Transforming chemistry through computation

What changes and what remains when we move to a new place?

Black sheep and White knights

Emerging needs for eInfrastructure in Climate Science
CONTENTS

What changes and what remains when we move to a new place? p : 4

Black sheep and White knights p : 8

Fighting for Subatomic Independence p: 12

Transforming chemistry through computation p : 14
  Molecules in strong magnetic fields p: 16
  Modeling of chemical reactions and catalysis p: 19
  Predicting hydrate stabilities p: 22
  Molecular dynamics simulations of the molecular Calcium pump p: 24
  Antimicrobial peptides emerge as a new class of drugs p: 26

StoreBioinfo – NorStore Life Sciences storage infrastructure p : 29

Emerging Needs for eInfrastructure in Climate Science p: 34

Commoditizing Norwegian eInfrastructure: Clouds p: 38

Cover picture: © Anders Lervik and Signe Kjelstrup

Number 3 : 2011
A magazine published by the Notur II project – The Norwegian metacenter for computational science

Editorial Director: Jacko Koster
Contributing Editor: Vigdis Guldseth
Subscription: An electronic version is available on www.notur.no. Here you can download a
PDF-file of the magazine and subscribe or unsubscribe to the magazine.
ISSN: 1890-1956

Reproduction in whole or in part without written permission is strictly prohibited.
e-mail address: sigma@uninett.no. Phone: +47 73 55 79 00. Fax: +47 73 55 79 01.
Print: Skipnes kommunikasjon AS

Disclaimer: UNINETT Sigma is not responsible for errors, opinions or views expressed by the authors or the contents of external information referred to in this magazine.
EDITORIAL

This issue of META contains contributions from scientists from a wide range of domain areas. The contributions show that usage of eInfrastructure is both broadening and deepening. Increasingly more user communities need access to infrastructure for computing and for managing scientific data, while communities that already adopted eInfrastructure as a core instrument in their work make increasingly advanced use of it.

eInfrastructure must be made attractive for scientific communities to develop long-term collaborative research in a national and global setting. The need for sustainability is obvious. This not only requires long-term and predictable funding. A sustainable eInfrastructure must also be able to make long-term commitments to maintaining and continuously improving a skills and competence base. Attention must be given to support and training, community development, mechanisms to couple eInfrastructure to research and development, and to strength, flexibility and capability to respond to new areas of science. Sustainability also requires the development of policies and mechanisms to encourage the integration of eInfrastructure in national and international research infrastructure. eInfrastructure must never be a goal in itself, but must be integrated in all research infrastructures that can make use of it.

Despite its growing popularity, there is still a need to increase awareness for eInfrastructure for modern research and science. eInfrastructure provides a broad infrastructure for the creation and support of data, information and knowledge. Besides large computational capacity, eInfrastructure must support the entire life cycle of data, support massive amounts of created and collected data, support the reuse and reusability of data, in particular for purposes other than what the data originally was created for, support preservation of data and include means to discover and use data through search and navigation across institutional, disciplinary and national boundaries.

This issue of META includes contributions from climate sciences, computational physics, life sciences and computational finance. The contributions show the diversity in which eInfrastructure is used and needed in the various scientific disciplines. The focus is on computational chemistry, the branch of chemistry that uses computer science to assist in solving chemical problems. Computational chemistry uses the results of theoretical chemistry, incorporated into efficient computer programs, to calculate structures and properties of single molecules, a group of molecules, or solids. In all cases, computer time requirements as well as the demand for memory and/or disk space increase rapidly with the size of the system or the accuracy that is required.

Jacko Koster
Managing Director
UNINETT Sigma AS
What changes and what remains when we move
What changes and what remains to a new place?

The memory circuits
Through evolution, mammals have developed a brain structure, the Hippocampus, that is used to transform everyday experiences into new memories. Patients with a severe lesion in this area are unable to generate new memories, such as depicted in fiction movies like Memento or 50 First Dates. They live in a constant present and everything beyond 5 minutes in the past is completely lost for them. Apart from this radical situation, the memory circuits are related to a number of more widespread mental diseases, such as Alzheimer’s Disease, Epilepsy or Post-Traumatic Stress Disorders.

Space and memory
Ever since the early studies of Ramon Y Cajal in the XIX century, understanding the Hippocampus has been one of the main interests of neuroscientists. However, the tools to study the electric properties of neurons, what we nowadays know as electrophysiology, were developed only in the second half of the XXth century. It was thus in the 70’s that researchers first noticed how important was the concept of space for the memory circuits. When rats are introduced into some new environment, neurons in their Hippocampus develop maps that can be later recalled and used for navigating the environment, even in the darkness when most sensory cues are not available. These neurons, called place cells, have the capacity to learn to associate their electrical activity to some coordinate (x,y) in space. As a result they become active only when the rat passes by that particular coordinate, different for every cell and for every environment. Place cells have been reported also in the human Hippocampus, although that interpretation is still controversial. Spatial navigation remains nowadays one of the best correlates of neural activity in the memory circuits of rodents and humans, making it an ideal paradigm to study how information is processed in these brain areas.

Is space computed in the Hippocampus?
The Kavli Institute for Systems Neuroscience and Center for the Biology of Memory, in Trondheim, has positioned itself during the last decade in the forefront of a worldwide community of neuroscientists seeking for the source of spatial information in the mammalian brain. In a succession of high impact publications, the Professors Edvard and May-Britt Moser showed that, although the Hippocampus uses spatial information to construct memories, it probably does not compute space on its own, but rather in collaboration with a neighboring structure, the Medial Entorhinal Cortex. It is there that they found neurons with perhaps an even more striking behavior: the grid cells [1,2]. One such neuron is active not in one but in many preferred positions – or grid fields – in a given environment. Strikingly, the spatial distribution of these grid fields forms a triangular grid that covers all the explored space. Due to this geometrical periodicity, the maps of grid cells can adapt to environments of all shapes and sizes, much like a Cartesian coordinate system.

Some parameters need to be specified to describe the map of a grid cell. In first place, the orientation of the three major axes, which are separated by an angle of approximately 60 degrees. In second place, the phase, or the displacement needed to reach the closest field starting from the center of the room. It turns out that all grid maps recorded simultaneously in a rat share the same orientation, but their phases are randomly distributed so that their maps cover homogeneously the whole space.

