Centrum voor Wiskunde en Informatica

ANNUAL REPORT

'96

Kruislaan 413, 1098 SJ Amsterdam, the Netherlands
P.O.Box 94079, 1090 GB Amsterdam, the Netherlands

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.

P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
Telephone +31 20 592 9333
Telefax +31 20 592 4199
URL: http://www.cwi.nl

CONTENTS

This Annual Report is complementary to the Jaarverslag SMC (in Dutch), which concentrates on the Institute's management including financial and social aspects. A complete overview of CWI's research activities, is also available.
This Annual Report and the other reports can be ordered at
Mrs. D.C.M. Amende; Phone +31 20 592 4128, e-mail: tamende@cwi.nl

OVERVIEW

For CWI, 1996 was a year of looking ahead as well as looking back. On the one hand, anticipating future needs, some drastic adjustments were carried through in the organization of its research, making considerable demands on the staff's time and attention. These adjustments were made in order to warrant optimal functioning of SMC (Foundation Mathematical Centre) and its research institute CWI also in the future. On the other hand, the fact that SMC was founded fifty years ago (on February 11, 1946, to be precise) was celebrated with a range of activities (see pages 8 and 9). The celebrations once more underlined that SMC's mission fitted societal demand remarkably well throughout the years, and that the past always inspired the future. In addition, CWI's profile as a leading research institute was reinforced, on the national level for example by active participation in the proposals for two Technological Top Institutes (Telematics and Logistics & Transport), on which a decision will be made in 1997, and internationally by a prominent role in some activities of the European research consortium ERCIM, including its Digital Libraries Initiative. Last, but certainly not least, CWI maintained its high level of research output – after all our core activity –, more so than might be expected in view of the many side activities occupying a more than average share of the staff's time and energy in the jubilee year. Our research is getting better and better known outside professional circles, to the public at large as well as to potential research partners. This indicates that CWI amply fulfils its task of operating in the midst of society. These, and many other aspects of our activities are treated at some length below.

National

Trade & Industry

Philips
Shell
NAM
AKZO Nobel
DSM
Gist Brocades
KPN
NS
NOB
HSA
HCG
KONI
Elsevier
SKF
Nederland-Haarlem
ID Research
Getronics
Data Distilleries
CAP Gemini
Arcobel
ACE
Digicash

Finance

ING
ABN AMRO
Roccade/DPFinance
Mees Pierson
Coopers Lybrand

Government

WL
NLR
ECN
MARIN
TNO
KNMI
RIVM
RWS
CBS
NIOZ
FOM

Academia

(see opposite page)

International

ERCIM

INRIA (FR)
GMD (DE)
CLRC (UK)
INESC (PT)
CNR (IT)
FORTH (GR)
SINTEF (NO)
SICS (SE)
VTT (FI)
SZTAKI (HU)
SGFI (CH)
CRCIM (CZ)
DANIT (DK)

Universities

Cornell
Charles Prague
Leuven
Cambridge
Chalmers Göteborg
CUNY Brooklyn
Novosibirsk
Bordeaux
London School of Economics
Grenoble
ENS Paris
Yale
Carnegie Mellon
UM Ann Arbor
UC Berkeley
City U. Osaka
Tel Aviv
W. Australia Perth
U. College London
King's College London
Technion Haifa
Australian National, Canberra
Sydney
ETH Zürich
Stanford
Oxford
UPF Barcelona
Royal Holloway London
ENS Lyon
UPC Barcelona
...

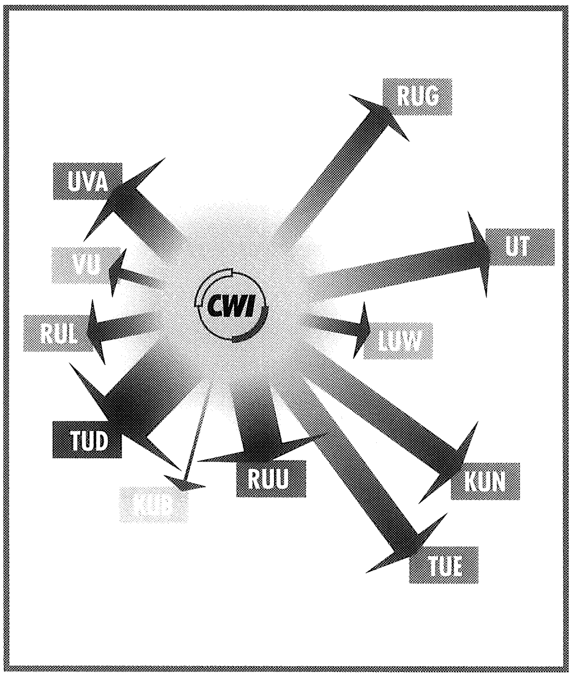
Industries/Research Centres

Bell Communications Research (US)
IBM T.J. Watson Research Center (US)
Verimag Grenoble (FR)
SKF (SE)
Xerox(US)
ICL (UK)
IFATEC (FR)
MacNeal-Schwendler (US)
Infratest (DE)
HP Labs Bristol (UK)
IFREMER (FR)
FINSTAT (FI)
Epsilon SA (GR)
Cartermill International (UK)
Comunicacion Interactiva (ES)
Egnatia Epirus Foundation (GR)
Cycnos Systèmes Ouverts (FR)
Cryptomathic (DK)
Eurocom Expertise (GR)
Europay International (BE)
FOGRA (DE)
IBM European Networking Center (DE)
Intracom (GR)
Otto-Versand (DE)
r3 security engineering (CH)
SEPT (FR)

CWI's research partners.

The 1992 internal policy document MOBILE still forms CWI's basic guide-line. Keywords in this document were: multi-disciplinary, thematic, flexible, society-oriented, innovative, knowledge transfer, and cooperation. The very same keywords subsequently appeared in several external documents, including *Kennis verrijkt* (Knowledge Enriches) and *Kennis in Beweging* (Knowledge on the Move), published in 1995 by the Netherlands Organization for Scientific Research (NWO) and the ministries of Economic Affairs and of Education, Culture & Science (OCW), respectively. Obviously CWI's course already very well fitted the developments wished for in these documents. (The influence on CWI's position of two other documents which appeared in the report year: the evaluation of NWO by a committee chaired by A. Rinnooy Kan, and the Science Budget 1997 by the minister of OCW, will become clear only later.) CWI also anticipated the expected further shift from basic funding to project funding, as is evidenced by its present relatively high percentage (30%) of external income. MOBILE's premises logically resulted in some radical organizational changes, which were effectuated in 1996: 'untwining' SMC's National Activities in Mathematics, and 'overturning' CWI's research groups.

SMC traditionally cherishes close links with academic research in mathematics and computer science. This was formalized in 1982, when NWO commissioned SMC to monitor its mathematics research projects carried out at the Dutch universities. For similar research in computer science – then a brand-new research discipline recognized as such in the Academic Statute – a separate foundation SION was created. During the 1990s it became clear that renewal of the operative management structure was indispensable. Hence, following an advice of the organization bureau &AEF, in May 1996 NWO's National Activities in Mathematics were accommodated in a new foundation SWON (Stichting Wiskunde Onderzoek Nederland). Its office remains located at CWI until further notice. SMC maintains close relations with SION and SWON, both on the management and the research level.



Research contacts with universities.

In accordance with recommendations made by a visiting committee examining in 1995 CWI's mathematics research and the aforementioned bureau &AEF, in 1996 Operation FIT (Flexible, Interdisciplinary, Thematic) was carried through. The objectives were to increase flexibility and interdisciplinary cooperation in the organization of CWI's research. Taking up the challenge CWI prepared an 'overturning' operation, which resulted in a new, theme-oriented structuring of its research, grouped into four 'clusters':

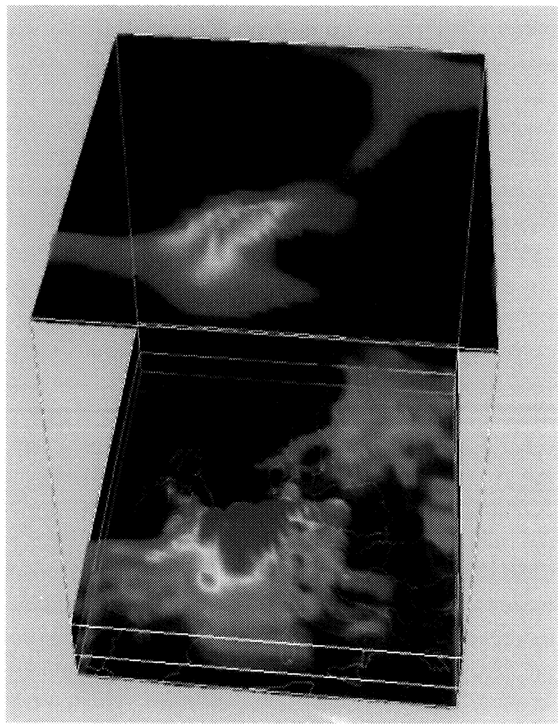
- Probability, Networks and Algorithms;
- Software Engineering;
- Modelling, Analysis and Simulation;
- Information Systems.

Each cluster is subdivided into a few research themes, totalling to eleven themes. In addition there are some 'pilot' themes, which have the potential to develop into a new research theme. Besides, more attention will be given to Human Resource Management and flexibility. For example, all managerial scientific positions are temporary. The new structure enables the desired further shift to interdisciplinary and application-oriented research in a restricted number of carefully selected themes.

Cooperation with Trade & Industry

CWI maintains a multitude of contacts in the world of science and beyond. Of course this is nothing new and goes with the tradition of a research institute. However, in recent years a – still ongoing – extension of contacts is noticed, in particular in the field of trade & industry. A survey of these contacts is given on page 4 (only cases of concrete cooperation are mentioned).

Smog prediction
Reliable smog prediction is an important tool in public health protection. Realistic models including all relevant physical and chemical processes require sophisticated mathematics to solve these models sufficiently fast to be useful in an early-warning system. Research at CWI, carried out in close cooperation with the National Institute of Public Health and Environmental Protection (RIVM), has resulted in a usable model for both winter and summer smog. Nowadays the term 'smog' refers to any excess of health damaging pollutants in the atmosphere. In The Netherlands, winter and summer smog are characterized by an excess of sulphur dioxide and ozone, respectively. The latter is far more difficult to model: we have to include a large number of compounds reacting at widely different time scales. As a consequence standard techniques for summer smog prediction were too slow for practical use. CWI researcher Maarten van Loon developed and implemented new numerical methods using a more flexible grid, based on the method of Local Uniform Grid Refinement. Together with some other improvements the overall result was an efficient computation method for air pollution models in aid of regional smog forecast on the European scale. The work is described in Van Loon's Ph.D. Thesis of June 1996 and is continued in CWI's Environmental Modelling group.



Four-layered simulated ozone distribution over Europe.

Annually, on the first Friday in October, CWI organizes its manifestation *CWI in Bedrijf* (CWI in the Market Place). This year's presentations, including the subjects datamining, fractal image coding, WWW, and transport phenomena in the underground, once again drew considerable attention from industrial, governmental and academic circles. On such occasions it is clearly felt that CWI's more business-like approach – another source of inspiration for rather than an impediment to fundamental research – helps narrowing the mental gap between science and industry. Apart from this event at home CWI participated in some external activities, initiated by industry. This included the R&D Platform of the two-day exhibition Software Automation Nederland, where CWI presented results on the ASF+SDF language prototyping system, the coordination language Manifold, datamining, and multimedia authoring.

Industrial involvement in its projected research is strongly emphasized in the national HPCN Programme, initiated in 1995 by the ministry of Economic Affairs. CWI scored exceptionally high in this programme, with participation in four out of the six projects granted in the first round of calls. Here CWI researchers are involved in a variety of problems:

- development of 3D models for long-term ozone simulation;
- the same, for surface water phenomena;
- coordination of shallow water modelling applications;
- development of an interactive visualization environment for HPC;
- applications of database techniques to financial services.

Utilization is also a keyword in the success formula used by NWO's foundation for the Technical Sciences STW throughout the years for the allocation of projects. CWI has been successful with its project proposals almost from the outset – clear evidence that industrial interest for CWI's research is not a recent phenomenon. Participation in current STW projects concerns interactive books (ACELA), wavelets, parameter identification, and circuit analysis.

Contacts with CWI's recently created spin-off companies (DigiCash, Data Distilleries, General

Design, NLnet) were not only maintained, but even intensified due to their successful operation. In particular Data Distilleries (DD) has, though still small in size, excellent growth potential, one of the reasons being a cooperation agreement effectuated in 1996 with the American company Tandem, which uses DD technology for datamining.

Mathematicians and computer scientists, traditionally applying their skills in activities like production and engineering, extended their attention in the 1990s increasingly to areas hardly considered before, but highly relevant for the economy: transport, publishing, financial services, and the like. These areas not only offer many possibilities for applications of fundamental research, but several fundamental questions still await their solution as well. CWI always was an early actor in the exploration of such new areas. The institute already had considerable expertise in problems of traffic and transport (physically as well as in computer and communication systems). Recently it has started research on interactive books and digital libraries and initiated in 1996 a working group on financial mathematics, intended to grow into a new research spearhead. CWI's extensive knowledge in software renovation found in 1996 an application in a commission from the banking world to find a practicable solution for conversion problems including those caused by the 'millennium bug' and the introduction of the Euro.

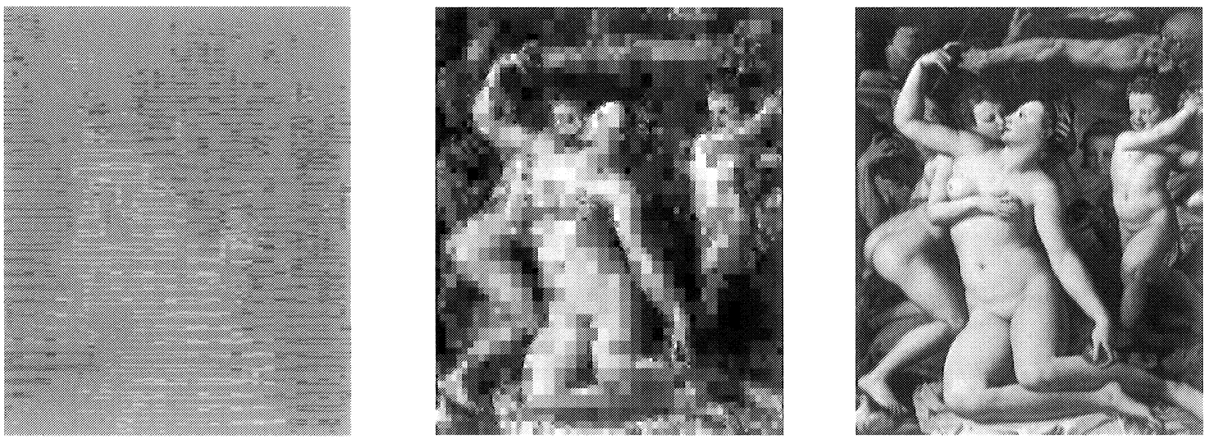
Furthermore, in twenty odd projects research is carried out in cooperation with and/or commissioned by large companies like Shell, Philips, KPN, NS, NAM, and HSA. In addition, five projects are underway with involvement from the service industries, including banks.

Cooperation with Academia

A multitude of cooperation frameworks and contacts – several extending over many years – are evidence

Fractal image compression

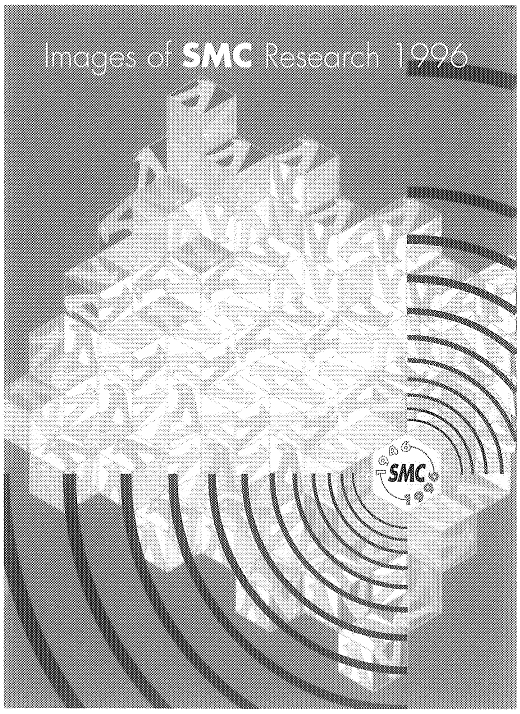
The explosive increase in possibilities of electronic data processing has induced a corresponding need for fast transmission of images. Without compression the transmission is far too slow and even current compression techniques should be improved where possible, because this can lead to further savings in communication time (and thus in money) and reduce the load on a network (WWW is an obvious example). Fractal image compression is a promising, relatively new technique, in which CWI started research in 1996. Whereas a standard method like JPEG can compress images with a factor up to 20, with fractal techniques a compression factor of 100 may be reached, depending on the features of the original image. (On the average a factor of 35 seems feasible.) The method, proposed in the eighties by Michael Barnsley, is based on the observation that fractals can generate deceptively realistic images. Then, conversely, it should be possible to store any natural image in the form of just a few basic fractal patterns, together with the prescription how to restore the image. Starting from a group of pixels of the original image one searches the image for similar groups of pixels which can be mapped more or less on the first group. The compressed image consists of the 'code book' of all these mappings. Fractal image compression is a 'lossy' method, i.e., information is lost during the process, which is no problem for applications like moving images. Images can be restored to any desired resolution, but compression is time-consuming. Hence, the method is well-suited for, e.g., presenting images on Internet. CWI addresses several unsolved mathematical questions, including a precise definition of 'more or less similar' images (mathematical modelling of human criteria) and statistical aspects of the method.



Decompressing an image encoded by fractals.

