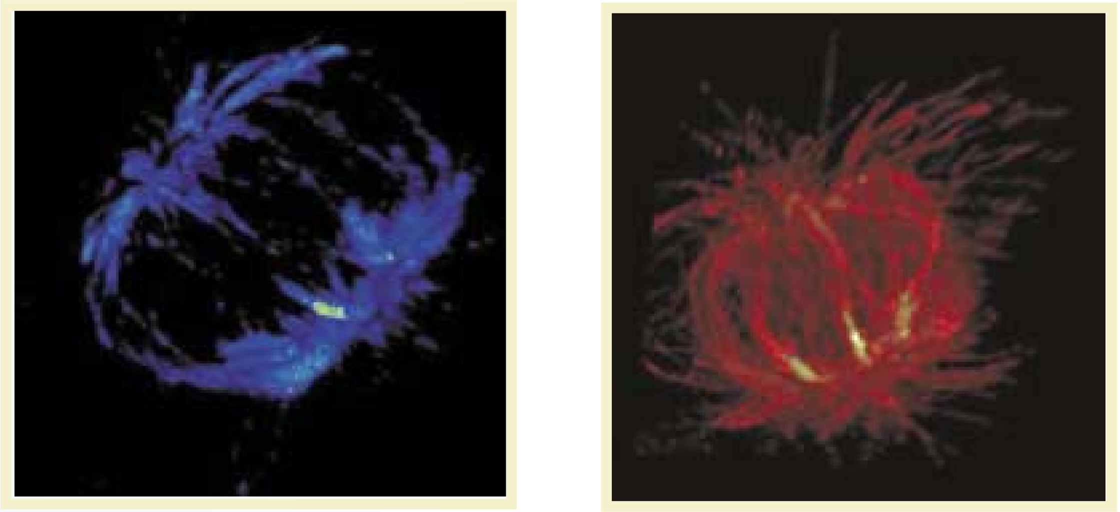
molecular biology	Algorithmic and combinatorial methods for molecular biology
whales	Database of photographic material of cetaceans (whales)
bio-mining	Data mining techniques in bio-informatics and medicine
living cell	The Silicon Cell (visualization, 'living cell' initiative)
gene-product analysis	Interrogation engine for gene-product analysis

CWI participates in the Telematics Institute project ASTA, which addresses, e.g., adaptive markets where customer relationships should be flexible.

Internet and mobile communication. Michel Mandjes joined this group and an exchange agreement with the University of Twente was concluded in 2000. Similar agreements already existed with Eindhoven University of Technology and the Free University in Amsterdam. Finally, the theme Interactive Information Engineering, after having run its natural course, was dissolved. Part of the acquired knowledge is now being exploited in the new spin-off company Epictoid. Another part of the activities is continued in the theme Multimedia and Human/Computer Interaction in the fields of Social User Interfaces and E-learning.

Networking

Participation in large-scale networks is an increasingly important condition for research institutions to maintain a leading role. Vital strategic information becomes available more directly and reactions on international developments can be faster. CWI has recognized this already for years, and participates in several organizations on the national and international level. CWI is a research partner in six projects of the Dutch Telematics Institute, and participates in the Amsterdam Science & Technology Centre WTCW. The company WTCW was created in 1999, in order to coordinate activities resulting from the allocation of 14 million Euro from the ICES-KIS fund to reinforce the knowledge infrastructure at the scientific institutions located at the WTCW



Visualization of cell division (mitosis). The coloured part represents chromosomes being drawn apart.

Domain Specific Languages

The challenge of software engineering is to create flexible, reliable, scalable and maintainable systems. To achieve this, the software engineer must apply a wide variety of tools and techniques. One of these is to take advantage of the underlying application domain, by means of a domain-specific language (DSL). A Domain Specific Language provides a notation that can be used to compose applications from a set of concepts tailored towards a specific application domain. Details of the underlying implementation platform are captured by knowledge built into the DSL compiler. This compiler can translate the DSL into, for example, executable code, calls to library routines, database transactions, or HTML forms. The Risla language is a case in point. It was developed by Cap Gemini and MeesPierson, in cooperation with CWI. Given a description of the essential ingredients of an interest rate product (loan, swap, financial future, ...), the Risla compiler can generate data structures in VSAM format, data entry screens in CICS format, and COBOL routines for registering modifications in the product or for yielding management information. A DSL compiler relieves the user from unnecessary details and machine dependencies. As a result, changes in the ICT infrastructure have to be accommodated only once in the DSL compiler. After re-compilation of all DSL programs, the desired infrastructure-related changes have been effectuated without making a single change to the DSL programs themselves. Contrast this with the traditional situation that domain-related knowledge and infrastructure-related knowledge are intermixed in the program code. The use of DSLs thus leads to greater flexibility and less maintenance. The DSL project is carried out under the auspices of the Telematics Institute (TI) and has as goals to develop methods for selecting suitable DSL domains, and for capturing domain knowledge into a DSL and its compiler; to develop meta-tools for the rapid prototyping of domain-specific languages; and to gain more experience, via case studies, with the use of domain-specific languages in a commercial setting. Legacy applications are often a valuable (but hard to exploit) source of domain knowledge. Therefore, the DSL project takes legacy systems and software understanding tools as one of its point of departure. So far, we have gained experience with a variety of analysis and renovation problems, including documentation generation for (mostly COBOL-based) legacy systems (in cooperation with the Software Improvement Group, ABN AMRO, Roccade and others), and analysis of C and SDL systems (both in cooperation with TI-partner Lucent). In addition, we gained experience with a variety of DSLs: RISLA for the financial domain (in cooperation with TI-partner Cap Gemini), Sophus for writing partial differential equation solvers (in cooperation with University of Bergen, Norway), used for instance in the area of seismic modelling for oil exploration, and a router description language (in cooperation with TI-partner Lucent). Last but not least, major progress has been made with the development of tools and meta-tools as incorporated in the ASF+SDF Meta-Environment. Results of the project can be found at www.cwi.nl/projects/dsl.

Seismic modelling and other computational modelling software benefits from the Sophus approach to software construction which is being developed at CWI in cooperation with the University of Bergen, Norway.

After 36 years service with CWI, Piet van der Houwen went on early retirement in October. Van der Houwen, characterized as 'Accurate, Civilized and Creative' in the Liber Amicorum presented to him on that occasion, studied at the University of Amsterdam. He received his PhD degree at the same university in 1968 on a thesis Finite Difference Methods for Solving Partial Differential Equations. *He became head of CWI's department of Numerical Mathematics in 1973, combining it two years later with a professorship in Numerical Mathematics and Informatics at the University of Amsterdam. In 1997 he gave up his position as a research leader, to become the first CWI Fellow. In this quality he still enjoyed some years of undisturbed research. At his retirement he was knighted (Knight in the Order of the Dutch Lion). Seventeen young researchers prepared their PhD thesis under Van der Houwen, who (co-)authored almost two hundred publications. During the years Van der Houwen developed a thorough knowledge and a deep love of wine. He feels a close relationship with Italy: scientific, culinary and touristic.*

premises. During the report year these institutions also took steps to shape their joint WTCW profile. WTCW offers an interesting mix of institutions with promising synergetic possibilities. CWI carries out research financed through ICES-KIS in cell biology and large distributed multimedia databases. CWI was the first institution in The Netherlands to move in August to a Gigabit Ethernet connection to SURFnet, the national academic computer network. Thus it prepares for optimal benefit from the installation of the 80 Gbit/s capacity SURFnet5 envisaged for 2001. The WTCW Giganet is connected to a national network to be realized within the national Gigaport project. CWI coordinated once again the annual WTCW Day for the Public, as a part of the National Science Week. Internationally, CWI maintains a pivotal position in the European Research Consortium for Informatics and Mathematics (ERCIM). In a document produced in fall ERCIM outlines its aims for the coming four years. Evidently ERCIM can play a prominent role in the creation of a 'European Research Area' advocated by Brussels. Actually, ERCIM is a blueprint of the proposed network organizations and is, hence, involved in preparatory discussions. Finally, CWI started participation in the Institute for Mathematics and its Applications (IMA) in Minneapolis, enabling its researchers to participate in projects and attend workshops.

A biography of David van Dantzig (1900–1959) appeared in 2000. Van Dantzig was one of the initiators of creating CWI just after World War II, and founding father of mathematical statistics as a scientific discipline in The Netherlands. The biography was written by mathematician and historian Gerard Alberts, who is seconded to CWI from the University of Nijmegen, and was presented at a well-attended symposium dedicated to Van Dantzig on the occasion of his 100th birthday.

MIA -- querying multimedia databases

The Multimedia Information & Analysis (MIA) project investigates the problem of access to multimedial collections. Finding textual information on the ever growing Web is already quite a challenge, let alone searching the forthcoming wealth of multimedia documents. Our research provides the basis for a new generation of search engines on a multimedia Web, and supports digital archiving in large companies. It is carried out with funds allocated to the Amsterdam Science and Technology Centre WTCW, in close collaboration with the University of Amsterdam (UvA). Digital multimedia documents are best stored centrally and distributed through broadband Internet, allowing remote access from anywhere anytime; shortly, even from your mobile phone. And, digital data are easily exchanged, also enabling the reuse of existing material. Furthermore, people at home create ever more digitized data: making your own compact discs with music has become extremely popular, and digital photo and film cameras become increasingly common. Technology developed in MIA focuses therefore on retrieving 'that song with that nice little melody', or 'that picture of our sailing, with the Boston skyline; when the weather was nice, was it 1997, or 1998?' MIA's database team at CWI focuses on two themes: scalability and query formulation. The target in the first theme is processing queries in a collection exceeding one million pictures. Common approaches based on the colour distribution of an example image breakdown for collections of this size, because often either all objects or none are found. Moreover, some 'back of the envelope' calculations show that the required time per iteration of the search process yields unacceptable response times as soon as the collection grows beyond several thousands of pictures. Yet, the toughest problem in searching picture archives with current systems, is that query results are hard to understand even for their software developers. The end-user cannot be expected to comprehend why an example picture of a sunset at sea results in (however beautiful) pictures of African savannas. Such inaccurate answers are an unwanted side-effect of the uncertainty inherent to posing the question, exposing the wide gap between our high-level perception and the (necessarily) much simpler methods for image analysis. We seek to reduce this gap with better formulated search requests, based on two different but complementary approaches. On the one hand, we assist the user with automatic query formulation; in our opinion, many improvements can be achieved by collecting more information about the user's interest in a carefully crafted, interactive query process. On the other hand, we attempt to give users more insight into the query process inside the system itself, enabling them to intervene in that process and eventually adjust it (query articulation). An interesting challenge in our research is to find the correct balance between these two approaches.

before:

after:

Figure 1: Initial query results from a photo collection can be improved by feedback from the user.

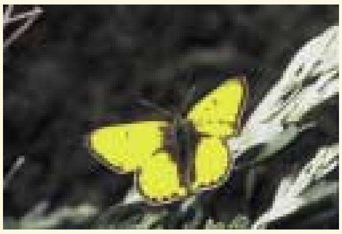


Figure 2: How to process queries like 'this butterfly but then in yellow' is an open problem.

WWW, IST

CWI's considerable share in the ongoing development of the World Wide Web remained clearly visible in the report year. The institute had a major role in the organization of the annual worldwide WWW conference (WWW9), which was held in Amsterdam in May, and attracted over 1400 participants. Iván Herman co-chaired the conference, Dick Bulterman (director of CWI's spin-off company Oratrix) presided the programme committee, and Lynda Hardman was vice-chairperson for Hypermedia. The World Wide Web Consortium (W3C) maintains a dozen offices around the world. CWI runs the Dutch W3C office, and organized some well-attended workshops on XHTML, SMIL, SVG, and Web Usability.

W3C Team Meeting at CWI on May 24 after the WWW9 Conference in Amsterdam (Tim Berners-Lee, inventor of the World Wide Web, with white shirt-sleeve, left below person in top position).

CWI has always been very successful in the European Framework Programmes for information technology, from ESPRIT in the 1980s to the present 5th Framework Programme IST (Information Society Technologies). It participated in a relatively large number of projects, frequently as coordinator. Present involvement in IST includes projects in the areas of Learning Systems, Quantum Computing, Video Coding, and Image Retrieval. CWI coordinates the last three projects.

Staff

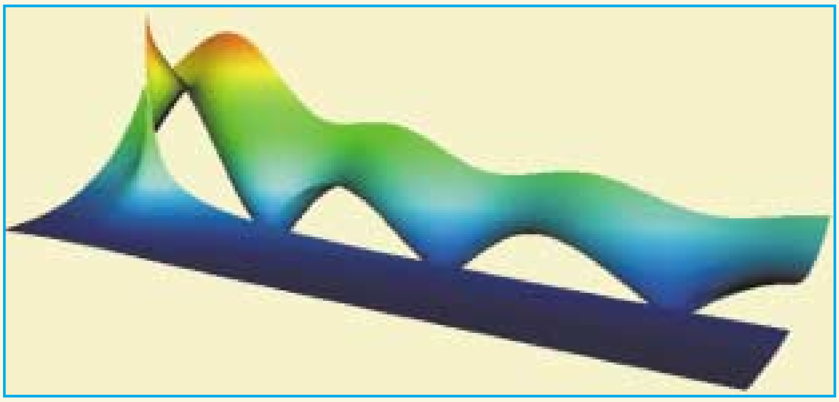
The scientific staff underwent some changes. The flow of talent to (mainly) universities is a natural process for CWI, and is a strong component of its knowledge transfer activities. Hans van Duijn, who headed the cluster Modelling, Analysis & Simulation, and Jan Friso Groote, leader of the theme Specification & Analysis of Embedded Systems, accepted a position as full-time professor at Eindhoven University of Technology. Previously Van Duijn was already part-time professor in Leiden and Delft.