What changes and what remains when we move to a new place?
If we compare maps of two different cities we will find that they differ in many aspects but also share some common elements. The spatial organization of streets, hospitals and parks, for example, are different. On the other
hand, the coordinate system and the Euclidean properties of space, such as the notion of distance between two given points, are the same. Thanks to this, we can estimate the time that it would take us to walk from one place to another even if we have never visited the city before. Does the architecture of the brain account for this dissociation between individual details that change and general properties of space that remain invariant? It seems so.

If a rat is exposed to two different environments, A and B, it will learn maps for both and later will be able to retrieve the one that is relevant for every situation. In the Hippocampus, the two sets of maps will be completely different, as if from A to B all place fields were shuffled randomly. In this way, a place cell that in room A fires in a corner may fire in the center in room B or not fire at all. In addition to this, each neuron seems to remap independently, without caring about what the others do.

In the Entorhinal Cortex the situation is very different. If we compare grid maps in A and B we note that they may rotate and change their phase with respect to the center of the room, but the rotations and phase shifts are the same for all neurons. In other words, while Hippocampus seems to represent details of individual environments, the Entorhinal Cortex seems to represent abstract properties of space.

Our model of grid cell formation

The work that we did was inspired on a previous model [3]. This model proposed a biologically plausible way in which grid maps could be learned in a neural network of modest size. In the simulations, there was a weak interaction among neurons. When one neuron was active, it inhibited the activity of all others, a competition mechanism that is present in practically every region of the brain. This was the main reason why in the simulations neurons took turns to fire in different places. Another element of the model was associative learning. If a neuron won the competition in some place at a given time, there would be a strengthening of the synaptic connections that helped it in doing so, and consequently an increase in its probability to win that competition again in future visits of the same location. The triangular symmetry of maps was an emergent property of the network, since it is the geometry that allows optimal packing of circular fields in 2 dimensions.

There were, however, several aspects of grid cells that this small network could not account for. The most important of them was that grid maps in our simulations did not have a common orientation. This was due to the fact that inhibition is a very unspecific and weak form of interaction between neurons. However, to include excitatory interactions, which in the brain are strong, plastic and specific for every pair of neurons, we needed a much bigger network. Luckily, we could count on the aid of NOTUR.

Grid alignment and global remapping

In the new version of the model, we included excitatory connections between grid cells, such as anatomical studies indicate. This form of interaction, much more specific than global inhibition, had the effect of aligning all grid maps into a common orientation, while keeping the phases randomly distributed. In addition, this happened also when we trained our virtual rat in two different environments. The grid maps of room B could be seen as a rotated and shifted version of those in room A. The crucial point was that all maps were rotated and shifted by the same amount, just as observed in experiments. This time our model was representing space in an abstract way.

Interestingly, to achieve this we needed to add one more element to the system: directional information, such as the one provided by a compass. This would have been a bit arbitrary if it was not for the fact that this kind of information is also found in some parts of the Entorhinal Cortex. Until now this was thought to be
Rate maps (in spikes per second) corresponding to the maps in the top Figure. Warmer colors indicate higher firing rate.

Example of the activity of a place cell (left) and a grid cell (right) while a rat explores a square environment (black trace). Each color dot corresponds to the position where the neuron fired a spike of activity.

A perhaps random superposition of functions in the network, but our work suggests that there is a need for this conjunction. Taking a close look at anatomical studies one finds that directional information and excitatory connections are either present together or absent, which makes our hypothesis more appealing.

Looking into the future
Our modeling work has showed that grid maps in the Entorhinal Cortex could be aligned through excitatory connections as long as they come in hand with directional information [4]. It is by virtue of these connections that they would form a coherent triangular coordinate system that can be used in multiple environments, in the same way in which the square Cartesian grid is used in different city maps.

An interesting question for future studies is what would happen in the system if either the excitatory connections or the compass providing directional information would fail. If individual neurons lost their ability to communicate properly with each other, their maps could be less structured and coherent. This effect relates to various possible experimental situations, but also to the sudden disorientation that is characteristic of patients suffering from mental diseases that affect their Hippocampus and surrounding areas. Answering these questions should not only improve our understanding of information processing in the brain, but also, in the long run, help us to design new methods for the diagnosis and treatment of a variety of diseases related to memory.

References
On September 15, 2008, Lehman Brothers, a large US investment bank, filed for bankruptcy protection. The event came to mark the onset of the financial crisis which recently spread to European sovereign debt. Since then, regulators around the world have implemented various emergency measures in an attempt to calm financial markets. Politicians became self-declared white knights, and short sellers were found to be the blackest sheep in the family of financial practitioners. Short sale bans and financial transaction taxes `to make speculators pay for their deeds' have been peddled as solutions to contain the crisis.
Only five years before the turmoil in financial markets started, the Nobel Prize Laureate Robert Lucas declared «...macroeconomics in this original sense has succeeded: Its central problem of depression-prevention has been solved, for all practical purposes...». With the gift of hindsight, nothing could have been further from the truth. Indeed, economics experiences its own crisis as it contemplates alternatives to the standard rationality paradigm in modelling economic phenomena.

Evolutionary finance
Evolutionary finance applies Darwin’s principle of natural selection to study trading behaviour and asset prices in financial markets. In this perspective, a financial market can be seen as a selection mechanism which transfers wealth to traders who are well adapted to the environment from traders who are less well adapted. The trading strategies of wealthy traders determine the prices of financial assets, first, because those strategies are backed by more wealth, and second, because the strategies of wealthy traders are more likely to be imitated by other traders. Wealth in financial markets is therefore the counterpart to fitness in biological systems.

Natural selection in financial markets is known to produce rational behaviour in the aggregate. The rationality assumptions that form the basis of the standard theory of financial markets can be seen as a proxy for this outcome. In many problems of interest, this proxy is excellent, in others it is not. One aim of evolutionary finance is to provide answers to questions on which the standard theory is silent, for instance how markets become efficient. Another aim is to study issues that cannot be analysed with the tools currently available within the standard approach. The recent financial crisis has raised a number of such questions.

Financial regulation
Since the outbreak of the crisis, financial regulators of many countries have considered various measures to limit the negative effects of the crisis. The two most popular ones have been short-sale bans and Tobin taxes.