50 years SMC

Then there was the fiftieth anniversary .... Activities to celebrate this milestone started right before SMC's official foundation day (February 11). The first event was a mathematical congress [1] with 200 participants on February 6 and 7. The invited speakers, including several celebrities such as L. Nirenberg, G. Faltings, and R. Graham, covered all areas of the National Activities in Mathematics administered by SMC. A policy symposium on the following day [2] marked the formal jubilee manifestation. Speakers included A. Rinnooy Kan, at the time still chairman of the Dutch Employers' Organization and – in his own words – a mathematician in the gutter, and the chairman of the NWO Board, R. van Duinen. This series of events was concluded on February 9 with a day for present and former SMC employees with their partners: in the day-time – of all days the coldest of the year – a puzzle tour on bicycle in the country [3], followed at night by a body- and heart- warming celebration downtown Amsterdam in Berlage's Exchange Building. Early in February the jubilee book [4] *Images of SMC Research 1996* appeared (430 pages, full-colour). Following four general articles, written by prominent Dutch experts in mathematics and computer science, thirty-four

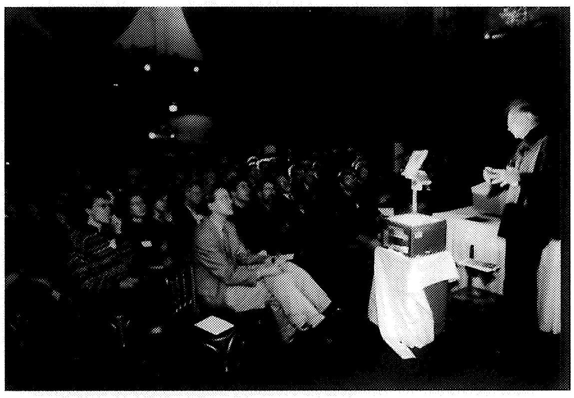


4. Jubilee book.

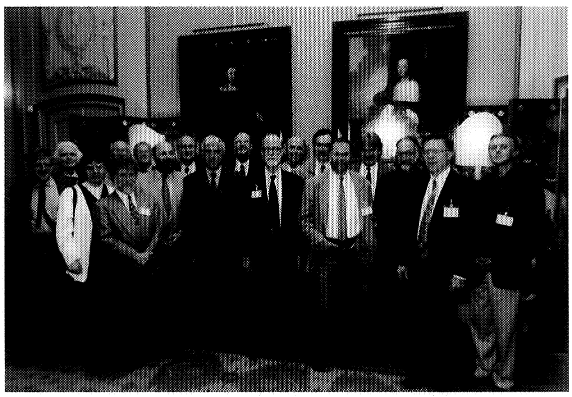
1. Invited speakers mathematics conference.

2. Jubilee address of A. Rinnooy Kan.

3. Brrrrrrr.... .



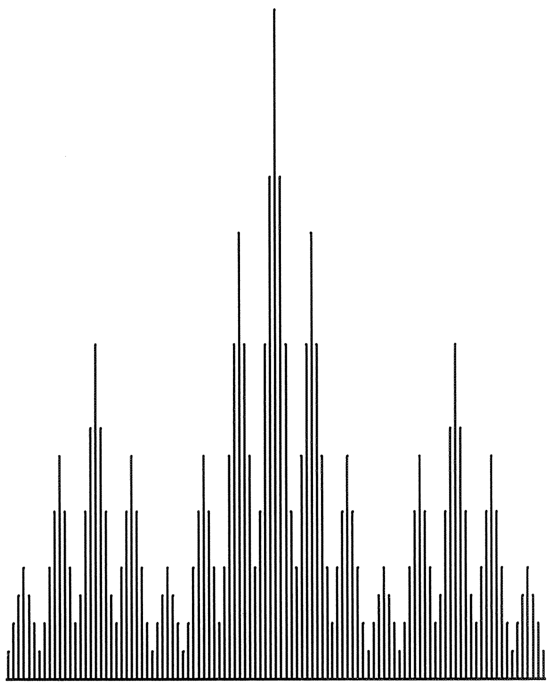
5. Donald Knuth lecturing.



6. Invited speakers computer science conference.

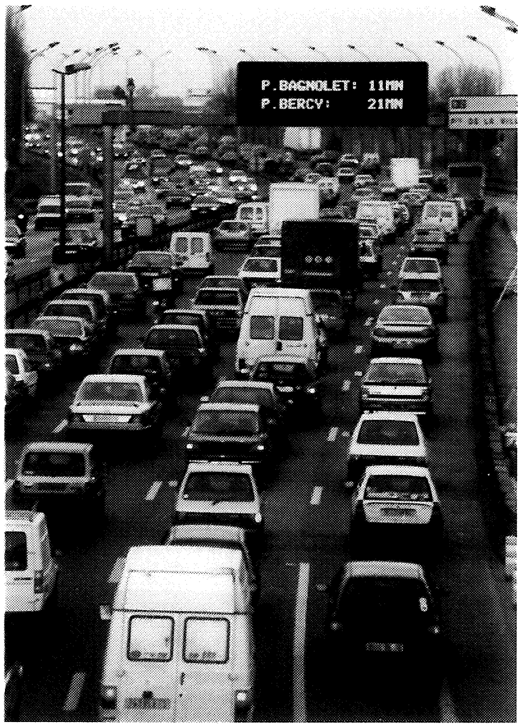
7. The national railroad puzzle.

contributions give an impression of the research carried out recently at CWI and in the framework of the National Activities in Mathematics. Four hundred mathematicians and computer scientists from all over the country attended in March a festive evening [5], organized by SMC in the heart of the city, with lectures by two leading exponents in computer science and mathematics: Donald Knuth and Benoit Mandelbrot. A computer science congress [6], jointly organized with the Academia Europaea, with among others R. Milner and A. Razborov as speakers, took place in Amsterdam the following month. Meanwhile preparations were made for a nation-wide puzzle competition [7]. The problem was derived from a research project commissioned by Dutch Railways in which CWI had investigated the optimal circulation of rolling stock. More than three hundred solutions were submitted. The meeting in November in Utrecht, where for almost two hundred attendants the prize winners were announced, was also the final jubilee event. Finally, worth mentioning is the vacation course for mathematics teachers organized by SMC. The course given in 1996 ([8], topic: Chaos) was the fiftieth in an uninterrupted series starting in the foundation year 1946, with only one exception when in 1954 the International Congress of Mathematicians was held in Amsterdam. This remarkable fact was given due attention.



8. Topological model of Julia set: from Moscow with love.

Control of motorway networks
The DACCORD project (1996–1998) aims at developing and testing coordinated control measures for motorway networks. It is part of the EU Telematics Application Programme – sector Transport. CWI contributes with theoretical research into integrated control and routing control of motorway networks. Given increasing traffic intensities of motorway networks, with capacities remaining limited, control measures are required to keep the networks operating properly. Facilitated by the relatively low costs of communication networks and computer hardware, by now several control measures have been installed: incident management, ramp metering, speed control, dynamic routing information, variable message signs, etc. These measures were installed incrementally, but in general they interact. In an integrated approach, where all control measures serve the same objective, the possibly negative effects of this interaction are prevented. Three major test sites are used for practical validation in real life situations: the Amsterdam network, the Paris network (including the ring road and the connecting motorways), and the Padua-Venice motorway. The twenty odd partners in the DACCORD consortium, located in eight countries, are mainly from Italy (5), France (4), and The Netherlands (8). Main contractor is Hague Consulting Group (NL).



Travel time display on the Corridor Périphérique of Paris. Photo: Peter Kunz.

of CWI's deep embedding in the academic world. These links with universities are charted on page 5. Dutch universities participate in the vast majority of CWI's externally financed projects, including European projects and contract research.

CWI-Academia cooperation also takes shape through the national Graduate Schools established during the 1990s. At present CWI has cooperation agreements with six Graduate Schools, the most recent being an agreement concluded in 1996 with the School for Information and Knowledge Systems (SIKS). Cooperation focuses on joint research, other activities ranging from participation on the management level to courses given by CWI staff. All graduate students at CWI are somehow embedded in a Graduate School.

Involvement of CWI in the academic world may also be measured by the rate of CWI staff taking permanent positions at universities. Lack of precise data does not prevent the conclusion that, taking only the last two decades, it is a matter of several dozens. At present eighteen staff members hold a part-time professorship at a Dutch university, including three full members of the Royal Netherlands Academy of Arts and Science (KNAW). Finally it is worthwhile to mention that, on an average over the last decade, ten young CWI staff members annually completed their research with a Ph.D. Thesis defended at a Dutch university (eleven Ph.D.s in 1996).

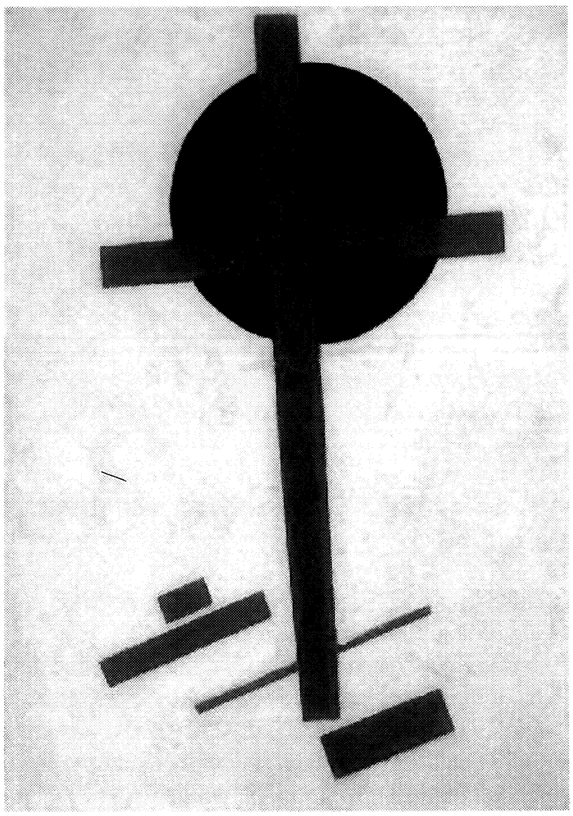
Contacts with governmental and other institutions
These contacts go back to the early days of the institute, a prominent example being the intensive research carried out by the Mathematical Centre (CWI's name before 1983) on behalf of the ministry of Public Works following the flood disaster of 1953. Throughout the years CWI carried out several research projects commissioned by or in cooperation with ministries and large governmental institutions. Contacts with the ministry of Public Works continued until the present day and now concern such diverse research issues as traffic control on motorways, the spread of (toxic) substances in shallow waters, and statistical analysis of oil spots in the North Sea.

Another group of natural partners for CWI, well-positioned as intermediaries on the trajectory from fundamental research to concrete application, is formed by the so-called Large Technological Institutes (GTI's), including Delft Hydraulics (WL) and the National Energy Centre (ECN), other large knowledge institutions like RIVM and KNMI, and the national Organization for Applied Research TNO. Collaboration with the latter is presently focused on visualization for HPC applications, large scale simulation models of the ozone distribution in the atmosphere (with participation of KNMI and RIVM), and transport phenomena in the underground.

The Royal Netherlands Meteorological Institute (KNMI) is not only a partner in the aforementioned ozone distribution models, but also in a research project of the NWO foundation for the Technical Sciences STW on wavelets (other partners being Shell and the maritime research institute MARIN). CWI's cooperation with Delft Hydraulics takes place in the framework of the HPCN Programme financed by the ministry of Economic Affairs and concerns flows in shallow water and 3D modelling. Finally, joint projects are underway with ECN (visualization) and CBS, the Central Statistical Office (datamining).

This is the appropriate place to mention the considerable efforts made in connection with CWI's participation in one or more of the Technological Top Institutes (TTI) – a joint initiative launched by the ministries of Economic Affairs, of Education, Culture & Science, and of Agriculture, Nature Management & Fisheries in the report year. In a TTI, industries and knowledge institutions closely cooperate in strategic fields considered crucial for the development of the Dutch economy in the 21th century. Out of several proposals six were selected for final examination, including two (Telematics and Logistics & Transport) with CWI participation. Decisions will be made in 1997.

International cooperation
CWI's share in research at the European level attained its peak in 1995, when it was involved in no less



Kazimir Malevich 1878–1935
Suprematism 1921/1927?, 72.5x51 cm

CWI developed the website of the Stedelijk Museum in Amsterdam. Integration of text, sound and images is studied in the Multimedia and Human-Computer Interaction research theme (see also page 31).

Digital libraries: automatic generation of thesauri
The key to information retrieval in large collections is metadata: extra information added to a document to make it retrievable, particularly key words and key phrases and classification information. Large thesauri – for mathematics a thesaurus of key phrases should comprise about 120,000 terms – should be generated automatically. By now there are programs to extract index and thesaurus terms from a corpus of electronic texts and to 'standardize' the raw list of key phrases thus obtained. The available data consists of a bipartite graph between terms and documents. CWI addresses several mathematical and computer science problems in this field. The thesaurus concept 'related term' can be quantified by defining a metric (Hamming distance) on the set of terms and documents, thus enabling for example neighborhood search. Mathematically much harder – and as yet unsolved – is the quantification of the concepts 'broader terms' and 'narrower terms'. By applying clustering techniques a 'bottom up' hierarchy on the set of terms can be defined. How to relate this hierarchy with the 'top-down' hierarchy of a classification scheme? Incompleteness of the set of terms generated as above can be thought of as missing centres of clusters. How to recover these missing centres? Generating one thesaurus all at once for a large area is not feasible. How to match the several thesauri (as charts of an atlas)? The research is part of CWI's Digital Libraries project, in which related work involves interactive books and multimedia.

RSA-130

1807082088687404805951656164405905566278102516769401349170127021450056662540244048387341127590812303371781887966563182013214880557

=

45534498646735972188403686897274408864356301263205069600999044599

x

39685999459597454290161126162883786067576449112810064832555157243

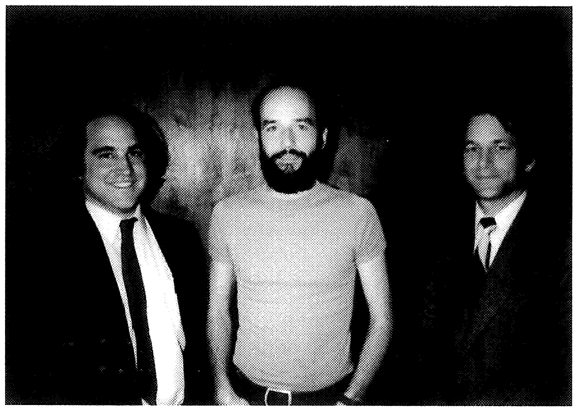
Factoring large numbers using Internet as a Supercomputer

RSA-130 is a 130-digit number in a list of prize-winning numbers issued by the American company RSA Data Security Inc. to be factored. The number was factored in April 1996 in a world-wide effort into two large, 65-digit prime numbers. This was achieved by combining a relatively new mathematical method, the Number Field Sieve (NFS), with the potential of World Wide Web for collecting the required data. The project was initiated by the Dutch mathematician Arjen Lenstra (Citibank, USA). Since December 1995, every workstation owner with an Internet-connection can contribute by collecting so-called relations, which are sent automatically to Lenstra. Thus Internet acts as the world's largest parallel supercomputer. Part of the work was carried out by Marije Elkenbracht-Huizing at CWI and the University of Leiden in the framework of NWO's National Programme in Mathematics. She collected some twenty million of the required seventy million relations, using sixty workstations' idle time. The final processing of the relations was also carried out at CWI, where some remaining problems, such as taking the square root of a large algebraic number, were solved by Peter Montgomery (USA). Further work along these lines might very well make the record short-lived, a fact of some importance for the protection of confidential data streams by RSA public key cryptographic techniques (Internet is protected by such techniques). The RSA-method, named after its inventors Rivest, Shamir and Adleman, is based on the difficulty of reconstructing from the product of two large prime numbers its constituent factors. For more information, see: http://www.npac.syr.edu/factoring.html .

than forty such projects. Not surprisingly, 1996 showed some decrease, mainly because EU programmes now tend to less basic research. Almost half of these projects were concluded. Some new projects started, including MERCURY (Performance Management of Commercial Parallel Database Systems), KESO (Knowledge Extraction for Statistical Offices, CWI project manager), DELOS (ERCIM Digital Library), STEM (Sustainable Telematics for Environmental Management), DACCORD (Development and Application of Coordinated Control of Corridors), COORDINA (from Coordination Models to Applications), and CONFER II (successor of the project Concurrency and Functions: Evaluation and Reduction, concluded in 1995). On aggregate, the current twenty-five projects underway still form a respectable number, certainly in view of the relatively small size of CWI's research staff.

CWI always considered active participation in the European Consortium for Informatics and Mathematics ERCIM to be of great importance (CWI was co-founder in 1989). Here too CWI's role is far more important than expected in view of its size. In the report year three new members joined: the Czech Republic, Switzerland and Denmark, whereas Spain and Portugal suspended their membership, forced by internal problems. The consortium presently consists of national research institutions in fourteen European countries.

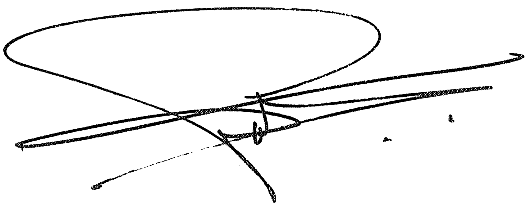
A. Bensoussan (INRIA) resigned in 1996 as an ERCIM vice-president because of his acceptance of the post of director at CNES. He was succeeded by CWI's director G. van Oortmerssen. ERCIM's Digital Libraries Initiative was boosted this year, for example by the start of the EU-sponsored DELOS project. Furthermore ERCIM's WWW Working Group (W4G), chaired by S. Pemberton (CWI), and representing ERCIM in the world-wide web consortium W3C led by MIT and INRIA, is actively involved in W3C activities such as the extension of web style sheets with facilities for visually impaired people. Also an ERCIM Working Group on Con-



From left to right: R.L. Rivest, A. Shamir, L. Adleman. Photo: RSA Data Security Inc.

straints was created by K.R. Apt (CWI), who also acts as chairman. Finally, the second Cor Baayen Award for young ERCIM scientists (named after CWI's former director and first ERCIM President) was granted to Dimitri Papadias (GMD).