Updating Abramowitz & Stegun

A project is underway at the National Institute of Standards and Technology (NIST) to develop a replacement for the *Handbook of Mathematical Functions*, commonly known as 'Abramowitz & Stegun', after its principal editors. This will be a major new mathematical reference source on the World Wide Web for special functions and their applications. The Handbook, published in 1964 by the National Bureau of Standards, is probably the most widely-used book on special functions, and perhaps even in mathematics at large. This book, whose sales number in the hundreds of thousands, may well be the most frequently cited source of all mathematical reference works. A paperback version has been photo-offset and sold by Dover Publications since 1965. Although Dover does not reveal sales data, it undoubtedly outsells the government edition many times over. The Handbook has never been revised, although there have been numerous advances in basic theory, computational methods, and domains of application since its publication. Its structure is that of a static reference volume; though still very useful, it does not meet the needs of modern users for information that can be conveniently exploited in a highly computer-oriented technical environment. Such needs have been communicated regularly to NIST (formerly, National Bureau of Standards). About five years ago it was decided at NIST that a successor to the NBS Handbook should be designed in the form of a knowledge base, called the *Digital Library of Mathematical Functions (DLMF)*. Via free access over the World Wide Web, using standard Web browsers, individuals will be able to obtain validated mathematical information in a semantic-based representation that incorporates metadata, interactive features, and extensive linkages. The Web server will be constructed, maintained and operated by NIST. The many tables in the existing version of the Handbook will not be present in the new version. They are inadequate to current needs, particularly when used to validate numerical software. Second, software packages exist that can compute vastly extended numerical tables. New chapters will be:

- Asymptotic Approximations
- Computer Algebra
- q-Hypergeometric Functions
- Painlevé Transcendents
- Asymptotics of Integrals with Coalescing Saddles
- Wavelets
- 3j, 6j, 9j Symbols

CWI is involved in this project. Nico Temme, a specialist on Asymptotics and Special Functions, is a member of the Editorial Board, and will write several chapters (on exponential integrals, error functions and parabolic cylinder functions). The Executive Committee for the whole project consists of the four Principal Editors: Daniel W. Lozier, Frank W.J. Olver, Charles W. Clark, and Ronald F. Boisvert. The Board of Associate Editors: R.A. Askey, M.V. Berry, W. Gautschi, L.C. Maximon, M. Newman, I. Olkin, P. Paule, Johannes Kepler, W.P. Reinhardt, N.M. Temme, and J. Wimp. The project is expected to be completed by 2003. (http://dlmf.nist.gov/)



$H_0^{(1)}(z)$: Hankel function of the first kind and of order zero.

In memoriam J.W. Cohen (1923–2000)

Professor J.W. Cohen passed away on November 12, 2000, at the age of 77. As an advisor of CWI (1982–1998) he played an important role in the development of CWI's research in queueing theory and performance analysis. For example, he initiated CWI's regular queueing seminar, and was always one of its most active participants. In 1986 he organized a major international conference at CWI on teletraffic analysis and computer performance evaluation. In the same year he received the AKZO prize for his whole scientific achievement. Cohen was a professor at Delft University of Technology until 1973, and then at the University of Utrecht until his retirement in 1988. He established his reputation as a scholar and world-leading queueing theorist with his book *The Single Server Queue* (1969), which has remained a classic ever since. He has contributed, more than anyone else, to establishing queueing theory as a mature mathematical discipline within Applied Probability. During his years as an advisor at CWI he published four books, one of which was translated into Russian, and a large number of papers. After retirement most of his research results were recorded in CWI Reports, the last of which appeared less than two months before his death. Cohen has made a lasting impression on those who have interacted with him. Behind the smoke of his ever-present pipe and his formal appearance – three-piece-suit plus sweater – was a warm and friendly personality. He was very caring and even protective for his students and friends.

Arno Siebes, leader of the theme Data Mining and Knowledge Discovery, accepted a full-time professorship at the University of Utrecht. Three staff members accepted part-time professorships: Harry Buhrman (Quantum Computing) and Jan Verwer (Numerical Analysis) at the University of Amsterdam, and Michel Mandjes (Stochastic Operations Research) at the University of Twente. Iván Herman and Steven Pemberton were seconded at the World Wide Web Consortium W3C, but maintain their office at CWI. Finally, Mike Keane was appointed member of the Chilean Academy of Science.

Visits

Several high-placed persons visited CWI. Prime Minister Kok was invited to visit the WTCW premises and was then also informed about CWI's activities. EU Commissioner for Research Philippe Busquin visited CWI with a delegation including Dutch EU parliamentarians Elly Plooij-van Gorsel and Dorette Corbey. They were informed about CWI's research, in particular Quantum Computing and Visualization, as well as its position in the research world, in particular ERCIM. Edsger W. Dijkstra, pioneer of computer science and CWI staff member in the period 1952–1962, visited his old

Gerard van Oortmerssen addressing Prime Minister Wim Kok during his visit to CWI.

Efficient computation of ship waves

A new solution method for the efficient computation of water waves around a ship was developed at CWI, in close cooperation with the Dutch maritime research institute, MARIN. The ability to predict the behaviour of flows is essential in many engineering applications, such as ship hydrodynamics. Although the physical laws describing fluid flow and their mathematical formulation (the full Navier-Stokes equations) are well understood for more than a century, due to its tremendous complexity the most accurate mathematical model is still unmanageable for engineering purposes. Therefore, those concerned with practical fluid-mechanics problems have always considered more tractable models. For practical computations by means of a computer, the unknown functions are approximated by tabular values, and the fluid-flow problem then translates into the problem of solving a very large system of equations in which the tabular values are the unknowns. This technique is called 'discrete approximation'. The approximation becomes increasingly accurate as the tabulation becomes finer.
On the other hand, however, the system of equations grows correspondingly and becomes more difficult to handle. The large system of equations is solved to arbitrary accuracy by a computer in an iterative (i.e., cyclic) process. Although the rapid development of computers has allowed the treatment of increasingly larger systems, it still remains a challenge (and a practical necessity) to obtain, within a given computational effort, better accuracy for more complex flow problems. One of the flow problems that have recently become tractable is: free-surface viscous flow in three dimensions, e.g., waving water flow around a ship hull. Here the problem is to determine both the (viscous) flow and the shape of the wavy water surface. Previously, only simpler models could be applied, and the surface and the flow problem were treated separately: either the wavy surface of an inviscid fluid-flow, or a viscous fluid-flow below a flat water surface was determined. However, simultaneous treatment is required to obtain more reliable results. The essential problem was the inefficiency of the available iterative methods for the viscous free-surface flow problem. As a consequence, in practice the accuracy of the approximation was limited, and the more sophisticated model gave less accurate results. Hence, an essentially more efficient iterative method had to be developed. The new method was shown to be efficient for a test problem.
Since October, computations for actual ship flows are in progress.

One challenge in the hydrodynamic design of ships is to find the optimal shape of the ship hull satisfying prior requirements of use, economy, safety, etc. CFD computations have led to improvements of the hull shape, so that the high amplitude of the long waves made by the ship could be strongly reduced. (Photo courtesy MARIN.)

EU Commissioner for Research Philippe Busquin (left photo, right) and Dutch EU parliamentarian Elly Plooij - Van Gorsel (right photo, with VR-glasses) during their visit to CWI, where they were received by director Gerard van Oortmerssen.

Edsger Dijkstra (centre) at CWI with his hosts Krzysztof Apt (left) and Gerard van Oortmerssen (right).

institute in October and delivered a lecture 'Combating complexity in problem solving'. Dijkstra is generally considered the father of structured programming. Several of his fundamental concepts are used in the development of programming languages. He won the Turing Award in 1972.

CWI itself featured prominently at a strategic conference on ICT, Knowledge and Innovation, organized by the ministry of Economic Affairs in The Hague. Here captains of industry were informed about developments in the ICT area. Furthermore, CWI as representative of the World Wide Web Consortium (W3C) in The Netherlands, consulted with under-secretary Vliegenthart of the ministry of Health on a project to increase Internet accessibility for the impaired.

Mathematicians and computer scientists are sometimes blamed for their supposed individualism and for considering team-work as a necessary evil. Of course, CWI is in no want of strong personalities – no creative research is done without them. Fortunately, the institute, supported by its strong network of relations, has always been able to maintain a research climate and an infrastructure doing justice to both collective and individual, while avoiding conflicts between the two. Together with its healthy mix of fundamental and application-oriented research, and considerable attention for knowledge transfer to trade & industry and to society at large, CWI is well prepared to keep playing an important role in research and society, of which it is – after all – a part.

Gerard van Oortmerssen

General Director CWI

No Minors Allowed

Research project: Networks and Optimization
Project leaders: A.M.H. Gerards and A. Schrijver
E-mail: Bert.Gerards@cwi.nl

Introduction

Consider a graph, consisting of a number of nodes and connecting edges. The graph obtained by removing some edges is a substructure of the original graph. It is an example of a 'minor' of a graph. Roughly speaking, a minor of a structure is a projection of a substructure. The notion is used for various objects of geometric nature, like graphs, polyhedra, and sets of vectors. These structures are widely used in modelling real-world problems in areas like transportation, production planning, telecommunication, chip design, and molecular biology. The problems are usually hard to solve, but become easier to handle when the modelling structure has certain properties. Minors play a prominent role in the analysis of such properties. For example, a graph is called 'planar' if it can be drawn in the plane such that its edges do not cross each other. A minor of a planar graph is planar as well. Planarity can then be characterized by listing all minors of graphs with the property that any further 'minorization' yields a planar graph. These minors play a crucial role in studying structures like graphs, etc. They are called 'forbidden' because, if by taking minors a structure can be reduced to such a minor, that structure cannot have the property in question (planarity in our example). Knowing all forbidden minors provide an easy way of showing that a particular structure does not have a certain property. The analysis of structures like graphs, polyhedra and sets of vectors is the main research theme of the CWI project 'Networks and Optimization'.

Minors of graphs – planarity

It is well-known – from children's puzzle books – that three cities cannot be connected to an electricity plant, a gas factory, and a water supply station such that the connections do not cross. Figure 1a shows connections with only one crossing, but avoiding also that one is impossible. One might wonder if cross-free connections would be possible if they were allowed to go via distribution stations, shown as black dots in Figure 1b. Of course not. If we had such cross-free connections via distribution stations, we could *contract* the connections between the facilities and the distribution stations as if they were elastic, leading to cross-free direct connections to the cities, and we know that to be impossible. Needless to say that cross-free rerouting the connections in Figure 1c, generated from 1b by adding two electricity lines, is certainly impossible: it is already impossible after *deleting* those two extra connections. We see that we can establish the impossibility of cross-free rerouting of a given network by constructing the network in Figure 1a as a 'minor', that is by deletions and contractions. This is the basic idea behind minors.

Like many practical problems the cross-free routing problem above can be modelled as a problem on a graph. Graphs are often visualized in a plane by representing its nodes as points and its edges as curves connecting pairs of points. In such a visualization typically some of the curves cross. If it is possible to avoid that we call the graph

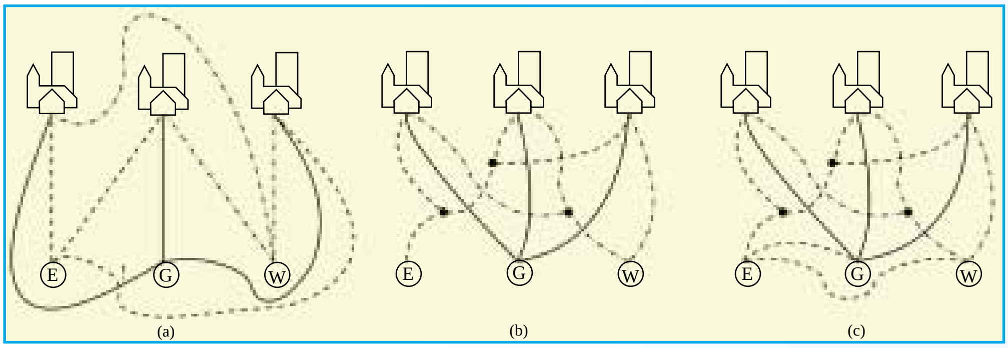


Figure 1: A puzzle.

planar. The above routing problem is just a question of establishing the planarity of a graph. Now a classical theorem of Kuratowski says that there are exactly two forbidden minors for planarity. They are shown in Figure 2. As soon as a graph has one of these as a minor, it cannot be planar. Many graph properties have been characterized by providing the list of their forbidden minors; each time it turned out to be finite. This prompted Wagner's conjecture: every graph property that is preserved under taking minors has only finitely many forbidden minors. In a long series of papers, Robertson and Seymour confirm this – without doubt the deepest result in graph theory (the Graph Minors Theorem).

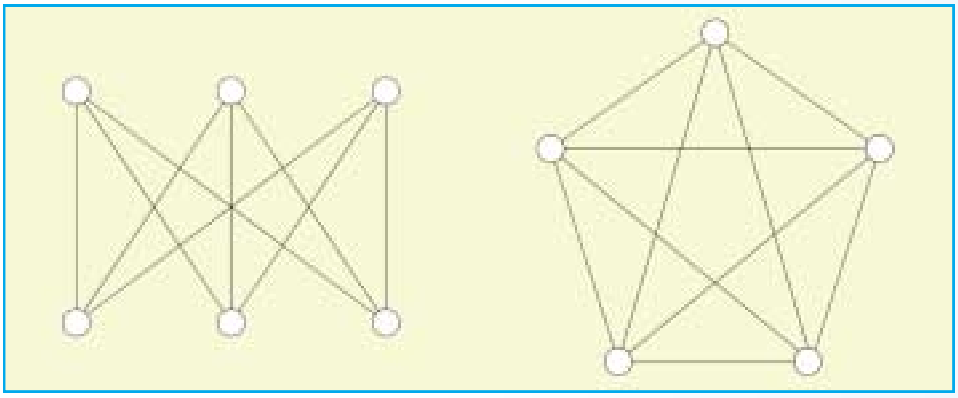
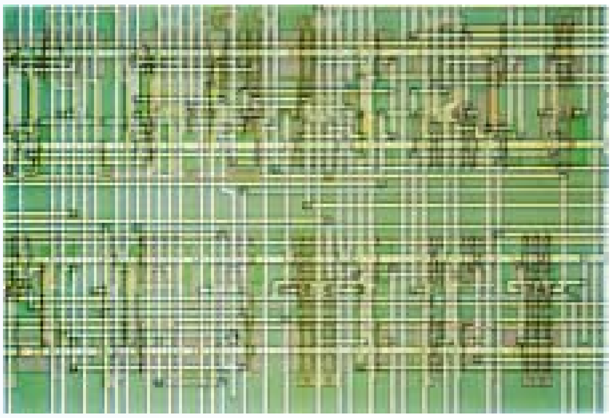


Figure 2: The forbidden minors for planarity.

Minors of polyhedra – integrality

The following problem, called the 'disjoint paths problem', has been intensively studied at CWI, particularly by Lex Schrijver. It turns up in, e.g., designing computer chips and transportation problems. In a graph, which we can imagine as a street map, we are given a number of pairs of nodes, called 'terminal pairs'. The task is to find for each terminal pair a path that connects its two nodes, such that all paths use different streets. An obstruction for the existence of such paths is the following. Suppose we partition the set of nodes into two sets, such that the number of terminal pairs with one element in each of these sets exceeds the number of edges connecting the sets. Then, clearly, the disjoint paths problem has no solution: the connecting edges form a 'bottleneck'. Identifying such bottlenecks is very useful: they show why disjoint paths do not exist and indicate where to enlarge the graph to make them possible. (Unfortunately, the nonexistence of bottlenecks does not guarantee the existence of the desired paths.)

