# CSC news

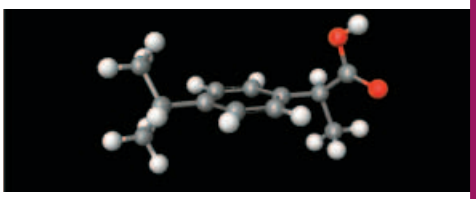

## Finnish radio astronomers

# First in world with ten-gigabit connection

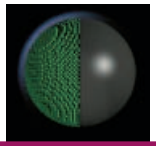
# contents 3/2006












CSCnews

Volume 18, Number 3

ISSN 0787-7536

CSCnews focuses on computational science and communicates ongoing research and research results to its readers in a way we hope will be both beneficial and entertaining. CSCnews is published four times a year.

**Editors-in-Chief**
Juha Haataja
Leena Jukka

**Editorial Board**
Pirjo-Leena Forsström, Earth sciences
Matti Gröhn, Visualization
Juha Haataja, Editor-in-Chief
Leena Jukka, Editor-in-Chief
Tiina Kupila-Rantala, Data storage systems
Per Harald Jonson, Bioinformatics
Katri Luostarinen, EU Coordination
Jura Tarus, Nanoscience

**Email addresses:**
first name.last name@csc.fi

**Graphic design and layout**
Maarit Kattilakoski
Unionimedia Oy

**Printed by**
Topnova Oy

**Subscriptions**
Subscriptions are free of charge.
Please use email:
tilaukset@csc.fi

**Publisher**
CSC



# editorial

## Building a European high-performance computing ecosystem

Kimmo Koski

"Ecosystem, *noun*: the complex of a community of organisms and its environment functioning as an ecological unit", *Merriam-Webster On-line Dictionary*

During the past few decades, supercomputing industry has become a battlefield of the biggest performance numbers. Ranking lists, such as the Top500, boost further the cold war directing the attention from the quality of the scientific results to pure hardware performance. Nations compete for the Top500 positions, as well as the computer vendors, all aiming at maximum publicity.

The systems are estimated to reach the next magic number, petaflop/s computing speed, during the next few years. Both, the USA and Japan, have already announced heavyweight projects probably nominated for world championships in computing power.

Should Europe join the competition for the greatest computer power? How can European computational science compete with the USA and Japan, who have a supercomputing industry of their own? These questions have lately puzzled the European decision makers and created initiatives to build European level resources with extreme computing power. But is pure computing performance the way to the success in science?

This key question was discussed during the European supercomputing conference, ICS2006 in Dresden last June. The answer was obvious—performance alone does not ensure superior science—but to compete with computational research results supported by extreme computing power Europe needs an innovative approach, efficient collaboration and competent people.

This discussion lead ten European countries to form a task force to define a strategy to develop HPC infrastructures in Europe. This HPC European Task Force, or HET for short, will publish its recommendations at the end of this year, after which it will be disbanded. These recommendations will include, for example, different scenarios for the European HPC ecosystem, funding models for extreme computing, models for boosting the collaboration and resource exchange over country borders, and also proposals on how to allocate the possible European-level computing resources to research groups.

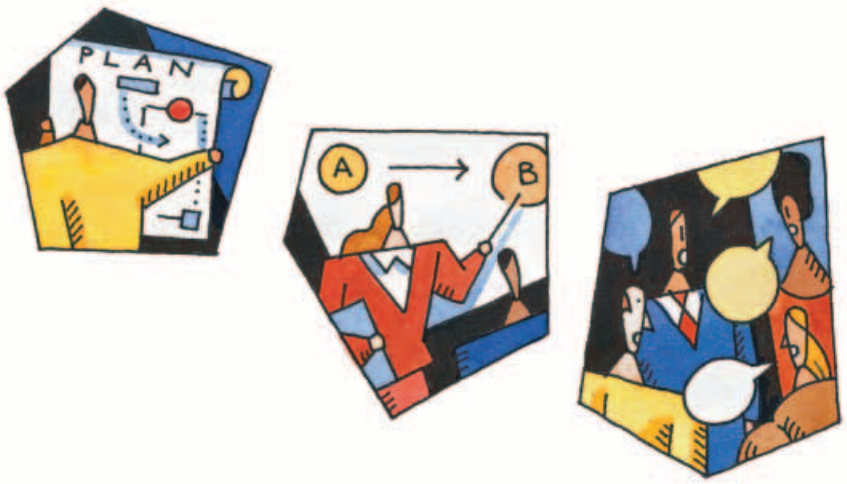
One of the key principles for HET is that it does not consider only the extreme computational power, but focuses to the whole pyramid of computational resources—from national systems to European-level resources. This is extremely important since success in computing intensive research requires much more than just the system: for example scalable algorithm design, skills for parallelization of software and code optimization competence can make a difference.

In general, competence in computational science is much related to individuals and is distributed widely in Europe. If we can build an efficient HPC ecosystem, in which collaboration between computational scientists and supercomputer centers of different size is intense and multi-scientific, we can close the gap in hardware performance through quality of service.

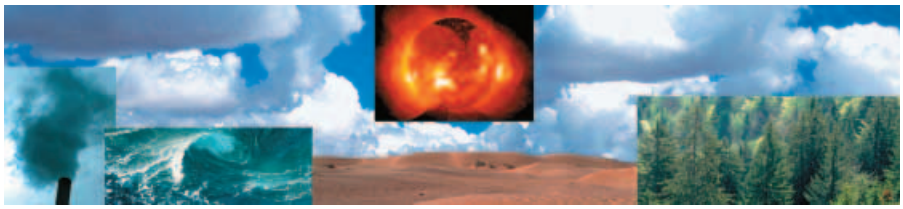
HET work is in its early stages, but already now I can see an enthusiastic atmosphere. Willingness to collaborate is obvious. This increases my confidence in the ability to develop a unique European model, which has sufficient performance for the most demanding computational research and grand challenge projects, but at the same time gives good, reliable service to the rest of the scientific community.

The next six months will be interesting—and hopefully rewarding through increased collaboration among the computational science domain of Europe.

The writer is Managing Director of CSC and chairman of HPC European Task Force. He can be reached via email kimmo.koski@csc.fi.




**"If we can build an efficient HPC ecosystem, in which collaboration between computational scientists and supercomputer centers of different size is intense and multi-scientific, we can close the gap in hardware performance through quality of service."**



## Earth System Models on the rise

The General Assembly 2006 of the European Geosciences Union (EGU) took place at the Austria Center Vienna in Vienna, Austria, at the beginning of April 2006. With a few thousand participants, it is by far the largest conference event in Geosciences in Europe.

"With countless parallel session, it is impossible for a single person to get an overview over more than one or two fields of the wide spectrum covered by Geosciences," reports Dr. **Thomas Zwinger**, CSC's application specialist focusing on fluid dynamics and glaciology.

"A newly emerging field is the application of Earth System Models to address certain climate related issues,especially in connection to global warming and climate change," says Zwinger. "Earth System Models combine different models describing the climate system. The coupled models change information and are thus able to mimic the interactions of climate system components." The interest in abrupt climate change has been rising. "These events, known as Dansgaard-Oeschger and Heinrich events, have occurred in the history of our planet's climate," says Zwinger and gives an example of **A. Ganopolski's** presentation at the EGU2006 meeting.

"Using combined atmosphere, ice sheet and ocean model (CLIMBER II), run on a millennium timescale with bi-directional coupling of Cryosphere and ocean, some of the principal mechanisms of these events could be elucidated. The theory explains the observed instability of the glacial climate as a result of rapid reorganizations of the ocean circulation, and a strong, bi-directionally interaction between North Atlantic climate and the ice sheets."

The abstract of this contribution can be found via the URL http://www.cosis.net/abstracts/EGU06/04712/EGU06-J-04712.pdf

The next European Geosciences Union General Assembly will be held in Vienna, Austria, 15–20 April 2007. The call for papers is from October to December. For further details go to http://meetings.copernicus.org/egu2007/ (CSC)

## Leif Laaksonen appointed chair of e-IRG

The e-Infrastructures Reflection Group (e-IRG) was founded by the European Commission to define and recommend best practices for European grid efforts. It consists of official government delegates from all EU countries, and the chair is reserved for the current EU president country.

The Ministry of Education appointed **Leif Laaksonen**, CSC's Director of Customer Support, as chair for the period of 1.7.–31.12.2006.

The main objective of the eInfrastructure initiative is to support the creation of a framework for the easy and cost-effective shared use of distributed electronic resources across Europe—particularly for grid computing, storage and networking.

Current important issues within the e-IRG are e-infrastructures in FP7, a policy for resource sharing, a registry/repository for European resources, coordination of new national and EU funding programs, and better links between Europe and other regions (*e.g.*, USA, Japan) engaged in similar activities

For more information, please visit www.e-irg.eu (CSC)

## ICTs to modernize health care services

The focus of future health care will shift from the treatment of illnesses to prevention. Indeed, health and welfare policy as a whole will need to be completely rethought, say experts from the Well-being and Health panelof the joint Academy of Finland and Tekes Finnsight 2015 project.

The health care sector is threatened by a shortage of labour, funding problems in local government and staff overload. There are also problems with the availability of specialised health care and emergency care arrangements in local regions. The structures and mechanisms of the service system can be completely overhauled by means of ICTs.

With the continuing ageing of the population and rural depopulation, there is a growing demand for home care and telecare services produced by health technology. Telecare systems also raise new kinds of professional and ethical questions.

New ICTs are also paving the way to improved customer management. With each patient's consent, health care staff can easily access relevant patient information - by virtue of wireless terminals even without time and place constraints. A national health portal is needed as a source of general information and guidance on the use of services and on how to apply for benefits. Information system services are important not only for improved health care delivery, but they are also valuable export products. Finnish businesses are exceptionally well placed to make use of the electronic health record and the national system architecture in their own product development.

Research in biomedicine can pave the way to breakthroughs in health care The impacts of research in biomedicine extend from basic research through to the health care sector. Genomics and proteomics knowledge is opening up new avenues for the pre-

© Tekes, Matias Uusikylä

vention, diagnosis and treatment of diseases. Practical applications are related first and foremost to the diagnosis and treatment of cancer.

Finland's competitive advantage is based upon public health care, an atmosphere that is favourable to research, well-researched and well-described population and patient datasets and the country's high standard of research.

According to the panel experts, Finland should invest in molecular medicine, nanoscience and nanotechnology and in practical applications in these fields. Finland is also strongly placed in the field of stem cell research, which in the near future may be bringing changes to treatment practices in many disease groups.

Brain research is making rapid progress and producing applications for the diagnosis and treatment of brain disorders and psychiatric illnesses. Practical applications have emerged most particularly in the field of imaging. (Academy of Finland)

## Teaching was a new topic at the Electromagnetics seminar 2006

The Electromagnetics seminar is an annual Finnish meeting, which provides a one-day forum for scientists to present their studies on electromagnetic topics and to meet each other enjoying an informal atmosphere. The 15th meeting on August 29, 2006, took place in the new building (Dynamicum) of the Finnish Meteorological Institute and the Finnish Institute of Marine Research at the Kumpula campus in Helsinki. The number of participants was about 50 coming from about 10 different institutes.

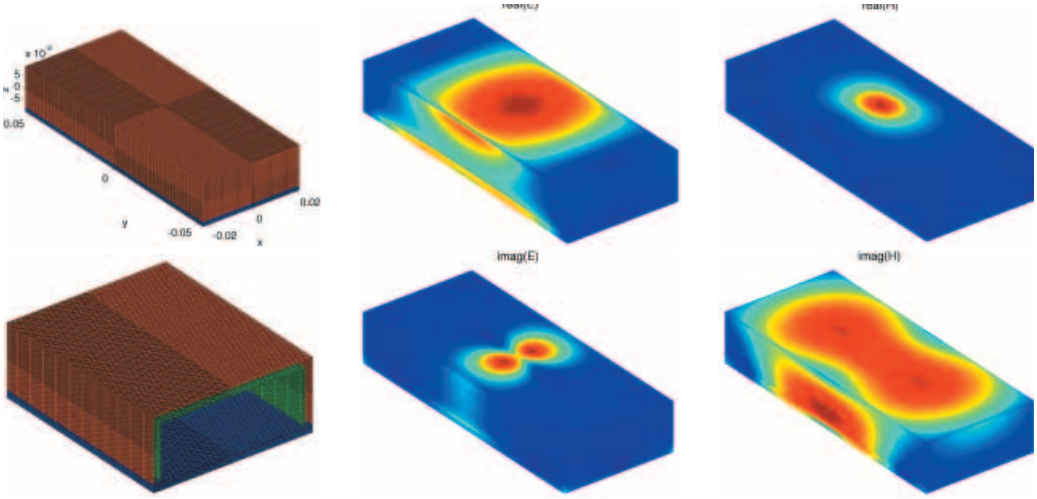
There were 15 presentations giving an extensive snapshot of the present electromagnetics research in Finland. The spectrum of studies is very wide containing new ideas in the basic theory and the rapid development in numerical methods as well as diverse applications from ground-based technology to space exploration. As a new topic in the seminar, there were two presentations dealing with teaching.

The activity of young scientists, both doctoral students and recently graduated doctors, to present their studies was very delightful. Participants were clearly satisfied with the format of the meeting, so the seminar series will continue. The next organiser is the Finnish Defence Forces Technical Research Centre and the place will be Riihimäki.

The abstract book of the seminar is freely available on the WWW server of the Finnish Meteorological Institute: http://www.fmi.fi/kuvat/smg2006_netti.pdf

More information of the seminar series is available on http://www.csc.fi/seminaarit/Sahkomagnetiikka/
(FMI)

© Matti Taskinen, TKK




Matti Taskinen from the Helsinki University of Technology presented an example of discretization and calculated electromagnetic fields of a plastic cover on top of a metallic plate. The cover and the metallic plate are 2 mm thick and the relative permittivity of the cover is 2. The source is a 4 GHz x-polarized electric dipole inside the box.
The fields are modeled with a new type of a frequency domain surface integral equation, Current and Charge Integral Equation formulation.
The discretization has more than one hundred thousand triangles, half a million unknowns and the matrix size is larger than four terabytes. The iterative solution converged in 21 iterations with a relative tolerance of 1/100.

## Scalable parallel applications at DEISA training

The 1st DEISA training organized at the beginning of July in Paris gathered together 30 participants interested in HPC and Grid computing. In this session, first a global introduction to the DEISA infrastructure and its general middleware services were given.

Then, the focus of the training was put on highly scalable parallel applications. First, HPC specialists presented tools and techniques for achieving optimal scaling on a large cluster of shared memory nodes. These presentations covered all the main aspects of optimization work, from some algorithmic aspects to the usage of performance analysis tools to detect the bottlenecks in the applications, and to various techniques to optimize especially communications and I/O.

Scientists from the fields of cosmology, climate modeling, fluid dynamics, and plasma physics reported on their experience on highly scalable parallel applications on the DEISA infrastructure through use case presentations. For instance, one speaker explained how he succeeded to reach nearly optimal scaling properties up to 4000 processors, switching to a completely different decomposition domain strategy then working on the optimization of the new code.

According to the feedback received, the participants were very satisfied with the training session. The scientists felt that the training had had a positive contribution in support to their research activities. In particular they appreciated the lectures on new techniques in

## Grids and Web Services in bioinformatics

Grids and Web Services offer interesting possibilities for the bioinformatics community. However, the usage of these fast-developing technologies is hampered by the difficulty of obtaining up-to-date, bioinformatics-focused information.

To remedy the situation, a symposium "Grids and Web Services in bioinformatics" was organized at CSC in June, with over 50 bioinformaticians from 25 countries attending. The symposium provided an update on grids and Web Services, and presented grid user experiences in bioinformatics.

The program covered the technology and benefits of Web Services, which can be accessed programmatically and combined to workflows using systems like Taverna. Also the newly developed grid infrastructures EGEE and DEISA were presented, with several bioinformatics user experiences. As grid and Web Services technologies are approaching each other, the importance of interoperability, standards, and best practices was emphasized.

The symposium was organized by the EU-funded EMBRACE Network of Excellence, which aims to build a bioinformatics service grid in Europe. Several speakers were also provided by the EMBnet, a worldwide network of bioinformatics service providers.

For the symposium presentations and further information, please see:
http://www.csc.fi/molbio/opetus/embrace/gridWS.html
http://www.embracegrid.info
http://www.embnet.org/

supercomputing and the opportunity to exchange ideas with researchers from different scientific disciplines.

The next DEISA training session will be organized in Jülich, Germany, on October 23–25, 2006. More information on this session is available at www.deisa.org/training. (CSC)

## HD video over IP Internet's next killer application

Multimedia, transmitting the live high-definition (HD) video and audio over IP will be the next killer application of the Internet. That was the main message of Terena Networking Conference 2006, which was held 15–18 May in Catania, Italy.

Terena is the organization for European research networks and its purpose is to promote cooperation and development of new network technologies. Its main event is a yearly conference, which gathers network specialists round the world to present new ideas and to discuss them.

**Egon Verharen** from SURFnet, Netherlands told about experiments to deliver HDTV programs over network. As a part of this experiment SURFnet distributed all 64 games of World Championship Football in HD over IP. Because of copyright restrictions the programs were only available on university campuses.

Another example of HD over IP was Vision 05, a research vessel mission of University of Washington. HDTV cameras take pictures of undersea earthquakes or volcanoes that emit black smoke 2500 meters below the Pacific surface and the video is readily immediately made available across IP networks **Mike Wellings** showed these impressive videos as part of his presentation.
http://www.terena.nl/events/tnc2006/programme/
(CSC)

## CSC facilities presented to Minister of Education and Science Antti Kalliomäki

**Antti Kalliomäki**, Minister of Education and Science, visited CSC on August 29, 2006. CSC's Managing Director **Kimmo Koski** introduced CSC's mission statement to the minister: As part of the national research system, CSC develops and provides state-of-the-art IT services for science.

The specific feature of the Finnish model is that all services are provided by a centralized national center. CSC is state-owned and governed by the Ministry of Education.

CSC's services comprise computation, Funet data communication services, multidisciplinary specialist services, and data management services.

Kimmo Koski introduced the currently active project to acquire a new supercomputer. The objective is to develop the supercomputing environment to an internationally competitive level. The decision about the equipment will be made in September–October 2006, and the installation will be implemented one phase at a time over the period 2006 to 2008.

"The national infrastructure must be sufficient to grant carefully chosen researcher groups access to international state-of-the-art resources," Kimmo Koski informed the minister. In Koski's opinion the essential challenges deal with financing infrastructures.

## Nanoscience Days 2006

This multidisciplinary, annual two-day seminar is organized on October 26-27, 2006 at the Nanoscience Center (NSC) of the University of Jyväskylä.