Short-selling means selling a stock that one does not own, and the popular opinion holds that short-sellers amplify market crashes by increasing the downward pressure on stock prices. Accordingly, a few days after the Lehman collapse, more than 20 countries introduced a temporary ban on the short-selling of financial stocks «to protect the integrity and quality of the securities market and strengthen investor confidence», as the US Securities and Exchange Commission (SEC) put it. In Europe, the temporary ban was reintroduced by Germany in May 2010, and by France, Belgium, Italy and Spain in August 2011, in attempts to stem the accelerating sovereign debt crisis.

A Tobin tax is a tax on financial transactions which is intended to curb speculation by making it more expensive to trade in financial markets. This tax has been used from time to time in the past, and has received fresh attention from regulators during the present crisis. Both the Nordic and the European Councils have called for the introduction of such a tax, and in August 2011, the German Chancellor Angela Merkel and the French President Nicolas Sarkozy said they would make such a tax a priority for Europe.

One trouble with both these regulatory measures is that their effects are neither well documented nor well understood. This is mainly due to a lack of adequate theoretical models and a shortage of historical data. After the US short-sale ban was lifted in October 2008, SEC Chairman Christopher Cox told Reuters that «While the actual effects of this temporary action will not be fully understood for many more months, if not years, knowing what we know now, I believe on balance the commission would not do it again. [...] The costs appear to outweigh the benefits».

Computational approaches to the fore
Computational evolutionary finance employs techniques from computer science to carry out controlled experiments, for instance on the impact of market design or regulation. This is useful for regulatory bodies who would like to know more about

![Figure 1 Simulation results of an order book market for equity in a firm with random earnings and unobservable business cycle regime.](image)
the effects of new regulatory measures before implementing them in the market place. The typical model in this field consists of a detailed description of the market microstructure, and a large number of individual economic agents who make portfolio and trading decisions within that microstructure. The agents’ trading strategies are represented as computer programs, and the model is solved by subjecting these programs to natural selection until the aggregate price process converges to a stationary process. Data from the model can then be analysed and compared with data from other experimental treatments with different market microstructure.

A model along these lines is illustrated in Figure 1. It contains two main parts: A financial sector and a real economy. The financial sector consists of 10,000 traders, who can invest in stocks and bonds issued by the firms in the real economy. They do this by submitting orders to a central stock exchange which maintains an order book and executes matching orders. The real economy is represented as one aggregate firm, whose main feature is an earnings process which subject to short-run fluctuations and medium-run business cycles. It is modelled as an Ornstein-Uhlenbeck process with a Markov-switching attractor that can take on two values: A high value, representing a boom, and a low one representing a recession. The earnings process is illustrated in Panel (1) of Figure 1, where the yellow bars represent recessions and the white areas booms. Debtholders receive fixed interest payments, represented by the red, horizontal line, and stockholders receive the difference between earnings [green curve] and interest payments.

The traders can observe the earnings process, but not the state of the Markov process governing the business cycle. However, they do have access to a Bayesian estimate of the probability that the economy is in a recession. This probability is illustrated in Panel (2) of Figure 1.

In order to make money, the traders must use their information about current earnings, the Bayesian state probability, the order book and their own portfolio to identify profitable investment or trading opportunities. Over time, the trading strategies of poor traders are replaced by copies or genetic recombinations of the trading strategies of more wealthy traders. After 3 million trading days, the gains from changing trading strategies are more or less exhausted. Then the simulation is terminated, the population is saved, and the model is restarted to generate data. Panel (3) of Figure 1 depicts the market price of one share [dark blue curve] along with the expected present value of the future cash flow per share [light blue curve]. Traders in the model turn out to be risk averse: the stock trades at a discount relative to the expected present value of the cash flow it generates. This discount, shown in Panel (4) of Figure 1, is called the equity premium. It is higher in recessions that in booms, meaning that only investors with a large appetite for risk buy stocks ‘when the cannons roar’.

**Experimental results**

Controlled experiments use simulations from a benchmark case with no short-sale ban and no Tobin tax, a case with a short-sale ban in place, and a case with only a Tobin tax. Each specification of the model is run for 100 different sample paths of the earnings process, corresponding to 100 scenarios of the real economy. Data are collected for each of the three experiments over 10,000 trading days for each of the 100 paths for the earnings process. Unlike in empirical research, we can re-run history with the benefit of creating a unique dataset of matched observations. This allows for comparisons of different regulatory regimes for an identical path of the real economy — the ultimate comparative analysis.

Figure 2 contains an overview of the results.

The radar plot provides information on the impact of financial regulation on order book markets along eight dimensions, all of which are central to the debate on the costs and benefits of regulation. Efficiency of the market is measured by liquidity [bid/ask spread] and price discovery; the strength and characteristics of price fluctuations are quantified using volatility, negative skewness and kurtosis; and the dynamics of long swings over the business cycle is captured by price bubbles and the depth of market crashes.

Short-sale bans were imposed with the goal of reducing price fluctuations, especially the depth of market crashes. Our results confirm that a short-sale ban indeed does have that effect. But the model also highlights the drawbacks. A short-sale ban increases the frequency of bubbles and leads to a generally overvalued stock market. The efficiency of the price discovery process is reduced, and so is liquidity.

Proponents of a Tobin tax have argued that it will reduce speculation and improve financial stability. These views are not supported by the model: A Tobin tax has no effect on the depth of market crashes, but large negative effects on price discovery and liquidity. The consequences are less trade and less informative prices; neither effect is positive.

Although the news seems to be one that will be met with delight by professional investors [and annoyance by politicians], there is more in store using our computational evolutionary finance approach. The model allows studying other modes of taxation such as a levy on market orders but not on limit orders or the introduction of a progressive tax scheme with an annual deduction.

**Computational issues**

Access to high-performance computing resources is a prerequisite for carrying out this type of experiments. The model sketched above has two main features which consume computing time. The first one is the order book, which, to capture the usual price-time priority of orders, must be represented as a tree with a branch for each stock price, followed by a sub-tree consisting of the orders submitted at that
price. Since the price has priority over order submission time, the priority of orders at one price is independent of their priority at other prices. One can therefore speed up the order book maintenance by implementing it as an array of sub-trees, one for each price.