The disjoint paths problem can also be modelled along the lines of linear programming – a well-known technique in the field of operations research to solve large sets (possibly millions) of linear inequalities. With two variables, for example, such inequalities delimit in the plane a convex polygon as the solution space. In higher dimensions the solution space is a polyhedron. Identifying a bottleneck for the disjoint paths problem turns out to be equivalent with finding among all points with integer coordinates inside the polyhedron those points for which the sum of the coordinates is minimal. This is *an integer linear programming problem*. Many real-world optimization problems can be modelled as such. In general, integer linear programming problems are hard to solve (they are 'NP-hard'), contrary to ordinary linear programming problems where the variables are allowed to take fractional values as well. For these problems efficient algorithms are available (efficient in the sense that the running time is bounded by a polynomial in the size of the input). Now, if the polyhedron is 'integral', i.e., all its corner points have integer coordinates, we may drop the integrality condition on the variables in the underlying equations, and thus turn their solution into a linear programming problem. This is a major issue in polyhedral combinatorics – another important research area at CWI.



Excerpt of a chip (2170-fold magnification). The disjoint paths problem, that shows up in chip design, is intensively studied at CWI. (Forschungsinstitut für Diskrete Mathematik, Rheinische Friedrich-Wilhelms-Universität, Bonn.)

So, turning to the disjoint paths problem we would like to know when its associated polyhedron is integral. Since in this case all coordinates are nonnegative, this polyhedron has the property that if it is integral, so are its intersection with any tangent coordinate hyperplane $\{x|x_i = 0\}$and its orthogonal projection onto such a hyperplane. The result of any series of these intersections and projections is called a 'minor' of the polyhedron. Recently Guenin confirmed a long standing conjecture of Seymour, stating that the integrality of the polyhedron associated with a disjoint paths problem has exactly one forbidden minor: the polyhedron corresponding to a disjoint path problem with only six edges and four terminal pairs (see Figure 3). Guenin's proof is long and complicated. Early 2001, Schrijver provided a very simple and short proof of this prominent result. The polyhedron modelling the disjoint paths problem also models the 'max-cut problem'. This problem is relevant in solid state physics and statistical physics, in particular for calculating the ground states of 'spin glasses'. Guenin's result ensures that for a large class of graphs the max-cut problem can be efficiently solved by linear programming techniques. Much research on the

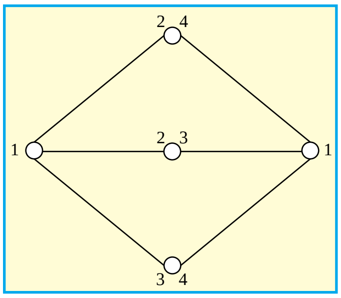


Figure 3.

polyhedral aspects of the max-cut problem has been carried out at CWI by Monique Laurent.

Minors of vector configurations – representability

We have seen how important it is to establish the integrality of a polyhedron. Another approach to this question is to study collections of vectors ('configurations') with binary coordinates, taking the values 0 or 1. Here a crucial role is played by the 'Fano configuration', see Figure 4. It consists of seven points and seven curves (six of them straight, one a circle). Each curve covers a triple of the seven points. Now try to reallocate the seven points such that for each of the seven curves only the corresponding triple ends up on a straight line. This turns out to be impossible. One explanation uses the binary vectors labelling the points in Figure 4. Note that the three vectors corresponding to a triple of points on a curve add up to 0, that is, if we use as addition rule: 1+1= 0 (in mathematical terms: if we consider the entries 0 and 1 as elements of the field with two elements). Actually, a triple lies on a single curve if and only if the three corresponding vectors are linearly dependent over that field. Now, if the triples on curves (and only these) could be reallocated to triples on straight lines, then the seven binary vectors could be replaced by real vectors in 3D preserving all linear independencies and dependencies among the vectors (just consider the plane with the reallocated points as a plane in 3D, not containing the origin). We say that the configuration is 'representable' over the reals, or 'regular'. Turning thus the problem of redrawing the configuration in Figure 4 with straight curves into a linear algebra problem, one can easily show that the Fano configuration has no straight line drawing.

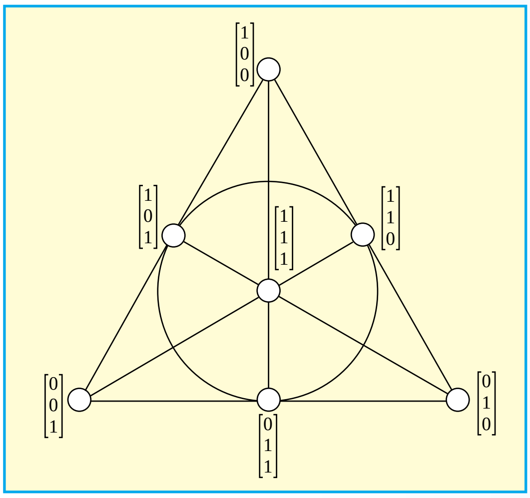


Figure 4: The Fano configuration.

Representability of a configuration is preserved by deleting elements from the configuration as well as projecting the configuration in the direction of one of its members onto a hyperplane not containing that member. Any image of a configuration under a series of these operations is called a 'minor' of the configuration. Thus the above result can be stated as: the Fano configuration is a forbidden minor for regular configurations. Tutte proved in 1958 that regularity has exactly one other forbidden minor: the 'dual Fano configuration'. (Gerards gave a short proof for this result in 1988.) In 1980 Seymour gave an other characterization of regularity. Every configuration of binary vectors consists of 'bricks' pasted together according to a composition rule which we do not detail here. A configuration is regular if and only if all its bricks are regular. Seymour's result says that each regular brick is either 'isomorphic' to a special regular configuration with ten elements, or is (co)graphic. Graphic and cographic configurations are related to graphs and always regular. Moreover, they can be efficiently recognized. Hence, Seymour's theorem is not only a deep explanation of regularity, but also implies an efficient algorithm for recognizing this property.

There is a direct link between regular configurations of binary vectors and integral polyhedra. The link is laid by 'totally unimodular matrices' – matrices with all subdeterminants equal to –1, 0, or 1. On the one hand, Hoffman and Kruskal showed already in 1956 that if A is a matrix with integral coefficients, then the polyhedron $\{x \mid Ax \leqq b, x \geqq 0\}$ is integral for each integral vector b if and only if A is totally unimodular. On the other hand, establishing the total unimodularity of a matrix has been shown by Tutte to be equivalent to establishing the regularity of a certain configuration of binary vectors. Hence, once again purely mathematical considerations turn out to be relevant for practical problems: abstract questions concerning regularity – and representability in general – of configurations lead to a better understanding of computational issues in integer linear programming. A detailed treatment of these issues is given in the book *Theory of Linear and Integer Programming* by Lex Schrijver (1986, John Wiley & Sons, Chichester).

As mentioned Tutte showed that there are (up to isomorphism) only two forbidden minors for regularity of configurations of binary vectors. Let us now turn to configurations of vectors over arbitrary fields and consider the general issue of forbidden minors for representability of such configurations over a specific other field. Tutte also showed that for such general configurations there is only one forbidden minor for representability over the field with two elements. Moreover, in the early seventies, Reid announced the forbidden minors for representability over the field with three elements; again up to isomorphism there are four of them. However, he never published his proof, the first published proofs were by Bixby and Seymour. Motivated by these results and the at the time existing forbidden minor characterizations for graphs, Rota conjectured in 1972 that for any finite field there are only finitely many forbidden minors for representability over the field. Rota's conjecture is still open. Actually, it lasted till the nineties before the issue was resolved for the next special case. CWI researchers Jim Geelen, Bert Gerards, and Ajai Kapoor proved that up to isomorphism there are only seven forbidden minors for representability over the field with four elements. For no other finite field the issue has been resolved.

In view of Robertson and Seymour's Graph Minor Theorem and Rota's conjecture one might wonder if all properties of configurations that are preserved under taking minors can be characterized (up to isomorphism) by finitely many forbidden minors. The answer is no. However, it could be true for properties of configurations over finite fields. A first small step in this direction has been made recently by Geelen, Gerards, and Whittle: any family of configurations over the same finite field, all with the same 'width', contains two members such that one is isomorphic to a minor of the other. This result implies its counterpart for graphs with bounded 'tree width' by Robertson and Seymour, which actually was the first step in their proof of the Graph Minors Theorem. Many computational problems that are hard for general graphs become efficiently solvable for graphs with small tree width. For this reason, tree width as defined by Robertson and Seymour receives much attention in the computer science community.

MODELLING, ANALYSIS AND SIMULATION OF BIOLOGICAL PROCESSES

Research project: Applications from the Life Sciences (cluster MAS)
Project leader: J.G. Verwer
E-mail: Jan.Verwer@cwi.nl
URL: http://dbs.cwi.nl:8080/cwwwi/owa/cwwwi.print_projects?ID=5

Introduction

Many biological processes can be modelled using systems of (partial) differential equations. The development and (computational) analysis of these models is not only of increasing importance for the understanding of the biological processes and for the verification of hypotheses about the underlying biology, but also for the development of new medicines. At CWI various applications from biology are studied, resulting in mathematical research on diffusion-reaction systems, integro-differential equations, and mixed parabolic-gradient systems. The research spans the full spectrum from basic to applied and theoretical, comprising modelling and simulation, analytical, numerical and computational topics. Other aims are to promote interaction between biology and mathematics and computer science, and to give advice.

Development of the research area

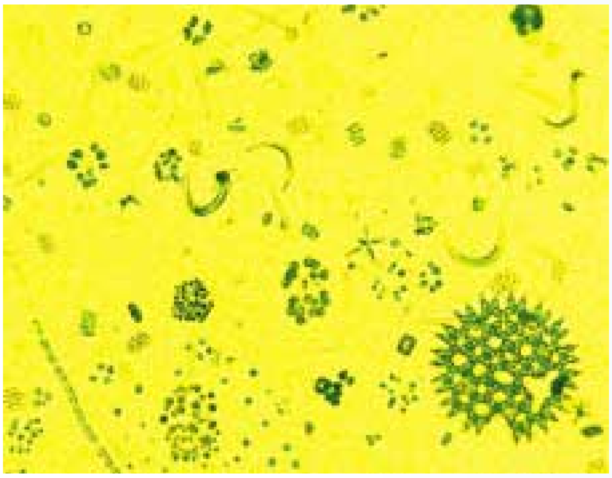
During the last decade, research in the life sciences has become increasingly quantitative as new technologies facilitate more and comprehensive experiments, especially on a microscopic scale. A good example is the complete genomic sequencing of organisms. As a consequence, mathematics and computational science have become crucial for the study of complex models of biological processes. Internationally, this has recently led to the new SIAM Activity Group on the Life Sciences to provide a cross-disciplinary forum for catalyzing mathematical research relevant to the life sciences. The aim is to facilitate rapid diffusion of new mathematical and computational methods in the life sciences. In 1999 CWI decided to invest in research related to important life science applications. A major effort concerns the mathematical research in the cluster MAS of various biological processes in close cooperation with both theoretical and experimental biologists from groups working in cell biology, neurobiology and microbiology. The combination of the expertise and strengths of the participating groups, which almost all have their basis in Amsterdam, is rather unique in The Netherlands: it combines excellence in the areas of quantitative biochemistry, microbiology, cell biology, aquatic microbiology, and neuroscience with the expertise at CWI in mathematical analysis and computational science. The current research activities are detailed below.

Macro-molecular crowding in Cell Biology

Living cells are densely packed with macromolecules (proteins). Hence, larger molecules may diffuse much slower than smaller ones. The influence of this 'crowding' effect on the uptake of, e.g., glucose by a cell is largely unknown. This process is modelled as a system of partial differential equations (PDEs) of the reaction-diffusion type, with the concentrations of the proteins as variables. With a finite diffusion rate the model yields concentration gradients for the proteins (some high near the membrane and low in the kernel, others vice versa). Since the concentration values determine the flux through the system, these gradients theoretically influence the uptake of 'food'. CWI addressed the question whether the intracellular environment delays or perhaps even prohibits a signal to travel from membrane to genes via a metabolic pathway (i.e., a chain of chemical reactions regulating the production of some component or its transport through a membrane). We studied two chains of reactions: the Mitogen-Activated Protein (MAP) kinase cascade, and the phosphotransferase (PTS) system. Both are well-known chemical reaction schemes with well-measured reaction rates. For the rate of diffusion for the different proteins we used values based on *in vivo* measurements. Extensive simulations revealed that diffusion does not limit the uptake of glucose by *E-coli*, but it does for larger organisms (say 10–30 times as large as *E-coli*). This could explain why the PTS system is only found in bacteria.
CWI's research is conducted in the framework of the recently formed Silicon Cell Consortium. The other partners are: the Swammerdam Institute for Life Sciences and the Institute for Informatics (both University of Amsterdam), and the Institute for Biomolecular Sciences (Free University Amsterdam). The long-term goal of the Consortium is the computation of Life at the cellular level. Its present, more modest ambitions are focused on: (i) computational modelling of specific modules of the cell metabolism, and (ii) development of specific tools and approaches for modelling living cell behaviour.

The dynamics of phytoplankton blooms

Phytoplankton plays a crucial role not only as the basis of the food chain in lakes and oceans, but also in climate modelling. The oceans take care of a reduction of approximately 15% of the greenhouse gas CO_2 in the atmosphere. The underlying mechanism behind this is CO_2 absorption by specific types of algae that sink to the bottom of the oceans. Since algae require light, the question is how these sinking



Phytoplankton. (Courtesy Dr Petra Visser, IBED-Microbiology, University of Amsterdam.)

species can survive. The model developed to study this phenomenon is an integro-PDE of the reaction-drift-diffusion type. The integro-part mimics the growth of the algae as a function of the light intensity. The PDE part consists of a turbulent diffusion term and adrift term describing the vertical movement caused by sinking. Analysis and computer simulations reveal that for intermediate values of turbulence there is a balance between population growth and death by sinking or mixing, so that the population can survive near the surface even if the individuals die at the bottom. Presently two extensions of the model are investigated. First, an ocean is better modelled with two water layers, an upper well-mixed warm one and a lower cold one with negligible mixing. Preliminary results show that the above conclusions stay unchanged under this modification, and the quantitative carbon transport through the thermocline is subject of further numerical study. Second, the dynamical processes in the competition for light between several phytoplankton species with different growth and sinking rates will be studied numerically. The research is carried out in cooperation with the Microbiology group at the University of Amsterdam.

Computational Neuroscience

Axon development. (Courtesy: Dr Arjen van Ooyen, NIH.)