Registration deadline is October 13. For more information: http://www.nano.jyu.fi/NanoscienceDays2006.html

## Human Genome Meeting 2006 in Finland

Human Genome Meeting, held at Helsinki Fair Center at the turn of the month, gathered together about 720 delegates and 44 company or institute exhibitors. Opening words given by Professor **Leena Peltonen-Palotie**, President of HUGO, and **Tuula Haatainen,** Minister of Social Affairs, primed the excellent scientific program that covered topics of current interest, such as large scale genomic projects and population studies, and in-depth analysis of human genome. Finnish research was well represented, since almost all sessions had a Finnish representative either as a speaker or as a chair. CSC delegate, **Jarno Tuimala**, gave a speech and presented a poster in the genetic epidemiology workshop.

## New services for videoconferencing

There is a growing interest in videoconferencing among the higher education and research community. To make the use of videoconference tools easier for newcomers the Funet community organized a VideoFunet project to gather basic information about video technology and videoconferencing, learning material and contact information.

The project published a www-site, where all the information is freely available. Most part of information is available in Finnish but essential information is in English as well. The address of the English version is http://www.video.funet.fi/index.php?lang=en.

### National Gatekeeper—CSC Gate

CSC offers for universities, polytechnics and companies a national H.323 reachability service. The aim of this service is to associate video conferencing equipment and video bridges at higher educational institutes with a uniform number space.

All the video conferencing devices and videoconferences will have a contact number of their own to be used instead of the IP-number of the terminal unit. CSC Gate will ease the organization of videoconferences and decreases the burden of bridge operators. To start a videoconference will become simpler and is it not anymore necessary to send IP-numbers back and forward.

To join the service an organization must have a gatekeeper server. The service is aimed to the producers of videoconferences and it is charged separately from the Funet membership fee. It is also open for commercial customers, whose system meets the compatibility requirements.
For more information, please look at www-pages: http://tv.funet.fi/videoneuvottelu/cscgate/index.jsp.en

### VideoParkki—a reachability service for end-users

A service called the VideoParkki is available to all Funet users who use H.323 videoconferencing. The VideoParkki gives the user a possibility to access the national H.323 service with a personal contact number instead of the IP-number of the terminal device. The use of the service is recommended if your home organization is not delivering gatekeeper services. Neighboring gatekeeper managers may also use it for testing and problem solving.

For more information, please look at www-pages: http://tv.funet.fi/videoneuvottelu/videoparkki/index.jsp.en
(CSC)

**Kimmo Koski elected chairman of HET**

## A strategy for European HPC in preparation

HET or the HPC European Task Force was formed as a temporary taskforce during the ISC conference in Dresden, June 2006, by participants from ten European countries interested in shaping the Continent's HPC infrastructure.

**Kimmo Koski**, managing director of CSC, was elected unanimously by the meeting to chair the Task Force until it is disbanded at the end of 2006.

"By that time, we will have a proposal for European HPC strategy ready to be published," says Koski, who sees this as a unique opportunity to increase HPC collaboration among European countries, to raise the visibility of computational science, and provide more resources for high-end computing.

The strategy is hoped to help the participants, all major HPC countries, in their infrastructure planning, and it will also be delivered to the EU bodies responsible for infrastructure development. HET will report to e-IRG.

"All this is hoped to benefit those scientists who need extreme computing resources beyond national systems. This applies to Finnish groups as well as scientists from other European countries," says Koski. "These scientists are working with computationally challenging problems such as those related to climate change or high-end physics, and in order to make the best possible use of the computing power available in Europe in the future, we need a strategy that most European HPC countries can commit to." (CSC)

## Millenium technology prize to professor Shuji Nakamura

Finland's 2006 Millennium Technology Prize, worth one million euros, has been awarded to **Shuji Nakamura** from the University of California, Santa Barbara, USA. Professor Nakamura has developed a new revolutionary source of light—bright-blue, green and white LEDs and a blue laser.

The Millennium Technology Prize is awarded by Finland's Millennium Prize Foundation for a technological innovation that significantly improves quality of human life and promotes sustainable development. Professor Nakamura's invention of new sources of light fulfils these criteria in a splendid manner. Applications that have already been developed using his innovation indicate that the technology can reduce global energy consumption, bring environmentally-sound and energy-efficient lighting to developing countries, sterilise water in more efficient and cheaper ways, and be used to store data much more efficiently.

The Millennium Technology Prize is funded by public and private actors, including the Academy of Finland and the Finnish Funding Agency for Technology and Innovation Tekes. Shuji Nakamura will receive the prize at a ceremony to be held in Helsinki in September 2006. (Academy of Finland)

## Baltic Sea research

The eight Baltic Sea EU Member States and the Russian Federation are preparing a joint Baltic Sea research programme to be launched in 2007.

The aim of the programme is to produce knowledge and tools, which improve the predictive capacity of the Baltic Sea ecosystem's response to human-induced changes and support better management of the environmental problems of the Baltic Sea. Funding for this five-year programme is planned to come from national research funding agencies and from the European Commission.

The planned research programme will include research aiming at solving the most severe environmental problems in the Baltic Sea, namely combating eutrophication, achieving sustainable fisheries, protecting biodiversity and preventing pollution. It will also take into account conditions of the socio-economical and governance structures, which are necessary for sustainable development of the sea. The preparation of the science plan of the programme is underway and it has involved over 800 scientists in all nine Baltic Sea states.

For more information, please visit. www.bonusportal.fi (Academy of Finland)

## IST2006 in Helsinki on 21.–23.11.

The IST 2006 Conference Programme looks at one of the central questions facing European competitiveness today from two different angles: "Creating a virtuous cycle between ICT research innovation and socio-economic benefits". IST 2006 is being held as the Commission launches FP7, its Seventh Framework Programme for Research and Development, so one of the main themes of the event will be FP7's ICT objectives and procedures.

In today's EU, however, ICT research does not take place in isolation. Reflecting the central importance ICTs play in stimulating innovation and competitiveness in all areas of industry, the conference programme will also contribute to the creation of a virtuous cycle between ICT research innovation and socio-economic benefits.

Purely scientific and technical subjects, on the other hand, will be addressed in the networking and workshop sessions.

The on-line registration for the IST2006 conference in Helsinki is now open. More information on the conference can be found at the web site www.ist2006.fi (CSC)

## Academy of Finland to pin down health effects of nanotechnology

At present, not enough is known of the health effects of nanotechnology. The Academy of Finland has, therefore, selected the consortium Nanohealth to spearhead the Research Programme on Nanoscience (FinNano). The objective of the project is to produce nanoparticles, define particles that affect working environments and assess exposure to them and their health effects. The project is headed by Professor **Kai Savolainen** of the Finnish Institute of Occupational Health.

Apart from health promotion, FinNano also provides funding for molecular electronics research and research into mechanical characteristics of nanostructures as well as for basic research in the field. In addition to the scientific goals the programme aims to advance responsible development of nanotechnology by taking into account ethical challenges, *i.e.* matters related to safety, health and the environment.

The Academy of Finland will provide nine million euros in funding for the programme. (Academy of Finland)

# Stem cell research, tissue engineering, and a few words on banking

In the cleanroom, every particle in the air counts, since the scientists work on tissue and cell products that are planned to be used on patients.

Tissue engineering is a beautiful term that captures the idea of combining life sciences with engineering to build something new and complex. In this case, the ultimate goal is to build new human tissue to replace or replenish tissue that has been damaged or turned dysfunctional. The building blocks are human stem cells grown in vitro and coaxed to differentiate into specialized tissue cells.

Leena Jukka

Imagine having banks where surgeons could have custom-made body parts, tissue or organs to draw on when patients need them. Or stem cells, either the patient's own, or if they carry a hereditary disease, a healthy donor's cells, that could be used to generate healthy and functioning specialized cells, which could then replace the dysfunctional cells.

Researchers believe that it is possible to use stem cells to engineer tissue products that could mimic any tissue in the human body—the time span, however, is better left untouched, but it is a question of years, not decades.

© Leena Jukka, CSC

Professor Outi Hovatta, left, heads Regea's stem Cell Unit, professor Riitta Suuronen is the director of Regea, and Susanna Miettinen works as a researcher. Regea is one of the few cell and tissue technology centers worldwide to combine expertise in stem cells and biomaterials with tissue bank operations.

### Tissue in a freezer

"Today, most tissue banks are just freezers in the back rooms of operating theatres," says Professor **Riitta Suuronen**, Director of Regea in Tampere, Finland. The Regea Institute for Regenerative Medicine was founded in 2003, and in the summer of 2004 it started its Tissue Bank operations, which together with stem cell research and the related fields of biomaterial applications and cleanroom facilities form the core of Regea's actions.

Currently, the Regea Tissue Bank provides fresh frozen bone and within a month also cornea and amniotic membrane. The tissue selection will be enlarged later with processed bone, heart valves, and skin, according to customer needs. The tissue is used in transplants: bone in the reconstructive surgery of joint replacements, cornea for eye diseases, and amniotic membrane, *e.g.*, in the treatment of chemical injuries to the eye and in plastic surgery. Says Suuronen, "Tissue banks are mainly for clinical use, but we here at Regea can use tissue bank products for research in the field of regenerative medicine, if we have the doners' permission and the tissue can not be used for clinical purposes." Regenerative medicine is used as a synonym for tissue engineering, and is expected to form the third major clinical treatment modality in addition to traditional medicine and surgical treatments.

**"Tissue engineering relies on stem cell research."**

Even though Regea continuously collects tissue and organs from hospitals, (the donation is anonymous and based on written consent), these transplants are in short supply. Suuronen wishes, "that as many people as possible would carry an organ donor card stating that they wish their organs and tissues to be used for medical purposes after their death."

Regea Tissue Bank's goal is to be the first in Finland to meet the new EU tissue directive requirements. Having come into effect on 7.4.2006, the new directive sets strict quality requirements for those tissue banks storing human tissue and dealing in tissue donation. According to Riitta Suuronen, "As a result of the new directive, tissue banks will organize their activities in a new way. Many will terminate their operations altogether, but some will move on and comply with the directive requirements." She estimates that the UK now hosts two bone banks, and Finland will probably host five, one for each hospital district.

### An educated guess —and then miracles happen

Tissue engineering relies on stem cell research. Stem cells are the undifferentiated cells of the body, starting from the fertilized egg, the mother of all cells, which is totipotent. When the egg divides and forms an embryo, then a fetus and finally a grown individual, the

stem cell count diminishes and they lose their potency. But surprisingly, and in contrast to previous belief, even an adult body harbors stem cells in its tissue.

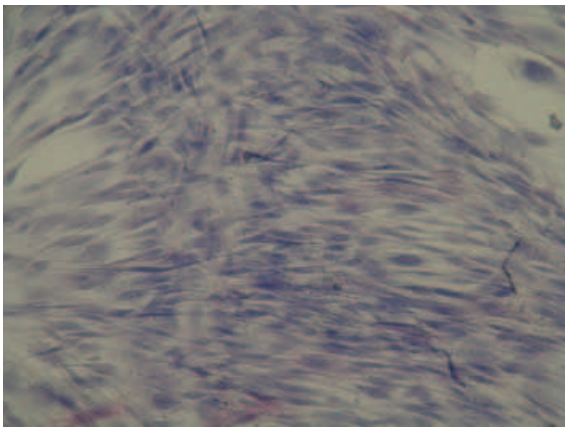
Stem cells in small quantities have been located in most tissue types of the adult body," says **Katriina Aalto-Setälä**, Medical Director of Regea. In the body, the stem cells either reproduce themselves as stem cells or differentiate into specialized tissue cells, like skin or bone. The role of stem cells in tissues like heart and brain is currently unknown, since they do not repair tissue damages in those organs.

The exact mechanisms that regulate a stem cell into a differentiated tissue cell are still a bit hazy, and in the laboratory, a lot is filed under miracles. In the laboratory, stem cells are grown in culture dishes with varying culture solutions. The stem cells are grown in an undifferentiated state, and after several divisions, the cell generations form a stem cell line. Says **Outi Hovatta**, who heads the Stem Cell Unit, "These cells need each other; they grow only in a population and in contact with one another." Quoting the 27 July issue of *Nature,* Finland has four stem cell lines and Sweden 55. At present, stem cell lines cultivated around the world are available only for research or animal testing, with the exception of few clinical trials in skin grafting.

One research interest is the culture solution; the solution has to be totally synthetic, when stem cells destined to be used in human cell therapies are grown, since animal substances can cause infections or immunogenity. This is a great challenge, according to Hovatta. Research is also done on the factors that induce the differentiation of cells: "The decision on which growth factors to use in the laboratory is based on an educated guess," says Hovatta, "We combine knowledge of the fetus' development with the gene expression of certain types of cells."

**"Research is also done on the factors that induce the differentiation of cells."**

But stem cell therapy is developing fast, and its rewards are expected to be colossal. "These cells will benefit dentistry, neurology, patients with bone and cartilage damage, stroke patients, heart disease," lists Hovatta. And expects to see clinical applications especially in the field of bone restructuring and neurology within a few years.



Left: Fat stem cells grown in a culture solution that keeps the cells in an undifferentiated state.

Right: Fat stem cells that have begun to differentiate into bone cells because they have been grown in change-inducing solution.

### The start of a heart

"Now we can differentiate stems cells so that they form a beating colony," says Katriina Aalto-Setälä. She shows us a film of a group of rhythmically beating cells, and continues: "The genes express all the right proteins of cardiac cells."

Adds Hovatta, "The electrical properties of the cells look like they should, and no other cells in the human body beat." Whether or not these cells would beat in harmony with the recipient heart, if ever transplanted, remains unanswered, says Hovatta, "they could start arrhythmia, or maybe not, we don't know."

Previously, scientists have succeeded in inducing human stem cells to differentiate into bone marrow cells, muscular and nerve cells, fat cells, skin, insulin-producing cells like the ones in the pancreas, for example. One landmark in tissue engineering was reached in April 2006, when the *Lancet* and *New Scientist* reported on the engineered bladder transplant work of **Anthony Atala** of Wake Forest University in Winston-Salem, North Carolina, US. Seven patients had been given lab-grown blad-

**"The immediate goal at Regea is to optimize the growing conditions for stem cells so that the cell lines cultivated will comply with the GMP requirements and be suited to treating human patients."**

## Coaxing a stem cell to change

**Susanna Miettinen** is a researcher at Regea studying mesenchymal stem cells and their use in a variety of tissue engineering applications. Mesenchymal stem cells can be found in the bone marrow, but are very rare: existing at an estimated frequency of about 1 in 100,000 bone marrow cells. Mesenchymal stem cells are capable of both self-renewal and differentiation into bone, cartilage, muscle, tendon, and fat. According to Miettinen, the recipe for differentiation goes in three steps:

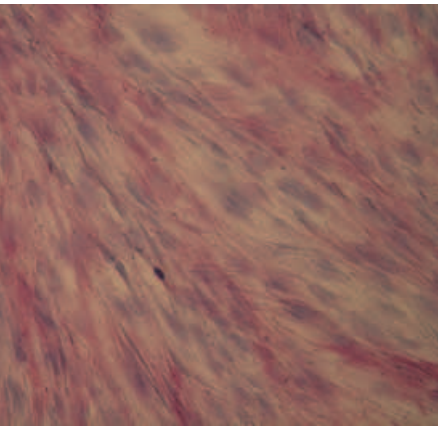
1) Add the growth factors, hormones, and adequate frequency of stem cells to the culture dish or combine the cells with the appropriate biomaterial, or scaffold, so the cells can be transferred into the patient.

2) Incubate in a cell culture incubator at 37°C for a couple of weeks.

3) Then stain the cells with bone cell specific histological staining. For example, if you have tried to coax the stem cells to differentiate into bone cells, you look for the cells that have produced the protein collagen type I. Instead of histological staining this can be done with the polymerase chain reactions, a standard tool used in bioinformatics.

This last stage kills the cells, so even though it is routine in research work, it is not done on the cells that are used to treat patients.

© Elisa Saarnio and Suvi Haimi, Regea



der transplants, engineered from their own cells, and the transplants had been functioning for an average of four years. It is pioneering work, but on the other end of the stick is the routine clinical practice of transplanting healthy bone marrow to leukaemia patients, so that the stem cells in the healthy tissue will replace the blood cells of the patient.

### A cleanroom that makes a surgery look like a dump

"The immediate goal at Regea is to optimize the growing conditions for stem cells so that the cell lines cultivated will comply with the GMP requirements and be suited to treating human patients," says Hovatta. Regea is aiming to be the first in the world to succeed in this endeavour, and, this will take some work, since the GMP or good manufacture practices "are really demanding. They are the same standards that are set for the pharmaceutical industry."

Good manufacturing practices (GMP) has been defined as the part of quality assurance, which is aimed at ensuring that products are consistently manufactured to a quality appropriate to their intended use. Good laboratory practices (GLP) are the guidelines for quality control and quality assurance in testing laboratories and thereby part of GMP itself. This international scheme is intended primarily for laboratories carrying out analysis and evaluations of substances for regulatory purposes.

Without displaying any demureness, both Suuronen and Hovatta explain how the working conditions at Regea and its GMP-standard laboratory and cleanroom facilities are of global excellence, and they were designed specifically for cultivating stem cell lines suited for human cell transplants.

Says Suuronen, "Our cleanroom makes the cleanest of the surgical theatres look like a dump", and explains that just to get in, you have to change dress three times, you cannot take any materials or equipment with you, since they enter through separate paths, and, preferably, you have to be female to work there. "Males secrete nine times more particles than women. And in the cleanroom, every particle counts."

The writer is a science editor at CSC, and can be reached by email: leena.jukka@csc.

## Regea Institute for Regenerative Medicine

Regea personnel work on the 6th and 7th floor of this Finn-Medi 5 building in Tampere, Finland. Nearby is the Tampere University Hospital and several research institutions, plus the ubiqituous Finnish nature to complete the research environment.

The Regea Institute for Regenerative Medicine is an independent institute founded at the University of Tampere in 2005. "This means that my boss is Rector of the University of Tampere," explains Director Riitta Suuronen. According to Suuronen, the staff of Regea is very research-oriented, Suuronen herself holds the post of Professor in Medical Biomaterials, at the Medical School of the University of Tampere. Outi Hovatta divides her time between Regea and the Karolinska Institute in Stockholm, Sweden, doing similar research.

Now Regea has 30 employees, but "when I started, in 2003 or 2004, we were a project of four people within the Institute of Medical Technology in Tampere," says Suuronen.

"In 2004, our new facilities, about 500m$^2$, were finished, and in 2005 Regea became an independent Institute of the University of Tampere."