Our annual reports traditionally focus on CWI's research results, some examples of which are given in the chapter Research Highlights. In reports like this related background activities are taken for granted. Nevertheless CWI researchers spend, year after year, considerable time in reporting on their work by writing articles and books, giving lectures, (co-) organizing workshops and conferences, drafting plans and proposals, serving on committees of all kinds, and devoting themselves to various matters not directly related to their research, but nevertheless considered highly relevant for a flourishing institute. (With respect to the latter, the report – jubilee! – year in particular drew heavily on our staff's endurance.) These activities, though probably less prominent than the achievements mentioned in this report, remain the backbone of a healthy research environment. I still feel privileged to work in such a positive, cooperative atmosphere on the continuing expansion of CWI as a leading research institute in mathematics and computer science, in The Netherlands and beyond.



G. van Oortmerssen

General Director

ORGANIZATION

CWI (Centre for Mathematics and Computer Science) is the research institute of the Foundation Mathematical Centre (SMC), which was founded on 11th February 1946. SMC is funded mainly by the Netherlands Organization for Scientific Research (NWO). The organizational structure of CWI is shown on the opposite page. CWI's mission is twofold:

- to perform frontier research in mathematics and computer science;
- to transfer new knowledge in these fields to society in general, and trade and industry in particular.

CWI's research is carried out in theme-oriented research groups, which are grouped in four clusters.

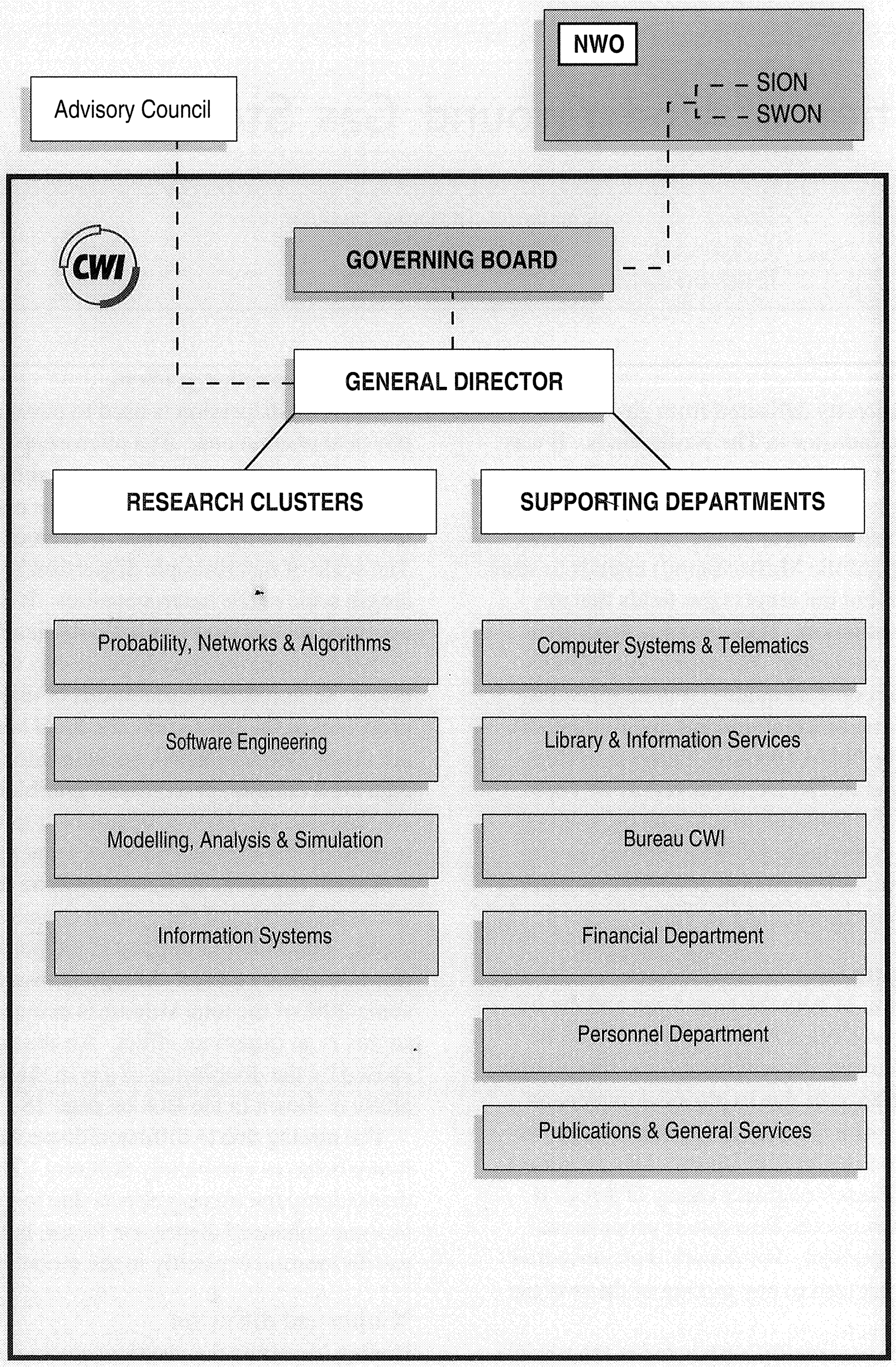
Cluster - Theme	**Cluster leader** Theme leader
Probability, Networks and Algorithms	**O.J. Boxma**
- Networks and Logic – Optimization & Programming	A.H.M. Gerards
- Traffic and Communication – Performance & Control	J.H. van Schuppen
- Stochastics	M.S. Keane
- Signals and Images (pilot)	M.S. Keane
Software Engineering	**J.W. de Bakker**
- Interactive Software Development and Renovation	P. Klint
- Specification and Analysis of Embedded Systems	J.F. Groote
- Coordination Languages	J.J.M.M. Rutten
Modelling, Analysis and Simulation	**C.J. van Duijn**
- Environmental Modelling and Porous Media Research	J.G. Verwer
- Industrial Processes	P.W. Hemker
Information Systems	**M.L. Kersten**
- Data Mining and Knowledge Discovery	A.P.J.M. Siebes
- Multimedia and Human-Computer Interaction	D.C.A. Bulterman
- Interactive Information Engineering	P.J.W. ten Hagen
- Quantum Computing and Advanced Systems Research (pilot)	P.M.B. Vitányi

The researchers are supported by state-of-the-art computer facilities, e.g., an internal high bandwidth glass fibre ATM network, and a well-equipped library of national importance. Hence, CWI is well-prepared to handle the dynamic and interdisciplinary demands of present day research.

SMC's responsibility for monitoring the research projects in mathematics, financed by NWO at Dutch universities (National Activities in Mathematics) ended in May with the creation of a separate foundation, SWON (Stichting Wiskunde Onderzoek Nederland), for these activities. CWI can obtain project funding from SWON, as it did already from SION – SWON's counterpart for computer science research.

SMC, now coincident with its institute CWI, replaced its Board of Trustees by a – considerably smaller – Governing Board, chaired by L.A.A.M. Coolen (director KPN Research). The other Board members are: P.M.G. Apers (University of Twente, chairman SION Board), J.H.A. de Smit (University of Twente, chairman SWON Board), K.M. van Hee (Eindhoven University of Technology, director Bakkenist Management Consultants), and H.A. van der Vorst (University of Utrecht). The chairpersons of the SION and SWON Boards are members *ex officio*, in order to keep SMC's research policy well-geared to the SION and SWON programmes.

SMC's management is delegated to the General Director, who is supported by a Management Team consisting of CWI's four research cluster leaders and the controller. An Advisory Council will be set up in 1997, with representatives occupying leading positions in science and industry in The Netherlands.



CWI organizational chart.

RESEARCH HIGHLIGHTS

Simulation of Underground Gas Storage

Research Project	:	Partial differential equations in porous media research
Researcher	:	J. Molenaar
E-mail	:	hansmo@cwi.nl

Introduction

Natural gas is directly delivered from gas fields to households and industry in The Netherlands. It was estimated that in the winter period of 1997 the pressure in active gas fields will become insufficient to meet peak demand. Therefore the Gas Unie/NAM (Nederlandse Aardolie Maatschappij) intends to store gas in depleted (but not empty) gas fields that are no longer in production. These old gas fields then act as an enormous buffer. Because it is necessary to deliver gas of constant quality, it is important to control the mixing of the stored gas and the initially present gas. The NAM sponsors a project to study this mixing in detail. At Delft University of Technology (TUD) the physical effects causing mixing are studied. It is the task of CWI to deliver an efficient, state-of-the-art numerical code for computer simulation of underground gas storage.

Physical mixing

The displacement of both the stored gas and the initially present gas is ruled by four physical effects: convection, diffusion, dispersion and gas dissolution. Convection is transport due to physical movement of the mixture. At a macroscopic scale this convective transport is described by Darcy's law, proposed in 1865 by the French engineer Henry D'Arcy. It states that the volumetric flow rate is proportional to the pressure gradient. We remark that convective transport does not lead to any mixing of the two gas components.

Mixing is conventionally introduced in the model by the processes of diffusion and dispersion. The mathematical difference between these two processes is that diffusion is independent of the convective flow rate, whereas dispersion is proportional to the convective flow rate. Diffusion is the result of random motion of gas molecules. It is governed by Fick's law, which says that the diffusive flux is proportional to the concentration gradient.

The term dispersion is used to cover a variety of physical phenomena. At a microscopic level it is mechanical mixing due to velocity differences within tiny capillaries. Dispersion at the macroscopic level is caused by variations in the rock properties. The scale of macroscopic dispersion is related to the length scale of the heterogeneities. It is generally accepted that on the field scale (typically hundreds of meters) macroscopic dispersion is the dominant effect. In numerical simulations of displacement processes at the field scale, the local heterogeneities are usually not modelled, because either insufficient data are available, or too many grid blocks are needed. Instead, the concept of macroscopic dispersion is used to model the total effect of local heterogeneities.

Recent experiments have shown an increased amount of mixing if the system contains immovable water. This water is trapped in the porous medium due to capillary forces. In typical reservoir rock some 20% of the total volume is occupied by water, so this is an important effect. An example of mixing caused by the dissolution of gas in the stagnant water phase is shown in the Box on page 18.

The mixing due to diffusion/dispersion and dissolution behaves completely different. Therefore we do not lump the mixing effects due to gas dissolution in some enhanced dispersion factor, but we include gas dissolution explicitly in the model.

Numerical diffusion

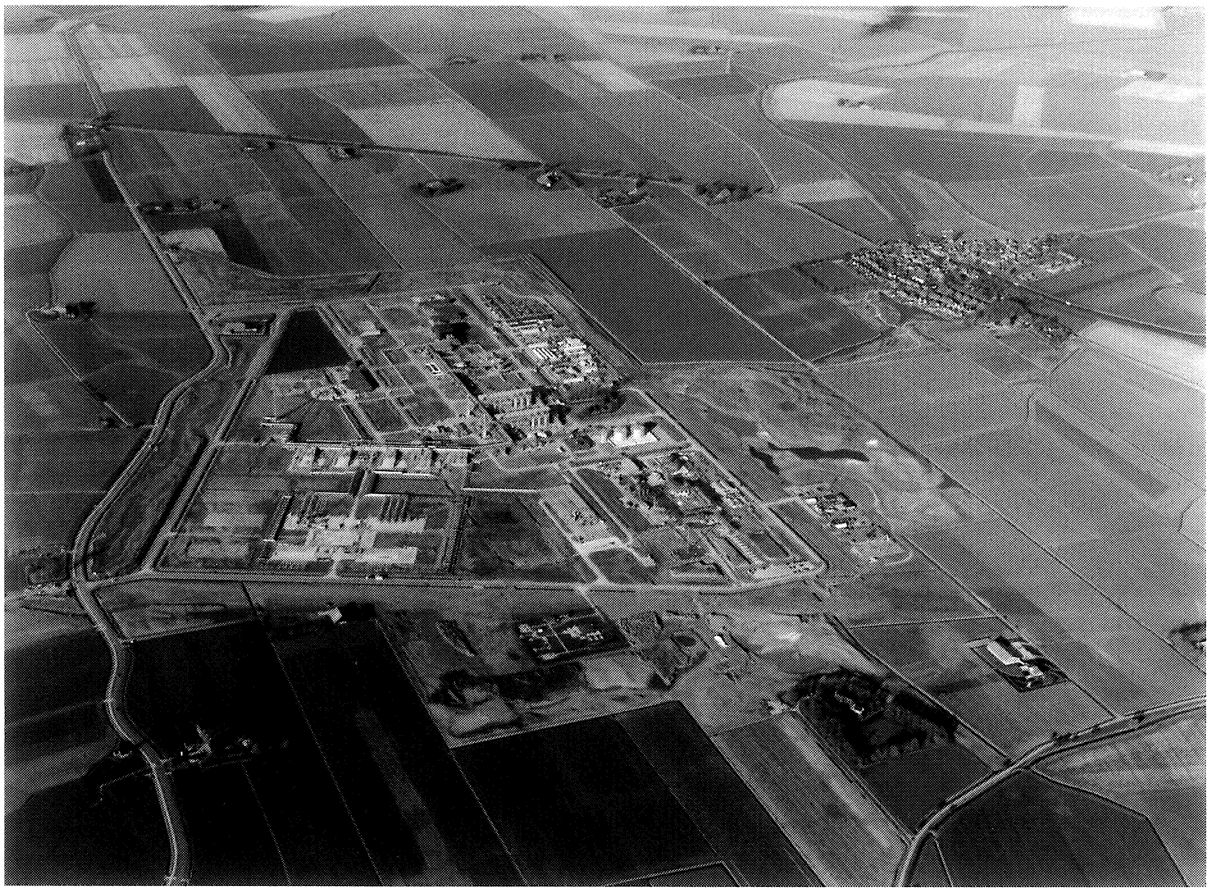
Having identified the physical mechanisms that cause gas mixing, it remains to design an algorithm for numerical simulation. Two requirements for the algorithm are obvious: it should be accurate and efficient.

The first point is not trivial. To demonstrate this we consider the convection-diffusion equation, which is a relevant simplification of the transport equation:

$$(1) \quad \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = \frac{1}{\mathrm{Pe}} \frac{\partial^2 u}{\partial x^2}.$$

In the case without diffusion ($\mathrm{Pe} = \infty$) the solution is trivial: a simple shift in time. Given this simplicity of the analytical solution, the design of good numerical schemes is embarrassingly complicated.
Straightforward application of the finite element method leads to a scheme that is notoriously ill-behaved. The problem is that in the finite element method (and *a fortiori* the finite difference method) some degree of smoothness of the solution is assumed. For the model problem without any diffusion this assumption is simply not valid: initial discontinuities persist.

Better numerical schemes are obtained by introducing more physics in the discretization. The model problem describes a flow in the positive coordinate direction. Using this direction of the flow in the discretization leads to the so-called first-order accurate upstream scheme. First-order accurate means that the numerical solution gets twice as accurate if the mesh width is halved. This is in contrast with the standard finite element discretization that is formally second order accurate, i.e., the numerical solution becomes four times more accurate if the mesh width is halved. However, the first-order upstream scheme is stable and preserves monotonicity of the initial data. The big disadvantage of this scheme is that it introduces a lot of numerical diffusion in the solution: the numerical solution is smeared out in a completely unphysical way. This difficulty has been studied in the past at CWI by the group of Hemker. New high-order accurate schemes have been proposed that still use upstream information and preserve monotonicity, but minimize the detrimental effects of numerical diffusion. These high-order accurate upstream schemes are not a true panacea. They are highly nonlinear, even for a simple linear model problem as Equation (1). The problem is now to construct efficient and robust solvers.

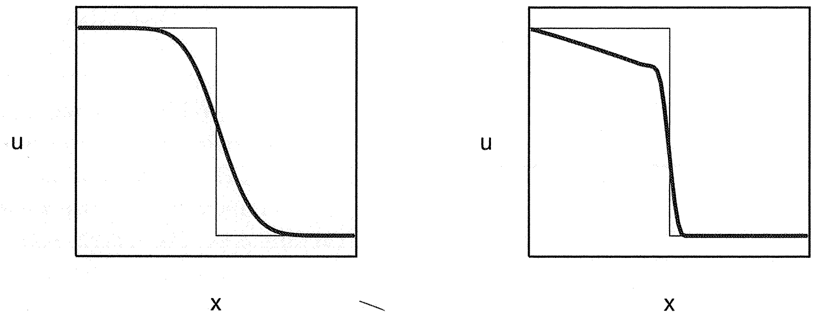


NAM gas field near Grijpskerk. Photo: Sky Pictures.

The combined effect of dispersion and gas dissolution can be modelled by the following convection-diffusion-reaction equation

$$\begin{aligned}\frac{\partial u}{\partial t}+\frac{\partial u}{\partial x} &= \frac{1}{\mathrm{Pe}}\frac{\partial^2 u}{\partial x^2}-\frac{\partial v}{\partial t}\\ \frac{\partial v}{\partial t} &= \mathrm{Da}(u-v)\end{aligned}$$

where u is the amount of flowing gas, v the amount of dissolved gas, Pe the Peclet number, and Da the Damköhler number that quantifies the speed of the dissolution process.



If there is neither diffusion (Pe $=\infty$) nor gas dissolution (Da $=0$) the solution has a sharp interface: there is no mixing. This solution is indicated in the figures by a thin line. In the left figure the solution is shown for the case with diffusion only. The solution is smeared out symmetrically, and the length of the mixing zone grows with the square root of time. If there is only gas dissolution there is also considerable (asymmetrical) smearing of the interface. The length of the mixing zone now grows linearly in time.