In this project CWI examines, in cooperation with the Netherlands Institute for Brain Research, a mathematical model for the development of neuronal connections in the nervous system. This model has been developed by brain researchers to describe the biochemical mechanisms steering axon guidance, bundling and debundling. A better understanding of these mechanisms is relevant for developing strategies to stimulate regrowth and reinnervation after nerve damage. The diffusion processes, the chemical interactions and the positions of the axons are modelled by systems of parabolic equations coupled with gradient-type equations for the positions of the axons. The model is solved numerically on a spatial grid, and the gradient equations are approximated by interpolation on this grid. CWI focuses on the numerical solution and the development of a software package to be used by brain researchers.

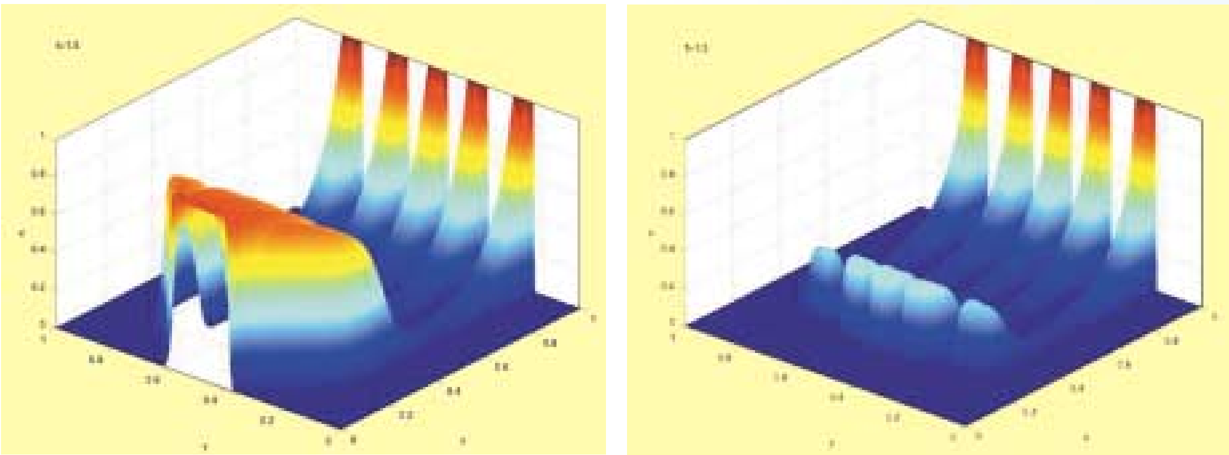
Numerical simulation of taxis-diffusion-reaction models

How far can tumour cells invade the surrounding tissue and under which circumstances does the initial cell mass break up into pieces? How does a tumour induce blood vessel development, in order to procure nutrients for further growth? Such questions are modelled by a taxis-diffusion-reaction system. Because model parameters and initial data are often only known as crude approximations, accuracy requirements are modest. Nevertheless, certain qualitative properties should be resolved correctly, and numerical solutions should be obtained within a fairly short computing time to enable extensive parameter studies. The solution of the taxis-diffusion-reaction equations starts with spatial discretization, resulting in a system of positive, nonlinear ordinary differential equations in the time variable. The solution of this system should be efficient and robust. Two (splitting) methods for the integration in time of these equations are studied. One is based on operator splitting, separating taxis and reaction-diffusion operators. The other is Approximate Matrix Factorization which can be seen as splitting on the linear algebra level. For this subject CWI cooperates with the Martin-Luther-Universität Halle-Wittenberg in Germany.

Many biological processes can be modelled by a system of partial differential equations of the reaction-convection-diffusion type

$$\frac{\partial \mathbf{c}}{\partial t} = \mathbf{R}(\mathbf{c}) + \nabla \cdot \mathbf{v}(\mathbf{x}, \mathbf{c}) + \nabla \cdot (D(\mathbf{x}, \mathbf{c})\nabla \mathbf{c}),$$

for $\mathbf{x}$ in some domain Ω and $t \in (0, \infty)$. This is a continuity equation for the components in a biochemical reaction system. The variables $\mathbf{c}$ in these equations are the volume concentrations of the components in the reaction system described by the reaction rate vector $\mathbf{R}(\mathbf{c})$. The reaction rate of a specific component is the sum of all production and loss terms for that component in the biochemical reaction system. The motion of the "particles" in space and time is a composition of a directed motion driven by an external force, the convection $(\nabla \cdot \mathbf{v}(c))$ and a diffusion term $(\nabla \cdot (D\nabla c))$. The last term is either Brownian molecular diffusion if the fluid is at rest, or describes the turbulence or mixing of the fluid. The equations in the system are in general only coupled through the reaction term. The mathematical model as stated above is not complete. Another set of equations is needed to describe the behaviour at the boundary of the domain and an initial solution on Ω has to be provided. Characteristics of the solution of such systems are positivity (negative concentrations do not arise) and mass conservation (the total amount of specific combinations of components does not change). It is important that a computational solution should mimic these characteristics. The usual numerical approach is to first discretize the spatial operators on a grid. The resulting (large) system of ordinary differential equations has to be integrated in time. This system is difficult to solve, since it is strongly nonlinear and stiff (the solution consists of components slowly and rapidly varying in time). The spatial operators of a continuity equation are discretized in a natural way by expressing them as fluxes through the boundary between grid cells. The diffusive motion is thus seen as driven by the concentration gradient. The resulting discretization is by definition mass-conservative ("what goes out of one cell enters the next"). How to integrate the equations in time is often less obvious. Keywords are preservation of positivity, stability, and efficiency. These objectives are often contradictory. The stiffness of the system is usually tackled by using (semi-)implicit methods. Such a method computes the next time step from the previous one by solving a set of implicit relations. It allows large time steps without causing instability, but is expensive per time step. Efficiency is obtained by splitting methods, either at the operator level ("first solve the motion, then the reaction"), or at the linear algebra level (Approximate Matrix Factorization).



Results of a simulation of tumour induced angiogenesis. The simulation assumes a circular tumour at the point (x=0, y=0.5) and a blood vessel along x=1, from where at the start of the simulation new capillary sprouts grow. The pictures show how the endothelial cell concentration develops without therapy (picture on the left) and with therapy (on the right). Without therapy the endothelial cells reach the tumour. With therapy no blood supply of the tumour can be established. If the simulation time is increased the already existing blood supply even degenerates. (Courtesy: Alf Gerisch, Institut für Numerische Mathematik, Martin-Luther-Universität, Halle, Germany.)

A dynamic model for biofilms on wet surfaces

Biofilms are thin layers of mucus and bacteria that line all wet surfaces. The mucus is secreted by the bacteria and serves as an external skeleton, shielding the bacteria from water currents. The bacteria feed on organic material caught in the mucus. In preparing drinking water, measures are taken to avoid biofilms, which are considered harmful. CWI advises the Aquatic Ecology and Ecotoxicology Group at the University of Amsterdam on the development of a dynamic model that will allow a better understanding of the exchange of solutes and particles between biofilms and drinking water. The advice concerns both the mathematical model and numerical topics. The aim is to build a numerical simulator for biologists to validate the model against experimental data.

Example of biofilm: human dental plaque. (Photo ASM Biofilms Online Manual.)

Towards Multimedia on the Semantic Web

Research project: Distributive Adaptive Hypermedia
Project leaders: Lynda Hardman, Jacco van Ossenbruggen
E-mail: Lynda.Hardman@cwi.nl, Jacco.van.Ossenbruggen@cwi.nl
URL: http://www.cwi.nl/ins2/

Introduction

The rapid growth of the World Wide Web, in quantity as well as quality, has induced considerable research effort in the field of distributed, adaptive hypermedia. Broadband Internet will soon enable high-quality multimedia presentations to be sent across the Web. To realize this in practice, however, several fundamental problems characteristic for multimedia should be overcome. Taking the SMIL language as a starting point, CWI has developed a flexible, interactive environment enabling the generation of multimedia presentations on the Web. Current research is directed to the incorporation of semantic features into such presentations. The bicycle, symbol par excellence *for the Dutch, is used throughout as an example.*

First and second generation Web pages

The first Web pages were all hand-made and, hence, very labour-intensive. They were made in the text-oriented markup language HTML. Such pages are still being created and remain an important source of information on the Web. However, the enormous growth of the Web has necessitated the move towards automatically generating Web pages. Moreover, user demands soon exceeded text-based applications. In the second phase Web pages could be generated on demand, for example by filling in templates with content retrieved dynamically from a database. This enabled the content provider (the Web equivalent of the traditional book publisher) to publish information on a particular topic in the form of a diverse set of Web pages, all providing the same basic information, to be used on a variety of devices, under varying network conditions, and intended for users with their own information display preferences.

This approach works with Web pages that are text-based, but may include pictures, animations, and occasional audio and video fragments, acting as decorations, i.e., inserted in some fixed place in the text. Multimedia features can, however, be employed more fruitfully and convey the content to the user more effectively. For example, when describing how the forces of the feet on the pedals of a bicycle drive the back wheel, a synchronized combination of animation, commentary and text labels could be employed. Such higher-quality multimedia presentations are beginning to take a hold on the Web. In particular, with the upcoming release of the Synchronized Multimedia Integration Language (SMIL 2.0) recommendation of the World Wide Web Consortium (W3C), multimedia will become a part of the established W3C integrated suite of languages. Research at CWI has contributed significantly to both the roots of the SMIL language and the development of the new version.

Multimedia document processing has requirements fundamentally different from text, which are more difficult to incorporate in a traditional electronic publishing environment. For example, multimedia document formatting cannot be based on text-flow layout rules, because these, for one, do not take into account temporal aspects. These differences make it extremely difficult to apply most commonly used solutions because of their bias to text-oriented solutions.

Present-day text processing enables individual users to tailor text documents to their own specific needs. The same should apply to multimedia. Current trends are that the Web is accessed using devices other than a standard PC, that the primary mode need not be textual, and that the contexts in which the information is used are more diverse. For example, using a mobile phone the primary mode is speech and the environment may be a market-place. Users are also becoming more demanding and expect content to be supplied in their own language and to take their cultural background into account. The research group at CWI has developed multimedia models and techniques that go beyond current text-oriented approaches and can be employed to cater for the required diversity.

Multimedia adaptation

Multimedia adaptation, such as creating a set of Web pages for presenting information in different settings, can be carried out both in the user device and by the content provider. SMIL already provides for the user device adaptation by including all the options for each eventuality in a single document. When the presentation is started up on the user device, information can be extracted from the user's environment and used to select the appropriate media items that should be downloaded for viewing. For example, if the device has no display, then the spoken commentary could be given with extra inserts detailing the information in the video. The problem with the SMIL solution is that it requires all options, and all combinations of these options, to be explicitly enumerated and included by the content provider in the document. Hence, an alternative more flexible approach is needed.

Such an approach is to generate individual, tailored documents on demand, based on the prevailing conditions, rather than requiring the content provider to predict these beforehand. The CWI group has constructed an environment which allows variations in end-user platform, network conditions and user preferences to influence the multimedia document generation process.

The Cuypers architecture

The Cuypers environment (see Figure 1), named after the Dutch architect Pierre

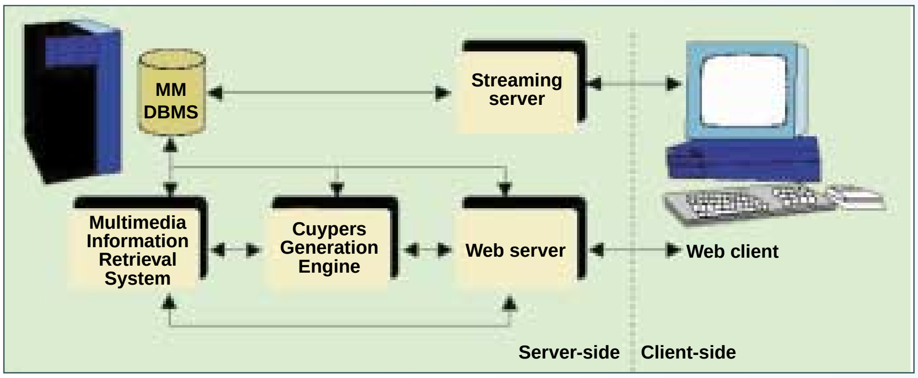


Figure 1.

Cuypers who designed the famous Rijksmuseum in Amsterdam, is a prototype hypermedia generation environment which is able to select media items from a database or the Web and combine them into a multimedia presentation. During the generation process, network conditions, models for discourse and interaction between user and content provider, user preferences and other varying conditions can be taken into account. The result of the process is the generation of a multimedia presentation that can be transmitted over the Internet and played on the user's device.

After the relevant media items have been retrieved, they need to be grouped into a coherent presentation that conveys the intentions of the author and is also adapted to specific user requirements. In order to achieve an adequate level of adaptation, presentation independent abstractions are required. There are various levels of such abstractions, ranging from very high-level abstractions, that focus on the semantics but say little about the presentation-oriented details, to the presentation models used by concrete multimedia file formats, that fully incorporate temporal and spatial aspects. Generating a readily playable multimedia presentation from a high-level semantic description involves bridging a large gap. To bridge this gap, the generation process is broken down into a number of steps (see Figure 2).

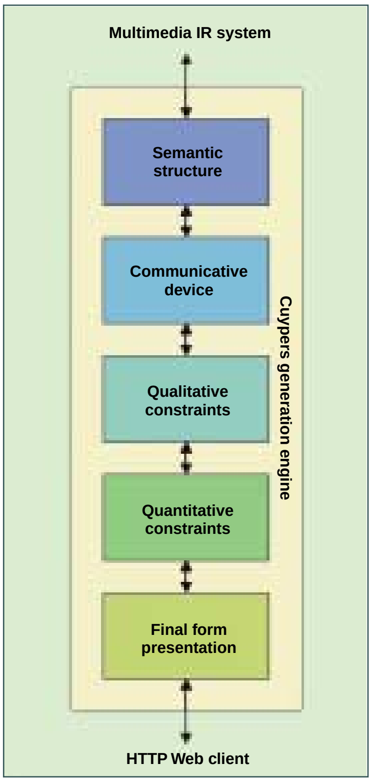


Figure 2.

The process starts with the high-level abstraction that describes the semantics underlying the presentation. An example of such an abstraction is the rhetorical structure of a presentation. (Rhetoric concerns the convincing aspects of a discourse. This can be effectively deployed in educational multimedia applications, for example how to learn riding a bicycle.) This structure is then transformed into high-level hypermedia constructs, termed communicative devices. These describe a presentation in terms of common hypermedia presentation design patterns. The communicative devices are then transformed into a set of constraints which describe the presentation in further detail by specifying the spatial and temporal relationships among the constituent media items. Constraints are used both on a high level (qualitative) and a low level (quantitative, specifying for example the number of pixels), in order to achieve an adequate balance between numeric precision, required for low-level formatting, and more global specifications, needed to effectively explore the hypermedia design search space.