The second computationally costly feature is associated with the genetic programming algorithm (GP) which is used to evolve trading strategies and solve the model. There are many variants of GP algorithms, we use a steady-state algorithm with tournament selection, which works along the following lines:

1. **Initialization.** Randomly generate a population of 10,000 trading programs.
2. **Trading.** Traders are randomly and repeatedly called upon to submit orders which are handled by the stock exchange.
3. **Tournament.** Randomly select 4 programs and rank them by (discounted) wealth.
4. **Reproduction.** Replace the two poorest programs by copies of the two wealthiest ones.
5. **Crossover.** With some given probability, swap random subsets of program instructions between the two copied programs.
6. **Mutation.** With some given probability, replace a randomly selected instruction in some copied program by a randomly generated instruction.
7. Go to 2.

The algorithm has a few additional features that are worth mentioning: (a) In step 3, traders are selected by first choosing the size of a local neighbourhood of traders, and then randomly selecting traders for the tournament from that neighbourhood. This form of local competition is known to improve the ability of the algorithm to generate innovative and potentially superior behaviour; (b) In step 2, a machine code representation of the programs is used for fast computation of trading decisions; and (c) in steps 5 and 6, a fixed-length byte code representation is used to simplify the genetic recombination. A fast built-in compiler translates byte code to machine code which is then typically used thousands of times before it needs to be recompiled. Our software also contains a byte code disassembler which generates C code that can be used for visual inspection of the evolved programs.

**Parallel processing**

Another feature of the GP model discussed so far is a tight interconnection between the individual agents of the population. It arises because the traders do business with each other on one common marketplace. This type of model therefore does not lend itself easily to parallelization due to the large amount of information that would have to be exchanged between the traders and the marketplace.  

On the other hand, engineering applications of GP are well-suited for parallelization. That includes financial engineering applications such as risk management and computerized trading systems. Developing such systems typically involves solving some ill-structured optimization problem, which is a natural habitat for GP. To solve these types of problems, one can use a large number of autonomous sub-populations deployed on separate worker nodes, and let them exchange good candidate solutions with their neighbours from time to time. A master node can be added to collect information about the progress of the worker nodes and perform other administrative tasks. Such a structure eliminates waiting time on the worker nodes and yields a high capacity utilization rate for the whole system.

**Lessons learned**

Our research applies the new field of computational evolutionary finance to study the effect of short sale bans and transaction taxes on financial market stability. The approach provides an unparalleled detailed model of order book markets and offers new insights into their dynamics. Neither financial transaction taxes, nor the emergency actions taken in 2008, and more recently in 2011, by imposing short sale bans, are capable of delivering what politicians desperately seek; calm and quiet markets. In the figurative nutshell: Short-sellers and active traders are not black sheep – no matter what self-declared white knights want to make believe.

**Outlook**

The approach combines natural selection with models of market interaction. But its potential for applications goes well beyond the specific model of financial markets presented here. Our current research looks at disparate issues such as the pricing and hedging of exotic options, and the impact of market fragmentation on the evolution of the market ecology and traders’ investment skills.

---

1. In real markets, this information flow is routinely handled by exchanges and other market venues. For example, the Chicago Mercantile Exchange processes some 13 million contracts per day. By comparison, the exchange in our model handles an order flow of that magnitude in less than 20 seconds of wall clock time. Obviously, that is possible only because all communication takes place on the same chip.
Protons and neutrons typically enjoy a peaceful coexistence, bound together in the nuclei of all matter. In the nuclear shell model, which has been very successful in explaining how nuclei are structured, there is an assumption that this peace comes with a certain level of detachment: each proton or neutron is an independent spirit, moving on its own within the collective force (or mean) field they generate as a whole. Deviations from an independent motion may reveal important features of how nucleons communicate under both normal and more extreme conditions. In a very recent article “Quenching of Spectroscopic Factors for Proton Removal in Oxygen Isotopes” that appeared in Physical Review Letters, Dr. Gaute Hagen and his colleagues attempt at understanding the role of correlations in very neutron rich nuclei, particularly in terms of the underlying nuclear forces, so that they can probe the limits of stability of matter.

Correlations play a crucial role in physics, and science in general. A proper understanding of correlations conveys important information about the underlying laws of motion. In quantum mechanical systems, the concept of independent particle motion has played, and continues to play, a fundamental role in studies of complex many-particle systems. Within such a picture, the various constituents in a complicated many-particle system are assumed (as in the nuclear shell model) to move in an average mean field set up by the other interacting particles. When scientists observe any kind of deviation from that model, they expect those observations to reveal important features of both the structure and the dynamics of a many-particle system. Unfortunately, traditionally there have been rather few such “observables” from which they can extract clear information. The consequences of correlations in many-particle systems are very difficult to measure experimentally and to interpret theoretically.

Working with Øyvind Jensen and Morten Hjorth-Jensen from the University of Oslo and B. Alex Brown and Alexandra Gade of Michigan State University, Gaute (an adjunct assistant professor of physics and member of the Oak Ridge National Laboratory Physics Division), studied the chain of oxygen isotopes in a sophisticated numerical simulation, made possible by state-of-the-art high performance computing facilities at both the National Center for Computational Sciences at ORNL and the Notur project in Norway. This work also enjoyed support from the UNEDF SciDAC project.

As nuclear physicists, the group is keenly interested in understanding the role of correlations, particularly in terms of the underlying nuclear forces, so that they can probe the limits of stability of matter. This takes their studies into the territory of nuclei at high densities and temperatures, as well as into the realm of nuclei rich in neutrons or protons, the so-called “dripines,” where adding just one more nucleon robs a nucleus of the stability it needs to stay intact. For years, nuclear theorists and experimentalists have worked together to understand the strongly correlated behavior of protons in the neutron-rich nuclei at the limit of the nuclear chart. For these nuclei, experiments suggest that the deeply-bound protons behave less independently as more neutrons are added to the system.

In a typical quantum mechanical fashion, the research team’s simulation of oxygen isotopes allows selected neutrons to leave the nucleus as free particles and at the same time stay inside to occupy bound orbitals. The results demonstrate a surprising, intimate relationship between the enhanced correlations of the outnumbered protons and the freedom of the abundant neutrons to partly escape from the nucleus: if the neutrons are forbidden to enter the free continuum, the protons will fight harder for their own independence. In Figure 1 the proton knockout of the various oxygen isotopes are shown together with the influence of the continuum on the neutrons in the nitrogen nuclei.

While further studies will be necessary, this work provides researchers with new insight into the dynamics of many-particle systems.