Efficient solvers

The solution of the high-order accurately discretized equations has been studied in depth at CWI by Hemker's group. The focus was on the steady-state Euler equations. Direct application of Newton's method is not attractive because of the amount of computational work involved, and the sensitivity to round-off errors. Therefore a defect correction method has been proposed. In this method we obtain the high-order accurate solution, while only solving first-order discretized problems. For the time being it is taken for granted that these simpler, first-order discretized problems can be solved easily. Suppose that the high-order accurate discretization is of the form

(2) $\mathcal{N}_2(u)=f,$

where $\mathcal{N}_2()$ is the nonlinear operator that corresponds to the high-order accurate discretization, u the numerical solution, and f the appropriate right-hand side. Let $u^{(i)}$ be some approximation to the solution u of problem (2), then an improved approximation $u^{(i+1)}$ is obtained by solving

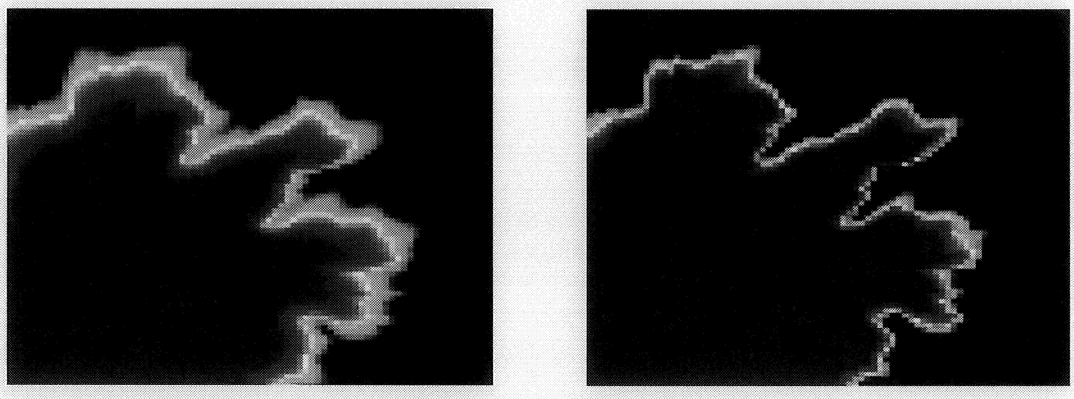
(3) $\mathcal{N}_1(u^{(i+1)})=\mathcal{N}_1(u^{(i)})+f-\mathcal{N}_2(u^{(i)}),$

where $\mathcal{N}_1()$ refers to the first-order discretization.

So one step in the defect correction iteration involves the solution of a first-order discretized problem for $u^{(i+1)}$. The second-order operator $\mathcal{N}_2()$ only enters in the right-hand side of Equation (3): we only have have to evaluate $\mathcal{N}_2(u^{(i)})$, but we do not need to invert $\mathcal{N}_2()$, which is the hard part in Equation (2). Upon convergence of the iteration (3) we have $u^{(i+1)}=u^{(i)}$. If this is the case it follows that $u^{(i)}$ is the solution to the original problem (2). In actual calculations we only want to perform a small number of defect correction steps.

This method has been applied successfully for solving steady-state flows in computational fluid dynamics. However, we are not really interested in steady states, but in evolutionary problems. Therefore we analyzed this defect correction approach for time-dependent problems. Convergence analysis indicates that the defect correction method is an efficient solver for the type of time-dependent model that we are interested in. Next we tried it for our full gas storage problem. In tests using realistic field data the defect correction method turned out to be an efficient and robust solver also in this harder case.

A last point to discuss is the solution of problem (3), which is necessary for making a step in the defect correction iteration. Because $u^{(i+1)}$ is an intermediate approximation to the solution u of problem



Here we have simulated the injection phase of a gas storage summer-winter cycle. The permeability field is highly heterogeneous. In this example no physical mixing effects are assumed, so ideally there is a sharp interface between the injected gas (red), and the initially present gas (blue). On the left the solution is shown for the first-order upstream scheme; there is significant mixing due to numerical diffusion. In the solution obtained by the high-order accurate scheme (right) there is hardly any unphysical mixing.

(3), it is not necessary to solve this problem exactly. Therefore we apply only a single Newton step to approximate $u^{(i+1)}$. The resulting linear system is solved by a linear multigrid method that has been developed at CWI by Molenaar for multiphase flow in porous media. Again we do not solve the linear problem exactly, but approximate its solution by a single multigrid sweep. Surprising as it may seem, this cascade of approximations yields a very efficient iterative solver.

In 1997 this algorithm will be implemented in an easy-to-use gas reservoir simulator. It will be handed to the Center of Technical Geosciences of TUD and to the Shell International Exploration and Production B.V. (SIEP). At TUD/CTG it will be used to simulate and guide further laboratory experiments, and at Shell/SIEP it will serve as a benchmark for their own developments in the field of numerical reservoir simulation.

System Theory and System Identification for Environmental Problems

Research Project	:	Control and system theory
Researchers	:	J.M. van den Hof and J.H. van Schuppen
E-mail	:	jmhof@cwi.nl, J.H.van.Schuppen@cwi.nl

Introduction

Government standards for maximal allowable daily intake of toxic substances are set after extensive investigations. In these investigations use is made of mathematical models in the form of compartmental systems for the flow of substances in the human and animal body. The determination of a compartmental system in a selected model class from experimental data leads to problems of system identification and of system theory. In the period 1992–1996 a research project has been carried out at CWI to develop system identification and system theory for compartmental systems. Research results include conditions for structural identifiability from input-output signals, realization theory for positive linear systems including theory for positive linear algebra, and a novel approximation approach including observers for positive linear systems.

Motivation and development of the field

It is well known that the environment is changing to a state that may be harmful to humans, animals, and plants. At many places the water, ground, and air are polluted, with all its consequences. One of the problems concerns the effect of the pollution on human beings. Which substances are toxic and to what extent are they toxic?

At RIVM (Rijksinstituut voor Volksgezondheid en Milieuhygiëne/National Institute of Public Health and Environmental Protection), Bilthoven, The Netherlands, investigations are performed to support policy makers in the field of public health and environment. Government decisions concerning storage of nuclear waste, control of sulphite-emissions, and control of toxic substances in food and industrial products are studied here. These decisions may be based on the evaluation of measures by mathematical models, for example models of the transport and diffusion of radioactive material in layers of the earth, of transport and diffusion of SO_2 in the atmosphere and its effect on the biosphere, and of the effect of toxic substances on human beings. To determine the effect of for example toxic substances on human beings, it should be investigated how these substances are processed by the body. Therefore a need exists for mathematical models in public health and environment protection, for procedures to obtain the parameters of such models from scarce data, and for procedures to use these models for evaluation of planned governmental measures.

Compartmental systems are mathematical systems that are frequently used in biology and mathematics. Such systems consist of several compartments with more or less homogeneous amounts of material, which interact by processes of transportation and diffusion. Compartmental systems consist of inputs (inflows), states (amounts or concentrations of the substance in the compartments), and outputs (observations). Because these variables are positive, compartmental systems belong to the class of positive systems. (Actually they form a subclass, since they also have to satisfy mass balance laws.) The dynamic behaviour of the concentrations involved may often in a first approximation be taken as linear. The resulting models then form a subclass of the positive linear systems.

Questions posed by public health and environmental protection can be translated into questions for the class of compartmental systems. Possible questions are: How to obtain from possibly scarce data a system in the class of compartmental systems that is both realistic and not too complex? How can such a system be used for prediction and control?

A special class of compartmental systems is formed by the Physiologically Based Pharmaco Kinetic (PBPK) models. These models are used to describe the effect of toxic substances on human beings and animals. The body is modelled as an interconnection of several compartments. The relations between these compartments are described physiologically such as by flow rates and absorption coefficients. The behaviour of the compartments can be measured only indirectly. The parameter values have a natural variability between organs within a subject and between subjects. A characteristic of these models is that the parameters have a direct physiological interpretation. This class of models allows extrapolation of conclusions from animals to other animals and to human beings. It is highly relevant for the investigation of the effect of the environment on public health.

An example of a PBPK model is that of dioxins in the body. The dioxin model describes the dispersion of dioxin in a rat. The intention of RIVM is, by extrapolating the conclusions of the research from animals to human beings, to determine the maximal tolerable daily intake (TDI) of dioxin, which is defined as the exposure level (g/kg/day) to which the *general* population may *chronically* be exposed without developing adverse toxic effects. It is already known that a very small amount of dioxin is carcinogenic for animals. For human beings it is only assumed to be carcinogenic. For adults the exposure to dioxins occurs mainly via the food. Taking variation in consumption habits into account the TDI is not to be exceeded at any age starting at one year.

CWI's research contributions

The problem of system identification is to construct from input-output observations a model belonging to a selected model class that approximates the observations in a specified way. An example is given in the Box, where such a model is constructed for the dispersion of dioxin in a rat. The construction procedure involves the following steps (here the selected model class consists of finite-dimensional systems).

First specify a phenomenon to be modelled, together with the purpose of the model.

1. Model class
Select a class of mathematical models from which a model for the phenomenon may be chosen.

2. Experimentation
Design an experiment, select inputs, and, if possible, perform the experiment, and collect data on the phenomenon.

3. Realization and parametrization
Describe the classes of observationally equivalent models and select a parametrization.

4. Selection
Select a model in the model class on the basis of the experimental data and a criterion such that this model is realistic and not too complex.

5. Evaluation
Evaluate the selected model according to additional criteria (different from the one in the selection step) and if the model is not satisfactory, adjust the options of this procedure such as the model class and repeat the necessary steps of this procedure.

Below attention is restricted to the Steps 3 and 4 of the above procedure because these relate directly to CWI's contribution.

The concept of parametrization of a class of dynamic models in Step 3 is the major point for the system identification problem. The issue is to construct a parametrization of the selected class of dynamic systems such that all systems are described and systems with identical input-output behaviour have the same parameter values, in other words, the input-output signals uniquely determine the parameter values. This is called the identifiability problem. Environmental modelling requires consideration of dynamic systems structured by physical laws, in which case we speak of structural identifiability. It was introduced by R. Bellman and K.J. Åström in 1970 and has been studied since then by many authors. Often structural identifiability from the impulse response function is considered. This function may be obtained from the input-output signals but this requires free and long-term experimentation with animals and human beings, which is often not feasible. Therefore the estimation and identifiability of the parameters of systems for human beings and animals involve more problems than for, e.g., electrical or mechanical systems, due to the scarcity of data. Because of that, for structural identifiability also a test on the data is necessary. CWI's contribution to system identification has been to provide such a test in the case of linear compartmental systems. Also sufficient conditions for structural identifiability of structured positive linear systems are presented.

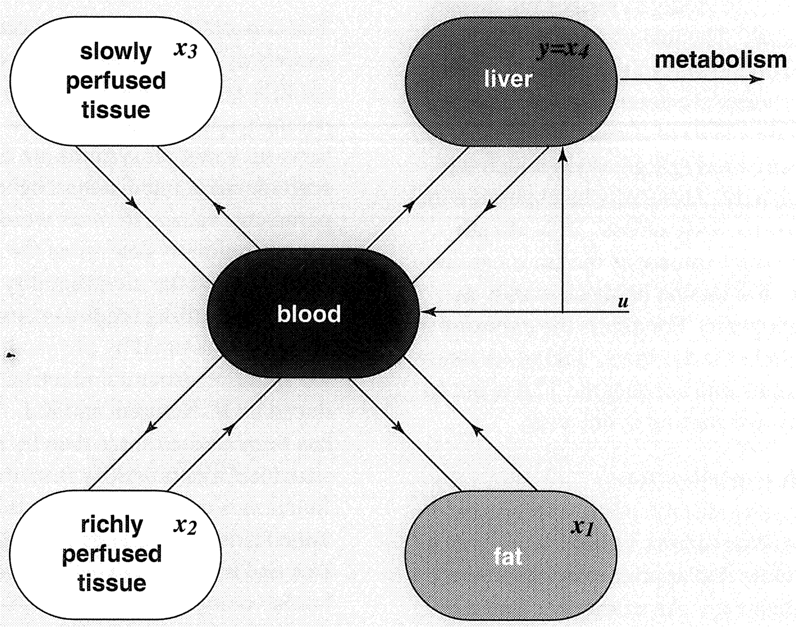
Identifiability conditions for dynamic systems follow from a solution of the realization problem: given a pair of input-output signals, represent it as the input-output of a system in a selected class, called a realization, and classify all minimal such realizations. (A realization in the class of linear systems is called *minimal* if the dimension of the state space is minimal.) Instead of a realization of a pair of input-output signals one has in the literature mainly considered a realization of the associated impulse response function. System identification of compartmental systems requires the solution of the realization problem of positive linear systems. CWI's contribution

Case study: PBPK modelling of dioxin

Step 1: Selection of Model Class

The Physiologically Based Pharmaco Kinetic model of the most toxic member of the large dioxin-family (to be precise, 2,3,7,8-tetrachlorodibenzodioxin or TCDD) describes the dispersion of TCDD in a rat, see [2] for more details. PBPK models form a special class of compartmental systems, used to describe the effect of toxic substances on human beings and animals.

The model of the dispersion of TCDD in a rat consists of five compartments: blood, liver, slowly perfused tissue, richly perfused tissue, and fat, as shown below.



PBPK model of dioxin.

Step 2: Experimentation

Step 3: Parametrization and Identifiability

The model is described by a set of five linear differential equations in the state vector x, its components x_1, x_2, x_3, x_4, and x_5 representing the amounts of dioxin in the five compartments mentioned above, respectively. Metabolism takes place in the liver. The amount of dioxin in the liver can be measured, i.e., observations are available for x_4. Taking these observations as output y, the system can be stated in the compact form of a structured linear system as follows:

$$\begin{aligned} \dot{x} &= A(p)x + B(p)u, \quad x(t_0) = x_0(p), \\ y &= C(p)x. \end{aligned}$$

The scalar function u denotes the rate of uptake of dioxin. In the model it is assumed that the dioxin is exchanged between the different other organs through the blood stream. The elements of the 5x5 matrix $A(p)$ and the 5-vector $B(p)$ depend on eleven unknown parameters p, five of which are the components of the initial state vector $x(t_0)$, the remaining six being related to physiological factors like blood partition. The 5-vector $C(p)$ is the unit vector along the 4-th axis.

The parametrization of this system is structurally identifiable from the Markov parameters and from the initial parameters as a positive linear system. If a rank condition on the input signals is satisfied, then the Markov parameters and the initial parameters can be uniquely determined from the input-output signals.

Steps 4 and 5 have not been considered at CWI.

to system theory concerns this problem, in which four steps can be considered: existence, characterization of minimality, classification, and the relation between realizations. For the existence of a realization in the class of positive linear systems, necessary and sufficient conditions are presented in terms of polyhedral cones. For the characterization of minimality a sufficient condition in terms of the positive rank of a matrix has been proposed.

The approximation problem is to select a compartmental system in a specified class such that the external behaviour of this system fits the given data within a certain approximation. This is the topic of Step 4. If the model class of systems has been selected, if the class has been parametrized using a parameter vector, and if the parametrization is structurally identifiable, then the selection of the best system within the set becomes a problem of estimating the parameter vector. This estimation is performed taking into account an approximation criterion. For time-invariant finite-dimensional linear systems the least squares criterion is commonly used. For system identification of positive linear systems the least squares criterion seems in general not appropriate. CWI's contribution to system identification was the development of a special nonlinear criterion for this class of systems. The criterion is based on an analogy with a stochastic system with an inverted gamma distribution for the observation noise. An algorithm is presented based on this criterion and on the use of a positive linear observer for positive linear systems. For additional information see [1, 2].

REFERENCES

1. J.A. JACQUEZ (1985) *Compartmental Analysis in Biology and Medicine*, The University of Michigan Press, Ann Arbor.
2. J.M. VAN DEN HOF (1996). *System Theory and System Identification of Compartmental Systems*, Ph.D. thesis, University of Groningen, Groningen.

Software Renovation

Research Project	:	Software renovation
Researchers	:	A. van Deursen and P. Klint
E-mail	:	arie@cwi.nl, paulk@cwi.nl

Introduction

While information technology is becoming more and more important for doing business today, a growing number of companies is being confronted with their so-called *legacy systems*. These systems hamper further organizational and business development as required by the ever changing environment in which they have to operate. Legacy systems can be characterized as being large, administrative, mainframe-oriented, often poorly documented, systems. Maintenance costs for these systems keep growing, thus further blocking the business potential to migrate to more open and flexible architectures.

Companies are therefore in search of effective ways to migrate their legacy systems to target architectures which preserve the existing system investments but are able to cope with future business requirements.

The field of *software renovation* addresses the problems just mentioned. There are, however, several factors limiting progress:

- Available commercial solutions are limited in terms of being very rigid, technology-specific and based on poor program analysis techniques.
- Existing academic knowledge in relevant fields of computer science is not known to or applied by many companies.
- System renovation as a research discipline is not recognized as such in academia.

The aim of the software renovation activities at CWI is to contribute, together with several industrial partners, to the solution of renovation problems by increasing and building up knowledge which can be applied in actual system renovation projects.

Background

It is the aim of a system renovation to restructure a legacy system such that it becomes easy to modify. In general, this is a complicated task, since documentation – if at all available – will be obsolete, the modular organization will be far from optimal, and software components will have many unclear interrelationships. The research at CWI tries to find techniques that can help to carry out a system renovation in a (semi)-automated manner. The research is based on two foundations: language technology and coordination technology.