Implementation techniques: backtracking and constraint solving

While the conceptual steps described above might suggest a straightforward sequential process, this is not the case during the actual generation process. High-level design decisions influence lower-level formatting, but the reverse might also be the case. When low-level rendering fails to meet the specified criteria, higher-level design decisions might need to be reconsidered. For example, if a choice has been made to display an animation together with a text label and it transpires that in order to make the font large enough for legibility the animation cannot be seen, then the decision to display the text label would be changed. Our environment combines constraint solving and backtracking techniques to define the rules that govern this complex processing in a declarative style.

Multimedia on the Semantic Web

Future Web pages (the third generation) will make use of rich markup, based on XML instead of HTML, along with metadata, e.g. encoded in RDF (W3C's Resource Description Framework), making the content not only machine readable, but also machine processable. This will allow not only a more directed search of documents, but also generates applications that utilize the semantic information in documents. E-commerce applications, for example, can make use of agreed-upon semantic markup to produce integrated information from several sources, e.g., if a customer wishes to buy a bicycle with certain features such as light frame, wide tyres, cushioned saddle, etc.

Figure 3. Presentation explaining the clair-obscur technique as used by Rembrandt, generated by the Cuypers system.

The work on the Semantic Web is only starting, and a number of fundamental problems have yet to be solved. For example, in the traditional knowledge representation approach, a domain description, or ontology, such as a standard treatise on the bicycle, can be assumed to be complete, consistent and reliable. On the Web, however, where openness is one of the most important considerations, ontologies may be fragmented, inconsistent and unreliable. These problems apply to all ontologies on the Semantic Web, but for multimedia other problems also hold: some means of relating the terms in the ontology to a media item itself is needed. For example, 'wide tyres' could be associated with an image of a wheel, a spoken commentary 'all terrain tyres' or a Dutch text 'brede banden'. This requires dealing with the diverse data formats of the media items themselves (like *jpeg* or *gif*). In addition, multimedia-specific ontologies may have to be developed to describe notions of the media themselves, for example: a film can consist of scenes, which are made up of sequences, which are made up of shots. It is not as yet clear as to whether the notions of time and space, fundamental characteristics of multimedia, need to be included in the primitives of ontological languages for describing multimedia. For example, is it necessary to be able to use expressions like 'this shot should come before that one', or can we do without using the notion 'before'?

CWI research focuses on semantic markup as a means to facilitate the presentation generation process. When media items are selected from databases, we need information, for example, on the potential role the item may play in the integrated presentation. We investigate to what extent this is possible, and what types of categorizations actually improve the generation process.

Verification of Shared Data Space Systems

Research project: Distributed Systems
Project leader: J.C. van de Pol
E-mail: Jaco.van.de.Pol@cwi.nl
URL: http://www.cwi.nl/~vdpol

Introduction

A distributed system is a collection of autonomous components, connected in a network. In order to accomplish some common tasks, these subsystems must coordinate, for instance by exchanging information or sharing scarce resources. Such systems occur in many safety critical applications, like air traffic control and railway interlockings, or in economy critical systems, like teller machines and e-commerce applications. Such applications impose high demands on correctness and availability. In particular, they must be robust against crashes of subsystems or failing network connections (fault tolerance). Due to the intrinsic complexity of such systems, various coordination models have been proposed over the last decade, of which the shared data space architectures form an important class. A particular example is SPLICE, introduced at Thales (formerly Hollandse Signaalapparaten) for the development of large-scale distributed control systems, such as traffic regulation and command-and-control systems [1]. Recently, Sun Microsystems introduced JavaSpaces™, which provides a coordination model based on a shared data space in order to program distributed Internet applications [2]. CWI research aims at developing mathematically rigorous methods for the correct construction and analysis of distributed systems, based on the shared data space coordination paradigm. In this way, we can scale up our successful work on distributed systems to large-scale control systems and Internet applications.

Shared data space architectures

Direct communication between components increases their interdependence, which has a negative effect on fault tolerance and flexibility. If for instance component *A* wants to exchange information with component *B*, then a crash of *B* causes a problem for *A*. Similarly, if *B* is replaced by another component *C*, we have to change component *A* as well. In a shared data space architecture components don't communicate with each other directly. Instead, they write items to and read items from a shared data space. So both *A* and *B* can write their information to this space, unaware of possible interested components. Similarly, *A* can read its information from the space, independent of whether this information was written by *B* or by *C* (see Figure 1).
Several implementations of these ideas exist, with subtle differences. Especially the 'read' operation comes in several flavours. With 'blocking read', a process that wants to read a value from the data space is blocked until a value is found. Thus, if the process continues, it is certain that a value has been found. A 'non-blocking read' allows a process to continue (after a time-out), even if no value was found. Using a 'destructive read', the read value is removed from the data space. This ensures that the value is read only once. This read is used for modelling resources, or for jobs to be done.

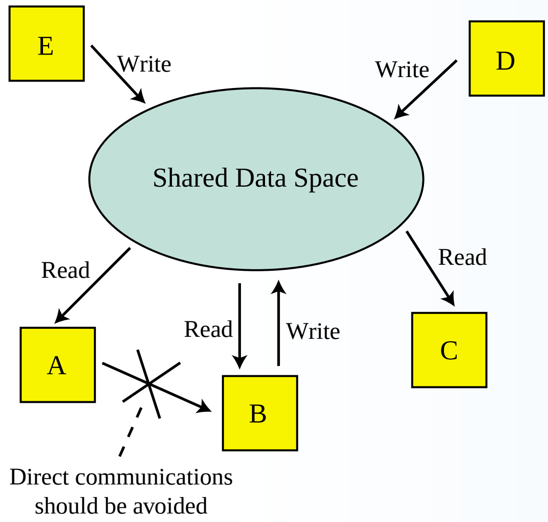


Figure 1: Communication via a shared data space.

Verification of distributed systems

In general, distributed systems are hard to analyze, mainly because the various components are autonomous, so that their activities occur in an arbitrary interleaved fashion. The order in which events occur is not fixed, but depends on unpredictable details such as clock differences and delays in the network. In shared data space systems we must in addition take into account that communication between components is very asynchronous: if two values are put in the data space, it is not guaranteed in which order they are read.

Due to these complexities, the only means for a reliable analysis is a mathematically rigorous approach. At CWI we use a verification methodology, supported by a toolset, which proceeds as follows (see also Figure 2).

First, model the distributed system in the formal specification language µCRL, based on process algebra (3), and developed at CWI in the early nineties. Compile this into a 'linear process equation' (LPO) – a compact internal representation of all possible behaviour of the system. Analyze and simplify this equation by, e.g., compiler optimization, static analysis, and theorem proving. Now unfold the symbolic representation to the full state space. This results generally in a very large graph. We have developed a compact file format for such graphs in the Telematics Institute project SVC (Software Validation Centre). The vertices of such a graph denote the states of the system, its directed edges (arrows) are labelled with an event. The size of the graph is greatly reduced by the preceding simplification. Then minimize the state space using graph algorithms, partly developed at CWI. Finally inspect the resulting state space by simulation, visualization and model checking. Visualization and simulation help to quickly spot errors in small systems. Model checking – an automatic technique to verify several requirements for all possible behaviour of the system – is used for large systems.

Application to shared data spaces

We will illustrate our verification approach with a simple example. Suppose we want to add many numbers (these constitute the shared data space) by means of an arbitrary number of coordinating processes, all running the same program: they successively read (destructively) two numbers (first m, and then n), add them, and write the sum $(m + n)$ back to the data space. If one of the processes can read only a single number, it concludes that this is the final answer, and reports accordingly. Hence, the following code should do the job:

repeat as long as possible:
  read number *m*
  read number *n*
  write number *m* + *n*
if only *m* could be read: write result *m*

The question is: is this code correct, i.e., does it do what it is supposed to do?

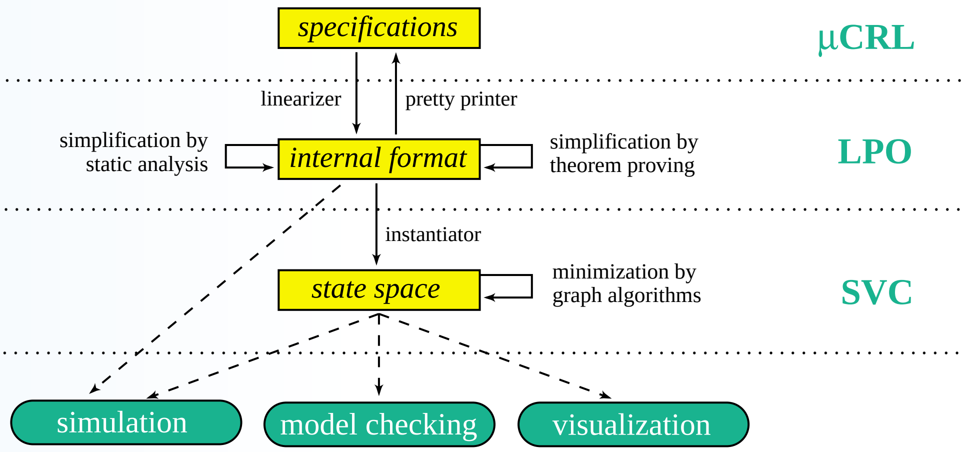


Figure 2: An overview of the µCRL toolset.

For the verification, we must first specify the shared data space (SDS) and the individual programs (ADD). In µCRL this can be done at a high abstraction level. Both kinds of processes can be modelled in a few lines. The complete system is obtained by putting everything in parallel: *SDS* ||*ADD* ||...|| *ADD*. Next, given the initial set of numbers, and the number of cooperating ADD processes, we can generate a picture showing all possible behaviour of the system. Figure 3 shows this for the initial set {2, 3} and two adders. This picture is generated automatically, using the technology of Figure 2. The initial state is labelled 0, and the final state 8 (there are no outgoing arrows). Going through the states (circles) 1, 5, 6, 3, 9, 8, we see that the correct answer is reported: 2 + 3 = 5.

Unfortunately, the system contains an error, because there is a possible interleaving where the wrong answer is reported. Follow for instance the path 0, 1, 7, 2, 10, 8. A close inspection reveals what happens. The first adder takes away element 2. Then the second adder takes away element 3, before the first has had the chance to do so. When the first adder continues, it discovers that it has the last item, so it will erroneously report its value 2. Similarly, the second adder thinks that it is ready, and reports value 3 as the final result. This program can, of course, be corrected in various ways, which we leave as a puzzle to the reader.

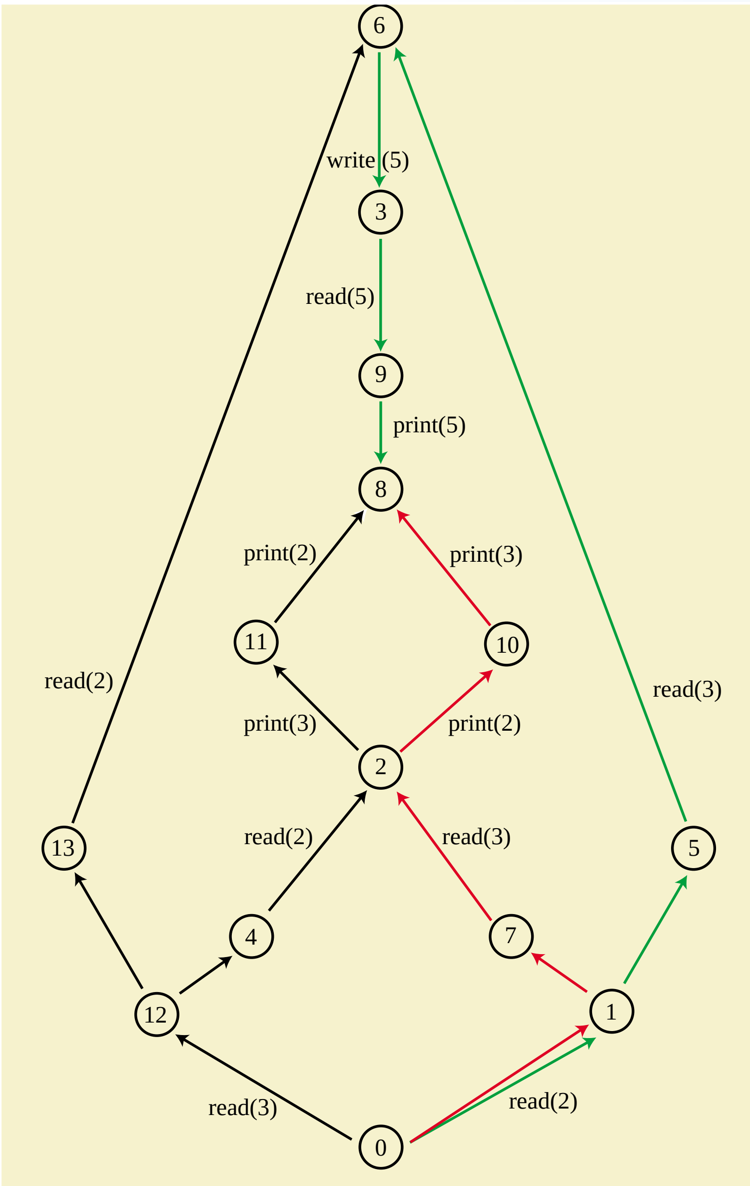


Figure 3: A correct path (green) and a wrong path (red) of the system with two adders on the set {2, 3}. Starting state: 0. Final state: 8.

Some successes

Our verification methodology and the toolset have been developed over the last ten years. We have applied it to many interesting examples, ranging from complicated distributed algorithms to large industrial applications. The project on shared data spaces started in 1999 and is seen as a new challenge for scaling up the capabilities of our techniques and tools. We conclude with an overview of the achievements in industrial case studies during the report year.

Analysis of an industrial lift controller
An extendible system of lift-legs was verified. Several errors were revealed, some of them unknown to the original designers.