More information

- National Center for Computational Sciences www.nccs.gov
- The Notur Project www.notur.no
- UNEDF www.unedf.org
Figure 1: Proton knockout from Oxygen-16, Oxygen-22 and Oxygen-24. The thickness of the arrow showing the removed proton reflects the spectroscopic factor for this process, or in other words, how "free" the protons can be considered in neutron rich oxygen isotopes.
Computational science plays an increasingly important role in finding solutions to challenging problems in chemistry, and research in chemistry has been significantly changed since August Comte in 1830 stated that “every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational. If mathematical analysis should ever hold a prominent place in chemistry [...] it would occasion a rapid and widespread degradation of that science”. Instead, computations are today an integral part of virtually all branches of chemistry.

Since the start of the first supercomputing program in Norway, computational chemists have been using large fractions of the available computational resources. In the early periods, the computational methodology and available supercomputing resources only allowed very small molecules to be studied, limiting the use of computational methods to theoretical chemists. Today, realistic computational models can easily be built using both in-house and commercial software packages, accessible to all chemists.

Because of the developments in chemistry software and in HPC resources, computational chemistry still accounts for about 30% of the use of supercomputing resources in the NOTUR program. This has to a large extent been fueled by the increasing use of computational methods by experimental chemists, who obtain from these computations insight that cannot easily, if at all, be extracted from experimental observations alone.

The Norwegian supercomputing programs, including the NOTUR programs, have allowed Norwegian theoretical and computational chemistry groups to reach a high international level. Indeed, the 2008 evaluation of basic chemistry research in Norway concluded that: “the Norwegian activity in Theoretical and Computational Chemistry is good, and in some cases considerably better than that”. This would not have been possible without good access to state-of-the-art HPC resources.

The diversity and vitality of computational chemistry in Norway is well illustrated in this issue of META. The contribution by Prof. Helgaker and his coworkers highlights the development of new novel computational methods, allowing for the study of molecules in neutron stars where the magnetic fields are so strong that they cannot be realized in a laboratory. Understanding the chemistry of these stars is thus not possible without computation.

Prof. Jensen shows how the clever design and exploration of molecular reactivity through computation allows us to go beyond chemical intuition, leading both to a better understanding of how we think about chemical reactivity and how catalysts work, but in a few cases also leading to the prediction of more efficient catalysts.

Prof. Kjelstrup and her co-workers demonstrate how computational studies of dynamical processes allow us to gain new insight into highly complex systems, such as how the calcium pump works and insight into the formation of gas hydrates, the latter of high relevance in the processing of natural gas.

The contribution of Prof. Brandsdal demonstrates how computational chemistry may add value to the development of new drugs by reducing the cost, time and efforts in experimental synthesis. This is achieved through the use of computations for obtaining detailed mechanistic insight into the processes that govern the efficiency of potential drug candidates.

Even though these are all impressive examples of the powers of computational chemistry, this is not the end. Computational chemists will continue to develop more and more realistic models, requiring increasing amounts of computer time. Through a continued focus on HPC infrastructure and support, Norway can continue to hold an internationally leading position in the field of computational chemistry.
The United Nations have declared 2011 as the International Year of Chemistry (IYC2011)

The Norwegian Chemical Society is responsible for the organization of the International Year of Chemistry 2011 in Norway, and this work is led by Prof. Einar Uggerud from the University of Oslo.

The goal of the IYC2011 is to celebrate the achievements of chemistry and its contributions to the well-being of humankind. Under the unifying theme “Chemistry - Our life, Our future”, IYC2011 will in particular highlight the importance of chemistry for reaching the Millennium-goals of the UN, such as sustainable energy for all, clean water and the eradication of deadly diseases such as HIV/AIDS and malaria.

The IYC2011 will also celebrate the 100th anniversary of the Nobel Prize in Chemistry to Marie Sklodowska-Curie and the role of women in research.

In Norway, the main target for many of the activities has been children and youth, for instance by providing training programs for elementary- and secondary-level school teachers.

Increasing the appreciation for chemistry in the general public has also been a goal, and the Science Days 2011 (Forskningsdagen 2011) contributed to this through its selection of chemistry as the main topic for the events organized during these days.

For more information about IYC2011, visit www.kjemi.no/iyc2011 or www.chemistry2011.org
Molecules in Strong Magnetic Fields

In quantum chemistry, we apply the methods of quantum mechanics to the study of chemistry. This field of chemistry can trace its origins to the early decades of the twentieth century, when it was discovered that molecules consist of positively charged, heavy nuclei and negatively charged, light electrons in relative motion, governed by the laws of quantum mechanics.

At the same time, it was also recognized that the underlying many-particle problem was computationally intractable as even a moderately sized molecule consists of hundreds of interacting particles. Indeed, in 1927, P. A. M. Dirac, one of the founding fathers of quantum mechanics, famously stated that [1] “The underlying physical laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known and the difficulty is only that the exact application of these laws leads to equations that are much too complicated to be soluble.”

However, at the time Dirac could not foresee the spectacular emergence and development of electronic computer in the second half of the twentieth century, which revolutionized much of chemistry by making molecules amenable to accurate quantum-mechanical simulations. As a result of Moore’s law, first-principles simulations of chemical systems and processes have now become commonplace and are today being performed by non-specialists more often than by specialists, in support of experimental activities and measurements. Indeed, a perusal of the most general journal of chemistry, Journal of American Chemical Society, reveals the ubiquity of computation in modern chemistry: about 40% of its articles are today supported by computation, mostly quantum mechanical. At the Department of Chemistry, University of Oslo, about one third of the scientific staff have authored publications supported by quantum-chemical simulations. This is an astonishing transformation of a science that only a few years ago was considered to be archetypically experimental and empirical in nature. Nowadays, computation is an integral part of chemistry and is widely perceived as the “third way”: simulations not only play an important role in the interpretation and prediction of experimental observations, they are in fact more and more often viewed as an alternative to experimental work, which may be dangerous, expensive, difficult or even impossible to carry out.

For example, in molecular clouds of the interstellar medium, a wide variety of molecules have been observed by spectroscopic techniques. Many of these molecules are too reactive to be studied in laboratories on earth, making their identification and characterization impossible by experimental techniques alone. In such cases, advanced quantum-mechanical simulations may be performed on a set of candidate molecules and their identity confirmed by comparing observed and simulated spectra. Once identified in this manner, further calculations may be carried out to establish the chemistry and properties of the detected molecules.