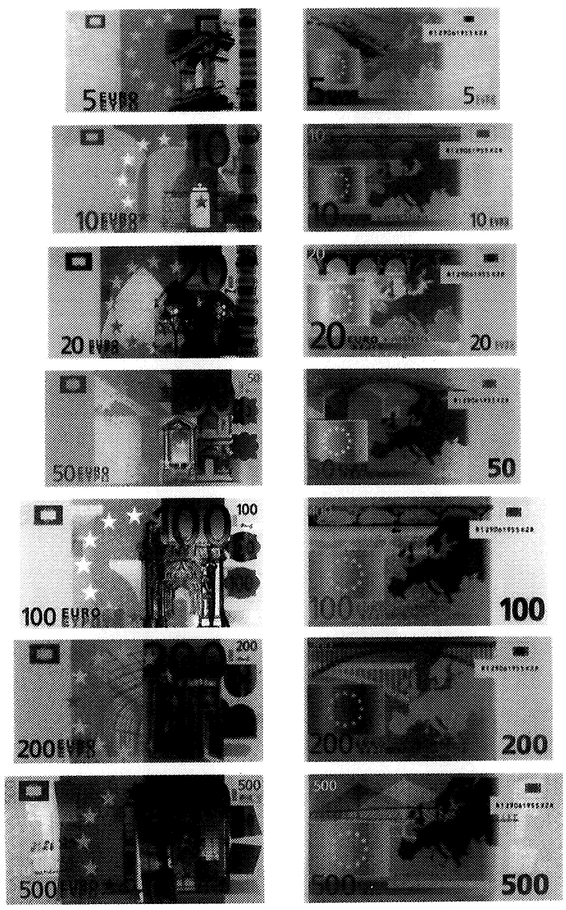
Language technology

The only reliable information on what a system really does is contained in the source code. Therefore, the first step in a renovation is the analysis of the sources of an old system. To facilitate this, tools are necessary that can help to identify clusters of coherent components, predict the impact of changes, visualize the analysis results in a way intelligible to human renovation engineers, and carry out the required modifications.

For system renovations, which usually involve a range of different languages and dialects it is of immanent importance that language-specific tools are easily obtained. During the last ten years, CWI together with the programming research group of the University of Amsterdam have developed the 'ASF+SDF Meta-Environment', a generic environment for constructing language-specific tools. Given the grammar of a language (say COBOL) and a formal specification of the rules for type checking, data flow propagation, or program transformations, the Meta-Environment generates a set of language-specific tools for parsing, analyzing and modifying programs. ASF+SDF has been used to build a range of tools for a wide variety of languages.

Coordination technology

Essential for making software easier to adapt is a flexible architecture, in which properly isolated components communicate in a well-defined manner. To facilitate this, CWI and the University of Amsterdam have developed the 'ToolBus', an architecture for coordinating the interaction between distributed software components ('tools'). One of the important properties of this ToolBus is that the communication between the various tools is described formally using process algebra. The ToolBus moreover is heterogeneous in nature: tools written in all kinds of



Euro banknote designs. (European Monetary Institute)

languages can be connected to it. Per language this only requires a simple 'adapter' for exchanging data and events between the tool and the ToolBus. In this manner, it becomes possible to write each component in the language that is most appropriate: for example, the user-interface part is written in Tcl/Tk, the efficiency-critical parts in C, and the AI parts in Prolog, while the communication between these components is handled by the ToolBus.

Steps in the renovation process

The figure illustrates the relation between language and coordination technology in a renovation context. Starting with a black box, the legacy about which little is known, language technology is used to extract individual components and their interfaces (step 1). Subsequently, the interfaces between these components are made explicit, formalized and structured using a ToolBus script (step 2). Once the communication is structured, individual old components can be replaced one by one and as far as needed by new improved components (repeated application of step 3).

The Resolver project

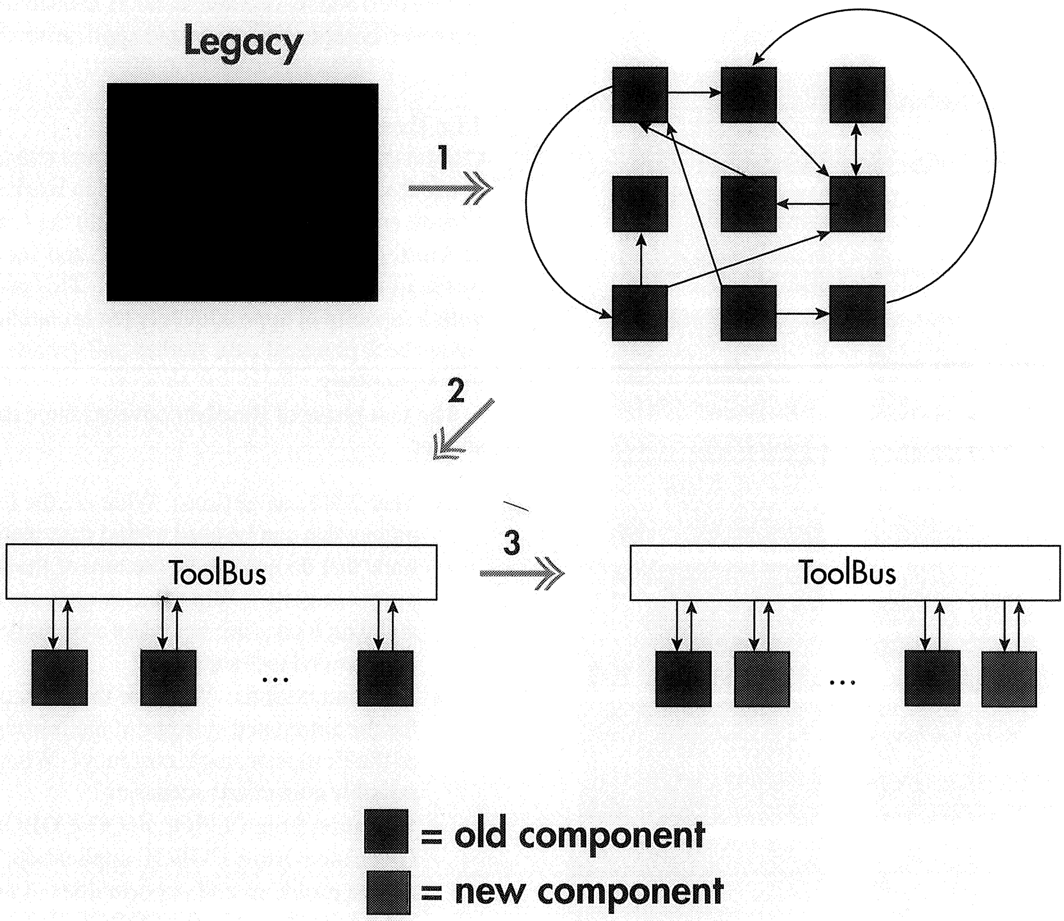
Commissioned by bank ABN Amro and software house Roccade, CWI started in 1996 to work on the 'Resolver' project, in cooperation with the University of Amsterdam and ID Research B.V., and subsidized by the ministry of Economic Affairs. This project, with a capacity of approximately ten researchers, covers both practical case studies and generic renovation technology.

The first phase of Resolver covered three case studies:

- Year 2000 conversions: What are the techniques that can be used to find date manipulations that do not handle the turn of the millennium correctly? Which techniques are used and which ones are not taken advantage of by commercial tool suppliers?
- Euro conversions: What are the consequences for the automated systems of the introduction of the European single currency? What are the possible conversion scenarios?
- Migration from COBOL to OO-COBOL: Given an existing COBOL application, what are the problems and opportunities of migrating it to object-oriented COBOL-97?

In addition to these three case studies, Resolver emphasizes the importance of setting up a language-independent renovation architecture. Currently, techniques have been developed, amongst others, for syntactic querying (what is the best way for a renovation engineer to formulate syntax-specific questions on sources?), data flow analysis (how do variables influence each other's values, taking into account goto's, aliases, procedures, etc.?), and program clustering (given a set of procedures and cohesion properties, what is the best modularization?). Moreover, work has been done to instantiate the generic results for COBOL, thus obtaining, e.g., COBOL-specific parsers and dataflow analysis tools.

The results from the case studies are used to identify problematic and open issues in the generic renovation architecture. These are translated into research questions that are addressed by the Resolver research team. The solutions found are used in subsequent project phases in further case studies.



Steps in the renovation process: (1) System analysis using language technology; (2) Global restructuring; (3) Incremental renovation per component.

The near future

System renovation is a subject of great practical importance, which at the same time raises a range of fundamental questions. CWI, in combination with the Resolver partners, will continue to work on this subject. In the near future, further effort will be invested in elaborating the generic renovation infrastructure, instantiating it for selected languages (such as COBOL), using it to carry out further industrial case studies, and intensifying the contacts with researchers and practitioners active in software renovation.

REFERENCES

1. A. van Deursen, J. Heering, P. Klint (1996). Language prototyping: An algebraic specification approach. *AMAST Series in Computing* **5**, World Scientific Publishing Co.
2. M.G.J. van den Brand, P. Klint, C. Verhoef (1997). Reverse engineering and system renovation – An annotated bibliography. *ACM Software Engineering Notes* **22**(1), 57–68.
3. I. Peterson (1995). *Fatal Defect: Chasing Killer Bugs*, Random House.

Physics of Computation and the Quantum Computing Challenge

Research Project	:	Quantum computing
Researchers	:	P.M.B. Vitányi, H. Buhrman, W. van Dam, B. Terhal, T. Tromp
E-mail	:	paulv@cwi.nl (project leader)

Introduction

New computation devices increasingly depend on particular physical properties rather than on logical organization alone as used to be the case in conventional technologies. The laws of physics impose limits on increases in computing power. Two of these limits are interconnect wires in multicomputers and thermodynamic limits to energy dissipation in all computers. Quantum computing is a new computational technology which promises to eliminate problems of latency and wiring associated with parallel computers and the rapidly approaching ultimate limits to computing power imposed by the fundamental thermodynamics. It opens the prospect for unlimited parallelism and has been shown to work fast for some standard problems like factoring, where all known classical methods use exponentially exploding time. Quantum search algorithms for unstructured databases turn out to be quadratic faster than classical algorithms can ever be. There are tremendous economical and societal interests involved in the ultimate realization of quantum computing.

Here we sketch the short history, basic principles, and implications of this emerging area. In The Netherlands CWI has been one of the first to be involved.

In nonsequential (parallel or distributed) computation one can hardly ignore the physical aspects of the underlying computer system. Complexity analysis may stand or fall with the account taken of physical reality. Moreover, nonclassical or nonstandard physical realizations of computers may have totally unexpected properties.

Nonconventional technologies which may yield novel computation opportunities, include quantum cryptography [1] and quantum coherent computation. Because the latter, if physically realizable, allows breaking most commonly used cryptosystems, quantum cryptography may be the only safe principle for public cryptography, in the sense that its safety rests on the validity of quantum mechanics, rather than on unproven cryptographic assumptions.

The spaghetti problem

Parallel computers that allow processors to randomly access a large shared memory, or rapidly access a member of a large number of other processors, will necessarily have large latency. If we use n processing elements of, say, unit volume each, then the tightest they can be packed is in a sphere of volume n. Assuming that the units have no 'funny' shapes, e.g., are spherical themselves, no unit in the enveloping sphere can be closer to all other units than a distance of radius $R = (3 \cdot n/4\pi)^{1/3}$

Because of the bounded speed of light, the lower time bound on *any* computation using n processing elements is at least order $n^{1/3}$.

Representing the processing elements and their interconnecting wires as a graph, we have derived a general theorem giving lower bounds for the average total edge length that are optimal in the sense of being within a constant multiplicative factor of an upper bound for several example graphs of various diameters. The fact that longer wires need larger drivers and have a larger diameter, that the larger volume will again cause the average interconnect length to increase, and so on, may make embedding in Euclidean 3-space altogether impossible with finite length interconnects [2].

The cooking problem

The ultimate limits of miniaturization of computing devices are governed by an unavoidable heat increase through energy dissipation. Such limits have already been reached by current high-density electronic chips. The question of how to reduce the energy dissipation of computation determines future advances in computing power. Since battery technology improves by only twenty percent every ten years, low-power computing will similarly govern advances in mobile communication and computing. Over the last fifty years, the energy dissipation per logic operation has been reduced from 10^{-3} joule in 1945 to 10^{-13}

joule in 1988. Extrapolations of current trends show that the energy dissipation per binary logic operation needs to be reduced below the thermal noise level (3×10^{-21} joule at room temperature). Even at that level, a future laptop containing 10^{18} gates/cm^3 operating at 1 GHz dissipates 3 MW/sec. Cooling down the computing device to almost absolute zero costs at least as much energy as it saves for the computing.

An alternative to cooling is to develop reversible logic that computes (almost) without energy dissipation. An operation is *logically reversible* if its inputs can always be deduced from the outputs. Around 1960, R. Landauer concluded that only logically irreversible operations must dissipate energy. All computations can be performed logically reversibly, at the cost of eventually filling up the memory with unwanted garbage information, because erasure of information is not reversible. Hence, reversible computers with bounded memories require in the long run irreversible bit operations. The minimal possible number of irreversibly erased bits to do so is believed to determine the ultimate limit of heat dissipation of the computation [3].

Quantum coherent parallel computation

To counteract the problems attending further miniaturization of parallel computing devices current research considers quantum mechanics based technologies. Here we consider one such technology: quantum coherent computing, or QCC. This has recently acquired great anticipated economic value, because QCC can break the universally used public key cryptosystems by being able to factor and do the discrete logarithm in polynomial time [4]. However, there are apparently formidable obstacles to surmount before a workable technology can be obtained.

The QCC approach, advocated by P. Benioff, R. Feynman and D. Deutsch in the 1980s, exploits the accepted theory that quantum evolution of a system consists in a superposition of (potentially infinitely) many simultaneous computation paths. It is theoretically possible that through the specific quantum mechanical rules of interference of the different paths one can boost the probability associated with desirable evolutions and suppress undesirable ones for certain algorithms. Upon observation one of the states in superposition is realized. The desired outcome can theoretically be observed or computed from the observed data with arbitrarily high probability.

The QCC approach will partially alleviate the wiring problem because an exploding number of different computation paths will be simultaneously followed by the same single physical apparatus requiring but a tiny amount of physical space. Of course, the different paths of a quantum computation cannot communicate, as is often a main feature in a parallel distributed computation. Moreover, the quantum evolution in a computation, if unobstructed by observation and decoherence, is reversible. Hence, the pure form of QCC, apart from the irreversible observation phase, is energy dissipation free. How much energy gets dissipated in the observation phase, seems to require a quantum Kolmogorov complexity analysis based on 'qubits' – the quantum analogon to the classical bits of information theory. Through the work of Benioff, Feynman and Deutsch, there has emerged a Turing machine model of quantum coherent computing.

Background: probabilistic Turing machines

In a probabilistic Turing machine, at each step the machine makes a move by flipping a coin. At each such probabilistic move the computation splits into two distinct further computations, each with probability $1/2$. A computation involving m coin flips can be viewed as a binary computation tree of depth m with 2^m leaves.

Now suppose the probabilistic machine is hidden in a black box and the computation proceeds without us knowing the outcomes of the coin flips. Suppose that after t coin flips in the computation we open part of the black box and observe the bit at the position of the Turing machine tape which denotes the truth assignment for variable x_5 ($5 \leq t$). Before we opened the black box all 2^t initial truth assignments to variables $x_1, \ldots, x_t$ were equally possible, each had probability $1/2^t$. After we observed the state of variable x_5, say 0, the probability space of possibilities has collapsed to the truth assignments which consist of all binary vectors with a 0 in the 5th position, each of which has probability renormalized at $1/2^{t-1}$.

Quantum Turing machines

In a quantum Turing machine we have the same computation tree, but now with amplitudes α_i associated with each state $|i\rangle$ of an observable of the system. The amplitudes are complex numbers satisfying $\sum ||\alpha_i||^2 = 1$, ($|| \cdot ||$ denotes the absolute value) where the summation is taken over all distinct states of the observable at a particular instant. (In a probabilistic Turing machine we have a probability $p_i \geq 0$ associated with each node i, such that $\sum p_i = 1$, the summation taking place over all possible states of a computation at a particular time instant.) Transitions are governed by a matrix U which represents the program executed. If $\underline{\alpha}$ is the vector of amplipossible machine states before a step, then $U\underline{\alpha}$ is the same after the step concerned. Quantum mechanics requires U to be *unitary*, that is, $U^\dagger = U^{-1}$ ($U^\dagger$ is the conjugate transpose of U). This constraint on the evolution of the quantum computation ensures that at all steps probability is conserved, and that the computation is reversible.

The common example here is a simple computation on a one-bit computer. The quantum superposition of states of the computer is denoted by

$$|\Psi\rangle = \alpha\ |0\rangle + \beta\ |1\rangle,$$

where $||\alpha||^2 + ||\beta||^2 = 1$. The different possible states are $|0\rangle = \binom{1}{0}$ and $|1\rangle = \binom{0}{1}$. Our unitary operator will be

$$U = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

It is easy to verify that

$$\begin{aligned} U\,|0\rangle &= 1/\sqrt{2}\ |0\rangle - 1/\sqrt{2}\ |1\rangle \\ U\,|1\rangle &= 1/\sqrt{2}\ |0\rangle + 1/\sqrt{2}\ |1\rangle \\ U^2\,|0\rangle &= -\,|1\rangle \quad U^2\,|1\rangle = |0\rangle \end{aligned}$$

If we observe the computer in state $U\ |0\rangle$, then the probabilities of observing states $|0\rangle$ and $|1\rangle$ are both $1/2$. However, if we observe the computer in state $U^2\ |0\rangle$, then the probabilities of observing states $|0\rangle$ and $|1\rangle$ are 0 and 1, respectively. Similarly, observation of state $U^2\ |1\rangle$ yields probabilities of observing states $|0\rangle$ and $|1\rangle$ of 1 and 0, respectively. Therefore, the operator U inverts a bit when it is applied twice in a row. This is the square root of NOT: $\sqrt{\mathrm{NOT}}$. It is *not* possible with probabilistic or other classical computation.

In a probabilistic calculation, flipping a coin two times in a row, we would have found that the probability of each computation path in the complete binary computation tree of depth 2 was $1/4$, and the states at the four leaves of the tree were $|0\rangle,\ |1\rangle,\ |0\rangle,\ |1\rangle$, resulting in a total probability of observing $|0\rangle$ and $|1\rangle$ being both $1/2$.

The principle involved is called *interference*, like with light. In a quantum computation, if the quantum state is

$$|\Psi\rangle = \alpha\,|x\rangle + \beta\,|y\rangle,$$

then for $|x\rangle = |y\rangle$ we have a probability of observing $|x\rangle$ of $||\alpha + \beta||^2$, rather than $||\alpha||^2 + ||\beta||^2$, which it would have been in a probabilistic fashion. For example, if $\alpha = 1/\sqrt{2}$ and $\beta = -1/\sqrt{2}$, then the probability of observing $|x\rangle$ is 0 rather than $1/2$, observe $|x\rangle$ with probability 1.