© Jonna Roine, Regea

Regea is a consortium of five research and education institutes: the founder-members of Regea are the University of Tampere, Tampere University of Technology, the University Hospital in Tampere, Pirkanmaa Polytechnic, and Coxa Hospital for Joint Replacement

Regea is one of the few cell and tissue technology centres worldwide and the only one in Finland combining expertise in stem cells and biomaterials with tissue bank operations.

Regea operates on non-profit basis, and mainly on public money and grants. Suuronen regales us with a story of a grant that started with a surprising phone call: "The caller was interested in giving Regea a gift of money, but was unsure as how to go about it, " she says. Referring to Regea's status as part of the University, Suuronen explained they could only accept grants, to which the caller courteously inquired, "What would you consider to be the appropriate amount? "

Suuronen had no idea what amount they would be talking about, so she solved the problem by elucidating that grants were usually allocated for research toward a doctorate and that a month's salary of a Ph.D. student was usually such and such. "By this time the caller had made up his mind, and said, 'Well, I will pay for one dissertation', and donated 144 000 euros for this purpose."

Tekes, the Finnish funding agency for Technology and Innovation, has also seen it fit to fund tissue engineering and especially the related stem cell research. To this end, it last year committed a total of three million euros to be spent over the next four years to finance this work.

This is a "considerable amount", even by Tekes' standards, according to **Teppo Tuomikoski**, Senior Technology Advisor at Tekes. But like he says in the Tekes press release of December 2005: "Stem cell research offers promising opportunities to develop new forms of cellular therapy for diseases formerly difficult to treat. Because of their important potential applications, stem cells have become a key focus of interest in biotechnology."

## Data integration and visualization system

# Enabling con

The omics revolution combined with advances in information technology has empowered us with new tools to study the biological systems by measuring a large number of molecular components in parallel, therefore enabling the systems approach. The wealth of new information, combined with existing repositories of knowledge dispersed across numerous databases and literature, demand new solutions for management and integration of life science data.

Matej Orešič, VTT Technical Research Centre of Finlan, matej.oresic@vtt.fi

**"Continuous frustration of researchers is the fact that it is not easy to map identical entities across multiple databases due to different naming conventions."**

Knowledge management and data integration are recognized bottlenecks in healthcare and drug discovery domains and current solutions are not yet capable of taking the full advantage of the information delivered by the modern omics technologies. More fundamentally, the ability to collect molecular information from biological systems in parallel is also challenging the ways we represent the biological systems and related knowledge, as well as the ways we design experiments to address specific biological questions.

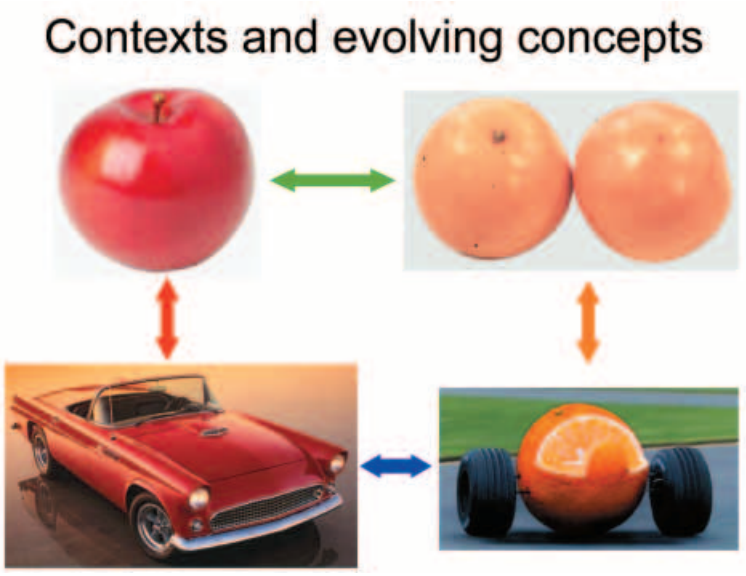
### Life science data integration and visualization solutions

Several approaches for biological data integration have been developed. In parallel, progress has been made to organize biological knowledge in a conceptual way by developing ontologies and domain-specific vocabularies. The emergence of XML and Semantic Web technologies has fostered the ontology-based approach to life science data integration. In this context, data integration comprises problems like homogenizing the data model with schema integration, combining multiple database queries and answers, transforming and integrating the latter to construct knowledge based on underlying knowledge representation.

However, the ontology-based approach alone cannot resolve the practical problem of evolving concepts in biology, and its best promise lies in specialized domains and environments where concepts and vocabularies can be well controlled. Neither can the ontologies alone resolve the problem of context, i.e., what may appear closely related in one context, may be further apart or unrelated in another.

Biological systems are characterized by the complexity of interactions of their internal parts and also with the external environment; integrating such information may result in a huge and heterogeneous network of biological entities. The visualization of such complex multi-level information is not only an issue of esthetics; it has also an impact on decision-making in sciences as well as in biomedical domain.

In order to address the growing need for visualization solutions in biomedical domain capable of handling and combining diverse sources of information, we established a project VISUBIOMED (supported by VTT Service Beyond theme). The VISUBIOMED approach combines multiple levels of biomedical information, from images, to molecular profile data, via a complex network of structured and unstructured data as well as statistical models.




Illustrative example of how similarities between the entities may depend on the context. The concept of redness links the car and the apple, while the roundness connects apple and orange. New concepts may emerge by combining the existing ones, such as the "orange-car". Due to rapid advances in life sciences, the knowledge management system in the domain has to be flexible enough to deal with evolving concepts.

# ceptual biology

### Handling structured data

Much of biological information, in particular at biomolecular level, is stored in databases such as genome databases (*e.g.*, GenBank), pathway databases (*e.g.*, KEGG), protein databases (*e.g.*, UniProt) or small ontologies like Gene Ontology (GO).

Even at this level, a continuous frustration of researchers is the fact that it is not easy to map identical entities across multiple databases due to different naming conventions. Therefore, any attempt at data integration should start with identifying the "atoms of information" and creating solutions to resolve the names. XML and RDF are useful technologies for creating such identity-mappings across multiple data sources.

There are ongoing efforts to standardize the life science data formats in order to facilitate the exchange of information and knowledge (*e.g.*, W3C Semantic Web Life Science). For every database (either containing annotations or information about entity relationships) we create a simple XML schema that enables mapping to other databases.

### Handling textual data and concepts

Most of life science information is still available only in textual form. This is particularly true for information on relationships between the molecular level events and more complex concepts such as events related to multifactorial diseases.

Text mining solutions are therefore needed to sift through the semi-structured disease-related data such as OMIM, literature (*e.g.*, PubMed) and patent databases. Established vocabularies and concepts such as Unified Medical Language System (UMLS) may be of help, but one should be aware that with rapid progression of life sciences and medicine today, new terms and concepts are rapidly emerging.

As part of the TRANSCENDO project (*In silico* models of disease pathogenesis and therapy) supported by the Tekes Modelling and Simulation Program (MASI), we have been developing an Ontology Aided Text miner system (OAT) for extraction of information on relationships between different biological entities or concepts.

We extract (subject, predicate, object) triplets from the raw text mass. We use a seed vocabulary of terms—at least two terms of the triplet must be found in the vocabulary—and the vocabulary is then augmented with the third term, if needed. The extraction is based on shallow parsing: instead of forming complete parse trees from each sentence, we rather extract noun phrase and verb phrase blocks that are further processed to produce the triplets.

In the end, the combination of all these triplets will form a conceptual graph that tells about the relations between the entities of interest.

### Everything is connected!

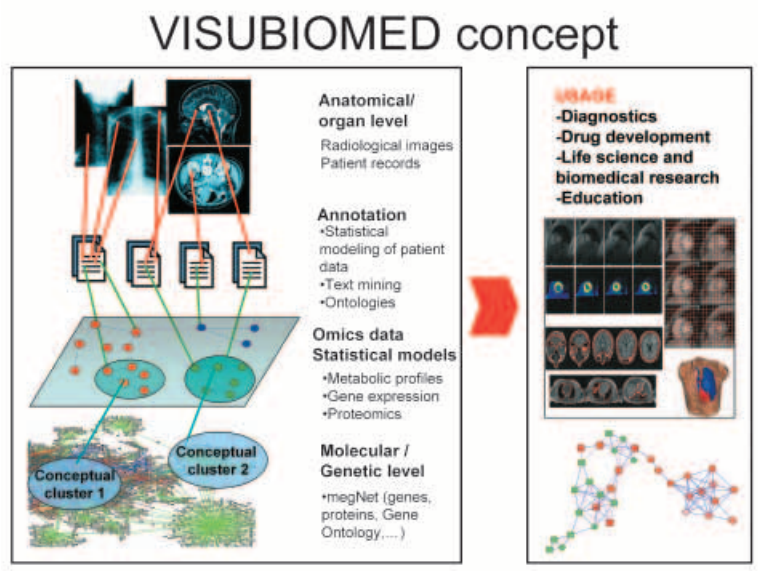
We have already integrated data across multiple life science databases and repositories and developed a three-tier database system that affords parallel retrieval across multiple databases, including metabolic pathways, protein-protein interactions, signaling and regulatory networks.

The results are then visually displayed as a network using the megNet© software environment developed as part of the VISUBIOMED project.

Edge attributes contain information about type of relationship, possibly quantitative or semantic information (such as "is located in" in case of linking a protein with a complex entity such as cell organelle, with information obtained by text mining).

The entities (nodes), which can include molecular entities as well as more complex concepts such as insulin resistance or diabetes, are connected via relationships (edges), which can either

**"We have already integrated data across multiple life science databases and repositories and developed a three-tier database system that affords parallel retrieval across multiple databases, including metabolic pathways, protein-protein interactions, signaling and regulatory networks."**




VISUBIOMED project, part of VTT Service Beyond theme, combines expertise of two VTT teams: Quantitative biology and bioinformatics led by Matej Orešič(project coordinator) and Signal and image processing led by **Jyrki Lötjönen**. The objective of the project is to develop a visualization tool utilizing state-of-the-art 3D techniques, modern mathematical modeling techniques, contextualization etc. for diagnostics, drug target discovery, drug development and education by linking information from macro to nano scales. Project web site: http://sysbio.vtt.fi/visubiomed/

be direct physical interactions, or more complex relationships. Nodes and edges may be clustered and mapped to an ontology-type structure.

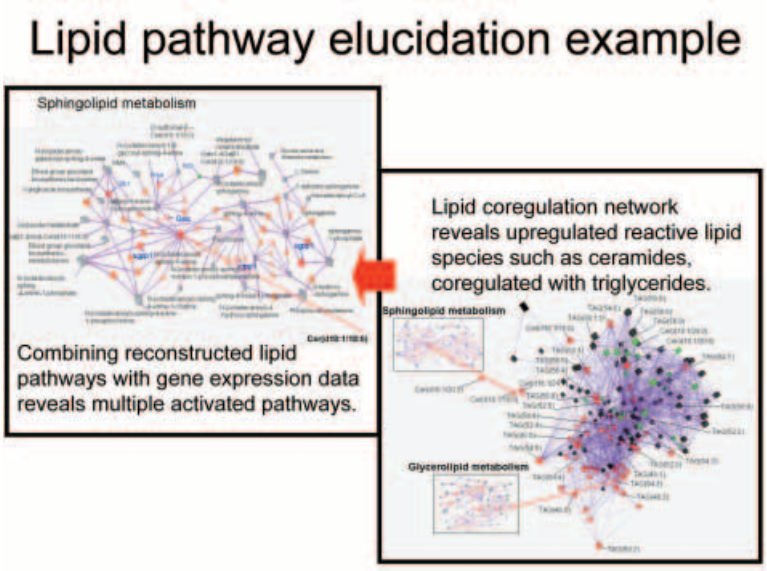
Specifically, we are also interested in retrieving quantitative information on relationships that can be used for predictive modeling in the future.

### Real life application

As part of the European Integrated Project HEPADIP (Hepatic and Adipose Tissue in the Metabolic Syndrome), we have studied the pathways related to insulin resistance in animal models of diabetes, obesity, and metabolic syndrome.

The megNet© software environment enabled us to study the coregulation networks of liver metabolic profiles from obese and insulin resistant mice, elucidating surprising relationships between the triglycerides contributing to fatty liver and the potentially toxic ceramide molecular species.

The database system then enabled us to reconstruct the liver lipid pathways that in combination with gene expression and metabolomics data enabled us to identify the upregulated pathways leading to synthesis of ceramides in livers of obese mice.



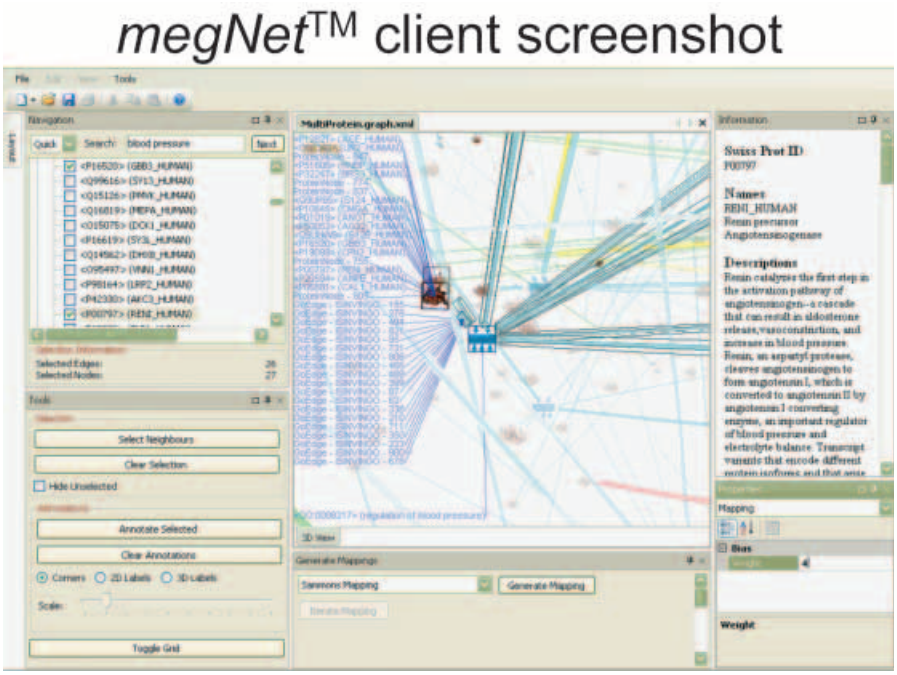

Elucidation of liver lipid pathways in insulin resistant obese ob/obn mice. Correlation networks help identify the clusters of coregulated lipid molecular species. The pathways for specific molecular species are then instantiated in context of available gene expression and metabolomics data.

This work is part of collaboration with group of Antonio Vidal-Puig from University of Cambridge and supported by HEPADIP project (http://www.hepadip.org).

### Concluding remark

While the megNet© system is currently in prototype stage, our data integration and visualization strategy already showed that it can facilitate discovery of novel or unexpected relationships, formulation of new hypotheses, design of experiments, data annotation, interpretation of new experimental data, and construction and validation of new network-based models of biological systems.

Docent Matej Orešič, is group leader at VTT Technical Research Centre of Finland. More information on his group's work can be found at: http://sysbio.vtt.fi




The megNet© client screenshot. User can select databases to search and combine, as well as generate queries that combined structured (i.e., identifier) and unstructured text. The importance of different relationships can be weighted in context-dependent manner and nonlinear mappings are utilized to cluster the nodes. The pathways shown were retrieved in context of regulation of blood pressure.

## TRANSCENDO project
**(Tekes MASI Program)**

The overall objective of the project is to build a comprehensive modelling and biosimulation framework for studying human pathophysiology in silico, with the initial focus area of diabetes.

The consortium assembles expertise in domains of computational systems biology (Matej Orešič, VTT Technical Research Centre of Finland, project coordinator), semantic modelling (team led by **Antti Pesonen** at VTT), probabilistic modelling (**Samuel Kaski**, Helsinki University of Technology), and pharmacology and animal models of metabolic disease (**Eriika Savontaus**, University of Turku).

The modelling goals of the project are:

- couple different modelling levels (cellular pathway-level knowledge, animal physiology, human physiology) together,
- predict the disease, its properties, and responses to interventions and
- explain the experimental findings such as clinical data.

Project web site:
http://sysbio.vtt.fi/transcendo/

# opinion

## The new supercomputer
## An opportunity to do something really new or just more of the same?

Jan Åström

Finland is investing a huge sum of money to update its computing capacity. Similar investments take place all over the world, but will this give the scientists a chance to do something really new and uncharted or is it just more capacity to do the same research as done before.

Finland has recently decided to invest 10 million euros in new computational hardware. The procurement is well on its way even though there are still many technical, legal and administrative problems to be solved and lots of negotiations to be carried out.

The new hardware will probably consist of an enlargement of the pc-cluster and a new high-performance supercomputer. Both are hopefully up and running early next year.

This kind of process is not unique to Finland but similar procurements take place all over the world. With such a perspective in mind the obvious question to be asked is: What kind of impact on science comes with a new supercomputer? Is there a real leap forward or do the scientist just do more of the same?

To put the question in other words: are the scientists eagerly waiting for the new supercomputer to attack the most relevant computational problems or are they already busy solving them using available computers?

There is no easy answer to this question. On one hand, there are several rapidly advancing fields of science for which computational peak performance is a clear bottleneck. Such fields include, *e.g.,* climate modeling, nanotechnology, complex fluids, soft matter, and fusion plasmas. On the other hand, even within these fields a lot of relevant research can be done on rather modest computers.

Taking an example from my own experience: the numerical simulation of clusters of suspended solid particles in a flowing fluid is a very challenging problem. The quality of such simulations has improved considerably with high-performance computers, but equally important has proved to be the development of algorithms and performance optimisation of the codes, and perhaps most important for the true understanding of the problem has been old-fashioned pen-and-paper mathematics.

It is quite clear that a modern supercomputer opens up new possibilities for computational research, but it is not decisive for the quality of the outcome of the research. Even computational science relies heavily on the scientific intuition, enthusiasm and hard work of the researchers. A new supercomputer is nothing but a tool and a tool is only as good as the hands (or maybe brains in this case) of its users.

With new and improved computer capacity it is tempting to just increase the amount and size of earlier simulations without ever stopping to ask if it is really scientifically motivated. Such a way of acting leads to scientific results of doubtful quality. The new supercomputer will soon be here. Now is the time to ask: what should we really do with it?

Jan Åström started to burn CSC's CPU-hours in 1988 as a student at Åbo Akademi and worked for 10 years as a group-leader in computational physics at the University of Jyväskylä. In 2003 he started at CSC as an in-house physicist. He works in close contact with CPU heavy-users and specializes in code optimization. His email address is jan.astrom@csc.fi.