While the molecules inside interstellar molecular clouds exist under conditions of extreme low pressure and temperature, entirely different conditions are experienced by atoms and molecules in the atmospheres of fast-rotating compact stellar objects such as white dwarves and neutron stars, whose magnetic fields may be as strong as $10^5$ and $10^8$ T, respectively – that is, orders of magnitude stronger than the fields generated by the earth (typically 30–50 μT) or in laboratories (up to about 100 T). In such extreme magnetic fields, chemistry changes and becomes unfamiliar [2]. This ‘alien’ chemistry can only be explored by performing simulations, using the advanced methods of quantum chemistry developed since the 1920s. Such simulations provide information that may help in the interpretation of spectra from these stellar objects. Particularly relevant is the study of molecules consisting of hydrogen and helium, by far the most abundant elements in the universe.

We have recently undertaken an extensive quantum-chemical study of atoms and molecules in strong magnetic fields, for which a special computer program LONDON has been developed in our group [3]. Previous such simulations have either been carried out for two-electron systems or at lower levels of theory, revealing a less detailed picture of molecules in
FIELDS

strong magnetic fields. In particular, our calculations are unique in that they are not restricted to special orientations of the field relative to the molecules, being rigorously gauge-origin invariant. In the following, we describe some results of our investigations.

Our first simulations concern the changes that occur in the electronic energy as a magnetic field is applied to closed-shell molecules – that is, to molecules that have only paired electrons and therefore possess no permanent magnetic moment [4]. According to Lenz’s law, the applied external field induces a net rotation of the electrons about the direction of the field, thereby increasing their kinetic energy and generating a local magnetic field that opposes the applied field. As a result of this increased electronic energy, the molecule escapes from the field to lower its energy. Most closed-shell molecules behave in this classically expected diamagnetic manner but a few behave non-classically, entering into the field rather than escaping from it. In these paramagnetic closed-shell molecules, the induced rotation reduces the kinetic energy of the electrons and enhances the applied field rather than opposing it.

However, as we found numerically with our LONDON code, all such paramagnetic molecules become diamagnetic at a critical field strength \( B_c \), characteristic of each molecule [4] – see Figure 1, where we have plotted the energy of the cyclic \( \text{C}_{20} \) compound in a field perpendicular to the molecular plane. We here and elsewhere use atomic units, in which the magnetic field is measured in units of \( B_u = 2.35 \times 10^6 \) T and the energy in units of \( E_u = 2625.5 \) kJ/mol. As illustrated in this figure, the energy of \( \text{C}_{20} \) has a local maximum at \( B = 0 \) and a double global minimum at \( B = \pm B_c = \pm 0.013 B_u \). As the magnetic field is slowly turned on, the electrons begin to rotate against the field, reducing their energy paramagnetically until rotation comes to rest at the critical field strength \( B_c \). As the field is increased further, rotation occurs in the opposite direction, increasing the energy diamagnetically rather than reducing it. Previously, molecular paramagnetism had only been studied perturbatively, by expanding the energy in low orders in \( B \) around zero field strength, not revealing the full picture. Guided by our numerical simulations, we were able to construct an analytical model for the diamagnetic transition, predicting that it should be observable in laboratories for large paramagnetic closed-shell molecules such as \( \text{C}_{20} \text{H}_{20} \) [4].

The transition from paramagnetism to diamagnetism discussed above has not yet been detected experimentally but should nevertheless be observable in laboratories. We now turn to molecules of astrophysical interest, in fields far stronger than those obtainable in laboratories. We begin with \( \text{H}_2 \), the most abundant molecule in the universe, consisting of two protons and two electrons.

Like all molecules, \( \text{H}_2 \) exists in [infinitely many] discrete electronic states of different energy, between which it may perform transitions accompanied by the emission or absorption of radiation. The state of lowest energy is called the ground state; all other states are said to be excited. To understand the interaction of molecules with radiation in stellar atmospheres, it is important to know these states as precisely as possible. Here, our interest concerns the evolution of the electronic states with increasing magnetic fields [5].

The most important electronic states are those of lowest energy. In \( \text{H}_2 \), we consider two states: the lowest singlet and the lowest triplet state. These states differ in the arrangement of the spins of the electrons. Spin is a property of the electron that may be thought of as rotation about its own axis. Each electron may rotate in one of two ways: spin up (alpha spin) or spin down (beta spin). Moreover, since the electron is negatively charged, its spin generates a magnetic moment. In the singlet state of \( \text{H}_2 \), the two spins are oppositely directed and the electrons have no net magnetic moment; in the triplet state, the spins are parallel and the electrons have a net magnetic moment.

In Figure 2, we have plotted the energy of the lowest singlet state (red curves) and the lowest triplet state (blue curves) of \( \text{H}_2 \) as a function of the internuclear separation \( R \) (in units of \( a_0 = 52.9 \) pm) and in the two magnetic fields \( 0B_u \) and \( 2.25B_u \). A minimum in such a potential energy curve at a given distance \( R \) shows that the molecule is stable at this bond distance; the absence of a minimum shows that the molecule is unstable, dissociating into separate atoms to reduce its energy.

In the absence of a magnetic field, the singlet state (red curve) is the molecular ground state, with a minimum at \( R = 1.4a_0 \), whereas the triplet is dissociative and higher in energy. In the strong field \( B = 2.25B_u \), the situation is reversed and the triplet state has now become the ground state.

**Fig. 1.** The energy of the cyclic \( \text{C}_{20} \) molecule in a magnetic field perpendicular to the molecular plane.

**Fig. 2.** Potential energy curves of \( \text{H}_2 \) in the lowest singlet state (red) and the lowest triplet state (blue). Full lines represent the energy with the molecule oriented parallel to the field; dashed lines represent the energy with the molecule oriented perpendicular to the field.
state. In both states, the electrons rotate diamagnetically, raising their energy but in the triplet state this increase is more than compensated for by the alignment of its net spin magnetic moment with the field, lowering the energy of the electrons. No such lowering occurs for the singlet state, which has no net magnetic moment.

In a magnetic field, the energy of a molecule depends also on its orientation relative to the field. In the singlet state, the energy is lowest when the molecule is parallel with the field (full red line) and highest when it is perpendicular to the field (dashed red line). The shaded area between the full and dashed line represents orientations intermediate between parallel and perpendicular. We also note that the singlet H₂ molecule is shorter in the field than outside the field (in particular, in the perpendicular orientation) and that the energy minimum is deeper in the field, generating a more stable molecule. These changes can be rationalized by noting that the atoms from which the molecule is formed become smaller in a magnetic field (following the induced rotation) and slightly elongated in the direction of the field.