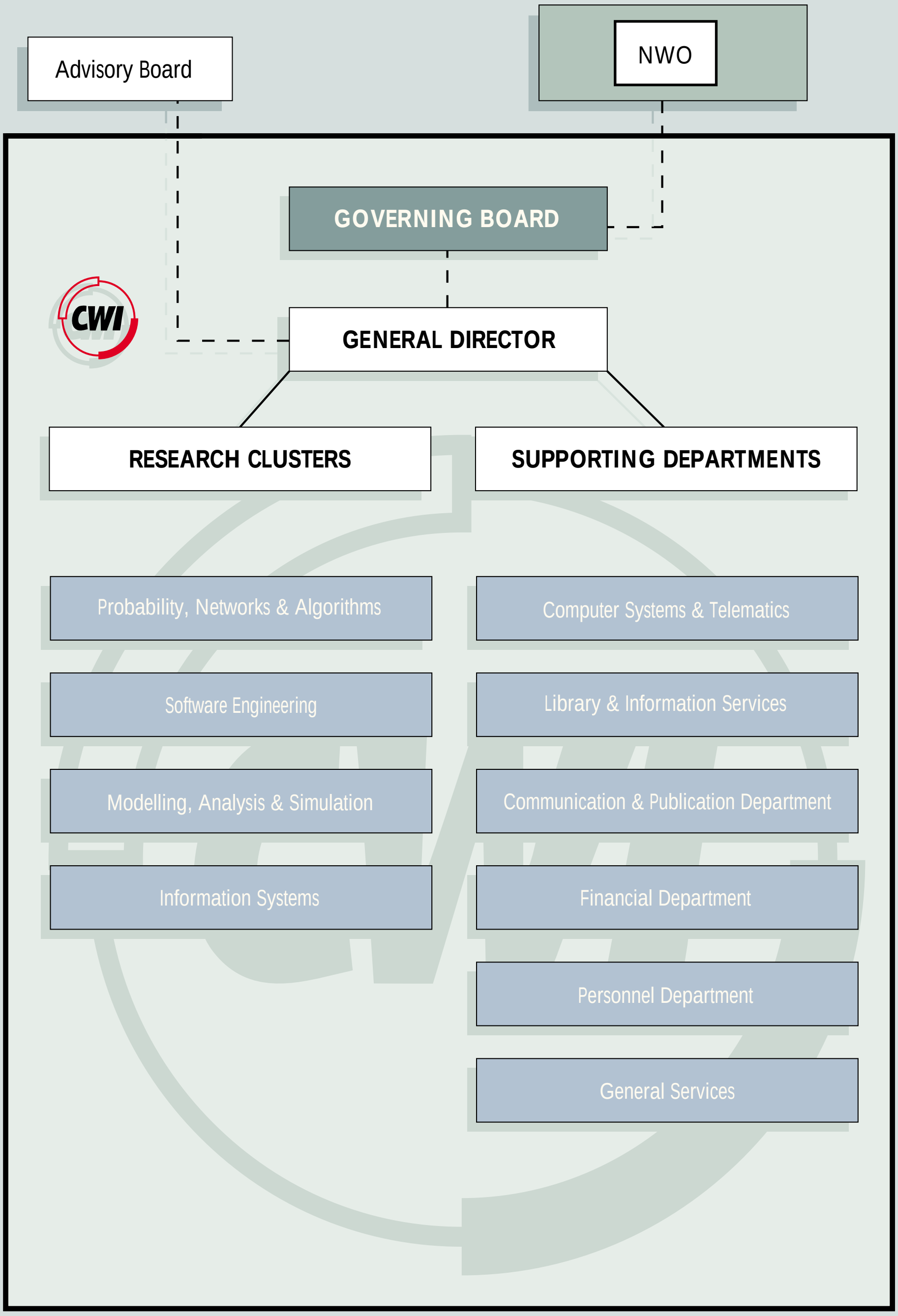
ORGANIZATION

Research

Cluster - Theme	Cluster leader Theme leader
Probability, Networks and Algorithms	**A. Schrijver**
- Networks and Logic – Optimization & Programming	A.H.M. Gerards
- Traffic and Communication – Performance & Control	J.H. van Schuppen
- Stochastics	J. van den Berg
- Signals and Images	H.J.A.M. Heijmans
Software Engineering	**J.W. de Bakker**
- Interactive Software Development and Renovation	P. Klint
- Specification and Analysis of Embedded Systems	W.J. Fokkink
- Coordination Languages	J.J.M.M. Rutten
- Evolutionary Systems and Applied Algorithmics	J.A. La Poutré
Modelling, Analysis and Simulation	**J.G. Verwer**
- Applied Analysis and Scientific Computing for PDEs	J.G. Verwer
- Computational Fluid Dynamics	P.W. Hemker
Information Systems	**M.L. Kersten**
- Standardization and Knowledge Transfer	M.L. Kersten
- Data Mining and Knowledge Discovery	M.L. Kersten
- Multimedia and Human-Computer Interaction	H.L. Hardman
- Visualization	R. van Liere
- Quantum Computing and Advanced Systems Research	P.M.B. Vitányi

Management

Management Team
G. van Oortmerssen (director)
J.W. de Bakker, J.G. Verwer, M.L. Kersten, A. Schrijver (cluster leaders)
D.G.C. Broekhuis (controller)
Governing Board
L.A.A.M. Coolen (director KPN Research), chairman
P.M.G. Apers (University of Twente)
F.A. van der Duyn Schouten (Catholic University Brabant)
K.M. van Hee (Eindhoven University of Technology, director Bakkenist Management Consultants)
H.A. van der Vorst (University of Utrecht)
Advisory Board
J. van Leeuwen (University of Utrecht), chairman
L.A. Peletier (University of Leiden)
J. Ridder (TNO-NITG)
G. Rodenhuis (Delft Hydraulics)
M.F.H. Schuurmans (Philips Research Laboratory, managing director)
M. Westermann
G. Wiederhold (Stanford University, USA)
B. Larrouturou (INRIA, France)
J. Gunawardena (Hewlett Packard Laboratories, Bristol, UK)

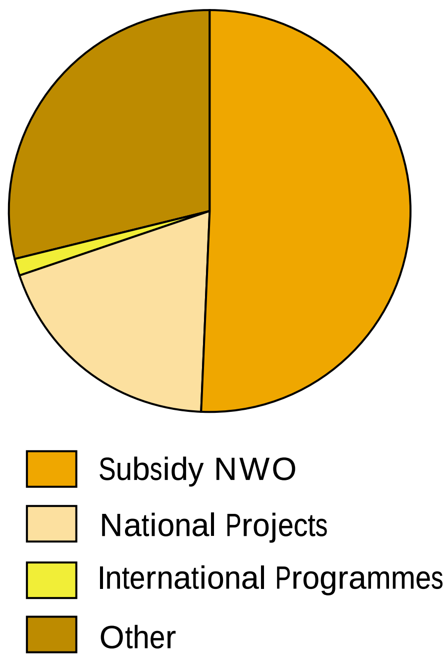

Advisory Board
NWO
GOVERNING BOARD
CWI
GENERAL DIRECTOR
RESEARCH CLUSTERS
SUPPORTING DEPARTMENTS
Probability, Networks & Algorithms
Software Engineering
Modelling, Analysis & Simulation
Information Systems
Computer Systems & Telematics
Library & Information Services
Communication & Publication Department
Financial Department
Personnel Department
General Services

FINANCES, PERSONNEL

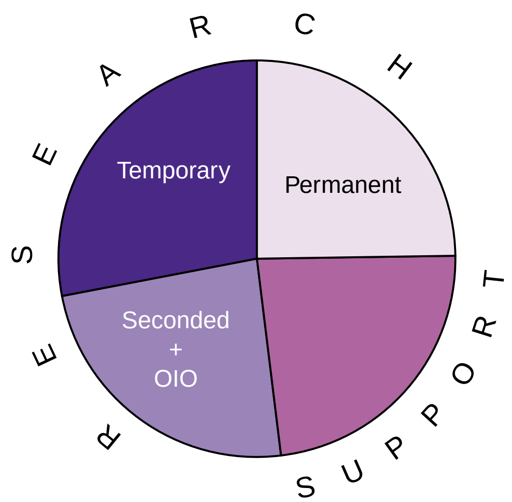
Finances 2000

In 2000, SMC spent Dfl. 35,42 million. The expenses were covered by a basic subsidy from NWO (Dfl. 18,63 million), by income from national projects and programmes (NWO, Telematics Institute, WTCW/ICES-KIS, total Dfl. 7,03 million) and international programmes (Dfl. 0,50 million), and finally by Dfl. 10,60 million as revenues out of third-party-services and other sources.

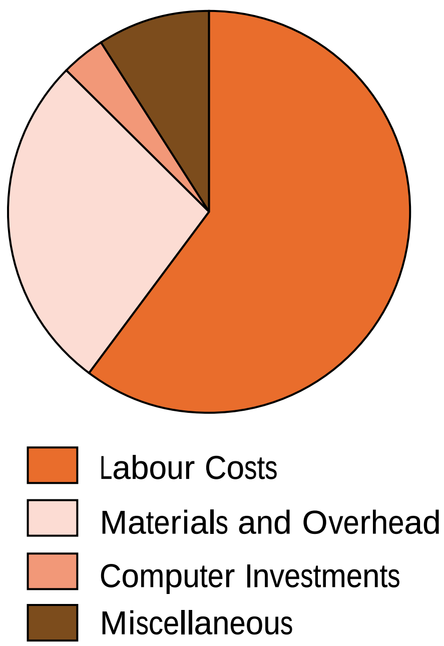
Income CWI

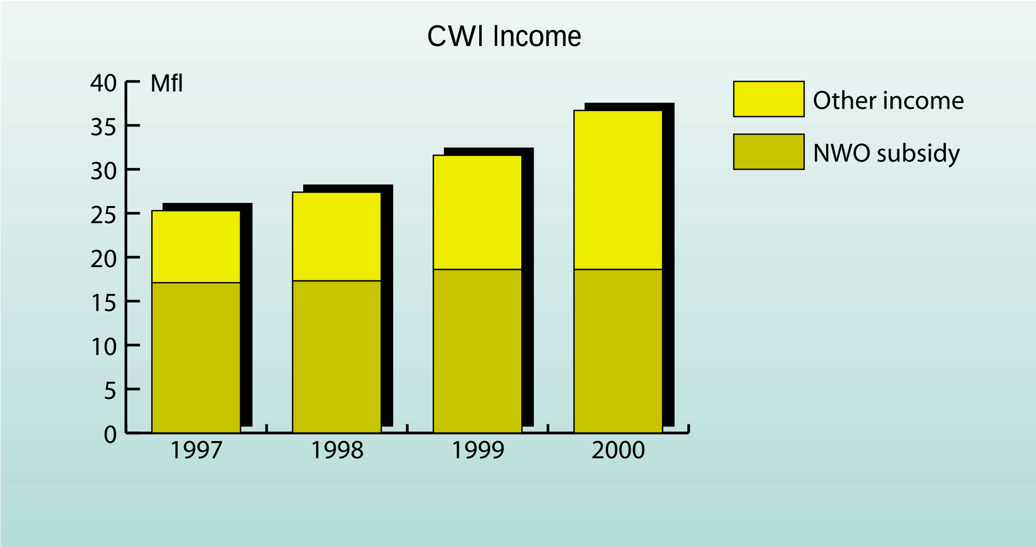


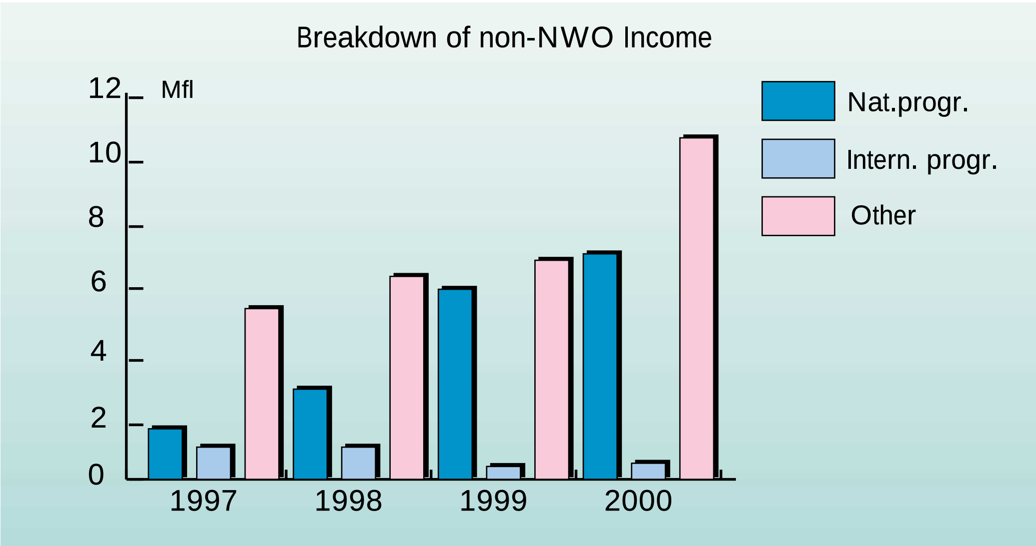
CWI Personnel:
170 fte + 41 fte seconded/OIO
(OIO = graduate student)

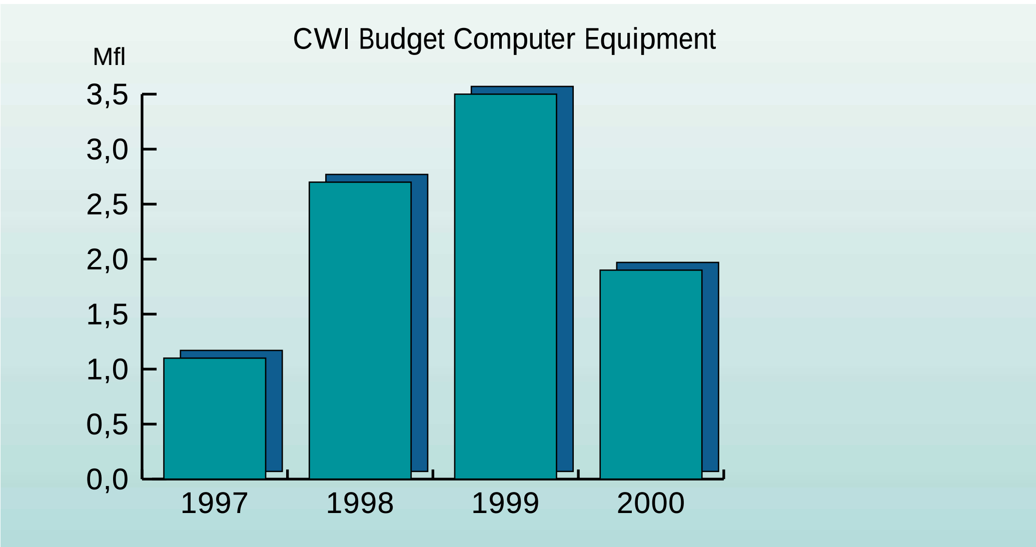


Expenses CWI




CWI Income
Mfl
40
35
30
25
20
15
10
5
0
1997
1998
1999
2000
Other income
NWO subsidy


Breakdown of non-NWO Income
Mfl
12
10
8
6
4
2
0
1997
1998
1999
2000
Nat.progr.
Intern. progr.
Other


CWI Budget Computer Equipment
Mfl
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
1997
1998
1999
2000

CWI PhD THESES

Author
Title
Thesis advisor(s) (for external advisors the university's name is added)

S.M. van Dongen
Graph Clustering by Flow Simulation
M. Hazewinkel, D.J.N. van Eijck

T. Fleiner
Stable and Crossing Structures
A.M.H. Gerards, A. Schrijver

J.E. Frank
Efficient Algorithms for the Numerical Solution of Differential Equations
P. Wesseling (Delft University of Technology), P.J. van der Houwen

W.O.D. Griffioen
Studies in Computer Aided Verification of Protocols
F.W. Vaandrager (University of Nijmegen)

J. Karlsson
Scalable Distributed Data Structures for Database Management
M.L. Kersten

N. Nes
Image Database Management System Design Considerations, Algorithms and Architecture
M.L. Kersten

R. Núñez Queija
Processor-Sharing Models for Integrated-Services Networks
O.J. Boxma (Eindhoven University of Technology), S.C. Borst

P. Olivier
A Framework for Debugging Heterogeneous Applications
P. Klint

P.J. Oonincx
Mathematical Signal Analysis: Wavelets, Wigner Distribution and a Seismic Application
T.H. Koornwinder (University of Amsterdam)

M.R. Vervoort
Games, Walks and Grammars: Problems I've Worked On
M.S. Keane

F.M. Waas
Principles of Probabilistic Query Optimization
M.L. Kersten

CWI RESEARCH PROGRAMMES

Probability, Networks & Algorithms

Cluster leader: A. Schrijver

Networks and Logic – Optimization and Programming

Theme leader: A.H.M. Gerards

Networks and optimization

Design, analysis and implementation of optimization and approximation algorithms for combinatorial problems with the help of methods from graph theory, topology, discrete mathematics, geometry, and integer and linear programming, with special attention to network problems (flows, routing and VLSI-design) and scheduling and timetabling.