**"With new and improved computer capacity it is tempting to just increase the amount and size of earlier simulations without ever stopping to ask if it is really scientifically motivated."**

# NIC.

NIC.FUNET.FI is the grand old dame of FTP servers. She was launched in 1990 and has recently gone through its fourth upgrade. One can still see in the FTP archive the legacy of the early Internet even though it has grown to meet the demands of today. Harri Salminen who has seen the growth of NIC.FUNET.FI from the beginning and researched the history of Funet tells here its story.

Harri Salminen, CSC
harri.salminen@csc.fi

**"Large and complex file collections are slow and cumbersome to download one by one through clicking web pages full of advertisements designed for users who only need to download one file at a time."**

## The history of NIC

FTP is one of the oldest Internet protocols developed in the early 1970s in the ARPANET, Internet's precursor. The RFC 454 document from year 1973 states that one of the primary objectives is to "to promote sharing of files (computer programs and/or data)".

ARPANET's Network Information Center furthered that goal by setting up an FTP server called SRI-NIC.ARPA from which anyone could download programs and documents by logging in as user anonymous. SRI-NICs most well known content was the Request For Comments (RFC) document series that still today is used to publish technical information and Internet Standards. Several other anonymous servers like Simtel were also set up in the early Internet to offer collections of freely distributable software.

In the 1980s, the Internet grows to connect the global academic and research community, and open-source software used and developed in many universities becomes popular content in many of the Internet's FTP archives, including a small one set up in a multipurpose funet.fi server in the late 1980s.

In the year 1990, when ARPANET was taken out of service and the Nordic countries shared a single 64Kbit/s US link, Funet launched a dedicated NIC.FUNET.FI server also called FTP.FUNET.FI, which not only distributed RFCs to Finnish Internet engineers but also provided a place where Finnish volunteers could collect all kinds of freely distributable files that they thought might interest other Finnish users as well, thus saving the precious international bandwidth.

It quickly became so popular that we were a net exporter of data in the early NORDUnet, unlike most other countries outside the US. We had to even develop our own FTP server to apply some speed limits for foreigners to protect the slow links of the early 1990s.

Most of the content maintainers were either CSC's staff or computer enthusiasts from Funet member institutions, thus the collection focused mostly on open-source or shareware programs but not totally forgetting arts and culture either.

To organize the collection, we created a /pub tree hierarchy that reflected the various interest areas of the different volunteers and created a contact address for each of them. As a shortcut to the collections in NIC.FUNET.FI's /pub tree we took in use a manually maintained /index directory. Most of the mirrors we placed under the /pub/mirrors/ directory with their original site names, which has now grown up to utilize about 95% of disk space of the archive.

Many of the files are still archived in their original directories created in the early 1990s even though some of them are no longer actively maintained. This is because we don't want to unnecessarily break references to them since there's currently no reliable way to inform all FTP users of the new location—since many browsers don't even show the messages an FTP server might be showing to them. Thus the archive is a reliable and stable repository to which new web interfaces and search engines can provide long working URL references for users.

## The launch of Linux

In 1991, a certain Finnish student offered a small software he called Freax to **Ari Lemmke**, who was at the time maintaining our /pub/OS area. He though that Freax was not the best of names and decided to create a subdirectory Linux based on the student's first name. The student, **Linus Torvalds**, also liked the name, and thus on 17th of September 1991 he posted a message to Usenet announcing that Linux version 0.0.1 is available from FTP.FUNET.FI.

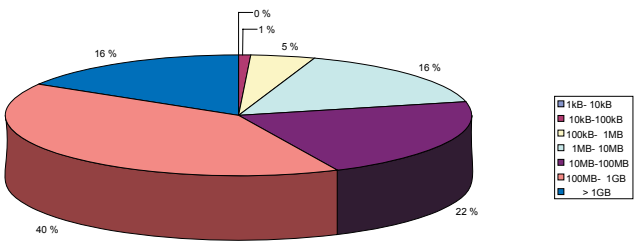
This single directory has since then grown to thousands of directories containing over four hundred thousand files that form the Linux kernel archive which doesn't include the millions of files included in different Linux distributions. Linus now maintains the Linux kernel archive in US where he now lives but it's still of course available via NIC as well.

After this small start Linux kernel was combined with GNU software tools and other open-source software to form Linux distributions that have become one of the most popular items at NIC. These distributions contain a large number of files needed for installing and upgrading a fully functioning Linux system on top of which all kinds of other software can then be installed or developed.

## Content maintenance

Many of the volunteers that have helped

File size distribution of the bytes transferred with ftp in 2005



# .FUNET.FI

NIC grow in the early days have left, but some of them still remain.

There's no longer a need to try collecting copies of all kinds of small files and documents because the fast lines provide good connectivity to millions of web sites. This has still not made large FTP archives like NIC unnecessary, though. They still have millions of active users for several reasons.

Currently a major service provided to the Finnish users is automatic mirroring of other anonymous FTP archives so that the users don't need to download the files from servers that may be far away or very loaded. Actually, the whole existence of open-source or other freely available software relies on the distributed infrastructure of mirrored archives. For example, all public OpenOffice.org download sites are mirrors to protect the master site from overload.

Without public FTP archives and non-commercial mirror cooperation, especially the large and popular software collections, like Linux distributions, wouldn't in practice be available to the users or even developers via the Internet. Actually, even the whole Internet or WWW wouldn't probably exist without a working open-source software distribution infrastructure built on volunteer effort, mutual mirroring agreements and open for all access.

### Focusing in on the future

A query sent to the Finnish users of NIC.FUNET.FI in 2003 resulted in 1,541 answers. Most of these expressed interest in open-source software and their updates. Many of the respondents said they valued a stable address were they could download their files quickly and reliably.

The survey proved that there's still a clear need for NIC.FUNET.FI's services even at the time of fast networks and dynamic web sites. Large and complex file collections are slow and cumbersome to download one by one through clicking web pages full of advertisements designed for users who only need to download one file at a time. For example, the popular Linux distributions Fedora, Ubuntu, Suse and Debian, which we mirror, consist of over million different files in total.

According to our statistics, 95% of files downloaded from NIC are over one megabyte in size and two thirds of the transferred bytes belong to CD images of various Linux distributions.

The old system distributed in 2003 over 150,000 CD images and the new system could in theory distribute even several millions when fully loaded, thus resulting in saving millions of euros to the users compared to buying the CD-ROMs via mail order, for example.

When a new CD or DVD image is released, there's often a load peak in all servers when thousands of users try to download the same large files. Especially to spread the load during these times we are currently investigating the various P2P technologies as an alternative access method to some of the large and popular files in our archive. For finding the right files among the millions in our archive we are currently considering different search and documentation techniques that we could use to enhance the web interface to our archive.

We are also planning to refocus major new resources and services so that they would better serve our constitution, primarily the Funet users and secondarily the Finnish society and open-source community as a whole. One of most important focus areas are of course the needs of the academic computing and networking community which means open source software as well as the RFCs.

In addition to mirroring, we also support the file distribution needs of local Finnish open-source, arts, science and higher education communities. For some of the old legacy content, only minor changes will be implemented. We will also try to preserve some of the historically important files and documents for future historians.

For more information

ftp://ftp.funet.fi/
http://www.nic.funet.fi/
http://www.nic.funet.fi/index/Funet/history/internet/

**"For example, the popular Linux distributions Fedora, Ubuntu, Suse and Debian, which we mirror, consist of over million different files in total."**

## NIC hardware

The original NIC.FUNET.FIi was a SUN Sparc-Server 630MP system with only 128MB of memory, few gigabytes of disks and a 10 Mbit/s shared Ethernet connection.

NIC has been upgraded three times. The latest upgrade was completed in spring 2006, when we deployed a new Fujitsu-Siemens PRIMEPOWER 450 server with four SPARC64 V processors, 16GB of memory and a fault tolerant EMC FibreCat CX300 fibre channel RAID storage subsystem initially configured with 15 X 300GB FC disks to give about 3TB of space for several millions of public files.

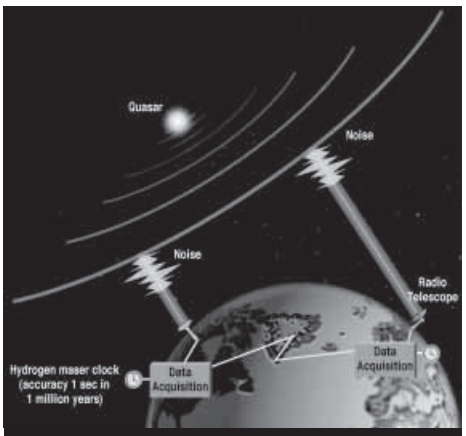
The operating system is the Solaris 10 Unix from Sun Microsystems that is configured to support a virtual read only zone for the public server processes. The server has a 1Gbit/ps dedicated connection to Funet. For auxiliary and support tasks we also utilize several other Solaris and Linux servers, tape archives and background networks.

©Metsähovi Radio Observatory TKK

# The first 10 Gbps customer link of Funet installed

Metsähovi Radio Observatory in Kirkkonummi, Southern Finland, is the first Finnish research institute to have a network connection whose bandwidth exceeds 1 Gbps. It is also the first radio observatory in the world connected at 10 Gbps. Data produced by the radio telescope is transmitted for processing to research centers in Europe and the USA.

Jouko Ritakari
Ari Mujunen
Metsähovi Radio Observatory, Helsinki University of Technology TKK
jouko.ritakari@tkk.fi
ari.mujunen@tkk.fi




In Very Long Baseline Interferometry (VLBI), the data received at each radio telescope is paired with timing information, usually from a local atomic clock, and then correlated with data from other telescopes similarly recorded, to produce the resulting image. In VLBI, the data is generated at locations geographically as far apart from each other as possible, since by the theory, the longer the distance, the larger the synthesized radio telescope, resulting in more accurate images of astronomical objects. International VLBI Service (IVS), NASA

The first 10 Gbps customer link of Funet was installed in July. Perhaps surprisingly, the new connection was not to one of the largest universities or to the telecom research labs, but to the Metsähovi Radio Observatory of Helsinki University of Technology TKK that is located in the woods of Kirkkonummi some fifty kilometers away from Helsinki.

### New possibilities for radio astronomy with high-capacity networks

As the only observatory in Finland operating a radio telescope dedicated to radio astronomy, the Metsähovi Radio Observatory routinely performs a wide range of radio astronomical observations, such as long-term monitoring of active galactic nuclei, mapping and tracking of flares and active regions of the Sun, and global collaborative Very Long Baseline Interferometry (VLBI) observations.

In VLBI, radio observatories around Earth effectively create a synthetic globe-sized giant radio telescope. The simultaneous recordings of radio signals of several telescopes located geographically apart are combined and correlated. For successful correlation processing, a true digitized representation of the wideband radio signal is needed—and the wider the bandwidth, the fainter and more distant objects can be mapped and imaged.

Routine data rates in today's VLBI observations lie between 256 and 1,024 Mbps, with the trials already on their way at 2 and 4 Gbps. "The unsatiable need for bandwidth makes VLBI one of the most interesting and demanding precursory multi-Gbps Internet applications", says laboratory manager **Ari Mujunen**.

### Nuclear scientists and radio astronomers need more bandwidth

Today, nuclear scientists and radio astronomers are among the largest users of Internet bandwidth. These experiments produce enormous amounts of data that must be transferred to processing centres that typically are located in central Europe or in the USA. The Joint Institute for VLBI in Europe (JIVE) in the Netherlands is the leading astronomy VLBI correlation center in Europe and is a regular partner with the Metsähovi Radio Observatory.

In VLBI, the data is generated at locations geographically as far apart from each other as possible, since by the theory, the longer the distance, the larger the synthesized radio telescope, resulting in more accurate images of astronomical objects.

"VLBI can easily outperform optical space telescopes like the Hubble by a factor of several thousands," notes the director of the Metsähovi Radio Observatory, Dr. **Merja Tornikoski**. However, to get the images out of the raw data, the signals of every station must be compared with each other.

Traditionally, to bring together several days worth of Gbps-class observation data (that is, tens of terabytes) from distant telescopes, physical courier shipments of tape or disk media have been used. This is rapidly changing, as network VLBI data transfer, dubbed "eVLBI", is increasingly replacing courier shipments.

Not only replacing tape and disk shipments, eVLBI enables new forms of VLBI research and new ways of correlation processing. Rapid response "target-of-opportunity" observations become possible, as well as distributed software processing of the recorded digital radio signals. Dedicated hardware-based correlator processors will be replaced by distributed Grid-based software correlators, and the Metsähovi Radio Observatory wants to take part in these developments.

### Dark fiber made this possible

Senior researcher scientist **Jouko Ritakari** explains, "We have longed for this connection for several years, but until now the cost was too high for a small laboratory that only has 20 employees. The telecom companies wanted to rent only connections that had their own active equipment."

The situation changed at the end of last year. Metsähovi Radio Obser-

**"Traditionally, to bring together several days worth of Gbps-class observation data (that is, tens of terabytes) from distant telescopes, physical courier shipments of tape or disk media have been used."**


vatory received funding from Helsinki University of Technology TKK to rent a dark fiber connection (fiber only, the customer provides the equipment) to the Funet hub at Otaniemi and the local telephone company was willing to provide it. Because the traffic bandwidth requirements will be greater than what all the rest of TKK typically uses, it made no sense to route the traffic through the TKK campus network.

Jouko Ritakari continues, "Originally we intended to have one or two 1 Gbps Ethernet connections. However, when we asked for tenders it became clear that we could have a 10 Gbps equipment for dark fiber at essentially the same price. We got very good co-operation from CSC, the operator of the Funet network, too. First they were willing to provide us a 1 Gbps connection although our traffic is quite heavy compared to what all the other users need. And when we saw that we could get a 10 Gbps connection, it only took a couple of email messages to **Juha Oinonen**, the head of Funet development, to upgrade that to ten times faster. Now we are in very good position to achieve new internationally interesting results in the field of eVLBI."

Although several radio observatories already have 1 Gbps connections and a few 2.5 Gbps links exist, the Metsähovi Radio Observatory was in fact the first radio observatory in the world to be connected at 10 Gbps.

### The demands of scientist lead to commercial applications

Installation of new 10 Gbps equipment went surprisingly smoothly, the equipment arrived at noon on 29.06.2006 and it was working four hours later. One of the first VLBI experiments transferred to JIVE for correlation was an experimental ESA Smart-1 lunar probe spacecraft position VLBI tracking observation. eVLBI transfers of joint observations of an Italian, a Dutch, and a Finnish radio telescope enabled quick and successful verification of a new software-based correlation processing technique developed at JIVE.

Of course going has not always been as smooth, "When we started using the Internet", notes Jouko Ritakari, "we very quickly noticed that the usual Internet protocols were not very well suited for long high-speed connections. We needed 512 Mbps to 1 Gbps speeds, we only got 10 Mbps to the Netherlands and 2 Mbps to Japan."

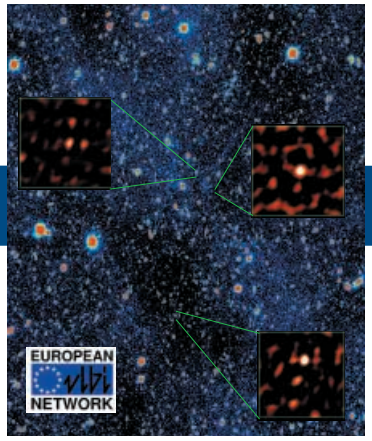
So the radio astronomers went on to work to develop new reliable UDP-based Internet protocols. The results were impressive in the sense that already in the first tests they achieved 640 Mbps speed from Finland to the Netherlands and 400 Mbps between the Netherlands and Japan. The first test quadrupled the traffic from the Nordic countries to central Europe, and the second one doubled the traffic in the US Internet2. Later, another protocol version was developed that is able to transfer realtime data through normal Internet at speeds approaching 1 Gbps.

"All in all, this is a nice example how pure science can lead to commercially important applications. For example, the work on realtime Internet protocols will come very handy when HDTV quality television broadcasts will be started in Internet", concludes Jouko Ritakari.

**"We very quickly noticed that the usual Internet protocols were not very well suited for long high-speed connections."**

The writers work at the Metsähovi Radio Observatory and have been involved in VLBI technology development since the early 1990s.

## An example of VLBI results




The most detailed images ever made of faint, distant radio galaxies, located billions of light years from Earth, reveal that many of them harbour central massive black holes, adding further support to the belief that super-massive black holes are inextricably linked with the way galaxies formed in the early universe. Because the radio images are three times sharper than those from the Hubble Space Telescope, they give fresh insight into what is happening in the centre of some of these galaxies. The images, generated by the European VLBI Network (EVN) are particularly valuable because they penetrate the dust that often blocks the view of even the most powerful optical telescopes.

The pioneering observations were conducted by an international team of radio astronomers from Europe and the USA. The radio signals were received by the giant 100-m telescope in Effelsberg, Germany; the 76-m Lovell Telescope in the UK, the 70-m NASA/DSN antenna near Madrid in Spain, and six other large radio telescopes located across Europe. Data at each of the radio telescopes were archived on high-speed magnetic tape recorders, generating almost 25,000 gigabytes of data in total. By means of a special, purpose-built supercomputer (operated by the National Radio Astronomy Observatory in Socorro, USA), the magnetic tapes were later played back and combined together to form a super-sensitive, giant radio telescope of continental dimensions.

Courtesy of Michael Garrett, JIVE, and the EVN

# comment

## Data-intensive science as a growing user for research networks

Juha Oinonen

Research community has for decades been provided with networking infrastructure, already in the time before Internet. Today, international networks like Géant2 and Nordunet in Europe connect the national networks, like Funet in Finland, to each other making it possible for academic users to utilize resources globally.

In experimental science, like astronomy in the case of Metsähovi, the research is based on measurement data collected with scientific instruments, here radio telescopes. Other examples of scientific instruments include the Large Hadron Collider (LHC) in CERN and scientific satellites, just to mention a few. Today, all these instruments generate digital information, which can be stored, processed and transmitted using computers, data storage devices, and telecommunication networks. Phrases like *data intensive science* or *data intensive research* are used for this kind of activity, and the word *Grid* is used for the technologies involved.