By contrast and somewhat counter-intuitively, the triplet H₂ molecule has its lowest energy in the perpendicular orientation (dashed blue line) rather than in its parallel orientation (full blue line). Interestingly, this triplet state now has a shallow minimum of about 38 kJ/mol at a bond distance of 1.75\(a_0\), unlike the dissociative triplet state in the absence of a field. Previously, only the parallel orientation of H₂ had been investigated and the molecule was assumed to be dissociative at this field strength. Only with the development of the LONDON program has it been possible to explore all orientations reliably and to identify H₂ as a bound molecule also at these field strengths.

The second most abundant element in the universe is helium. It is an elementary fact of chemistry that helium and all other noble gas atoms such as neon and argon do not form bonds and are therefore only very loosely attached to each other, by weak dispersion forces. Indeed, it is their inability to form chemical bonds that has given these atoms their name: noble gas atoms. However, our calculations have shown that the situation is very different in strong magnetic fields.

In Figure 3, we have plotted the potential energy curves of the helium dimer in various magnetic fields, in the three possible spin states of this molecule: the lowest singlet state (to the left) with no net spin magnetic moment, the lowest triplet state (in the middle) with a net spin magnetic moment equal to that of H₂, and the lowest quintet state (to the right) with a net spin magnetic moment twice that of the triplet state. As for H₂, in Figure 2, full and dashed curves correspond to parallel and perpendicular orientations in the field, respectively.

In the absence of a magnetic field, the singlet state He₂ is lowest in energy but only weakly bound. As the field increases, the singlet He₂ increases in energy, becoming smaller and more strongly bound (in the perpendicular orientation). However, this singlet state is no longer the ground state in strong magnetic fields. In fact, in the magnetic fields 0.5\(B_0\) and 1.0\(B_0\), the very strongly bound triplet state is the ground state (in the parallel orientation), whereas the less strongly bound quintet state becomes the ground state in fields stronger than 1.0\(B_0\) (in the perpendicular orientation). The helium atom is thus not so ‘noble’ in magnetic fields, forming strong bonds with other atoms.

Our calculations are the first studies of the helium molecule in strong magnetic fields, revealing a complicated evolution of the electronic states with increasing magnetic field strength. We will continue to explore the chemistry of these and other species in magnetic fields. Such calculations reveal a fascinating chemistry, different from that experienced on earth, where noble gas atoms form chemical bonds and molecules align themselves in unexpected ways relative to the field. In fields even stronger than those explored by us, chemistry becomes stranger still, molecules forming long chains of atoms parallel to the field, with all electrons in beta spin and all orbitals singly occupied.

**Bibliography**

2. D. Lai, Rev. Mod. Phys. 73, 629 (2001)
MODELING OF CHEMICAL REACTIONS AND CATALYSIS

Quantum mechanical calculations can give detailed insight into the whereabouts of the particles taking part in a chemical reaction, insight which is not readily accessible from experiments. However, due to the computational challenge involved quantum chemists have always had to strike compromises with respect to the size of the chemical systems studied and the accuracy of the methods applied. The good news is that, with the computational methods and hardware of our time, it is possible to make realistic simulations of a broad range of reactions and to help answer key questions of contemporary chemistry and biology. An example is catalytic reactions for which computational methods to an increasing extent are used also in the development of new and improved catalysts.

Chemical reactions and how to follow the electrons
Molecules are held together by bonds between the atoms and chemical reactions are, in general, characterized by the rupture of one or more of these bonds in the reacting molecules, followed by the formation of one or more new bonds to give the reaction products. The chemical bonds are made up of (pairs of) electrons. In other words, chemical reactions bring about changes in the electronic structure of the molecules involved and molecular-level simulations of such processes need to take account of the electrons, at least those directly taking part in the reaction. This is the reason why methods for such simulations are based on quantum mechanics and not classical (Newtonian) mechanics in which the atoms are the smallest units. The various methods of quantum chemistry are designed to arrive at approximate solutions to the Schrödinger equation and computational quantum chemistry may offer microscopic insight into properties determined by the electrons, be it the electronic transitions between energy levels in atoms and molecules, the band structure in semi-conducting solid materials, or the reactivity of the active center in an enzyme.

Moving in the energy landscape
Computational investigation of chemical reactivity faces particular challenges. Whereas most studies of molecular properties involve calculations on equilibrium geometries (i.e., the most stable geometries) only, studies of chemical reactivity and reaction mechanisms require consideration also of molecular structures far from their stable arrangements, including short-lived intermediates, that is, shallow minima, and transition states (first-order saddle points) along the reaction pathway on the multidimensional energy landscape (the potential energy surface (PES)), dramatically increasing the complexity of the problem; see Figure 1.

The study of reacting molecules involves, in addition to solving (approximately) the Schrödinger equation repeatedly to give the potential energy, optimization to reach minima and transition states on the PES, or in some cases even the tracing of the complete reaction path from the minimum of the reactants, via the transition state, to reach the minimum of the product geometries (Figure 1). Even if all of this may seem (and often is) challenging, the possibility to study any given point along a reaction also represents one of the most important strengths of quantum chemistry, the application of which offers results
which nicely complements the information typically obtained from experiments. In other words, computational quantum chemistry is particularly suited for studies of what much of chemistry really is about: chemical reactions.

Most of the CPU time invested in studies of reactions goes into geometry optimization of minima and transition states and their connecting pathways, whereas subsequent (often more accurate) calculations of the properties of these stationary points usually account for much less of the overall computer time. The same is true for the human time invested. In most cases, it still takes a lot of work, trial and error, and quite a bit of “chemical intuition” to map out the most probable pathway of a chemical reaction. There are usually a fair number of possibilities [the energy landscape is everything but simple] and each of these possibilities must be tested explicitly by model building the corresponding molecular structures. Next, the calculated energies will tell whether a given reaction route can be discarded as improbable or not. This is an important point since computations alone never can prove a reaction pathway. We can, however, by careful and systematic work, discard certain routes as improbable (due to the high energies of the transition states involved) and point at others as more probable and viable. Combining such computational results with experimental information it is sometimes possible to go a long way towards actually proving a reaction mechanism.