Quantum parallelism and realizations
The currently successful trick discovered by P. Shor and others, is to use a sequence S_n of n unitary operations S (similar to U above) on a register of n bits originally in the all-0 state $|\Psi\rangle = |00\ldots0\rangle$. The result is a superposition of

$$S_n\,|\Psi\rangle = \sum_{x\in\{0,1\}^n} 1/\sqrt{2^n}\ \ |x\rangle$$

of all the 2^n possible states of the register, each with amplitude $1/\sqrt{2^n}$ (and hence probability of being observed of $1/2^n$.) Now the computation proceeds in parallel along the exponentially many computation paths in quantum coherent superposition. A sequence of tricky further unitary operations and observations serves to exploit interference (and so-called entanglement) phenomena to effect a high probability of eventually observing a desired outcome. Physical realizations of QCC will have to struggle with the fact that the coherent states of the superposition will tend to deteriorate by interaction with each other and the universe (*decoherence*). W.G. Unruh recently calculated that QCC calculations using physical realizations based on spin lattices will have to be finished in an extremely short time. For example, factoring a 1000 bit number in square quantum factoring time we have to perform 10^6 steps in less than the thermal time scale $\hbar/kT$ which at 1 K is of order 10^{-9} seconds. Such a QCC computation would need to proceed at optical frequencies. Another problem is *error correction*: measurements to detect errors will destroy the computation. New methods using quantum information theory and quantum error-correcting codes seem most promising.

References

1. C.H. Bennett, G. Brassard, A. Ekert (1992). Quantum cryptography. *Scientific American*, 50–57.
2. P.M.B. Vitányi (1988). Locality, communication and interconnect length in multicomputers. *SIAM J. Computing* **17**, 659–672.
3. M. Li, P.M.B. Vitányi (1996). Reversibility and adiabatic computation: trading time and space for energy. *Proceedings Royal Society of London* A(452), 769–789.
4. P. Shor (1994). Algorithms for quantum computation: Discrete log and factoring. *Proceedings 35th IEEE Symposium on Foundations of Computer Science*, 124–134.

ELECTRONIC RACE TRACK AT CWI

The report year saw the conclusion of a major technological project at CWI: an essential upgrading of its internal electronic communication. A brand-new glass fibre network, in combination with ATM (Asynchronous Transfer Mode) switches, now enables CWI researchers to exchange information across the network with initial speeds of 155 Mb/sec. The complete installation, which extended over two years (1994–1996), was made possible by a NWO grant totalling 6 Mfl.

Main goal of this project was to enhance the local network infrastructure to a level of high bandwidth and low delay, and to upgrade the computer infrastructure accordingly. A new network and computer infrastructure was considered imperative for CWI's leading-edge research into such fields as multimedia systems, scientific visualization, vector and parallel algorithms, and statistical image analysis.

The new network environment anticipates the 'lego' principle, which is becoming in fashion among computer manufacturers as a basis for experiments. By coupling individual workstations with large servers, 'virtual' computers can be composed enabling researchers to tackle problems which are still beyond the reach of ordinary computers. Traditional copper cable networks cannot keep up with this development, caused by the ongoing increase in power and number of computers.

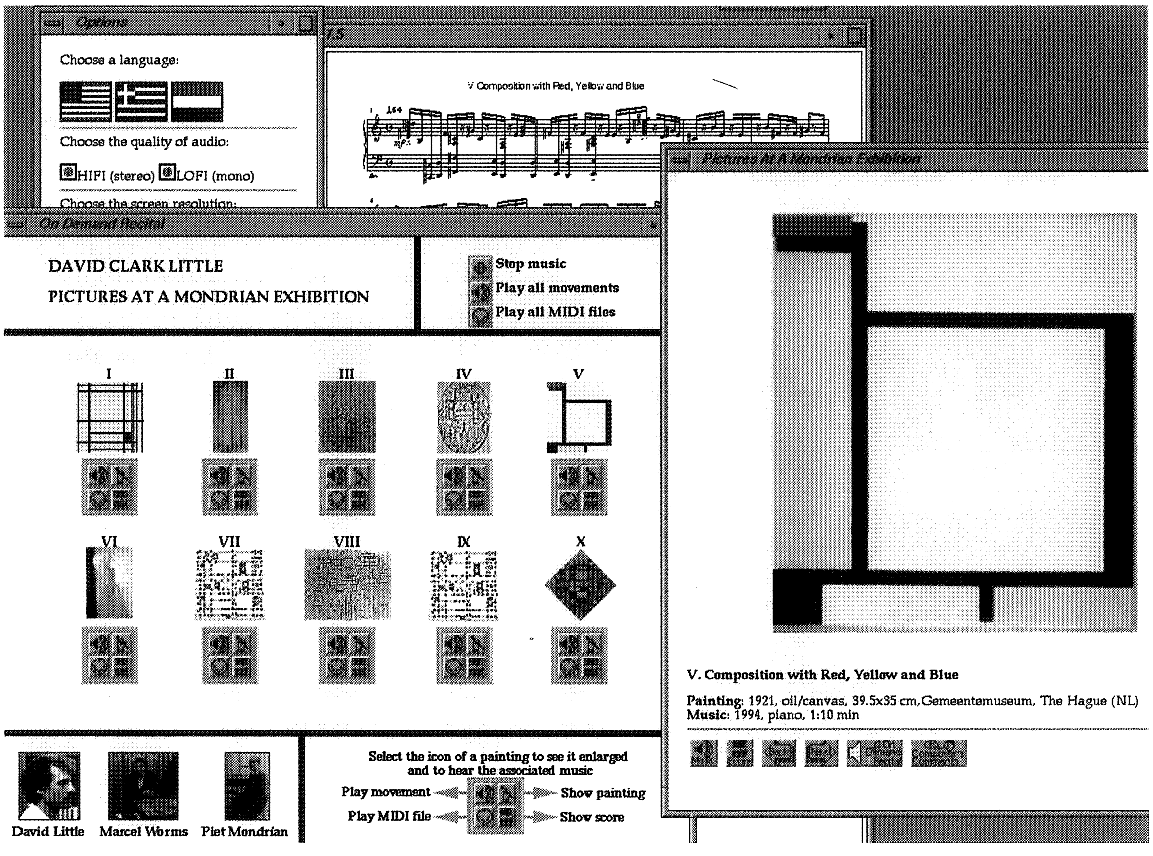
The glass fibre network, covering 200 working rooms in a star topology (more flexible than the traditional Ethernet structure), is built from fibres with a total length of 228 kilometres. At present there are two network connections per desktop. The theoretical maximum network capacity is several hundred Gb/sec. A comparison: in 1980 CWI's internal network could not yet accommodate a flow of 10 kb/sec. Another comparison: after the envisaged upgrading of the coupling of academic research networks in Europe, to be realized early in 1997 with EU funds, data exchange at a rate of 34 Mb/sec will be possible. The new network infrastructure, combining ATM and glass fibre, provides the flexibility required for advanced applications well into the next century. Converters will enable the use of standards other than ATM in the glass fibre network.

The ATM part of the network contains more than hundred ports distributed over, inter alia, two switches coupled through a 622 Mb/sec interconnect. Crucial servers are provided with multiple network connections for load balancing and redundancy. In the near future bandwidths of (multiples of) 1244 Mb/sec are expected. About half of CWI's workstations are now connected with ATM. Compared with CWI's previous (Ethernet) network the present capacity is about a thousand times larger. This is mainly due to the fact that each user of an ATM connection has the full network capacity at his disposal, instead of sharing it with other network users. Our choice of ATM transcends its function as a backbone, in which several machines and nodes are coupled. The goal is rather to provide the individual researcher with tools which are not – or only at high cost – available on his own workstation, for example connections with large computational power and/or memory, disk-arrays and CD-jukeboxes. Future developments are expected to include ATM connections outside the institute, thus further improving electronic communication with the academic world. Finally, end 1996 an investment was made in new workstations, where a 'unified memory' architecture enables the effective use of the ATM bandwidth and full exploitation of the considerable increases in processing power, memory and disk space over the past decade.

In particular two research fields have gained considerable momentum from the network upgrading at CWI: multimedia and visualization.

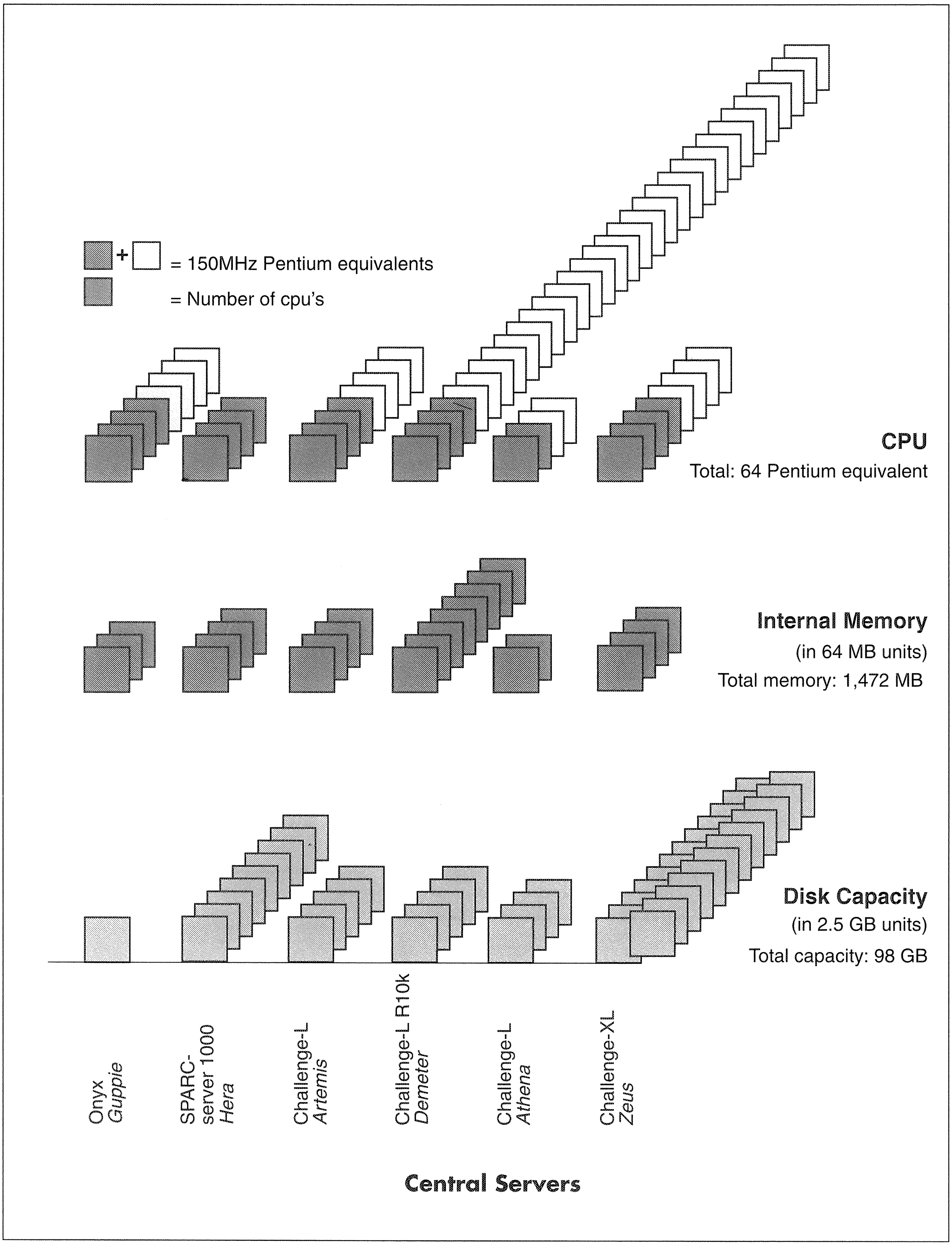
Multimedia research requires massive and fast data transfer at guaranteed transfer rates. For example, full-frame video requires the transfer of hundreds of megabits. In CWI's case the transfer is between collections of sources and workstations. The distributed algorithms involved must synchronize the data streams in (more or less) real-time and the data should arrive at relatively predictable intervals. These requirements went far beyond the extant Ethernet capability. Given CWI's research focus on distributed, heterogeneous multimedia systems, a network supporting concurrent high-throughput, low-latency data transfer at a rate in excess of 100 Mb/sec was considered essential. The development and utilization of CWI's multimedia editor CMIFed in the subsequent years, for example in the ESPRIT project CHAMELEON, would have been unthinkable without such advanced network facilities.

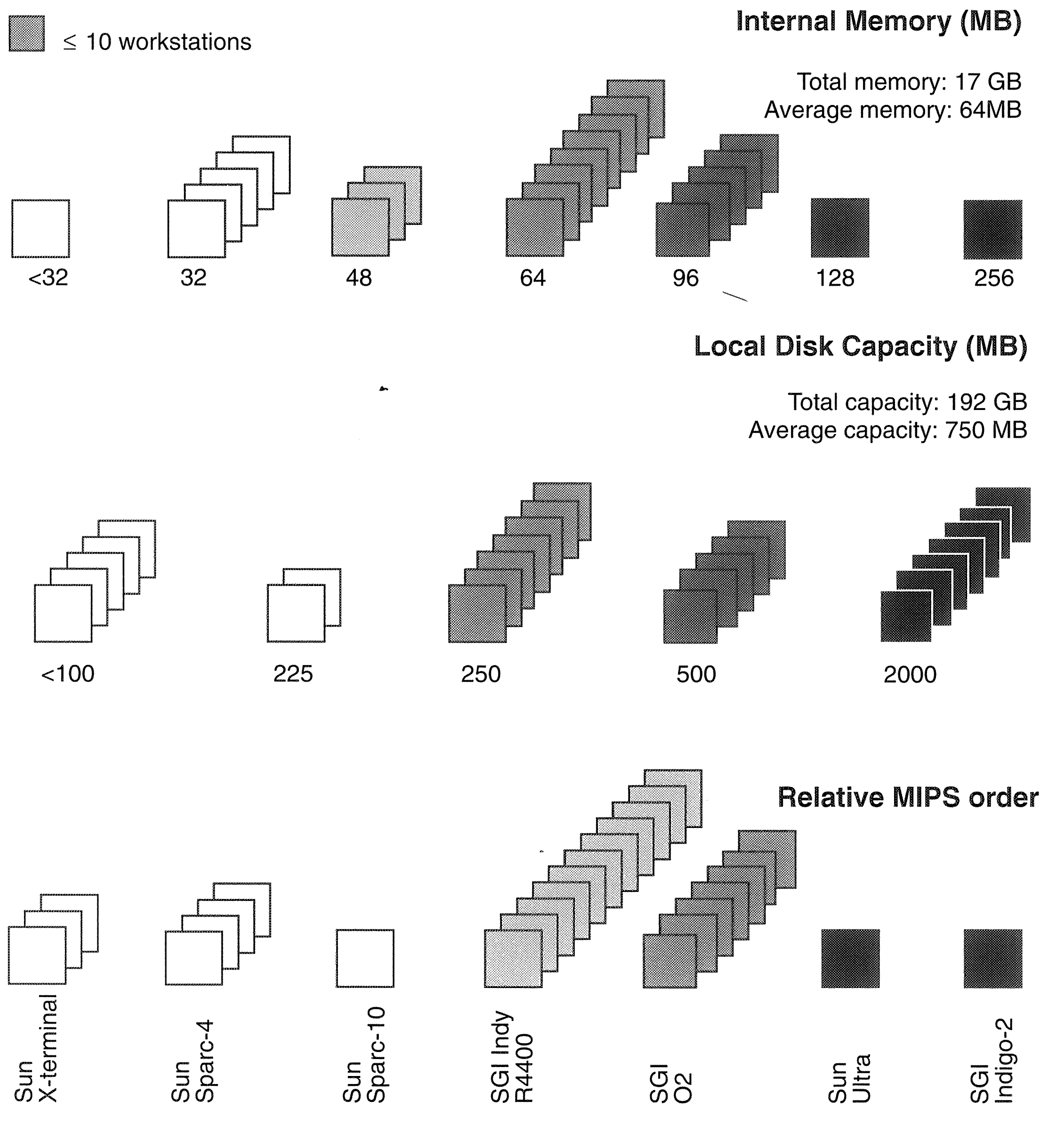
Scientific visualization has rapidly developed into an essential supporting tool in several disciplines, involving aspects of interactive systems, distributed data collection, data processing and analysis. It requires the combined skills of experts in visualization techniques and researchers bringing in their domain knowledge. The keyword here is communication: first, as a method of joint development of algorithms, in which interactive sharing of information is essential, and second, as a basis for transporting information among various processors during visualization experiments. One area of interest is the study of interactive visualization techniques, which require a large data transfer capacity plus a responsive network to allow for interaction among users involved in the visualization process. A result due to the network upgrading was the development over the past few years at CWI of a Computational Steering Environment editor and its application to several problems encountered by researchers in their practice. Also several applications in HPCN, including the application of parallel algorithms to the factorization of large numbers, were realized by CWI's up-to-date network facilities.



This image shows part of the 'Pictures at a Mondrian Exhibition' hypermedia presentation developed using CWI's CMIF authoring and presentation system. The application is used to illustrate the complex iterations that can exist in next-generation networked hypermedia systems. In the lower-right portion of the image, we see a collection of 10 Mondrian paintings that form the basis for a work by the American composer David Little, who lives in Amsterdam. Any of the paintings can be selected for presentation. Once selected, the user can experience both the artwork and the corresponding musical piece. In this illustration, the painting 'Composition with Red, Yellow and Blue' is being presented. The user has opted for Hi-Fi stereo output, and has decided to read the score as well as look at the painting. Comments from the composer are available, as is background information on the composer (David Little), the pianist (Marcel Worms) and the artist. Specifying and controlling delivery of such documents in a predictable manner over a computer network is the focus of CWI's multimedia research.