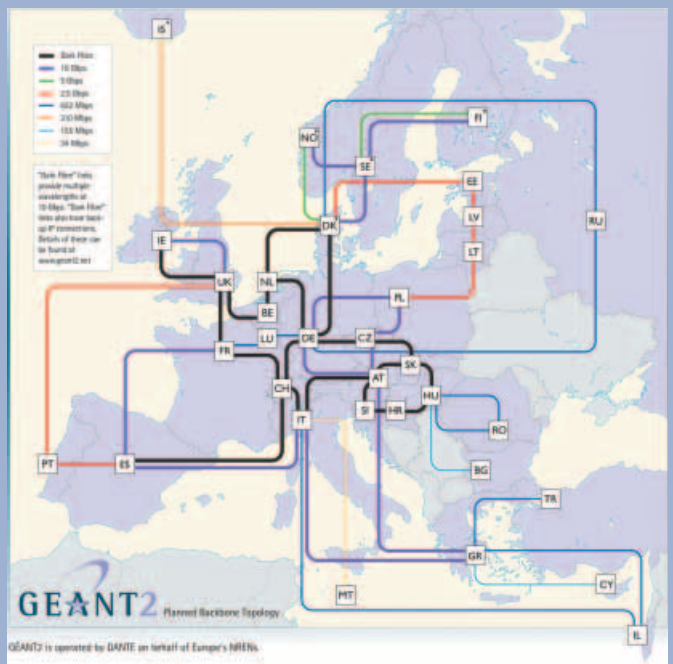
More and more research groups and disciplines are seeing a way to benefit from Grid technology to extend their research to questions that used to be out of reach, or to get results faster and more reliably. This means the research network infrastructure must be kept at an adequate level in order to guarantee it will be able to serve the increasing number of potentially high-capacity and high-reliability demanding applications and users.

Europe, partly because we are the home continent of CERN, has responded to this need by building the Géant2 backbone network, which will provide up to 10 gigabits-per-second (Gbps) connections for research users. National research networks will also need to be in place and up-to-date to provide high-capacity connectivity between the nearest Géant2 point-of-presence and the research institute involved. Finally, also the institutes' local networks must have similar capabilities. All this requires collaboration, coordination, and funding. And as networks cannot be built overnight, some sort of crystal ball would be nice so that the capabilities could be designed also for the relevant locations of future users and applications.

Due to the geographical and international nature of networks, no single party can set up its own network; neither can a network be operated from a single point in the same manner as, say, a supercomputer. Physical conditions, legislation, market situation etc. vary between countries and even between different areas inside a country. In order to optimize the financial and mental investment in the network, the same infrastructure must be shared by different applications. With modern, commercially available and widely used fiber optic based technologies this can be done without compromising network performance or security.

The research networking community has decades of experience and understanding on how to manage the operational, technological, and administrative challenges of international network infrastructure. Close collaboration with the telecom industry and standardization bodies makes it possible to provide services for the increasing needs of research. All this requires that correct political decisions are made in all necessary levels and in time.

Juha Oinonen has worked with Funet at CSC for almost ten years, and has lived through the many technological and political changes of research networks. He can be reached via email juha.oinonen@csc.fi.



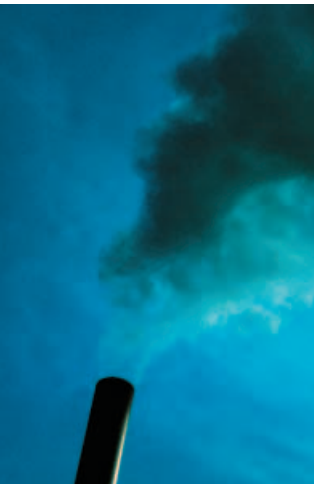

© Leena Jukka, CSC

**"Due to the geographical and international nature of networks, no single party can set up its own network; neither can a network be operated from a single point in the same manner as, say, a supercomputer."**

# Computing turbulence: law

Flow and turbulence are common phenomena we encounter daily. Whether you go to work on foot, riding a bike, or driving a car, a sudden whirl may snatch your cap, make the going heavy or splatter the windshield of your car. In spite of its daily presence, turbulence is a difficult phenomenon to understand and model, even though the equations describing fluid dynamics have been known for almost two hundred years.

Petri Majander, Helsinki University of Technology, petri.majander@tkk.fi



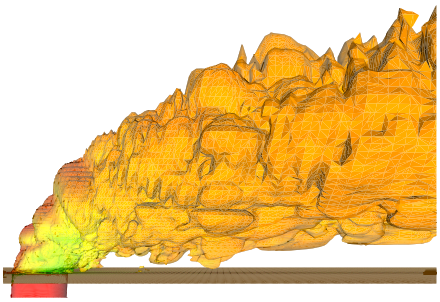
Nobel Prize winner **Richard Feynman** has said that the quantum-mechanical Schrödinger equation is sufficient to describe life as a whole. In a sense, this may be true but it is difficult to imagine that we could do without the approach of a biologist or psychologist to explain genes, cellular structures, or human behavior during Christmas shopping.

With turbulence the situation is rather similar, for the Navier-Stokes equations of turbulent flow, in their simplest forms, can be written in just a couple of lines. However, analytical solving of nonlinear equations is possible only in those few cases when the flow is smooth like, for example, the slow movement of oil along a pipe. Often the flow we observe around us is turbulent: it fluctuates and twirls in a seemingly random manner, even if in principle, the movement completely follows the above mentioned law.

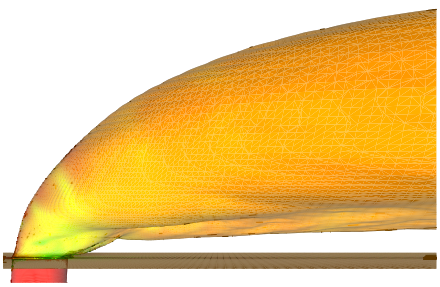
A turbulent flow consists of movements on several scales. A moving car leaves behind vortices that have a diameter of the same magnitude as the car. Close to the car body small vortices may appear with diameters within the range of one millimeter. If we want to solve the momentary dynamic flow on a car, we must take into consideration both the small and the large vortices. In addition, a small vortex has a very short lifetime compared to the axial flow rate of the main wake vortex.

To use today's trend word, we are dealing with multiscale modeling, though all phenomena are described following the continuous field theory of matter. Even a small movement involves a large number of molecules described as packages passing each other in transition. Direct Numerical Simulation (DNS) is a direct solution of the Navier-Stokes Equations (NSE) and covers all scales of turbulence calculated at the time point of every small vortex. It has been estimated that at the pace of development of our current computing capacity, to solve the NSEs for the flow over an airplane wing will be possible in 2080. During the third millennium, classic natural sciences are going to provide tangible challenges that exceed the nanometer scale.









The upper visual image depicts a jet in a cross flow. Turbulent pipe flow bursts out from the plate hole and bumps vertically into another flow. On the surface, substance from the pipe has mixed 5% with the mainstream flow. The surface and the plate colors are depicted with pressure; at the point of collision the pressure is higher (depicted as red). The lower visual image shows a time-averaged flow of the same surface at approximately 40,000 time steps.

### If there's a will, there's a way—engineers solving problems of turbulence

For decades problems such as an air flow over an airplane wing or a pipe flow have been solved with sufficient accuracy. A model is used that makes it possible to accomplish the computation within a reasonable time. In turbulent flow, transfer properties are increased. For example, in a pipe flow a parcel of fluid near a wall moves slower than the one in the core flow due to the friction exerted by the wall. In a turbulent flow the friction effect of the wall is increased when the slow near-wall fluid moves into the fast core flow and vice versa, as the fast core flow approaches the wall. In a laminar flow the parcels of slow and fast fluid do not collide or mix that much, but rather follow smooth streamlines side by side.

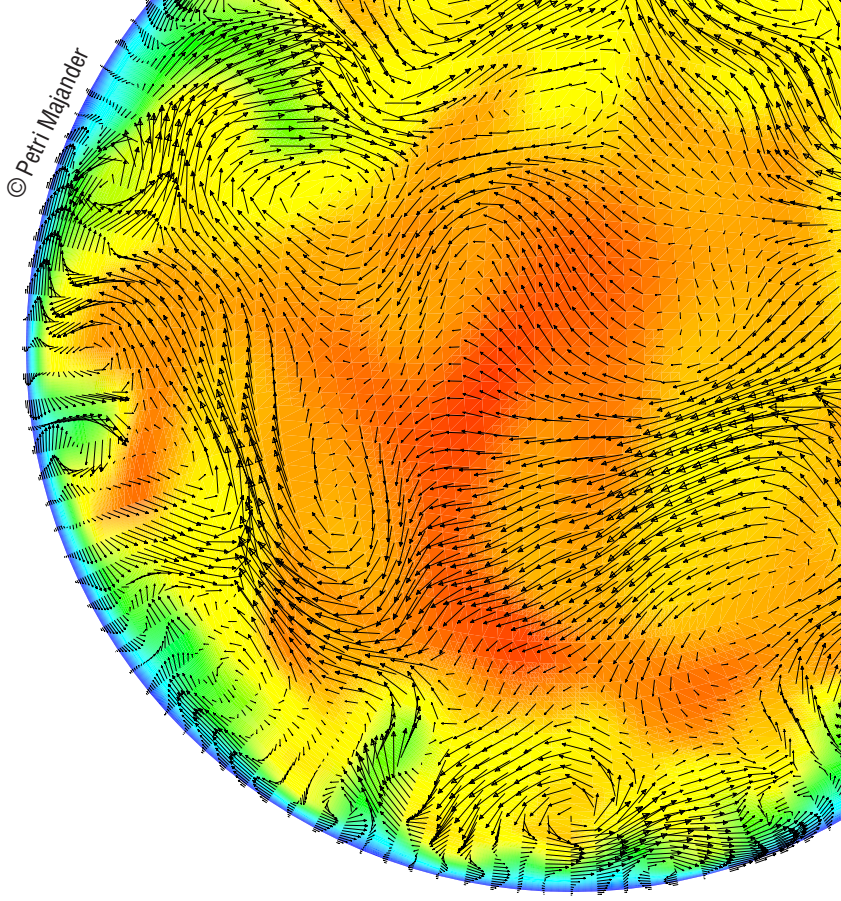
The most commonly used way to take turbulence into consideration is to average all time-dependent transition as a solution of steady-state Reynolds Averaged Navier-Stokes (RANS) equations. This time averaging produces stress terms that must be modeled. A simple way is to apply eddy viscosity. To describe the problem using laminar flow equations, viscosity must be increased and the fluid must be made syrupy where there are eddies. In other words, a solution is obtained for steady-state flow in which the impact of turbulence is described in full using a model.

The above ideas were presented more than a hundred years ago and they have been applied especially in the computerized era since the 1950s. Later the time-averaged RANS models have become quite complicated and, at least to some extent, case-dependent.

### Children crying out for computer game capacity

The complexity of models means that the size of computation is reasonable. Even a normal desk-top computer can solve fairly demanding problems. The turbulence model, however, may need more capacity. The model must be able

# and disorder




A momentary visual image of turbulent flow in a pipe when Reynolds number is 10,000. In the vertical section, liquid travels towards the viewer; the speed is faster (depicted as red) in the middle than at the boundary walls. The vectors depict vorticity projected at the sectional level.

to take into consideration the effect of large and small vortices on the final solution.

Let us consider two different types of situations. During a flight in normal conditions, the wing is aligned with the air flow and the flow is tightly attached to the wing leaving behind small wake vorticity. In a less desirable situation, during stalling, the wing is at a large angle to the air flow, which causes separation of the boundary layers leaving behind strong turbulence. These two flow types are strongly case-dependent, because the different positions of the wing cause different turbulence and vortices.

In the latter situation, a better result is achieved if the large vortices are solved time-dependently and by modeling the small vortices that are thought to be similar in different turbulences. This approach is called Large-Eddy Simulation, LES. Formally, the equations are locally averaged, in other words, filtered. The equations created contain a similar modeling term as the RANS equations. Normally, simple eddy viscosity models are used. Computation is considerably heavier, because it must be performed time-dependently. Additionally, computation must always be made three-dimensionally in space, because turbulent motion is hardly ever symmetrical. The computation mesh must be denser than when computing the corresponding case with the RANS model.

The turbulence of the computation is determined by the Reynolds number that can be interpreted as the ratio between inertial and viscous forces. When the Reynolds number exceeds a certain threshold value, the smooth flow becomes turbulent. When the Reynolds number is further increased, the smallest structures in the turbulent flow become smaller, while the larger ones remain the same. The result is that the computation time using Direct Numerical Simulation (DNS) is proportional to the Reynolds number cubed. In a pipe of 1 cm diameter, when the flow rate of water is 1.5 m/s, that figure is $10^4$. Filling a 10-liter bucket at this rate takes almost a minute and a half.

Calculating this pipe flow using direct numerical simulation is quite possible with current computers. If the bucket needs to be filled in less than ten seconds, the flow rate and the Reynolds number must be made ten-fold. The computation would become 1000 times greater and, as far as I know, such DNS has not been accomplished anywhere. The magnitude of computation becomes multifold, because the flow field must be calculated using a finer mesh (in more spatial points) and with a smaller time step (more frequently). In practice, the results are often presented time-averaged, as in wind tunnel measurements or when solved using turbulence models. The computation must be continued throughout the lifecycle of several large eddies in order to get a convergent result. Furthermore, it is possible to make animations of the momentary images of the flow fields and study the development of eddies.

Through Large-Eddy Simulation the smaller structures are modeled to reduce computation work. Additionally, scaling is not quite as ruthless as in direct simulation. In spite of this, Large-Eddy Simulation needs plenty of computing capacity, at the level not routinely used to solve flow problems common in industry. An individual desk-top computer has capacity enough to simulate channel flows that take place with a small Reynolds number within as little time as one day. This is a good start providing possibilities for a certain type of basic research, such as model development. For challenging computations the massive parallel computers will be needed for a long time to come for solving time-dependent simulations.

### Jet flows

At the Laboratory of Applied Thermodynamics Large-Eddy Simulations have been studied over several years. Simulations have been made with an in-house control volume solver based on a pressure correction method. The computation field can be divided into blocks that can be distributed to different processors. Communication with the neighboring blocks occurs through a Message Passing Interface (MPI) library. The majority of the computations have been performed with the IBMSC of CSC, the Finnish IT center for science.

Recently, computations have been made about jet flows in various cross-flow situations. This is the same phenomenon as smoke appearing from a chimney and bursting into the wind. We have studied a similar case, against experimental observations, using several different parameters. The flow field must be studied to find out, for exam-

**”The Navier-Stokes equations of turbulent flow, in their simplest form, can be written in just a couple of lines. Yet, the analytical solving of nonlinear equations is possible only in a few cases.”**

**“During the third millennium, classic natural sciences are going to provide tangible challenges that exceed the nanometer scale.”**

ple, how the jet is turned by the mainstream flow and what kind of wake stream is developed behind the jet. We solved the jet dispersion by solving a transport equation for its concentration of contents. The computation was shown to predict the experimentally observed phenomena quite well. Though the computation comprised 2.7 million control volumes (CVs), the differences, especially near the solid surfaces, were mainly due to insufficient resolution. The flow was calculated at more than 5,000 time steps for the mainstream to flow from the inlet to the outlet. The time averaging was done during seven flow-through times.

In the model reactor air is fed into the mainstream flow from several holes on opposite sides. This corresponds to additional oxygen being fed into a fluidized bed furnace to increase burning. The case has been modeled by calculating three opposite jet pairs using sectional boundary terms. The reactor's flow rate is slow compared with the jet velocity. In this case, the difference between the large and small structures is especially intensive. The jet diameters are only eight millimeters and the calculation field is approximately two meters. The computing mesh contains approximately four million CVs, placed densely close to the jets. At the beginning there is strong variation in time scaling, which necessitates a short time step and a long computation time.

### On we go, through hail and rain

As stated above, Large-Eddy Simulation is not often possible for problems of fluid dynamics we come across in everyday life. However, the solution would often be better if large eddies were solved accurately relating to time instead of averaged. Recently new methods have been developed in which the RANS equations and LES models are combined (hybrid RANS-LES). The idea is to solve the RANS equations within the boundary layer close to surfaces. In the boundary layer friction forces are significant and solving small vorticity is too expensive. LES can be applied to solve the equations for wake vortices further away. This makes it possible to predict both friction and time-dependent vorticity sufficiently accurately.

Ever since turbulence simulation became common practice, an increasing amount of physical phenomena have been included. For example, noise can be modeled by solving compressive flow equations by DNS to determine pressure fluctuations that travel long distances without damping. Pressure fluctuation can be used to solve the noise caused by the flow. Since noise must usually be solved away from the cause, the flow is solved only near to the source and the sound waves further away are solved from the simpler acoustic equations to reduce computation time.

Computation of turbulent flow using any method has turned out to be a challenge. CFD researchers are comforted by the knowledge that computation capacity is constantly growing. The simulations mentioned in this article would not have been possible without the parallel computation services provided by CSC. At the Helsinki University of Technology researchers work with turbulence simulation and modeling at the Laboratories of Applied Thermodynamics and Aerodynamics, with funds provided by the National Technology Agency of Finland (Tekes) and the Academy of Finland.

The writer works at the Laboratory of Applied Thermodynamics at the Helsinki University of Technology.

This article was originally published in Finnish in Tietoyhteys—CSC's customer magazine 1/2006. It was translated by Pirkko Huuskonen.



A momentary visual image of three opposite jet pairs at weak cross flows. The surface is a quantity that depicts strong turbulence, as seen next to the bursting jets.

# Petri Majander tackles

Leena Jukka

The writer is a science editor at CSC, and can be reached at leena.jukka@csc.fi

In the midst of all the talk about international collaboration and interdisciplinary research groups, **Petri Majander**, turbulence researcher, calmly reminds us that science still relies on the brainwork and intuition of an individual scientist.

Petri Majander, a research scientist at the Helsinki University of Technology recounts how he came to be involved in physics research, "The story is often heard," he says. "I liked physics and mathematics in school. When it came time to further my studies, I dared not choose Helsinki University, because I was haunted by a nightmare that I would end up a teacher at my old school."

So the Helsinki University of Technology, or TKK for short, seemed a safe alternative.

Fluid dynamics and turbulence came in through Majander's background, "having always lived close to water with hobbies like fishing, fluid motion fascinated me." Computational fluid dynamics is compute-intensive, but, says Majander, "I also like programming the algorithms to solve those equations, so computational fluid dynamics is probably the best match for me in terms of topic and method."

After finishing his MSc studies, Majander spent a period in an industrial laboratory, and then returned to TKK and started on the current research. Most of his group's funding comes from Tekes, the main public funding organization for research and development in Finland.

Referring to the discrepancy in the time spans of scientific research and current funding practices, Majander says, "Funding urges us to do research that is of interest to the industry. The risk is that long-term research is reduced into the normal activity of a commercial engineering office. At the same time we are required to produce scientific publications and degrees, which constitutes a contradiction to

# Careers in research

## turbulence with computers

the type of funding."