Constraint and integer programming

Study of the foundations and applications of constraint programming, in particular the design and implementation of an adequate programming environment for constraint programming, and the use of constraint programming for various optimization problems drawing on integer programming techniques.

Algorithmic and combinatorial methods for molecular biology

Mathematical analysis of molecular structures in biology and the design, analysis and implementation of algorithms for computational molecular biology. The methods come from combinatorics (graph theory and combinatorial optimization), computer science (constraint programming and computational complexity) and mathematical programming (linear, integer, and semi-definite programming).

Traffic and Communication – Performance and Control

Theme leader: J.H. van Schuppen

Communication and computer networks

Development of queueing theoretic models, methods, and algorithms for studying congestion phenomena in communication networks, and of concepts and theory for network control by discrete event and stochastic methods. Research in performance focuses on service integration in communication networks.

Traffic networks

Performance analysis and control of traffic in urban, motorway, railway, air traffic, and other networks, in particular performance aspects of congestion, reliability and availability. The focus is on control of piecewise-linear hybrid systems, system theory of hybrid systems, decentralized failure diagnosis and decentralized control of discrete-event systems, and coalgebra and control.

Control and system theory

Motivated by control and signal processing, the aim is to formulate classes of systems, focusing on Gaussian systems, finite stochastic systems, positive linear systems, and hybrid systems, and to develop theory for realization and system identification. The focus is on computer algebra algorithms for identification and control problems for linear systems, the approximation problem for stationary Gaussian systems, stochastic realization, realization of positive linear systems, and application of positive systems in the biosciences.

Stochastics

Theme leader: J. van den Berg

Probability

Research in probability theory and its applications, with emphasis on reinforced random walks on finite graphs, on mathematical models of biological and physical processes with self-organized critical behaviour, and on random spatial processes and elliptically invariant probabilistic measures, with applications to robotic control and disk scheduling.

Statistics

Fundamental and applied research in statistics, in particular change-point estimation, saddlepoint approximations, Poisson intensity functions, resampling, bootstrap calibration and the Stringer bound statistical methods for compound sums, with applications in finance and ongoing research on estimating the intensity of oil pollution in the North Sea.

Stochastic analysis

Fundamental and applied research, in particular statistics for random processes with applications to mathematical finance, statistical methods for dynamical stochastic models, and statistical inference for stochastic processes and fractal analysis of financial data.

Ergodic theory and dynamical systems

Fundamental and applied research, in particular rigid sets, classification of Bernoulli schemes, fractal analysis, and superexponential convergence.

Signals and Images

Theme leader: H.J.A.M. Heijmans

Content-based coding, indexing, and retrieval

Image coding, feature extraction, and content-based image indexing and retrieval, in particular spatial grouping, multiresolution approaches, integration and interpretation, and feedback-operated user interfaces for applications.

Wavelets and morphology

Research on wavelets and mathematical morphology and their applications, with special attention to the construction of nonlinear multiresolution decompositions, in particular the design of morphological pyramids and wavelets.

Stochastic geometry

Parameter estimation for random sets, spatial statistics, and image segmentation with applications to agriculture.

Software Engineering

Cluster leader: J.W. de Bakker

Interactive Software Development and Renovation

Theme leader: P. Klint

Software renovation

Development of new technology for the renovation and maintenance of legacy systems, including documentation generation and domain-specific lanquages.

Domain-specific languages

The aim is to develop methods for selecting suitable DSL domains, and for capturing domain knowledge into a DSL and its compiler, and to study the practical use of DSLs in various settings such as financial engineering and PDE solving.

Generic language technology

Redesign, reimplementation, and improvement of the ASF+SDF Meta-Environment, in particular the development of a flexible and extensible generic environment to be used in domain-specific language prototyping and software renovation. Specific aims are: compilation of ASF+SDF to C, unparsing, parser generation, global architecture, and application of developed components in other generic environments.

Specification and Analysis of Embedded Systems

Theme leader: W.J. Fokkink

Distributed systems

The study of specification, analysis and testing techniques for computer controlled systems, by developing and implementing algorithms for the analysis and verification of distributed systems with the µCRL toolset. Techniques and algorithms are assessed via case studies (communication protocols, embedded systems, hybrid systems, etc.).

Process theory and verification

Fundamental study of verification techniques, and the development of methods for proof checking as a means to improve the quality of mathematical proofs, thus establishing the correctness of programmed systems ‘beyond reasonable doubt’. Central issues are process theory, binary decision diagrams, automated deduction, and term rewriting.

Coordination Languages

Theme leader: J.J.M.M. Rutten

Formal methods for coordination languages

Development, on the basis of transparant semantic models, of formal methods for coordination languages, with special attention to UML and Java. Aims are to provide formal methods and tools for the compositional verification of embedded real-time systems within UML, a critical assessment of UML, to extend Java with a notion of component and to develop an accompanying programming environment.

Component-based models and software architectures

Development of formal models for components and component-based software that (1) capture the relevant semantics of a component, as well as its syntax, in its interface; (2) allow compositional derivation of the properties of a system from those of its constituent components; and (3) support notions of distribution and mobility.

Coordination and applications

Study and development of practically useful coordination patterns and protocols in various real-life applications. This includes coordination-based constraint solvers, coordination of parallel and distributed domain composition, and software engineering environments for distributed information systems.

Exploratory research: coalgebraic models of computation

Development of coalgebra as a unifying mathematical framework for (transition dynamical, probabilistic) systems. Continued study of coalgebraic logic, which generalizes modal logic, as well as various schemata for coinductive definitions, including behavioural differential equations.

Evolutionary Systems and Applied Algorithmics
Theme leader: J.A. La Poutré

Evolutionary systems

Study of evolutionary systems in management, economics and e-commerce, including: economic and commercial strategies, complex adaptive systems, adaptive agents, e-commerce, negotiation and trade, bounded rationality, interaction games, automatic programming, information filtering, optimization, dynamization, autonomous systems of trade agents in e-commerce, and evolutionary exploration systems for electronic markets.

Neural networks

Classification of data by several types of neural networks, concerning, e.g., benchmark classification problems, scaling, remote sensing, filtering, event prediction and decision support. Study of financial systems, including neural networks for prediction, event prediction, and decision support, computational methods, and mathematical modelling.

Discrete algorithms

Design of efficient algorithms for on-line optimization problems underlying various management and design problems in computer systems and net-works, as well as the use of quality of service in managing and optimizing on-line scheduling for multimedia processes.

Modelling, Analysis and Simulation
Cluster leader: J.G. Verwer

Applied Analysis and Scientific Computing for PDEs
Theme leader: J.G. Verwer

Atmospheric flow and transport problems

Numerical solution of time-dependent PDE problems from atmospheric circulation and air pollution modelling, addressing issues like linear and nonlinear advection, tailored stiff chemistry solvers, operator and time splitting, approximate matrix factorization, local refinement and parallelization. More in particular: development of numerical algorithms and parallelization of the global atmospheric dispersion model TM3, and numerical atmospheric circulation modelling.

Analysis, asymptotics and computing

Modelling and simulation with nonlinear partial differential equations, with emphasis on the analytical study of the equations governing subsurface transport, asymptotics and advanced numerical methods.

Applications from the life sciences

Mathematical modelling and numerical simulation for life sciences, in particular biology and medicine. Current projects concern: mathematical modelling of biochemical processes in living cells and study of signal transduction, numerical methods for mixed parabolic-gradient systems modelling the development of neuronal connections in the nervous system, numerical study of partial integro-differential equations modelling the growth of phytoplankton, dynamic modelling of the exchange of solutes and particles between biofilms and water.

Pattern formation and low temperature plasmas

Study of pattern formation using nonlinear PDEs, with applications including reaction-drift-diffusion models and free boundary problems in electric discharges and low temperature plasmas. Related fundamental mathematical research deals with analytical and numerical properties of 'pulled fronts'. Analysis of phytoplankton bloom and diffusion of macromolecules.

Computational Fluid Dynamics

Theme leader: P.W. Hemker

Computational fluid dynamics

Computation of fluid flows in gases, liquids, or combinations of these (multi-phase flows) for industrial applications. Current research includes advanced discretization methods for systems of nonlinear conservation laws, multigrid and sparse-grid solution methods, numerical methods for linear algebra problems, local grid adaptation, and parallel and distributed computing. The present emphasis lies on the development of numerical methods for the computation of free-surface flows, discontinuous Galerkin method for convection-diffusion problems, overset grid techniques for convection dominated problems, sparse-grid algorithms for 3D flow problems, and parallel solution of very large eigenvalue problems.

Computational number theory and data security

Application of new mathematical and computational techniques for the solution of problems in number theory with impact on cryptography. Triggered by the emergence of public-key cryptography, the project studies algorithms for factorization and primality testing, for computing discrete logarithms, and for the solution of large, sparse systems of linear equations over finite fields. In addition, classical conjectures like the Riemann hypothesis and the Goldbach conjecture are studied, as well as problems involving Euler's ϕ-function and the sum-of-divisors functions.

Information Systems

Cluster leader: M.L. Kersten

Data Mining and Knowledge Discovery

Theme leader: M.L. Kersten

Standardization and knowledge transfer

Research into applied logic, including dynamic logic, construction of electronic text-books for logic, and interactive information engineering; construction of thesauri for knowledge engineering, combinatorics and linear algebra; knowledge transfer and research on user-centered Web applications and evolving standards, participation in standardization activities and organizational support for W3C.

Data mining and knowledge discovery

Knowledge discovery from hidden relationships (correlations) in vast amounts of data which either prohibits human evaluation or makes it too tedious, with special attention to relational databases and sequential data, e.g., time series, mining in biology (DNA) and medicine (physiological time series).

Multimedia databases

The objective is to achieve efficient storage and retrieval of multimedia data, such as pictures, video and audio, in particular by using feature detectors to simplify and speed-up multimedia data query. This is complemented by an effective query articulation technique for image and video retrieval.

Database architectures

Research into the architecture of next generation database systems and dissemination of Monet database technology.

Multimedia and Human-Computer Interaction

Theme leader: H.L. Hardman

Distributed adaptive hypermedia

Development of models and tools for automatic generation of high-quality hypermedia presentations, taking into account user characteristics, platform-specific requirements and network conditions. At present research focuses on Hypermedia Authoring on Demand, and Structured Document Languages for Hypermedia.

Social user interfaces

Research into humanoid interfaces, such as discourse planning and management, multi-channel communication, and level of quality services of avatars, by applying AI-based methods. The focus is on constraint-based facial animation, exploration of the facial expression space, talking heads, scheduling multi-channel communication, and non-realistic rendering and animation of avatars.

Visualization

Pilot leader: R. van Liere

Data visualization

Study and development of methods for interactive scientific visualization of large data sets, and putting them into practice. The primary application domain is bio-informatics.

Information visualization

Design and implementation of techniques for navigation in multi-dimensional information spaces, with multimedia retrieval systems as primary application domain.

Immersive interfaces

Initial studies of novel techniques for the development of immersive user interfaces. The primary application domain is game technology.

Quantum Computing and Advanced Systems Research

Theme leader: P.M.B. Vitányi

Quantum computing

Investigation of quantum information and communication technology, quantum computer architectures, quantum algorithms, quantum communication complexity, quantum complexity classes, quantum information retrieval, quantum simulation of quantum mechanical physical systems at the elementary level (computational quantum matter) and quantum information theory.

MDL learning and evolutionary computing

Design, implementation, and comparative analysis of a series of practical applications of machine learning techniques. Applications include automatic grammar generation from large text corpora and comparative evaluation of predictive accuracy of MDL and new forms of stochastic complexity, and GP learning of neural network governed robot locomotion and general techniques improving speed and storage requirements of GP implementations. Moreover, basic mathematical requirements for performance guaranties of evolutionary programs.

Advanced algorithms and systems

Design and analysis of algorithms for distributed and parallel systems. Limitations and possibilities of future systems are identified by exploiting fundamental mathematical techniques of (Kolmogorov) complexity theory. A major item is descriptional complexity leading to the ‘incompressibility method’ and ‘learning by compression’. Research into mobile and nomadic computing and communication, focuses on the design, development and assessment of computational tools for the exploration of genomic data.

INTERNATIONAL AND NATIONAL PROGRAMMES

This appendix summarizes the major national and international projects in which CWI participates.
The following data are given for each project:
- title,
- period,
- cooperation with other institutes,
- CWI project leader(s).