COMPUTING EQUIPMENT RESOURCES



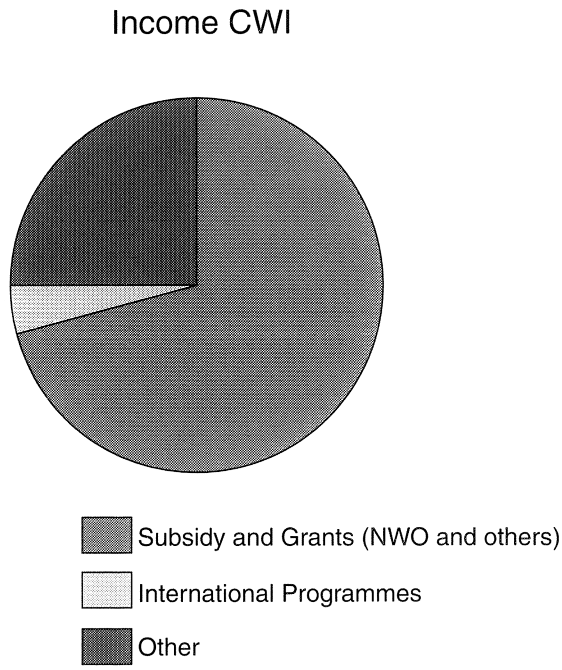


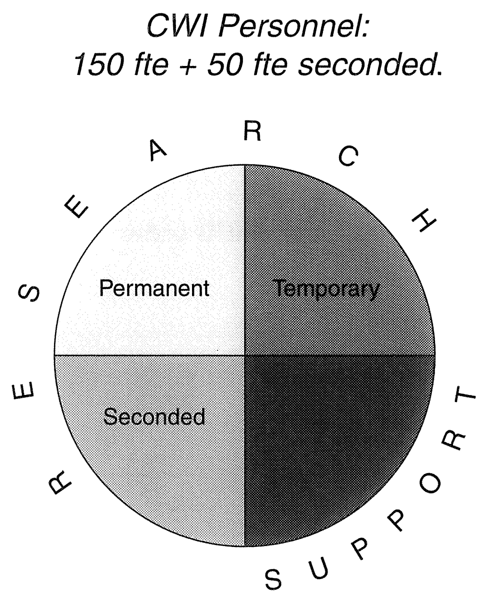
Workstation Distribution

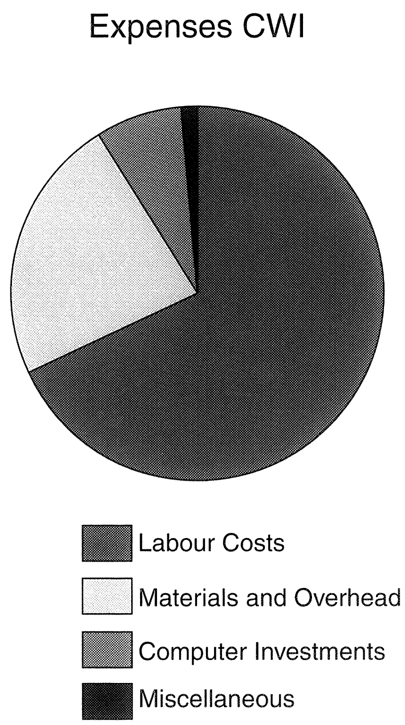
FINANCES, PERSONNEL, PH.D. THESES

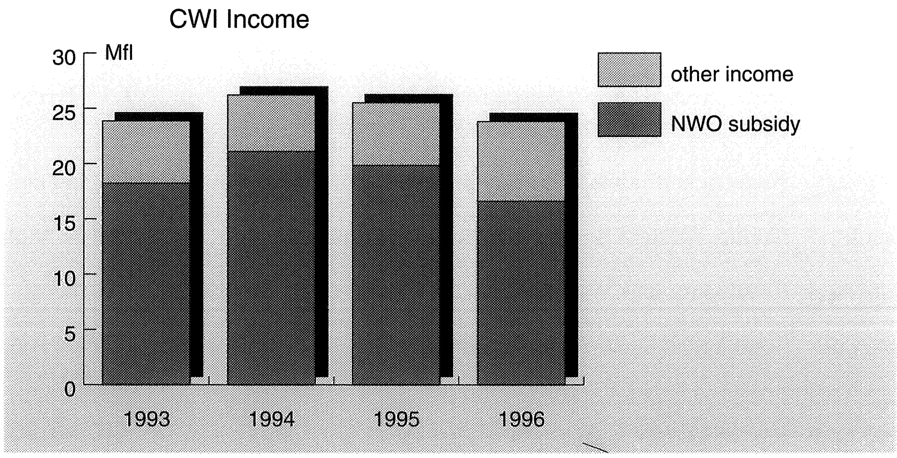
FINANCES 1996

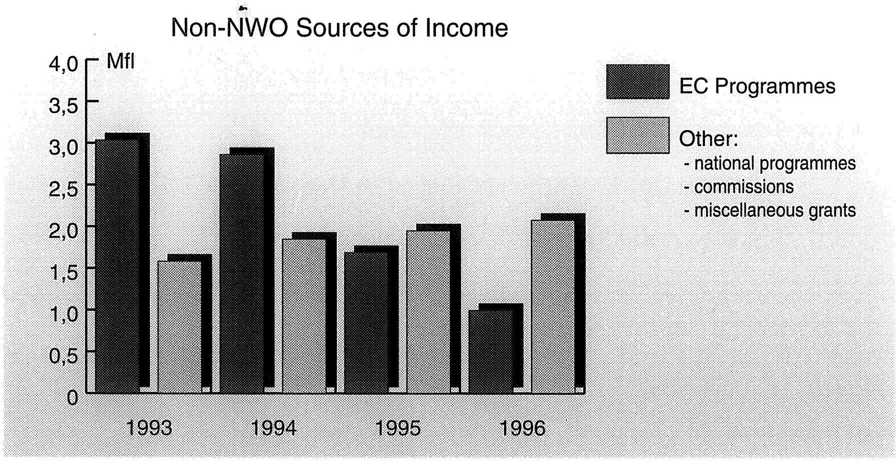
In 1996, SMC spent Dfl. 22.99 million. The expenses were covered by a subsidy from NWO (Dfl. 16.60 million), other subsidies and grants (Dfl. 0.34 million), and from the international programmes (mainly EC programmes, e.g., ESPRIT and HCM) (Dfl. 0.96 million). Finally, an amount of Dfl. 5.91 million was obtained as revenues out of third-party-services and other sources.

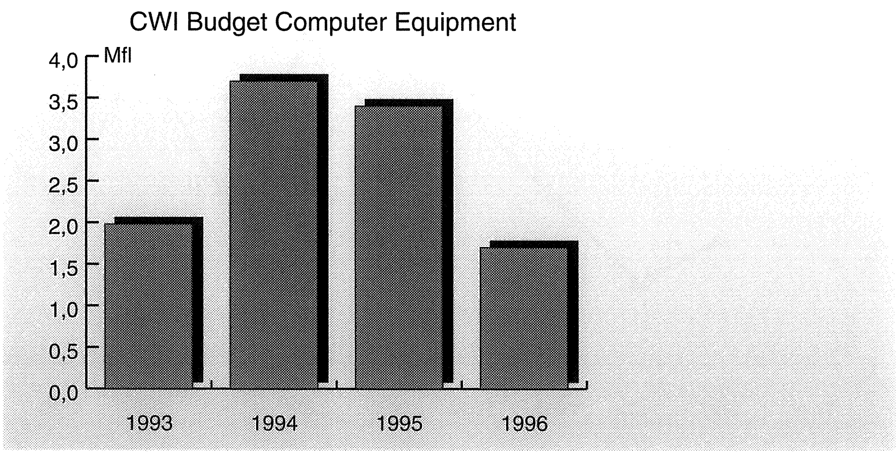







CWI Income
Mfl
30
25
20
15
10
5
0
1993
1994
1995
1996
other income
NWO subsidy


Non-NWO Sources of Income
Mfl
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0
1993
1994
1995
1996
EC Programmes
Other:
- national programmes
- commissions
- miscellaneous grants

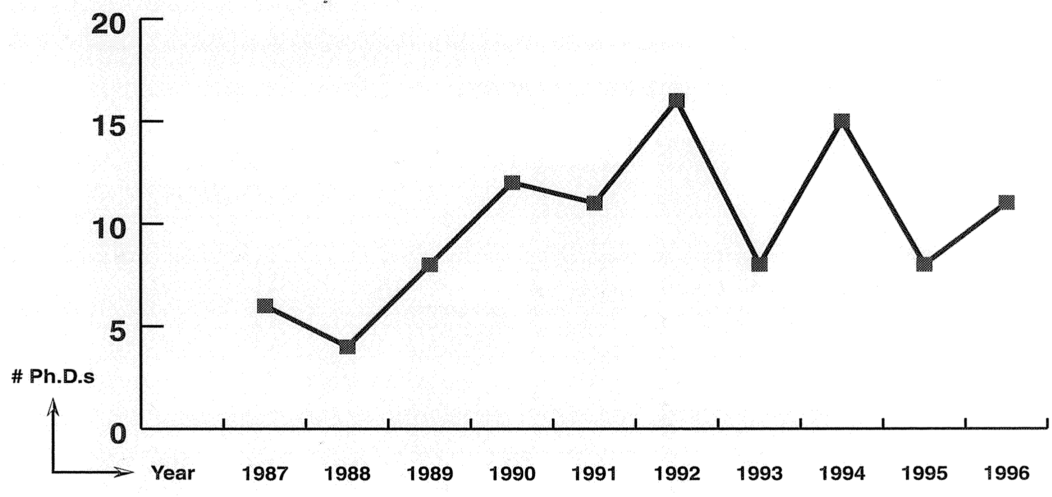

CWI Budget Computer Equipment
Mfl
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
1993
1994
1995
1996

CWI Ph.D. THESES

Author	Title	Thesis advisor(s)+)
A.A.M. Kuijk	On a Layered Object-Space Based Architecture for Interactive Raster Graphics	L.O. Hertzberger (UvA)
M. van Loon	Numerical Models in Smog Prediction	P.J. van der Houwen
J.H. Hoepman	Communication Synchronization Fault Tolerance	P.M.B. Vitányi
F. van Raamsdonk	Confluence and Normalisation for Higher-Order Rewriting	J.W. Klop
M.M. Bonsangue	Topological Dualities in Semantics	J.W. de Bakker, J.N. Kok (RUL)
A.A.F. Overkamp	Discrete Event Control Motivated by Layered Network Architectures	J.H. van Schuppen
J.M. van den Hof	System Theory and System Identification of Compartmental Systems	J.H. van Schuppen
F.J.M. Teusink	Logic Programming and Non-Monotonic Reasoning	K.R. Apt
J.F.T. Kamperman	Compilation of Term Rewriting Systems	P. Klint
D. Turi	Functorial Operational Semantics and Its Denotational Dual	J.W. de Bakker
H. van der Holst	Topological and Spectral Graph Characterizations	A. Schrijver

+) For external advisors the university's acronym is added:

UvA = University of Amsterdam
RUL = University of Leiden



CWI RESEARCH PROGRAMMES

Probability, Networks & Algorithms

Cluster leader: O.J. Boxma

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.

Parallel declarative programming

Study of the relation between two declarative programming paradigms: logic and lazy functional programming. The envisaged translation should lead to an alternative, parallel implementation of a class of logic programs.

Traffic and Communication – Performance and Control

Theme leader: J.H. van Schuppen

Communication and computer networks

Modelling, performance analysis and control of traffic streams in communication and computer networks, in particular queueing-theoretic methods for congestion phenomena and network control by discrete event and stochastic methods.

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, and the development of control theory for discrete event and hybrid systems.

Exploratory research: control and system theory

Control and signal processing require mathematical models in the form of a dynamic system or a control system. Research is directed at realization and system identification problems.

Stochastics

Theme leader: M.S. Keane

Probability

Research in probability theory and its applications, with emphasis on combinatorial probability and large random systems.

Statistics

Theoretical and applied research in statistics, with emphasis on resampling methods and spatial statistics: statistics for inhomogeneous spatial Poisson processes, bootstrap resampling, nonparametric regression, inference on rare errors.

Stochastic analysis

Theoretical and applied research, in particular applications to stochastic finance and statistics for stochastic processes.

Ergodic theory

Theoretical and applied research in ergodic theory and stationary processes: applications of ergodic theory to number theory, simplification of proofs of ergodic theorems, nonexpansive mappings and their asymptotics, mathematical methods for stochastic discrete event systems.

Signals and Images (pilot theme)

Theme leader: M.S. Keane

Image modelling and coding

Statistical analysis of images by way of quadtrees, modelling of bit-error bursts in transmission lines, and fractal image coding.

Wavelets
Analysis of seismic data, in particular the use of Slepian wavelets in earthquake analysis and of wavelet- and Radon transforms in data preprocessing, and the asymptotics of filter coefficients for seismic processing.

Morphological image processing
Research on mathematical morphology and its applications to image and signal processing, in particular connected morphological operators and morphological pyramid transforms.

Stochastic geometry
Research on random sets (weighted Boolean models) and spatial statistics.

Image generation for virtual environments
Fundamental algorithms and software tools for viewer centered interaction paradigms and visual display techniques, in particular the development of an adaptive image rendering environment and a software environment for visibility and view volume culling.

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. Present focus is on the SOS Resolver project, addressing in particular the development of a language-independent, re-usable renovation architecture as well as actual renovation problems such as the year 2000, the euro single currency, and the migration from COBOL to OO-COBOL.

Software development
Design and implementation of advanced software development tools using algebraic principles. Further development of the ASF+SDF Language Prototyping System.

Interactive visualization environments
Management of scientific data, visual geometry specification, techniques for high level input, and distributed visualization environments, with applications to HPCN as well as traditional design and engineering.

Specification and Analysis of Embedded Systems
Theme leader: J.F. Groote

Process specification and analysis
The study of specification and analysis techniques for process behaviour, in particular the search for a real time extension to the language μCRL. Applications include testing techniques developed in cooperation with Philips NatLab.

Data manipulation and term rewriting
The focus is on term graph rewriting and higher-order rewriting, as well as extensions such as infinitary rewriting to provide an operational semantics for term graph rewriting.

Proof searching and proof checking
The study of proof search in simple logical systems, viz., propositional logic, in order to increase the efficiency of symbolic verification techniques applied to verify requirements on processes. Furthermore, the development of proof checking methods to establish the correctness of programmed systems 'beyond any reasonable doubt'.

Coordination Languages
Theme leader: J.J.M.M. Rutten

Formal methods for coordination languages
On the basis of transparant semantic models techniques are developed for the evaluation of language design, program development, refinement, and verification.

Experimental testbed for control-oriented coordination
Design and implementation of an experimental testbed for practical control-oriented coordination programming on heterogeneous platforms and its programming support environment. This testbed is currently accessible through the coordination language Manifold, its utilities, and its visual programming interface, Visifold.

Coordination applications
Study and development of practically useful coordination patterns and protocols in various real-life applications, leading to program modules built on top of the above experimental coordination testbed system.

Exploratory research: coalgebraic models of computation
Further development of coalgebra as a unifying mathematical framework for (transition and dynamical) systems, with emphasis on the composition and coordination of coalgebras.

Modelling, Analysis and Simulation

Cluster leader: C.J. van Duijn

Environmental Modelling and Porous Media Research

Theme leader: J.G. Verwer

Numerical algorithms for air quality modelling
Numerical modelling of the long range transport and chemical exchange of atmospheric air pollutants, involving advection schemes, stiff chemistry ODE solvers, splitting methods, grid generation/adaptation and implementations on supercomputers and parallel computers.

Surface water quality modelling
Design of parallel numerical methods for the simulation of water pollution (calamitous releases), the marine eco-system, dispersion of river water, sediment transport, etc.

Partial differential equations in porous media research
The modelling of transport processes in the subsurface, with emphasis on the analytical study of the governing partial differential equations.

Exploratory research: analysis of PDE's and their discretization
Fundamental research – though possibly application driven – into mathematical questions and properties, currently focussing on initial-value problems for parabolic equations and on the Method of Lines applied to fluid mechanics.

Industrial Processes

Theme leader: P.W. Hemker

Computational fluid dynamics
Computation of flows in gases, liquids, plasticly deforming solids, or combinations of these (multiphase flows) for industrial applications. Current research includes advanced discretization methods for systems of nonlinear conservation laws, multigrid and sparse-grid solution methods, local grid adaptation and distributed computing.

Circuit simulation and control engineering problems
Development and implementation of new algorithms especially designed for use on parallel computer systems, in order to solve problems in circuit analysis and control engineering.

Plasma physics simulation
Development, implementation and analysis of new parallel iterative methods for the computation of eigenvalues and eigenvectors of generalized eigenvalue problems, where the matrices are very large, sparse and (complex) non-Hermitian. The algorithms are tested on problems coming from plasma physics and porous media flow.

Discontinuous dynamical systems
Research into the interaction of 'continuous' and 'discrete' dynamics, which arises when regime switches, occurring for example in electrical networks, in robotics and in processes with a discrete controller, are modelled as discrete events.

Computational number theory and data security
The study of algorithms for factorization and primality testing, for computing discrete logarithms, and (derived from these) for the solution of large, sparse systems of linear equations over finite fields. In addition, continued research into certain number-theoretic problems traditionally studied at CWI.

Modelling of processes in chemistry
Tuning mathematical models, given as systems of non-linear differential (-algebraic) equations, to sets of experimental data – currently mainly taken from (bio)chemistry – by identification (or estimation) of model parameters and the computation of the corresponding confidence regions in parameter space, and by studying the non-linear behaviour. Also optimal experimental design is included.

Mathematics of finance
In the present exploratory stage CWI expertise in relevant fields, including stochastic processes, numerical analysis, statistics and optimization, is linked to needs of the financial world (banks, pension funds, insurance companies), for example reduction of risk associated with the volatility of key indicators such as exchange rates, interest rates, and asset prices.