Lately, however, according to Majander, the industry has showed interest to invest in a larger project concentrated on fundamental phenomena. "We are hopeful to see a new kind co-operation where both parties obtain their goals."

When the talk turns to themes like international collaboration and interdisciplinary research groups, Majander reminds us that science has *always* been international and interdisciplinary. "It was so before there were nations or disciplines, as we understand them. Ideas have always filtrated from one culture, national or disciplinary, to another."

CFD is no different, "Computational fluid dynamics has been affected by engineers of different branches, meteorologists, mathematicians etc., and this process still continues."

"Cultural differences can be greater between laboratories located in the same building than those located on the opposite sides of the globe, says Majander. "For example, the traditions and problems in the departments of Mechanical Engineering and Technical Physics are probably quite different."

What we sometimes forget when talking about international and interdisciplinary groups, Majander points out, is that a lot of the research work is a very solitary activity and happens in the head of each individual. "Each research group consists of individuals, who create both its atmosphere and its success. Improved or different working models can be adopted both from the neighboring laboratory, department or from the opposite side of the world."

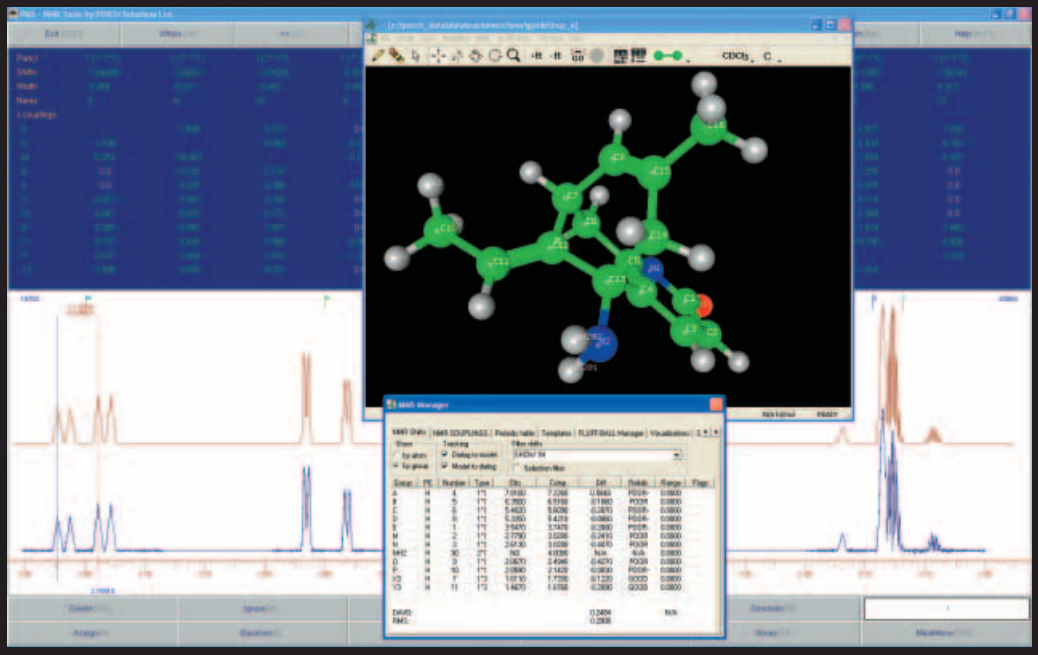
"I, like others in the CFD group, publish my work in international magazines and conferences," says Majander of the everyday international aspect of science. "Email is also a magnificent way to contact colleagues anywhere in the world but mostly I work with the other members in our group. "

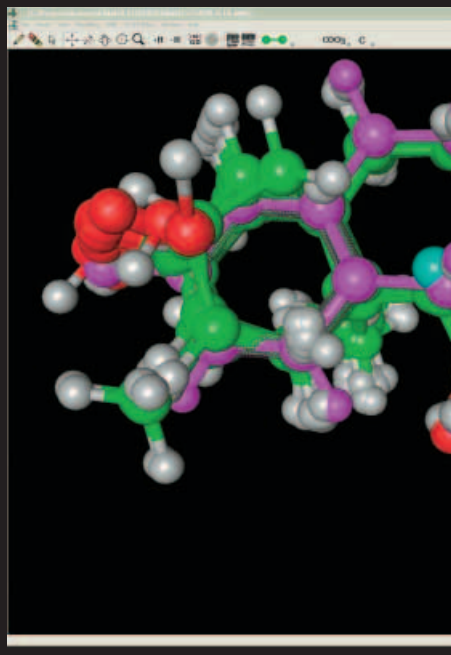
© Leena Jukka, CSC


### Petri Majander

- Born 1969 in Askola, Finland
- MSc degree 1995 from the Helsinki University of Technology, Dept. of Engineering Physics
- LicSc degree 1999 from the Helsinki University of Technology, Dept. of Mechanical Engineering
- Thesis research 'Large-Eddy Simulation of a Round Jet in a Cross-Flow' submitted to pre-examination
- Researcher at the Laboratory of Applied Thermodynamics, where his work consists mainly of Tekes-funded projects.
- Journal publications:
  1. Majander P., Siikonen T., A comparison of time integration methods in an unsteady low-Reynolds-number flow, *International Journal for Numerical Methods in Fluids,* vol 39, pp 361–390, 2002.
  2. Majander P., Siikonen T., Evaluation of Smagorinsky-based subgrid-scale models in a finite-volume computation, *International Journal for Numerical Methods in Fluids,* vol 40, pp 735–774, 2002.
  3. Majander P., Siikonen T., Large-eddy simulation of a round jet in a cross-flow, *International Journal of Heat and Fluid Flow,* vol 27, pp 402–415, 2006.

More information on the CFD group at the Laboratory of Applied Thermodynamics at TKK, where Petri Majander works can be found from http://cfdthermo.tkk.fi/.





The environment for NMR spectral analysis as combined to molecular modeling.

## Computerized chemistry at

# PERCH NMR Software,

The computerized chemistry team of the Department of Chemistry in Kuopio has developed interesting software for chemists: the PERCH NMR Software is an internationally well-known spectral analysis package with advanced molecular modeling features, FLUFF-BALL is a novel tool for ligand superposition and 3D QSAR analysis, and Ghemical an open-source molecular modeling package.

University of Kuopio, Department of Chemistry
Tommi Hassinen, hassinen@uku.fi
Samuli-Petrus Korhonen, samuli-petrus.korhonen@uku.fi
Reino Laatikainen, reino.laatikainen@uku.fi

University of Kuopio, Department of Biochemistry
Mikael Peräkylä, mikael.perakyla@uku.fi

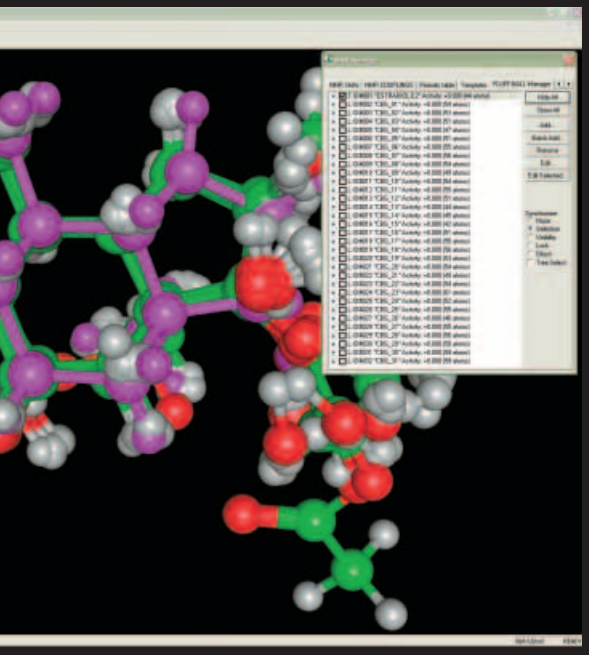
PERCH Solutions Ltd.
Matthias Niemitz, niemitz@perchsolutions.com

The tradition of computerized chemistry at the Department of Chemistry in Kuopio can be dated back to the early 80s and the days of computerized NMR spectral analysis. In early 90s, Prof. **Reino Laatikainen** and **Matthias Niemitz** started developing graphical tools for NMR spectral analysis in PC environment, nowadays known as PERCH NMR Software. The addition of a molecular modeling module, by **Tommi Hassinen**, completed the present concept of the software: a special spectral analysis and molecular modeling package for structural chemists and NMR spectroscopists.

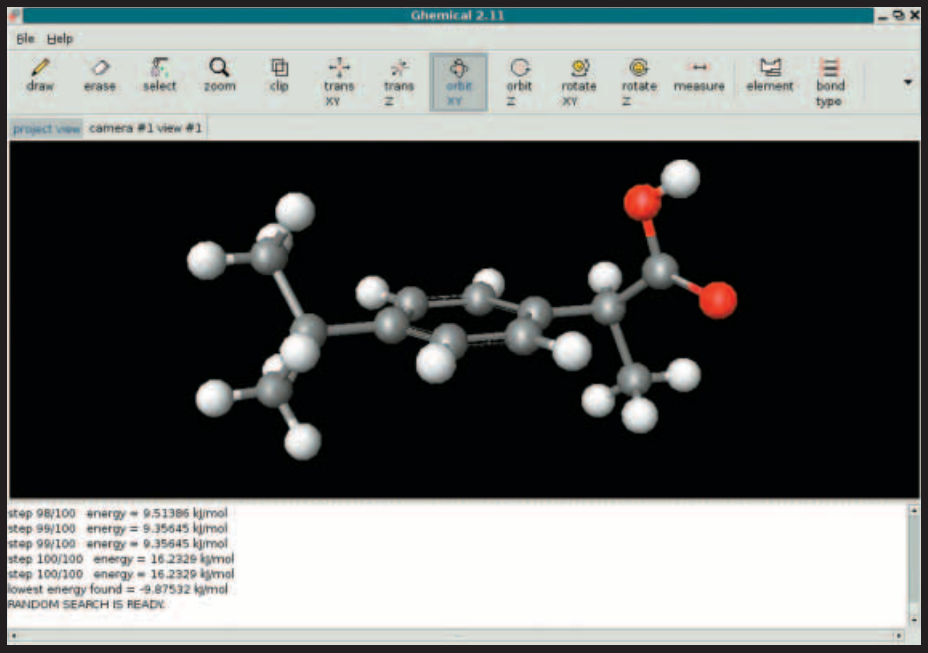
In 1998, Ph.D. **Mikael Peräkylä** brought his expertise in quantum mechanical methods and protein modeling to the team. Samuli-Petrus Korhonen took later responsibility of the modeling user interface and, when **Juha Jungman** started developing the Fuzzy Force Field (FFF), the PERCH chemistry team was in its present lineup.

### PERCH NMR software

The program development project was commercialized in 2002 when six chemists from the department founded PERCH Solutions Ltd. to handle the commercial contacts of the project. To-



The FLUFF and BALL algorithms, which are developed and implemented by Samuli-Petrus Korhonen, along with a set of necessary utilities, are incorporated into the PERCH MMS software, a molecular mechanics program running under Microsoft Windows originally developed for use with PERCH NMR.




A screenshot of Ghemical showing the most important elements of its' user interface: a menu and a toolbox, a graphics window, and a text output window.

**the Department of Chemistry, University of Kuopio**

# FLUFF-BALL and Ghemical

day dozens of laboratories around the world use the program and, as result of a co-operation with Bruker Biospin, some intrinsic features have been interfaced to the new Bruker TOPSPIN software released in summer 2006.

Since the end of 2003, PERCH NMR Software has been available to all academic institutions in Finland through CSC's scientist's interface (http://www.csc.fi/suomi/tutkija/index.phtml.en). This offers full access to the software's Professional Edition. A free teaching and evaluation version of PERCH NMR Software can be downloaded from www.perchsolutions.com. For Finnish schools there is also a free simplified modeling version available.

The PERCH NMR Software package features a wide range of applications going far beyond the software coming with the NMR spectrometer. One of the unique tools of the software, the spectral analysis based on special integral transforms, was published already in 1991. Later, the project expanded to quantitative NMR.

At this moment the emphasis is in biomedical applications and NMR metabolomics. This has activated the development of mathemathical tools like PCA, PLS and neural network methods in combination with the NMR methodology. The programs are under rigorous testing by the NMR group of the University of Kuopio.

### 3D-structure editor, molecular modeling and spectral prediction

An essential feature of PERCH is the algorithm for NMR spectral prediction based on the 4D-molecular structure. Unlike other approaches that rely on a 2D-connectivity table, PERCH's prediction is based on 3D-molecular structure and the conformational dynamics, spanning the 4th dimension. This enables not only the prediction of stereochemical effects but also provides higher accuracy of the prediction and thus allows deeper interpretation of the spectral information and molecular structure.

Structures can be imported or drawn and manipulated by a built-in 3D-structure editor, providing some unique features like the interactive tools for conformational adjustment under the "rubber-band-mode". In addition to the proton prediction, an improved prediction of coupling constants and a new $^{13}C$ chemical shift prediction were added to the ensemble in 2006.

The newest version also includes the FLUFF-BALL molecule superposition and QSAR tools, developed and

**"The name PERCH is an acronym that stands for PEak reseaRCH. It also reflects the similarity of the perch dorsal fin and NMR spectrum, and maybe the special attachment of PERCH's father to the aquatic vertebrate."**

implemented by Samuli-Petrus Korhonen, and preparation of input files for GAUSSIAN quantum mechanics program.

### FLUFF-BALL—ligand superposition and 3D QSAR technique

The Flexible Ligand Unified Force Field–Boundless Adaptive Localized Ligand, or FLUFF-BALL, is a matching pair of superposition and QSAR techniques especially designed to facilitatea rapid analysis of flexible molecule libraries with minimal user intervention.

Primary design emphasis has been to balance the computational simplicity necessary for fast screening while ensuring that the FLUFF-BALL remains easily tunable allowing the user to import any and all available a priori information. Special care has also been taken to ensure that the FLUFF-BALL takes into account the fuzzy nature of the molecular structure. It also uses a template molecule to generate a local co-ordinate system and thus avoids the problems inherent with a global co-ordinates system.

The FLUFF superposition algorithm is in essence a specialized force field based on a modified Merck Molecular Force Field (MMFF94). The superposition score is expressed as the total energy of the model, which can include repulsive terms to incorporate not-like-that type of "negative" superposition rules. The actual superimposition is usually accomplished by performing a geometry optimization using the superimposition force field. Alternatively, molecular dynamics (MD) or Monte-Carlo search (MC) can be utilized. The FLUFF superposition algorithm can be used to perform rigid, semiflexible, and fully flexible superposition.

In order to fully utilize the FLUFF flexible superposition, the BALL descriptor uses a local grid tied to the centers of the template atoms. This kind of grid can be interpreted as an extreme form of variable selection, as the BALL descriptor is extremely sparse and only a few hundred variables are needed to describe van der Waals and electrostatic fields. Also, as one moves away from the template, the descriptor terms generated become increasingly fuzzy and the exact orientation of long side-chains, whose position is usually very poorly defined, is not crucial which increases the robustness of BALL models when dealing with diverse sets.

During extensive validation runs utilizing many data sets FLUFF-BALL usually produced results comparable to the other QSAR techniques. However, for highly congeneric systems the BALL was slightly inferior to other techniques reported in literature, but on the other hand for highly diverse sets FLUFF-BALL met or exceeded the results of the standard 3D-QSAR methods.

### Ghemical—chemistry in a GNU flavour

The molecular modeling program Ghemical has grown from Tommi Hassinen's studies and interests in applying computers and 3D models of molecules to chemistry-related problems. The program has always been an essential tool in his research work of developing simplified molecular models of proteins, and also in his educational work. Starting from May 2000, Ghemical has been available for download in the net under an open-source license from http://www.uku.fi/~thassine/ghemical/.

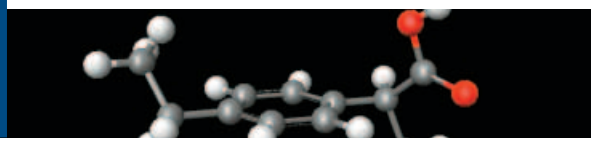
Ghemical is basically a molecular editor, allowing users to draw 3D molecular models using the mouse, and to refine the model so that the structure would be as correct as possible. The program is neither a simple molecular viewer (Jmol being better at this task since it works on web pages as well), nor it is (yet?) a replacement for professional modeling systems designed to cover all fields of computational chemistry. Calculations of the molecular mechanics type are handled by the program itself, but for quantum-mechanical calculations it is interfaced to other programs (MOPAC7 and MPQC at the moment).

At it's current state the program is approaching maturity in a sense that adding new features is a less important goal than making program's structure cleaner and making the program more portable to different types of computer systems. Also, making the program more accessible for users is an important goal for future.

The next big release of the program will be version 3, and it is a well-known rule of thumb in the software world that at this stage software usually becomes actually usable.

## Ghemical

- Ghemical homepage with a complete feature list, more screenshots, and the program's source code for download: http://www.uku.fi/~thassine/ghemical/.
- The program works in Linux, MacOSX and Windows systems. In Linux, the program must be compiled from source code.
- The MacOSX and Windows ports are created by users at the University of Iowa. They have also added a front-end module for the GAMESS quantum mechanical program. http://www.uiowa.edu/~ghemical/
- At bioinformatics.org website there is a mailing list where users can send their questions and comments about the program: http://bioinformatics.org/mailman/listinfo/ghemical-users



## More information


1. Laatikainen, R. "Automated Analysis of NMR Spectra." *J. Magn. Reson.*, 1991, **92**, 1–9.
2. Hassinen, T., Peräkylä, M. "New Energy Terms for Reduced Protein Models Implemented in an Off-Lattice Force Field." *J. Comput. Chem.*, 2001, **22**, 1229–1242.
3. Korhonen, S.-P., Tuppurainen, K., Laatikainen, R. and Peräkylä, M. "Comparing the Performance of FLUFF-BALL to SEAL-CoMFA with a Large Diverse Estrogen Dataset: from Relevant Superpositions to Solid Predictions." *J. Chem. Inf. Model.*, 2005, **45**, 1874–1883.
4. PERCH NMR Software, PERCH Solutions Ltd., http://www.perchsolutions.com

# VIEWPOINT

## From desktop to petaflops

Jifeng Yao

Desktop supercomputers and petaflops supercomputers have attracted a great deal of attention in modern high-performance computing, HPC, community, and this attention will probably be kept up for the next few years.

Desktop supercomputers, or personal supercomputers, are small clusters making scientific or industrial calculation for individual scientists, workgroups and departments. The emergence of desktop supercomputers relies mainly on two factors: the rapidly growing demand for HPC and the progress of computer technology.