European Programmes

ESPRIT

CONFER II (21836): Concurrency and Functions: Evaluation and Reduction
1996–2000
INRIA, ENS, CNET, ICL, KTH, Universities of Bologna, Cambridge, Edinburgh, Pisa, Sussex and Warwick
J.W. Klop

COTIC (23677): Concurrent Constraint Programming for time-critical applications
1997–2000
Universities of Utrecht, Pisa, Lisbon and Kent, SICS, CR&T
K.R. Apt

COORDINA (24512): From Coordination Models to Applications
1997–2000
INRIA, Xerox, U. Leiden, 8 European Universities, Signaal
J.J.M.M. Rutten

NeuroCOLT II (27150): Neural and computational learning
1998–2000
11 universities across Europe
P.M.B. Vitányi

VHS (26270): Verification and control of hybrid systems
1998–2001
U. Joseph Fourier, U. Dortmund, U. Nijmegen, U. Aalborg, INP Grenoble, Chr. Albrechts U. Kiel, Weizmann Inst. Israel, U. Gent, Nylstar Engineering, Sidmar NV, Krupp Uhde Gmbh
J.H. van Schuppen

DEDUGIS (28115): Deductive Constraint Databases for Intelligent Geographical Information Systems
1998–2001
CNR/CNUCE, U. Pisa, GMD – First Berlin, U. Würzburg, Sistemi Territorialli Pisa, DEBIS Berlin, INTECS Pisa, SISTEMA Grosseto
K.R. Apt

IST -- Information Society Technologies

Natural biofilms as high-tech conditioners for drinking water
2000–2003
U. Barcelona, Wasserforschung Mainz Gmbh, Czech Acad.Sc., U. Amsterdam
J.G. Verwer

TRIAL Solution
2000–2003
U. Koblenz-Landau, Heidelberger Akad. Wissenschaften, Trinity College Dublin, U. Nice-Sophia Antipolis, FIZ Karlsruhe, Ges. f. Wiss.-Techn. Information, Open University (UK), TU Chemnitz, U. Köln, Springer-Verlag, Harri Deutsch, Shang IT
M. Hazewinkel

QAIP – Quantum Algorithms and Information Processing
2000–2002
U. Oxford, U. Bristol, U. Aarhus, U. Paris Sud, Hebrew U. Jerusalem, Weizmann Inst., Technion Israel, U. Waterloo, IMCS, MGU Moscow, U. Calgary
H.M. Buhrman

TELEMATICS

EULER (LB5609): European Libraries and Electronic Resources in Mathematical Sciences
1998–2000
FIZ Karlsruhe, EMS, Documentaire Nationale pour les mathematiques, U. Lund, U. Göttingen, U. Degli Mari Group, U. Joseph Fourier
F.A. Roos/P.J.W. ten Hagen

TMR

DONET: Discrete Optimization: Theory and Applications
1998–2002
U. Leuven, London School of Economics and Political Sciences, U. Pierre et Marie Curie (Paris), Rheinische U. Bonn, CNR, U. Lisbon, Société Coopérative ALMA, DASH Associates Ltd, Ecole Polytechnique Fédérale de Lausanne
A. Schrijver, A.H.M. Gerards

ERNSI: Systems Identification
1998–2002
KTH Stockholm, TU Wien, CNR-LADSEB, U. Leuven, INRIA, U. Rennes, U. Cambridge, U. Linköping, U. Eindhoven, U. Delft
J.H. van Schuppen

INTAS

Mathematical methods for stochastic discrete event systems
1997–2001
U. Moscow, U. Novisibirsk, U. Cambridge, U. Braunschweig, INRIA
M.S. Keane

Numerical analysis of local and global bifurcations in ordinary differential equations
1999–2001
U. Gent, Russian Acad. Sciences, U. Nizhny Novgorod
M. Hazewinkel

Bilingual English-Russian thesaurus
in mathematics
1999–2002
Russian Acad. Sciences, Steklov Inst. of Mathematics, Yaroslav State U., U. Utrecht
M. Hazewinkel

JOULE

WELGEL: Polymer Gel Injection
1998–2000
U. Delft, U. Wageningen, NAM, TNO, U. Leiden
C.J. van Duijn

INCO

Dr. Tesy: Methods and Tools for Distributed Real Time Embedded Systems Design and Analysis
1998–2000
Moscow State U., GMD Berlin, RedLab Ltd, State Research Institute of Aircraft Systems (Gosnias)
J.F. Groote

DEVIEW: Designing and Developing the Viewer Centred Paradigm in Virtual Environments
1998–2001
U. Capetown, U. College London
P.J.W. ten Hagen

SEEDIS: Software Engineering Environments for Distributed Information Systems
1998–2001
Universities of East Anglia, Manchester, and Cyprus, Space Application Services
F. Arbab

RTN

AMORE – Algorithmic Methods for Optimizing the Railways in Europe
2000–2004
U. Konstanz, ETH Zürich, IT-DTU Lyngby, CTI Patras, DIS-DIE Rome, L'Aquila Italy
A.M.H. Gerards

DYNSTOCH – Statistical Methods for Dynamical Stochastic Models
2000–2004
Universities of Copenhagen, Amsterdam, Berlin, Cartagena, Freiburg, Helsinki, London,
Padua, Paris
K.O. Dzhaparidze, P.J.C. Spreij

Co-operation with GMD

Sparse grids and overlapping grids in LiSS
1998–2001
B. Koren

TM3 – Advanced numerical simulation for photochemical dispersion models
1998–2002
U. Utrecht/IMAU, INRIA/CERMICS, Imperial College London, KNMI, U. Twente
J.G. Verwer

Application of techniques from propositional logic for the verification of processes
1998–2002
U. Delft
W.J. Fokkink

Distributed Collaborative Virtual Environments
1999–2002
R. van Liere

Mining for groups with distinct behaviour
2000–2004
M.L. Kersten

National Programmes

NWO Council for the Sciences

Dutch-Hungarian cooperation
1995–2000
OTKA
J. van den Berg

LT – Performance analysis of communication networks (long-tailed traffic phenomena)
1996–2002
U. Eindhoven, Columbia U., Lucent Technologies, Eurandom, U. Wroclaw, U. Twente
S.C. Borst

A modular toolset for μCRL
1997–2000
U. Utrecht, U. Eindhoven, U. Amsterdam, U. Nijmegen, U. Groningen, U. Twente, Philips
W.J. Fokkink

Foundations of declarative programming
1997–2002
U. Amsterdam, Free U. Amsterdam
K.R. Apt

Dynamic algorithms for on-line optimization
1997–2001
Philips Research
J.A. La Poutré

AMIS: Advanced Multimedia Indexing and Searching
1997–2001
Data Distilleries, ICL, IFATEC, ING-Group, Tandem, Heriot-Watt U., U. Twente,
U. Eindhoven, U. Amsterdam
M.L. Kersten

Quantum Computing and advanced systems research
1997–2001

U. Amsterdam, U. Delft/DIMES, U. Oxford, IBM T.J. Watson Research Centre,
LRI-CNR Paris
P.M.B. Vitányi

Parallel computational magneto-fluid dynamics
1997–2001
U. Utrecht, FOM Inst. Plasma Physics
H.J.J. te Riele

Learning, cryptography and randomness
1997–2001
U. Amsterdam, U. Twente, U. Waterloo, McMaster U., U. Chicago, UPC Barcelona
P.M.B. Vitányi

CIP – Constraint and Integer Programming techniques
1997–2002
Partners in ERCIM WG on Constraints
K.R. Apt

PERS – Parameter Estimation for Random Sets
1997–2002
Eurandom, U. Utrecht, U. Berkeley
M.N.M. van Lieshout

PROMACS: Probabilistic methods for the analysis of continuous systems
1998–2003
U. Eindhoven, Free U. Amsterdam, U. Amsterdam, U. Nijmegen, U. Dresden, Indiana U.
J.J.M.M. Rutten

Monte Carlo and quasi-Monte Carlo simulation for efficient valuation and risk assessment of financial derivates
1998–2000
Universities of Groningen, Delft and Twente
M.S. Keane

Sparse grid methods for time-dependent PDE problems
1998–2001
UU/IMAU, RIVM, KNMI, TNO, U. Iowa
J.G. Verwer, B. Koren

Rigid sets
1998–2002
Free U. Amsterdam
J. van den Berg

Statistics for random processes with applications to mathematical finance
1998–2002
Free U. Amsterdam
K.O. Dzhaparidze

Protocols, reference models and interaction schemes for multimedia environments
1998–2002
W.J. Fokkink

LRS – Large Random Systems and combinatorial probability
1998–2000
Cornell U., Hungarian Acad. Sciences, U. Utrecht, U. Delft, U. Leuven, U. Cambridge, Chalmers U. Gothenburg, U. Rome
J. van den Berg

SPP – Overset grids and Singularly Perturbed Problems
1998–2002
LLNL Livermore, U. Nijmegen, U. Amsterdam, U. Dresden
B. Koren

GenTrans – Generation of Program Transformation Systems
1999–2001
U. Bergen, U. Utrecht
J. Heering

Spectral aspects of struggle for light
1999–2000
U. Amsterdam
J.G. Verwer

Component based framework for constraint solving
1999–2001
K.R. Apt

Quality of service for multimedia systems
1999–2003
Philips Research
J.A. La Poutré

Asymptotics and special functions
1999–2001
U. Delft, U. Wageningen, NAM, TNO, U. Leiden
N.M. Temme

ToKeN2000: Interaction between humans and information systems
1999–2003
Universities of Eindhoven, Maastricht, Delft, Leiden, Nijmegen, and Rijksmuseum Amsterdam
H.L. Hardman

SICA – System Identification with Computer Algebra
1999–2003
Free U. Amsterdam, U. Eindhoven, INRIA Sophia Antipolis, UCAM
J.H. van Schuppen

WA – Wavelets and their Applications
1999–2003
U. Groningen, U. Eindhoven, U. Twente
H.J.A.M. Heijmans

Evolutionary exploration systems for electronic markets
1999–2003
U. Amsterdam
J.A. La Poutré

MRA – Multi-Resolution Approaches
1999–2004
U. Delft
H.J.A.M. Heijmans

Field ionization instabilities in low current electric discharges
1999–2004
U. Eindhoven, U. Leiden, U. Amsterdam, FOM, CPS, St Petersburg (Ioffe Inst.), U. Münster,
U. Rome, U. Barcelona, U. Essen, Jülich Forschungsz.
U.M. Ebert

Numerical singular perturbation problems (network)
2000–2002
U. Nijmegen, MGU Moscow, Russian Acad. Sciences, POMI St Petersburg
P.W. Hemker

Algorithmic methods for special functions by computer algebra
2000–2003
U. Amsterdam, J. Segura (Madrid), Editors Abramowitz & Stegun
N.M. Temme

Coordination based constraint solvers
2000–2003
U. Nantes
J.J.M.M. Rutten

Dynamo – Semi-automatic hypermedia presentation generation
2000–2004
U. Eindhoven, Philips
H.L. Hardman

Spatial grouping
2000–2004
K.U. Leuven/ESAT
E.J. Pauwels

Special NWO projects

CAM: Number Field Sieve and related subjects
1997–2001
U. Oxford, Australian National U., Citibank New York, San Rafael, U. Groningen, U. Leiden, Macquarie U. Sydney, U. Bordeaux, U. Georgia, U. Giessen, IRI Toulouse
H.J.J. te Riele

SPINOZA – Logic in action
1997–2002
OZSL, U. Utrecht
D.J.N. van Eijck

MPR: Parallel solution of very large eigenvalue problems
1998–2001
U. Utrecht, FOM, U. Delft
H.J.J. te Riele

ALW – Bio-VR: Application of VR in cell biology
1998–2001
U. Amsterdam
R. van Liere

OPSS – Optimization Problems for Spatial Processes
1999–2000
U. Glasgow
M.N.M.. van Lieshout

STW (Foundation for the Technical Sciences)

XRAY – Preprocessing of seismic data: wavelet X-ray transform
1996–2000
Shell, U. Eindhoven, U. Delft, MARIN, U. Groningen
H.J.A.M. Heijmans

FASE: Facial animation
1997–2001
U. Delft, Philips, NOB, Institute for the Deaf, KPN Research
Zs.M. Ruttkay

MOBILECOM: mobile communication networks
1998–2001
U. Amsterdam, Free U. Amsterdam, U. Eindhoven, U. Delft, KPN, Libertel
R.J. Boucherie

Multiresolution image analysis and synthesis
1998–2002
Johns Hopkins U., TNO, AKZO-Organon, Signaal
H.J.A.M. Heijmans

Development of a state-of-the-art Navier-Stokes solver for water flows around moving ships
1999–2002
MARIN
B. Koren

Formal design, tooling and prototype implementation of a real-time distributed shared dataspace
2000–2003
Signaal
J.C. van de Pol

Improving the quality of embedded systems by formal design and systematic testing
2000–2003
Weidmüller
W.J. Fokkink

SENTER

RTIPA – Real Time Internet Platform Architectures
1999–2001
Philips, Oratrix, U. Eindhoven, EOLRING Int., France Telecom, GIP RENATER, Hitachi, INRIA, Italtel SpA, LIP6, Politecnico di Milano, Siemens AG, Telebit, Thomson-CSF
H.L. Hardman

NCF

Parallel simulation of the formation of neuronal connections in the nervous system
2000–2001
B.P. Sommeijer

EURANDOM

Reinforced random walks on finite graphs
1998–2002
J. van den Berg

Image segmentation with applications to agriculture
1998–2002
U. Wageningen
M.N.M. van Lieshout

ICES-KIS Programme

Molecular crowding – mathematical modeling of biochemical processes in living cells
1999–2003
U. Amsterdam
M.A. Peletier, J.G. Blom

MIA – Multimedia Information and Analysis
1999–2003
U. Amsterdam
M.L. Kersten

Distributed virtual reality for cell biology
2000–2001
R. van Liere

Telematica Instituut

DMW: Digital Media Warehouse Systems
1998–2002
CTIT, TICO, KPN, Syllogic
M.L. Kersten

SVC: Systems Validation Centre
1998–2002
CTIT, KPN, CMG, Lucent, TI
W.J. Fokkink

U-Wish: Web-based service for information and commerce
1999–2001
TNO-TM, CTIT
S. Pemberton

DSL: Domain Specific Languages
1999–2002
ING Bank, Cap Gemini, Lucent
A. van Deursen

DRUID: Multimedia indexing and retrieval on the basis of image processing and language and speech technology
1999–2003
TNO, CTIT
M.L. Kersten

Autonomous systems of trade agents in E-commerce
1999–2003
TNO, ING, KPN, IBM, Bolesian
J.A. La Poutré

KNAW

Statistical methods for compound sums
2000–2004
Gadjah Mada U.
R. Helmers

Statistical estimation of Poisson intensity functions
2000–2004
Gadjah Mada U.
R. Helmers