Exploratory research
The study of singular perturbation problems arising from the numerical computation of strongly convective flows, with emphasis on uniform convergence in 2D domains, in particular special non-uniform grid requirements and adaptive meshes.

Information Systems

Cluster leader: M.L. Kersten

Data Mining and Knowledge Discovery

Theme leader: A.P.J.M. Siebes

Data mining

For the data selection phase, research is focussed on structure in data, for example time-series, geographical data, multi-valued attributes, and non-universal relations, and on (non-)random samples. For data mining proper, the emphasis is on model-representation and search. An important aspect is the reformulation and generalization of well-known data mining algorithms in the KESO formalism.

Database architecture

Research on main memory and parallel DBMS architectures, emphasizing facilitation of data mining and efficiency; experiments with parallel data mining by porting MONET to systems like PARSYTEC and SP2; further development of performance monitoring and prediction software; (multi-)query optimization and support of special data types (e.g., time-series databases and picture databases), in connection with efficient data mining.

Active databases

The objective is to develop a model for autonomous databases. The emphasis is on the relationships and interactions between autonomous objects.

Multimedia and Human-Computer Interaction

Theme leader: D.C.A. Bulterman

Multimedia authoring systems

Based on CWI's CMIFed authoring system, durable, professional-quality multimedia documents are defined, which are quickly and easily adaptable to new environments, under control of the document's content owner, and research into system-level support for adaptive documents is conducted. Particular attention is paid to the author-controlled synchronization of independent data streams for presentation on heterogeneous (networked) user workstations.

Distributed multimedia applications

Presently the focus is on applications in distributed international electronic commerce, in particular the delivery of complex multimedia documents over high-speed (ATM) network infrastructures, and in remote access to publicly available information, for example the development of tools for network-based multimedia information sharing.

Interactive books

Software technology aspects of the development of a prototype for an interactive book on Lie algebras, emphasizing good design of the supporting architectures. In addition, more generic aspects of the construction of interactive books are considered, including the development of theory for the propagation of incremental changes across changes in data representation.

Specification of secure protocols

In order to bridge the gap between the theoretical considerations used to establish security and the actual semantics of implementations of such protocols, formalisms (notations and techniques) are investigated which are amenable to both theoretical analysis and practical implementation.

Interactive Information Engineering

Theme leader: P.J.W. ten Hagen

Information engineering framework

Building upon the experience gained in the MADE and PREMO projects, an information engineering workbench will be developed, for use in fields such as digital libraries and interactive books. Interoperability among presentation generators and players will be an important aspect.

Digital libraries

Development of a thesaurus based search facility in the vast on-line collections of mathematical and computer science research papers, of tools for creating dictionaries of keyphrases, through which user-friendly access to the real material is achieved, and of a standard dictionary for mathematics on behalf of authors and publishers. Another goal is the interactive enrichment of on-line mathematical texts.

Applied logic and interactive books

Application of logic, viewed as a science of information, to key issues in the processing and engineering of information, and the dissemination of this view of logic. Production of interactive textbooks where logic is presented as a core part of an emerging science of information processing and information flow analysis.

Facial animation

Production of a prototype system capable of capturing facial emotional expressions as enacted by a speaking performer and of reproducing usercontrolled transformations of those expressions as part of information presentations.

Quantum computing and advanced systems research (pilot theme)
Theme leader: P.M.B. Vitányi

Quantum computing
Exploratory study of algorithms and systems based on quantum mechanical principles (in particular previous work on quantum cellular automata and quantum information retrieval algorithms). The aim is to enable the realization of practical quantum computation.

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.

Advanced algorithms and systems
Design and analysis of algorithms for distributed and parallel systems. 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'. Also mobile and nomadic computing and communication are considered.

INTERNATIONAL AND NATIONAL PROGRAMMES

This chapter 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,
- special role of CWI (if any),
- CWI project leader(s).

European Programmes

ESPRIT

CAFE (7023): Conditional Access for Europe
December 1992 – February 1996
DigiCash, PTT, Cardware, Gemplus, SEPT, Ingenico, SINTEF-Delab, Institut für Sozialforschung Frankfurt, Institut für Informatik Hildesheim, Siemens, Universities of Leuven and Aarhus
Coordinator: R. Hirschfeld

NeuroCOLT (8556): Neural and Computational Learning
January 1994 – January 1997
Royal Holloway and Bedford New College, Univ. Mons, Rheinisch-Westfälische Tech. Hochschule, Univ. Pompeu Fabra, Techn. Univ. Graz, London School of Economics, Helsingin Yuopisto, Lab. de l'Informatique du Parallelisme, Univ. Milan
P.M.B. Vitányi

MERCURY (20089): Performance Management of Commercial Parallel Database Systems
January 1996 – December 1998
ICL, IFATEC, ING, Heriot-Watt Univ.
M.L. Kersten

KESO (20596): Knowledge Extraction for Statistical Offices
January 1996 – December 1998
National Statistical Offices of Finland, Greece (via FORTH), and The Netherlands, Infratest Burke (D), Data Distilleries (NL), GMD (D), University of Helsinki
Project manager: A.P.J.M. Siebes

CHAMELEON (20597): An Authoring Environment for Adaptive Multimedia Documents
November 1995 – November 1998
CLRC (UK), Epsilon SA (GR), Cartermill International (UK), Comunicacion Interactiva (SP), Egnatia Epirus Foundation (GR), Cycnos Systèmes Ouverts (F)
D.C.A. Bulterman

OPERA: Open Payments European Research Association
January 1996 – December 1996
Cardware Ltd, SIGMEW (a consortium of European banks)
R. Hirschfeld

DELOS (21057): ERCIM Digital Library
March 1996 – March 1999
Elsevier, Univ. Michigan, all ERCIM Institutes
F.A. Roos

SAGA (21871): Scientific Computing and Algebraic Abstractions
September 1996 – September 1997
Univ. Bergen, Univ. Swansea
J. Heering

CONFER II (21836): Concurrency and Functions: Evaluation and Reduction
November 1996 – November 1999
INRIA, ENS, CNET, ICL, KTH, Universities of Bologna, Cambridge, Edinburgh, Pisa, Sussex and Warwick
J.W. Klop

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

COORDINA : From Coordination Models to Applications
1997 – 2000
INRIA, Xerox, 8 European Universities, Univ. Leiden, Hollandse Signaalapparaten
J.J.M.M. Rutten

COST

Verification and Validation Methods for Formal Descriptions (247)
1993 – 1997
Philips, Universities of Utrecht, Eindhoven, Amsterdam, Nijmegen, Groningen, Twente + institutions in 18 European countries
J.F. Groote

MAST Marine Science and Technology

NOWESP (MAS2-CT93-0067): North-West European Shelf Programme
September 1993 – September 1996
RWS, Institut für Meereskunde, Univ. Leuven, NIOZ, Proudman Oceanographic Laboratory Bridston, Sir Allister Hardy Foundation for Ocean Science, Institute of Marine Research, Inst. für Ostseeforschung, Delft Hydraulics, BSH, IfBM, IFREMER, MUMM, Univ. Delft, Trinity College, Universities of Bordeaux and Liverpool
P.J. van der Houwen

MMARIE: Application of High Performance Computing Techniques for the Modelling of Marine Eco Systems
February 1995 – February 1998
Univ. Leuven, Delft Hydraulics, Univ. Southampton, IFREMER, CRS Cagliari, Univ. Hamburg, Univ. Liège, Univ. Delft, RIKZ, CETIIS, Univ. Bradford, Hydraulic Research Wallingford, Proudman Oceanographic Laboratory Bridston, UP de Catalunya.
P.J. van der Houwen

RACE/ACTS

SEMPER (AC026): Secure Electronic Marketplace for Europe
September 1995 – September 1998
Cryptomathic (DK), DigiCash (NL), Eurocom Expertise (GR), Europay International (B), FOGRA Forschungsgesellschaft Druck (D), GMD (D), IBM European Networking Center (D), Intracom (GR), KPN Research (NL), Otto-Versand (D), r3 security engineering (CH), SEPT (F), Univ. of Freiburg and Hildesheim (D)
D.C.A. Bulterman

TELEMATICS

STEM (EN1014): Sustainable Telematics for Environmental Management
January 1996 – January 1997
Univ. Edinburgh, Assynt Crofters Trust, Software AG España, CEAM, Univ. Karlsruhe, Implex Environmental Systems
L. Hardman

FRACAS (LRE 62-051): A Framework for Computational Semantics
January 1994 – April 1996
Univ. Edinburgh, Univ. Saarland, Univ. Stuttgart
D.J.N. van Eijck

DACCORD (TR1017): Development and Application of Coordinated Control of Corridors
January 1996 – January 1999
Hague Consulting, TU Delft, Univ. Lancaster, TNO, RWS, Univ. Naples, CSST, Autostrade Italia, INRTS, Ile de France, Ville de Paris, Univ. Crete, TCU
J.H. van Schuppen

VALUE

POWER (CTT-646): Performance Oriented Workbench Experiment on Real Information Systems in the Energy Field
1994–1996
IFATEC
M.L. Kersten

CAFE Exposure (CS-657): Promotional Activities related to the ESPRIT 7023 project CAFE
1994–1996
Cardware
Coordinator: R. Hirschfeld

SCIENCE/HCM

EXPRESS: Expressivenes of languages for concurrency (CT93-0406)
1994–1997
Univ. Utrecht, SICS, Univ. Genova, Univ. Rome (La Sapienza), Univ. Hildesheim, Univ. Amsterdam, INRIA, GMD, Univ. Sussex, Univ. Nijmegen
Coordinators: J.F. Groote/J.W. Klop

Statistical inference for stochastic processes (CT92-0078)
1993–1996
Universities of Paris VI, Berlin, Aarhus and Freiburg, INRIA
K.O. Dzhaparidze

EUROFOCS: European institute in the logical foundations of computer science (CT93-0081)
1994–1996
Univ. Edinburgh, INRIA, Universities of Pisa and Cambridge, ENS
J.W. de Bakker

The equations of fluid mechanics and related topics (CT93-0407)
1994–1996
CMAP, Univ. Paris VI, IX, XIII, Univ. Pisa, Univ. Ferrara, Univ. Nantes, IST (Lisbon), Universities of Trento, Pavia, Grenoble, Coimbra, London and Valladolid
J.G. Verwer

Algebraic combinatorics (CT93-0400)
1993–1996
Univ. Magdeburg, KTH Stockholm, Univ. Perugia, Univ. Cagliari, Univ. Bielefeld, Univ. Strasbourg, Univ. Bayreuth, Univ. Vienna, Univ. Paris VI, Univ. College of Wales, Universities of Copenhagen, Erlangen and Bordeaux I, Konrad Zuse Inst.
M.A.A. van Leeuwen

DONET: Discrete optimization and applications (CT93-0090)
1993–1996
Univ. Joseph Fourier, ZOR Bonn, Univ. Oxford
A. Schrijver

DIMANET: Discrete Mathematics Network (CT94-0429)
1994–1996
Universities of Bielefeld, Bologna, Cambridge, Montan, Lisbon, Madeira, Milan, Oxford, Paris 1, and Umea, Konrad Zuse Zentrum, Danmarks Tekniske Univ., Techn. Hochschule Darmstadt, Ecole Polytechnique Lausanne, Queen Mary and Westfield College, ENS Lyon, CNRS, Kungliga Tekniska Hogskolan
A. Schrijver

ERCIM computer graphics network (CT93-0085)
1993–1996
P.J.W. ten Hagen

ERCIM advanced databases technology network
1994–1997
M.L. Kersten

DIA: Digital Identification and Authentication (CT94-0691)
1995–1997
ENS, Univ. Salerno, Univ. Saarland, Univ. Aarhus
Coordinator: R. Hirschfeld

INTAS

Network Mathematical Methods for Stochastic Discrete Event Systems
1994–1997
O.J. Boxma

National Programmes

SION (Netherlands Computer Science Research Foundation)

Computational Learning Theory
1992–1996
P.M.B. Vitányi

Declarative and procedural aspects of non-standard logics
1992–1996
K.R. Apt

MathViews – Functional and architectural aspects of mathematical objects
1992–1996
A.M. Cohen

From ideas to reality – Implementing cryptography
1994–1998
L.G.L.T. Meertens

WINST: Themes for collaboration in mathematics and computer science
1994–1997
Universities of Nijmegen and Eindhoven
J.W. Klop, M. Hazewinkel

Design theory for autonomous databases
1993–1997
A.P.J.M. Siebes

Incremental parser generation and disambiguation in context
1993–1997

Univ. Amsterdam
D.J.N. van Eijck

MDL Neurocomputing
1994–1998
P.M.B. Vitányi

Equational term graph rewriting
1994–1998
J.W. Klop

Generic tools for program analysis and optimization
1994–1998
P. Klint

Checking verification of concurrent systems with type theory tools
1994–1998
Univ. Utrecht
J.F. Groote

Constraints in object-oriented interactive graphics
1994–1998
Univ. Eindhoven
P.J.W. ten Hagen

Higher-order and object-oriented processes (HOOP)
1994–1996
Universities of Eindhoven and Leiden
J.W. de Bakker

Constraint-based graphics
1994–1997
Univ. Eindhoven
F. Arbab

Cryptography, learning and randomness
1994–1997
Univ. Amsterdam
P.M.B. Vitányi

Classifying proof techniques for propositional logic
1994–1999
Univ. Delft
J.F. Groote

Scientific Visualization – from data visualization to interactive exploration
1995–1997
Univ. Delft, Free Univ. Amsterdam
J.J. van Wijk

Parallel declarative programming: transforming logic programs to lazy functional programs
1996–1998
K.R. Apt

Dynamic logic, artificial intelligence and information processing
1996-1998
D.J.N. van Eijck

SWON *(Netherlands Mathematics Research Foundation)*

Computational number theory: factorization of large integers
October 1992 – September 1996
Univ. Leiden, Oregon State Univ.
H.J.J. te Riele

Nonlinear convection and diffusion of contaminants in porous media
December 1992 – November 1996
Univ. Bath
C.J. van Duijn

Computer intensive methods in stochastics
September 1993 – December 1997
Cornell Univ., Math. Inst. Hungarian Ac. Sc., Univ. Utrecht, Delft, Leuven, Cambridge (UK), Chalmers (Gothenburg), Rome
M.S. Keane

WINST: Themes for collaboration in mathematics and computer science
October 1995 – September 1997
Univ. Nijmegen, Eindhoven
J.W. Klop

Lie theory and special functions
December 1995 – June 1996
Univ. Leiden, Amsterdam, Nijmegen, Twente, Utrecht
M. Hazewinkel

Statistical properties of movements in the plane
February 1996 – January 2000
M.S. Keane

Integer polyhedra and binary spaces
June 1996 – May 2000
A.M.H. Gerards

Discontinuous dynamical systems
August 1996 – July 2000
Univ. Groningen, Twente, Brabant
J.M. Schumacher

ALADDIN – Algorithmic Aspects of Parallel and Distributed Computing
1992–1996
Univ. Utrecht
P.M.B. Vitányi

NFI *(National Facility Computer Science)*

Intelligent CAD systems (MANIFOLD)
1986–1996
Univ. Eindhoven
F. Arbab

Special NWO projects

AIDA: Algorithms in algebra
1993–1996
Universities of Eindhoven, Groningen and Twente
M. Hazewinkel

Nonlinear systems
1993–1996
Universities of Groningen, Delft, Utrecht, Wageningen and Leiden, KSLA
C.J. van Duijn

Computationally intensive methods in stochastics
1993–1996
Universities of Leiden, Amsterdam, Rotterdam, Utrecht and Nijmegen
M.S. Keane

Computational number theory
1993–1996
University of Leiden
H.J.J. te Riele

Plasma simulation
1993 – 1998
Univ. Utrecht, FOM Inst. Plasma Physics
H.J.J. te Riele

Singular perturbation problems in strongly convective flows
1992 – 1997
Univ. Nijmegen, Inst. Math&Mech Ekaterinenburg, Novosibirsk
P.W. Hemker

Applied logic dissemination (Spinoza)
1997 – 2002
ILLC
D.J.N. van Eijck

STW *(Foundation for the Technical Sciences)*

Parameter identification and model analysis for nonlinear dynamic systems
1993–1997
Univ. Delft, Heidelberg, Beer Sheva, Nova Scotia
P.W. Hemker

Parallel codes for circuit analysis and control engineering
1993–1997
Univ. Amsterdam, Philips, AKZO-Nobel
P.J. van der Houwen

ACELA – Architecture of a Computer Environment for Lie Algebras
1993–1996
Univ. Eindhoven
L.G.L.T. Meertens

Wavelets
1996–1999
Univ. Delft, Eindhoven, Groningen, Shell, KNMI, MARIN
N.M. Temme

NCF

CIRK: Mathematical modelling of global transport and chemistry of trace constituents in the atmosphere
1994–1997
Univ. Utrecht, RIVM, KNMI, TNO, Univ. Iowa
J.G. Verwer

Cray-Long Term Ozone Simulation
1994–1997
Delft Hydraulics, Min. Public Works-RIKZ, Univ. Delft, IFREMER, Univ. Leuven
P.J. van der Houwen

ICES HPCN Programme

HPCN for Environmental Applications
1996–1999
Univ. Delft, Delft Hydraulics, TNO
J.G. Verwer

IMPACT – HPCN for Financial Services
1996–2000
ING, Univ. Amsterdam, Univ. Twente, Getronics, Univ. Delft, CAP Volmac, Data Distilleries, BIT by BIT
A.P.J.M. Siebes, M.L. Kersten

High Performance Visualization
1996–1998
ACE, CAP Volmac, Arcobel, TNO
R. van Liere

NICE – Computational Fluid Dynamics
1996–1999
TNO, Univ. Twente, MARIN, Delft Hydraulics, Univ. Groningen, Univ. Delft, NLR
F. Arbab

ELSIM – Electrotechnical Simulations
1996–1999
Univ. Utrecht, Philips, Universities of Leiden and Groningen
P.J. van der Houwen