Nowadays, HPC is affecting every vertical. In SSC, the Shanghai Supercomputer Center, we have seen a distinct increase of HPC users and application fields. Besides the traditional HPC realms such as physics, chemistry, biology and weather forecasting, industrial applications are becoming another large user group of HPC technologies. Shanghai Dredging Corporation, one of SSC's new customers, are now using some OpenMP-based parallel codes to simulate the moving of sand under water and are making their dredging plans according to the computation results, all of which can greatly decrease the amount and cost for dredging operations.

Many customers of SSC need no more than 64 processors. The cluster technology and multicore processors will make it possible for them to have their own supercomputers instead of scrambling for computation power from a supercomputer center.

Intel is going to release Clovertown, the first 4-core processor, by the end of this year. You can imagine a 4-way server based on a 4-core processor; it's a real supercomputer, not only because of its powerful computation capability, but also because it is a parallel computer. In order to take advantage of such a system, parallel computing technologies are necessary. Fortunately, all kind of parallel codes for both scientific and industrial computations are more readily available.

Only a small portion of our users use their own codes, most of them are using existing commercial or open-source software. The convergence of commercial and non-commercial parallel applications, affordable high-performance hardware and the increasing demand for high-performance computing are making supercomputing personal, from centralized supercomputer center to desktop.

Another trend is the petaflops computing. The demand for supercomputing not only focuses on quantity, but on quality as well. Scientists are always hungry for computation power, and the petaflops supercomputer is next on the list. We've already got a few 100 teraflops supercomputers listed on the Top500, but a petaflops supercomputer is not simply a 100T system times 10.

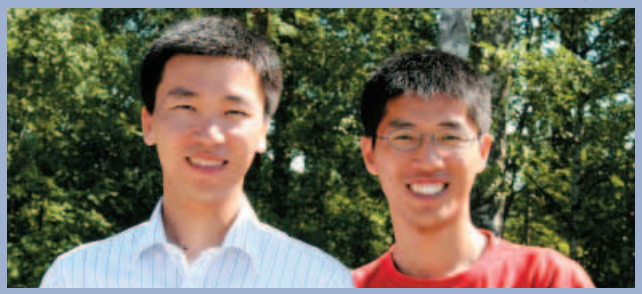
We need some new ideas, architectures and technologies. The United States and Japan have been in competition with each other for the first petaflops supercomputer for several years. In China, a plan has also been made to develop the first petaflops system before 2010. The problem lies more in the usage of such a system than in the building of it. What really counts is not only hardware, but also software, including system software, development software such as compilers, and parallel applications.

Parallel algorithms with poor scalability need to be re-designed. Parallel codes need to be re-tested for thousands of processors to ensure the correctness. And in order to get high-performance from a petaflops system, most applications need some kind of modifications or even the re-writing of some important parts. We may need new models and algorithms, powerful compilers, and effective programming languages. All these constitute challenges for computational scientists, computer scientists and also application scientists.

Currently, SSC's users tend to run small problems on their own PCs or workstations to test their models or codes, and use our teraflops system to perform large-scale computations and real simulations. Will this situation change to desktop supercomputer and petaflops supercomputer? Can we eliminate the gap between these two systems? That is another challenge we're facing now.

**"The convergence of commercial and non-commercial parallel applications, affordable high-performance hardware and the increasing demand for high-performance computing are making supercomputing personal, from centralized supercomputer center to desktop."**

© Leena Jukka, CSC



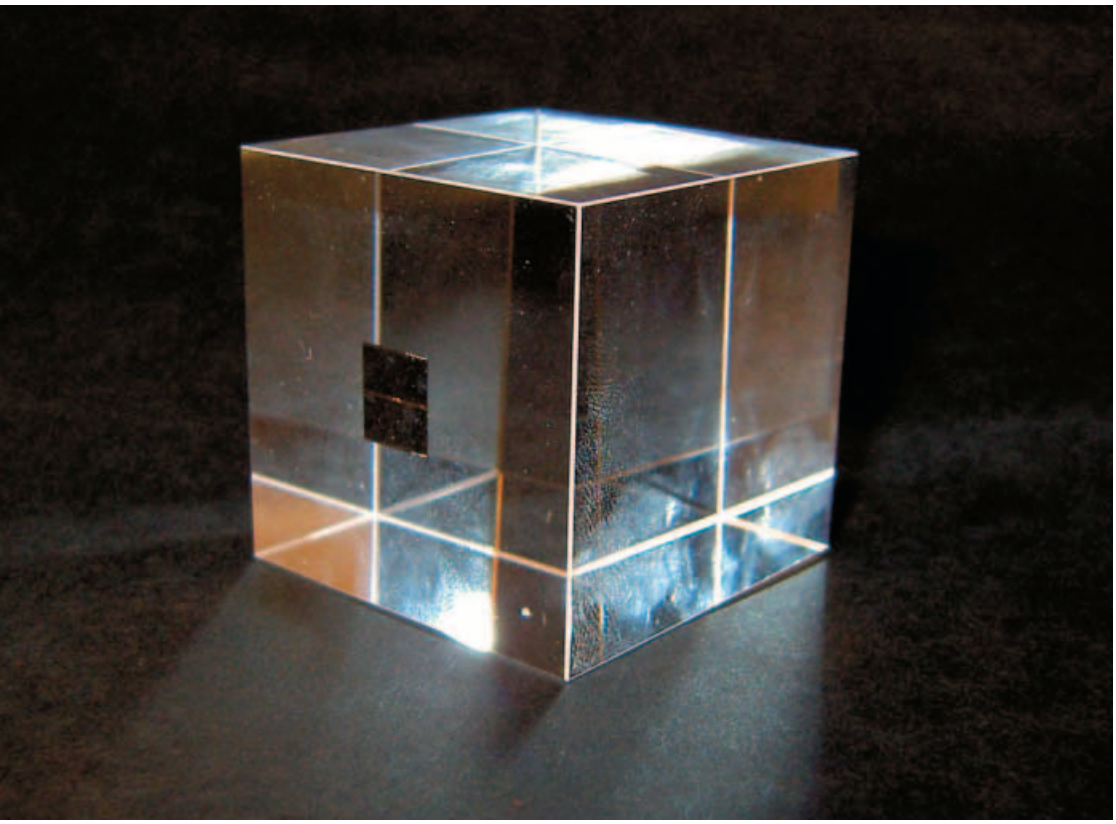
Jifeng Yao and Jiancheng Wu from the Shanghai Supercomputer Center visited CSC last summer to discuss possibilities for co-operation.

Dr. Yao is director of parallel computing and manager of R&D department at Shanghai Supercomputer Center, the open and largest supercomputer center in China. His own research interests include parallel algorithm and system & application benchmarking.

# Quantum fracture—detecting atomic scale fractures

The very smallest fractures are created by breaking only a small number of atomic bonds. Until today there has been no measuring technique that could detect fractures on this truly microscopic scale. By accident, it was recently observed that a type of detector developed for searching dark matter is an excellent tool for this purpose.

Jan Åström



A dark-matter detector from the underground laboratory at Gran Sasso, Italy. The detector is a pure sapphire crystal, which was accidentally clamped too tight during measurements, and detected fracture events instead of elusive dark-matter particles possibly included in cosmic radiation.

Fracture mechanics is usually described using the theory of elasticity. This means that the material is described as a continuum. On large scales, say millimeters, this is a good approximation, but on atomic scales, such a description breaks down. Fracture is no longer continuous, but shows quantized features.

The detector for searching dark matter was developed at the Max Planck Institute for Physics in Munich. It is based on a very pure sapphire crystal kept at a very low temperature. The actual detecting device is a thin metal film right at the temperature of its superconducting transition, attached to the surface of the crystal. When a fracture appears in the crystal, it creates lattice phonons, which appear as a temperature pulse, big enough to create a measurable change in the conductivity of the film. This method is so sensitive that it can probably detect the fracture of only a few atomic bonds.

The fracture data analyzed so far originates from accidental fractures within the detectors used in dark-matter search. These data contain registered fracture events as small as a few kiloelectron volts, which corresponds to the breaking of a few hundred atomic bonds. A detector designed especially for fracture research could be made more sensitive.

CSC has participated in the statistical analysis of the existing fracture data. The most striking feature of the data is that they display correlations and an energy distribution strongly reminiscent of earthquake fractures measured in tectonic faults on length scales of thousands of kilometers.

Jan Åström works as an in-house phycisist at CSC focusing on code optimization, and can be reached via email jan.astrom@csc.fi.

## Further reading:

1. J. Åström *et al.* Fracture processes observed with a cryogenic detector, *Physics Letters A, vol 356,* Issues 4–5, 14 August 2006, Pages 262–266

2. Fracture processes studied in CRESST, Nuclear Instruments and Methods in *Physics Research A 559,* **754** (2006).

3. Dark matter detector finds tiny cracks instead, *New Scientist,* 29 April (2006).

**"The most striking feature of the data is that they display correlations and an energy distribution strongly reminiscent of earthquake fractures measured in tectonic faults on length scales of thousands of kilometers."**

# column

Osmo Pekonen

## Math fiesta in Madrid

Osmo Pekonen

Some 3500 participants from all over the world attended the 25th quadrennial International Congress of Mathematicians (ICM) in Madrid, 22–30 August. Four Fields Medals were awarded by the International Mathematical Union (IMU): to the Russians **Grigory Perelman** and **Andrei Okounkov**, to the Frenchman **Werner Wendelin**, and to the Australian **Terence Tao**. The King of Spain was there to hand out the medals.

Perelman's achievements appear to be the greatest as he seems to have proved the time-honored Poincaré Conjecture, making sagacious use of the flow of Ricci curvature on 3-dimensional manifolds. Such an approach had been suggested by the American **Richard Hamilton** already in 1982, yet he lacked the muscle to push his idea through. Instead, Hamilton's dream was fulfilled by Perelman, a 40-year-old Jew and a former graduate of Leningrad's famous high school No 239. However, Russia's beautiful mind didn't become world famous because of his proof – which few people will ever understand anyway – but because he declined the award and refused to attend the royal ceremony, despite the frantic last-minute efforts of **Sir John Ball**, President of the IMU, to persuade him to come.

A rare scandal in the silent world of mathematics ensued. Suddenly, there was something of an American–Russian–Chinese struggle for scientific supremacy in the air. As described by **Sylvia Nasar** and **David Gruber** in the *New Yorker* dated 28 August, the Chinese mathematician **Shing-Tung Yau**, a past Fields medalist and Hamilton's former collaborator, organized murky maneuvers to appropriate some part of Perelman's glory for himself. Indeed, two of his students (**Cao** and **Zhu**) had written up a polished version of Perelman's result. In the Chinese press, their contribution was amplified to become the proof itself. **Jim Carlson**, President of the Massachusetts-based Clay Mathematics Institute, had to intervene in the debate to promise one of the coveted Clay Prizes, worth of one million dollars, to Perelman alone once his result has been checked.

(In the next act of the drama, the lonely genius of Saint Petersburg is likely to decline that prize, as well.)

Where does Finland stand in the global contest for mathematical excellence? Not precisely in the waiting room for Fields medals, I am afraid. **Lars Ahlfors** (1907–1996) won this distinction 70 years ago but even he made most of his career in the United States. Next year, the centenary of Ahlfors' birth will be celebrated with much pomp in a conference in Helsinki. However, excessive commemoration of past glories may not be healthy for the science of a small country.

We had a dinner-table discussion on this with my friend **Vagn Lundsgaard Hansen**, President of the Danish Academy of Natural Sciences. There was a time when Danish mathematics almost perished due to an overdose of almost periodic functions, the study of which was eagerly advocated by **Harald Bohr**. Yet, mathematical biodiversity survived, and Denmark developed flourishing schools of Algebraic Geometry, Algebraic Topology, and K-theory, to name a few.

Denmark is a nation comparable in size with Finland. Let us contrast the mathematics of Denmark and Finland in the light of the recent ICM's. For us Finns, the ICM 1978 of Helsinki marked the heyday. Lars Ahlfors appeared as a plenary speaker, and **Rolf Nevanlinna** was the honorary president. Leaving that special year aside, the sad truth is that besides Ahlfors and Nevanlinna there have *never* been any other Finnish plenary speakers in the entire history of the ICM, whereas there was a Danish plenary speaker both in the ICM 2002 in Beijing (**Uffe Haagerup**) and in the ICM 2006 in Madrid (**Ib Madsen**). This testifies about the vitality of Danish mathematics.

The problem with Finnish mathematics is its single-mindedness. As the Mathematics Genealogy Project website reveals, we have a single school of Real and Complex Analysis which has pervaded all Finnish universities and occupied nearly all the professorships. Nearly all doctoral students are recruited by this single school which has reached its sixth generation by now since **Ernst Lindelöf**. I don't find this state of affairs entirely healthy anymore.

Don't get me wrong. No one can doubt the expertise of the Finnish school of Analysis. Several of its exponents are quite outstanding; for instance, **Kari Astala** is a Salem Prize Laureate. Even so, Analysis is only one among the twenty sections in the ICM table of contents. My point is that, besides Analysis, there seems to be little interest into anything else in Finland.

There was a time in the Humboldtian university when professors were ideally required to display proficiency in at least two different topics of their discipline. This was – and still is – the idea of the German *Habilitationsschrift*. In Finland, however, most professors seem to know only one topic, and requiring more would be unrealistic. But could we require that a country of the size of Finland should have internationally recognized standing in more than one field of mathematics?

**"I would prefer not to."**
– Bartleby the Scrivener

The author is a mathematician and a science writer based at the University of Jyväskylä.He can be raeched via email pekonen@mit.jyu.fi.

# Big in Japan **ICQC 2006**

Every three years, the density of quantum chemists abruptly increases at a specific location of the earth.

Mikael Johansson

This phenomenon was first observed in Menton, France, in 1967.

That year saw the creation of the International Academy of Quantum Molecular Science (IAQMS). From the beginning, a main goal of the Academy was the organizing of a series of international meetings for the community. The 12th International Congress in Quantum Chemistry, ICQC for short, was held in May earlier this year. The setting for the conference was Kyoto, the heart of old imperial Japan; the topic, not surprisingly, the latest developments and advances in the field of quantum molecular sciences. Another main goal of the conference was to initiate interaction between theory and experiment. More than seven hundred participants from some forty countries attended the event.

Robert G. Parr, with a seventy-year long experience of the field and being one of the founders of IAQMS, gave a fascinating opening lecture about both the past and the present of quantum chemistry. Over forty lectures followed during the six days. It is noteworthy that all speakers performed in the same hall. The lack of parallel sessions was welcome; the high standard of the lectures would have made choosing between them nigh impossible. In addition to the lectures, quantum knowledge exchange took place during the four poster sessions. Carbohydrates fueling the brain through the two hundred posters per day were provided by ample servings of sushi and pizza. Topics covered included everything from reaction dynamics to innovative advances in methodology, both DFT and *ab initio*. A notable trend is the emphasis on efficient parallel implementations of algorithms.

The massive program, with its IAQMS Medal awards, interesting talks and posters, stretched the normal conference day to a twelve hours plus intense information flow. A well deserved rest for the brain, or rather, the left half of it, came in form of the conference excursion to three representatives of Early Japan. The inner tranquility of the shrines and temples could be sensed even through the heavy filter of the omnipresent tourist crowd.

ICQC'06 was without question the most flawlessly organized conference I have attended. The smoothness of operation was surpassed only by the prime reason of the event, the diffusion and exchange of scientific ideas. All participants left Kyoto with an, at least to some extent, greater knowledge of science, I'm sure. For me personally, the extent was significant.

Mikael Jøhansson works as an in-house chemist at CSC. He can be reached via email mikael.johansson@csc.fi.

Cultural exchange and correlated motion among quantum chemists.

## ICQC 2009—Finland

The settings for the 13th conference in the series will be bright white summertime Helsinki; ICQC'09 will take place between the 22nd and 27th of June, 2009. CMS, the Finnish Center of Excellence in Computational Molecular Science coordinates the congress organization, while the organizing committee has members from universities around the country. Preparations are well on their way, with, for example, the lecture hall already reserved. A number of satellite symposia, covering perhaps a more narrow area of quantum chemistry than the all-encompassing ICQC, have traditionally accompanied the congress. Symposia are expected again, also in neighboring countries.

The funding is still open, and collaborative partners are welcomed. The three previous congresses were held in Japan, Germany and France; "The 13th ICQC will help to put Finland on the Quantum Chemical World map", says Prof **Pekka Pyykkö**, head organizer of the congress and vice president of the IAQMS. The event is awaited to gather around five hundred quantum chemists from around the world. Says Pyykkö, "Finland is exotic and attractive, people will want to attend".

# SCIENTIFIC APPLICATION USAGE

SIRPA KOTILA
SIRPA.KOTILA@CSC.FI

KIMMO MATTILA
KIMMO.MATTILA@CSC.FI

CSC's specialists maintain the widest selection of scientific software in Finland, attending to software installations and database version updates. Approximately 200 scientific software packages and 70 databases are installed in the CSC computing environment. In addition, CSC is the national license holder and distribution center for several programs and databases.

## APPLICATION SOFTWARE AND DATABASE USE ON CSC SERVERS

Most of CSC's application programs are installed on the new SunFire 25K application server (corona.csc.fi) and on the HP ProLiant Linux cluster (sepeli.csc.fi). The applications on IBM eSuper Cluster 1600 (ibmsc.csc.fi) are mainly quantum chemistry and bioinformatics applications.

According to processor time in 2005, the most popular application programs were the four chemistry programs, Gromacs, Gaussian, TURBOMOLE, and Cerius2, and the BLAST and POY biosciences programs. In terms of search commands, the most frequently used programs were the following four chemistry programs: Ampac/Mopac, Gaussian, TURBOMOLE, and Dmol. According to active user ID logins, the most popular programs were Matlab for mathematics (261 users), EMBOSS for biosciences (225), and Lemmie2 for linguistics (214).

| APPLICATION | NO. OF USERS |
|---|---|
| **Biosciences** | |
| GeneSpring | 198 |
| InsightII | 94 |
| Ingenuity | 57 |
| KDE* | 23 |
| Genomatix | 17 |
| Quanta | 7 |
| **Chemistry** | |
| MDL/Crossfire* | 160 |
| PERCH | 58 |
| Materials Studio | 58 |
| CSD | 50 |
| Tripos/Sybyl | 61 |
| Gold | 54 |
| Cerius2 | 50 |
| MDL/Isis* | 15 |
| **Earth sciences** | |
| ArcInfo | 200 |
| MapInfo | 400 |
| Geographical information system (GIS) databases | 150 |
| **Structural analysis** | |
| Ansys, Adams and Abaqus | 215 |
| **Total** | **1867** |

* The service comprises locally installed customer software utilizing a CSC server.

**The 2005 statistics on the most important CSC-licensed, locally used software, databases, or services and their numbers of users. The table concerns only the use outside the computing servers of CSC and for some programs the user figures are estimates.**

The share of scientific application programs of the total processor time used at CSC's computing environment is approximately 25 percent. The rest of the computing time (75%) is consumed by projects which use computing codes developed by themselves or other programs freely available for academic research (such as Open Source applications).

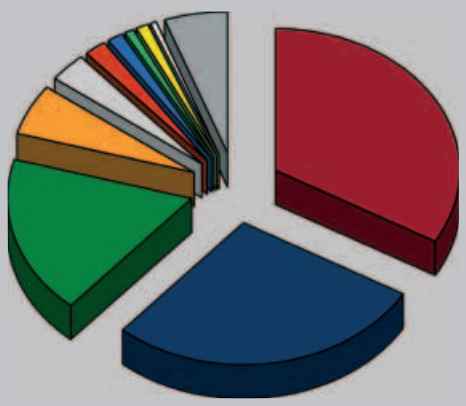
## USAGE OF APPLICATION SOFTWARE OUTSIDE CSC'S SERVERS

In addition to the customers using CSC's application servers, there is a growing number of researchers who use application programs or databases with a CSC license on their own computers. In 2005 almost 2000 researchers used scientific application user licenses issued by CSC.

CSC administrates campus licenses of the academic geographic information systems (GIS) programs (ArcInfo and MapInfo) and provides GIS database services for hundreds of academic users. National licenses and license consortiums are used also in biosciences, chemistry, and structural analysis. In all these disciplines, programs and database user interfaces installed on users' own computers represent a significant part of the services provided by CSC.

*Sirpa Kotila is CSC's customer administration manager and Kimmo Mattila is CSC's bioinformatics specialist.*

## USAGE OF APPLICATION SOFTWARE ACCORDING TO PROCESSOR TIME IN 2005



| SOFTWARE | PROCESSOR TIME* | DISCIPLINE |
|---|---|---|
| Gromacs | 307554 | Chemistry |
| Gaussian | 259055 | Chemistry |
| TURBOMOLE | 168393 | Chemistry |
| BLAST | 54251 | Biosciences |
| POY | 30121 | Biosciences |
| Cerius2 | 18348 | Chemistry |
| Fluent | 12678 | CFD |
| ADF | 8352 | Chemistry |
| NWChem | 8208 | Chemistry |
| Abaqus | 7197 | Structural analysis |
| Other | 50310 | |
| **Total** | **924467** | |

* cpu hours

# Open Workshop on *e-Infrastructures*

## CSC, the Finnish IT center for science, Espoo, Finland • October 4–5, 2006

e-Infrastructure Reflection Group will organize an open workshop on October 4–5 in Espoo, Finland. The purpose of the e-IRG Workshop during Finland's Presidency is to update the current Roadmap and White Paper.

The series of open e-IRG workshops support the e-IRG activities by enabling and stimulating the discussion of e-infrastructure related topics with the community across thematic and country borders. Experts on various areas will get together to explore open issues and outstanding questions for enabling easy and cost-effective shared use of distributed electronic resources across Europe and beyond, based on sustainable e-infrastructures.

**Speakers will include:**

- Sakari Karjalainen (Finnish Ministry of Education)
- Risto Nieminen (Finnish Grid)
- Timo Skyttä (Nokia, Liberty Alliance)
- Kimmo Koski (CSC, chair HPC European Taskforce – HET)
- Michael Grønager (Technical Coordinator NDGF, Nordic GRID)
- David Giaretta (CASPAR project on data access)
- Klaus Ulmann (DANTE)
- Christoph Witzig (SWITCH)

**The open e-IRG workshop is hosted by CSC, the Finnish IT center for science. For more information and registration, please visit http://www.e-irg.org/**

e-IRG e-Infrastructure Reflection Group

CSC

eu2006.fi

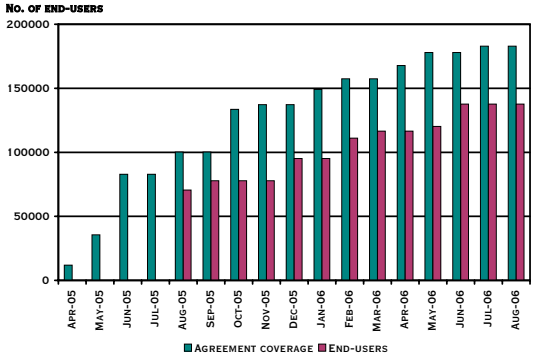
## Facts about Haka Federation

- the identity federation of Finnish universities, polytechnics and research institutions
- operated by CSC
- based on Shibboleth middleware version 1.2.1-
- Important dates:
  - May–June 2005: The first membership applications were accepted and service agreements signed. The first members of the Haka federation were: *the University of Helsinki, Helsinki University of Technology, Tampere University of Technology, Åbo Akademi University, and Tampere Polytechnic.*
  - August 2005: the first IdP servers were registered to Haka *by Helsinki University of Technology, University of Helsinki and Tampere University of Technology.*
  - May 2006: the first Federation Partners were accepted to Haka Federation
- End-user statistics (September 2006):
  - 19 contractual members (11 out of 20 Finnish universities and 8 out of 29 polytechnics have signed the membership agreement), 2 contractual partners
  - The membership agreement coverage of Haka is 60 percent of students and staff in Finnish higher education (80% of all university end-users and 28% of all polytechnic end-users are under signed membership agreements)
  - Ten members have registered their Identity Provider (IdP) server to the Federation. Others are in the process to take up the service.
  - 138 000 end-users are presently able to use Haka authentication and authorization services
  - 43 000 logins registered on services in April 2006, 58 000 logins in October 2005
- Identity Providers in the federation in August 2006:
  - University of Helsinki, University of Oulu, Lappeenranta University of Technology, Pirkanmaa Polytechnic, University of Tampere, Tampere University of Technology, Helsinki University of Technology, Åbo Academi University, and University of Turku
- Service Providers in the federation
  - Nelli portal of Finnish Electronic Library (based on ExLibris Metalib)
  - Three learning management systems – Optima at the University of Oulu, A&O and Moodle at Tampere University of Technology
  - 'Teemu' Library Database of Helsinki University of Technology
  - CSC's Scientist's Interface extranet
  - CSC's extranet for Funet member organisations
  - JOOPAS (electronic circulation of applications for visiting student in another university)

http://www.csc.fi/suomi/funet/middleware/english/index.phtml

Sirpa Kotila and Mikael Linden (CSC)

**End-users of the Haka Federation**



# Federated identity supports academic collaboration

Mikael Linden

Everyone of us is familiar with a computer prompting the user to log in to a service. Actually, most of us have way too many usernames and passwords to remember in our daily life in the networked world. More and more services want to provide personal tailored service to their users and, thereby, need to know who the user is.

During last few years computer science research and software industry has recognised this, and a new field of interest for scientists and practitioners, called digital identity management, have been born. Identity management tries to find ways to organise and manage the identity data of the users, in order to make services easier to use and maintain, without jeopardising users' privacy and security.

Because of the increasing networking and collaboration of organisations, services are often provided by several distinct organisations, and the user accounts should be shared between them. In federated identity management, collaborating organisations form a federation to share user data. From an end user perspective, the user is able to use a single username and password to sign in to any service in the federation, whoever organisation provides the service.

Universities and research institutions, connected with research networks, have been the place where many new networking technologies have entered daylight. For instance, Internet, electronic mail, web and multicast have been introduced at first to the academic community.

Universities now seem to be a community where identity federations are being formed. Universities and research groups have a lot of collaboration, the users are highly skilled and open-minded for new ways to use the network and they utilise the network for research and learning in daily basis.

## Haka federation had its first anniversary

Haka federation is the identity federation of Finnish higher education and research community, aiming at supporting them in network-based collaboration. The end users in Haka federation's member organisations can use a single set of credentials, provided by their home organisations IT service unit, to sign in to local and remote network services. When a user signs in to a service, her personal data is fetched from her home organisation, making maintenance of services' user data easier.

Haka federation has been operational since August 2005. Currently the federation has 18 members organisation with 183,000 end users. Eight services are available for federated login, with approximately 50,000 logins in a month. Haka federation is based on open standards (SAML) and their free open source implementations (Shibboleth).

Scientist's interface, CSC's extranet for researchers, is an example of the services in Haka federation. From the first, CSC's customers have been provided separate usernames and passwords to sign in to CSC's computing environment. Since spring 2001, Scientist's interface has been a convenient web based alternative for the traditional command-line use of CSC's services. In March 2006, Scientist's interface was registered to Haka for federated authentication. Instead of entering the CSC-provided username and password, researchers have been able to click the Haka button instead for subsequent authentication at the home university.

## Supporting dynamic collaboration a challenge

Haka and other federations rely on using the usernames and passwords provided by the universities' IT service units—typically each student and staff member gets one when they first enter the university. However, the nature of collaborations is very dynamic; new collaborative groups, often called virtual organisations, are formed and others fade away.

During their lifetime, virtual organisations often need supportive tools, such as mailing lists, wikis and document repositories. Designing federated services to support identity management of a virtual organisation and its collaborative tools is challenging. To keep the management of a virtual organisation flexible, it makes sense to detach it from the university's IT service unit. In forthcoming years, we will see this as one of the major development challenges of federated identity.

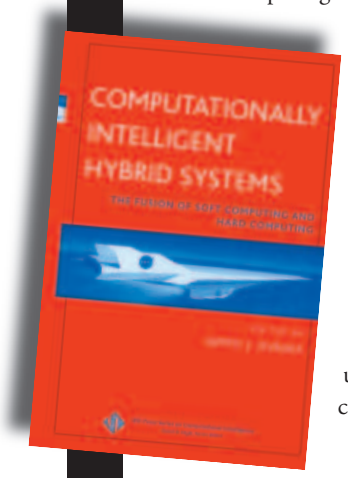
Mikael Linden has worked with Haka from its beginning in 2002, and can be reached via email mikael.linden@csc.fi.

# Imprints

Juha Haataja

## Both sides of the brain

Computationally Intelligent Hybrid Systems: The Fusion of Soft Computing and Hard Computing (Seppo J. Ovaska, ed.; Wiley-IEEE Press, 2004)




What are soft computing and hard computing? These two ways of using computers can be compared to the two hemispheres of the brain. Humans apply logical and mathematical reasoning to solving problems using the left hemisphere of the brain. However, humans also excel in pattern recognition, creativity, synthesis of ideas using the right hemisphere.

Today, computers can be used for both soft and hard computing. Integrating these capabilities is needed to solve real-world problems. But how to combine genetic algorithms and traditional optimization? What about neural networks and control systems? Or fuzzy logic and Bayesian data mining?

The book edited by Ovaska offers a valuable introduction to the fusion of soft and hard computing. The book consists of eleven chapters written by experts of the respective fields. However, the book is not just a collection of separate articles. The editor has written introductory texts to all chapters of the book, connecting the topics together and providing additional insights about the topics.

Seppo Ovaska, the editor of the book, is a professor in the Department of Electrical and Communications Engineering at TKK (Helsinki University of Technology). The dedicated work by the editor of the book is all too rare nowadays.

The chapters are application-oriented, and thus benefit readers interested in subjects like large-scale planning in a plant, adaptive flight control, intrusion detection in computer security, generating human-computer interfaces, and data mining in www usage.

The book succeeds in providing a link between the two methods of computing, both on the theoretical and practical level. Although the field develops fast, the theoretical background and the focus on systematic thinking about hybrid systems makes the book a valuable reference.

## It's all about software

Towards 2020 Science (Microsoft Research, 2006)

A group of distinguished scientists wrote a report titled Toward 2020 Science on the role and future of science over the next 14 years. The findings and conclusions of group are both provoking and insightful.

But there is a serious omission. All the 41 members of the 2020 science group were men. How was it possible for Microsoft to limit the group of internationally distinguished scientists to men only? Are women not worth listening to when discussing science and computing?

Despite the curious men-only approach to science, the report has a lot to offer to readers. The expert group tries to look beyond the current situation.

Today we are trying to apply computing to scientific research. This way of doing research is often called computational science, in contrast with experimental and theoretical science. However, in the future researchers need to integrate computer science directly into scientific research.

Moving beyond computational science is a big challenge, although on the surface the difference seems to be subtle. But computing offers new possibilities for researchers, such as new ways of thinking about research problems. Integrating computer science into scientific research can provide fundamental new developments for example in biology, biotechnology, and medicine.

The report is written for both non-experts and experts in computing and science. Thus the report skips over more technical details, but does provide an interesting overview of applying computing to various scientific research areas.

Most of the discussion focuses on bioscience and medicine, where new computational tools may be most needed. But many of the conclusions apply to other topics as well, such as nanotechnology and Earth science.

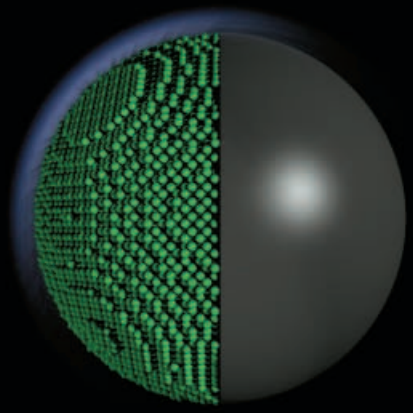
The report states that science and science-based innovation should be at the top of the political agenda. Today this is not so. Instead in Finland and elsewhere high-level planning centers on technology-based innovation. In the future decision-makers should focus on the contribution of science-based innovation to economic prosperity

The report also recommends an urgent re-think in educating tomorrow's scientists. In 2020, today's children will be starting their careers. Will they be sufficiently computer literate to benefit from the possibilities of integrating computing and science?

The key question on solving the research problems of the future is how to develop the technological tools needed in 2020. Computer science is needed to find solutions to scientific grand challenges. "It's all about software."

Will we have in 2020 the skills to solve the computational problems posed by the scientists of the future?

© Jura Tarus, CSC



# Supercomputer shopping at final stage

Paavo Ahonen
Juha Fagerholm

The supercomputer purchase project has been active nearly a year now. During the past months, the project group, as well as the vendors, have been busy in preparing data for the purchase. The benchmarks have been run and analyzed.

### Events up till now: more benchmarks, negotiations, and final requirements

Two additional benchmarks were prepared to ensure that the most demanding applications would also have a good platform in the future. The new benchmarks were a quantum chromodynamics code, SU3_AHIGGS and a large-scale weather forecast code XL HIRLAM. Although the weather code is not presently run on CSC's computers, these programs represent the requirements of fast interconnection and fast processors that the next supercomputer will have to meet.

The final invitation to tender, based on the negotiations carried out in May and June, was sent to shortlisted vendors in early July, and the final tenders were received at the beginning of August. In August and September, the project group has evaluated the data preparing for the decision.

### What is still ahead?

The final decision and the announcement on the next supercomputer are anticipated to be made shortly after this issue of CSCnews is published. The project group is confident that a good result will be achieved and the Finnish research community will get a competitive supercomputer environment.

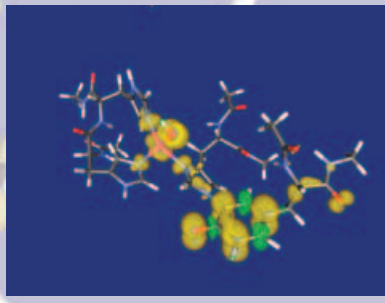
The supercomputer acquiring process will move to the installation phase. But that is another story.

Paavo Ahonen is a science editor at CSC following the procurement process of the new computer and Juha Fagerholm is an in-house physicist working on parallel computing.

## Computational Chemistry Day at CSC on 23.10.2006

The Computational Chemistry Day is intended for researchers in Finland who are involved or interested in computational chemistry. The program will consist of invited talks from different fields of computational chemistry and several contributed talks given by the participants.
Confirmed speakers:

- Doc. **Ari Paavo Seitsonen** L'institut de minéralogie et de physique des milieux condensés, Paris, France, **Ionic liquids from ab initio molecular dynamics: Case study of EMIM+AlCl$_3$**
- Doc. **Mikael Peräkylä,** University of Kuopio, **Molecular Dynamics Simulation Studies of Receptor Activity Modulation by Ligand Binding**
- Dr. **Maija Lahtela-Kakkonen,** University of Kuopio, **Computational tools in drug design: designing new drugs for sirtuins**



The last day for registration is 09.10.2006.
Please see http://www.csc.fi/suomi/koulutus/kurssi.phtml.en?id=322 for more information.

# DEISA TRAINING

Distributed European Infrastructure for Supercomputing Applications

## SPECIAL TOPIC: PERFORMANCE AND PORTABILITY

The next DEISA training session will be organized at Forschungszentrum Jülich GmbH, in Jülich, Germany on October 23rd - 25th, 2006. Scientists from all European countries and members of industrial organizations involved in high performance computing are invited to attend.

The purpose of the training is to enable fast development of user skills and know-how needed for the efficient utilisation of the DEISA infrastructure. The first part of the training will give a global description and introduction to the usage of the DEISA infrastructure. The second part of the training will be dedicated to the topic of performance and portability. The attendance is limited to 30 participants.

**The detailed agenda and the registration form are available at www.deisa.org/training. The registration will close on October 13th, 2006.**

**Monday, October 23rd**

| | |
|---|---|
| afternoon | DEISA general presentations: Description and usage of the infrastructure |

**Tuesday, October 24th**

| | |
|---|---|
| morning | DEISA general presentations (continued) |
| afternoon | Hands-on session "Access to DEISA" |
| | Performance and portability - tools and techniques |

**Wednesday, October 25th**

| | |
|---|---|
| morning | Performance and portability |
| afternoon | Use Cases |

egee

egee Industry Forum

# EGEE Industry Day

24th November, 2006, Helsinki, Finland

## Biomedical Challenges

The EGEE Industry Day will be a unique platform for Industry to interact directly with the EGEE project, the biggest Grid infrastructure in Europe, and will bring together decision makers, research heads, policy makers and CTOs to learn how industrial applications can be deployed on EGEE. The event will highlight where grid computing can create new industrial solutions and how organizations can benefit from sophisticated computing resources of grid, not available in traditional IT infrastructures. Interactive discussions will provide an opportunity to get an industry perspective and discover how EGEE can work towards a commercial grid.

**The EGEE Industry Day is hosted by CSC, the Finnish IT center for science. For more information and registration, please visit www.eu-egee.org/HelsinkiDay .**

CSC