Early studies of reactions

Early computational studies of chemical reactivity concentrated on reactions of relatively small organic (i.e., carbon-based) molecules and were usually based on semi-empirical and highly approximate solutions to the Schrödinger equation. The conclusions that could be drawn from such calculations were typically qualitative rather than quantitative, but were still often surprisingly useful. The Woodward-Hoffmann rules (first published in 1965) for the expected structural outcome of a sub-class of ring-opening and ring-closing reactions, for which the Nobel Prize in chemistry was awarded in 1981, represent an example in this respect. These and other rules derived from quantum chemical calculations and considerations are not only helpful in explaining why chemical reactions proceed as they do; they are also highly predictive and useful for experimental chemists targeting new compounds and reactions.

Contemporary studies of reactions and catalysis

In the decades since the seminal work of Woodward and Hoffmann, we have seen dramatic improvements in both the quantum chemical methods and the computer hardware. These dual developments have spurred a move to less approximate, more sophisticated and more accurate methods. Perhaps equally important has been a simultaneous and tremendous spread in the kind of chemical systems, in terms of both the number of atoms and kind of elements involved, that are subjected to computational investigation. Whereas much of the argument in Woodward and Hoffmann’s work was based on simple and qualitative calculations on organic molecules containing 10–20 atoms, carefully prepared computational studies may now yield results [e.g., activation energies] that can be compared with the experimental counterparts even for enzymatic reactions. A typical study of this kind may involve a quantum mechanical [QM] treatment of the active site of the enzyme counting up to a few hundred atoms (see Figure 2 for an example of a QM model of an active site), sometimes with parts of the surrounding protein, described by classical molecular mechanics [MM], being coupled to the QM region as is the case in the so-called QM/MM methods.

Enzymes mediate chemical reactions in the organism, and computational studies may offer fundamental insight into the workings of these biological molecules as well as contributing to development of pharmaceutical compounds. Their “cousins”, the non-biological catalysts (a catalyst is a compound which increases the rate of a chemical reaction without being consumed in the process) are widely applied to make industrial processes work efficiently, with an appreciable rate, low energy consumption, and little waste and side-products. It is hard to overestimate the importance of catalysts for the modern society and it is not surprising that catalytic reactions, biological as well as industrial, have been the subject of numerous computational studies over the last couple of decades and the importance of molecular-level simulations is still increasing rapidly in catalysis. Also in Norway are such studies performed in several groups, and they consume a fair share of our national computational resources.

Most of the time results from such computational studies are less predictive than Woodward and Hoffmann’s rules, and their effect on the chemical community is more indirect and manifests itself by influencing the way chemists think about their reactions. Calculated reaction mechanisms of existing catalysts represent a typical example of such results. The computational studies thus play a crucial role in explaining working chemistry. However, there are still relatively few examples where theory has contributed directly, say, to development of new and better catalysts.

In Bergen we are actively exploring the
potential of using molecular-level calculations directly in catalyst development. In these studies we start from computationally obtained mechanistic information (key intermediates or transition states) and correlate calculated molecular properties of a large number of catalysts with their corresponding calculated or experimental activities or selectivities. In other words, we work to obtain a clear understanding of the underlying factors governing the activity or selectivity of the catalysts. Such correlations are most efficiently uncovered by constructing multivariate quantitative structure–activity relationships (QSAR) models. Once established, a QSAR model also considerably speeds up the subsequent computational testing of candidate new catalysts and we have used such models in prediction of transition-metal based catalysts which subsequently have been synthesized either in our group or in other research groups.

Outlook

Thus, in our case the need for high-performance computing resources is not only connected to the use of more accurate computational methods and larger and more realistic chemical model systems, but also to the sheer number of calculations we perform and analyze, semi-automatically, in the screening-type prediction procedures (see Figure 3). Ultimately, it should also be possible to fully automate and integrate all the computational parts of the development (see Figure 3), and, in a project funded by the evITA program of the Norwegian Research Council, we are working with the group of Prof. Bjørn K. Alsberg, Department of Chemistry, NTNU, to achieve this. Once fully operative, such a scheme for in silico development should represent a very powerful tool for design of catalysts and other functional compounds. However, as the scheme is based on calculating the fitness (the scoring function) of large numbers (into the millions) of individual molecules, it also has the potential of swallowing large amounts of computer time, and, in general, compromises between the available computer resources and the quality of the fitness will have to be drawn. A far cry from the “back-of-the-envelope” calculations of Woodward and Hoffmann, but certainly a worthy goal if the resulting machinery turns out to be as useful in prediction as their rules have been.

Literature


Acknowledgments

The author would like to thank current and previous group members (in particular Giovanni Occhipinti, Elaine Olsson, Wouter Heyndrickx, and Manuel Sparta) and colleagues and collaborators (in particular Knut J. Barve, Bjørn K. Alsberg, Aurora Martinez, Knut Teigen, and Hans-René Bjørsvik) who have contributed to the work mentioned here, and also greatly acknowledges funding from a range of programs of the Norwegian Research Council: KOSK (projects 160072/V30 and 177322/V30), GASSMAKS (182536/I30), NANOMAT (158538/S10), FORNY (203379/I30), and eVITA (205273/F30), and also from the Nanoscience program of the University of Bergen. Finally, this research depends on computing resources made available via the NOTUR supercomputing program.
PREDICTING HYDRATE STABILITIES

Computers can be used to predict stability of gas hydrates, of importance for example for the processing of natural gas. Computer calculations give molecular insight for new process ideas.

Gas hydrates

Gas hydrates are the crystalline ice-like structures, which consist of water and some guest gas molecules (Fig. 1). The water molecules form cages, in which the guest gas molecules are "trapped". The presence of the gas molecules stabilizes the cages, which are otherwise unstable and will collapse to the liquid water or to an ice structure. Gas hydrates are found in the sea bottom and are a potential source of energy, since the trapped guest molecules are the typical components of the natural gas: methane, ethane, etc [1]. Gas hydrates are also found in outer space (in comets), and are important wherever water exists. Knowledge of hydrates as absorbents is important for several reasons. We concentrate here on their relevance for gas technology.

Many different gases can form hydrates, and carbon dioxide is also one of them. Thus, hydrates can serve as a potential storage of carbon dioxide. The question of trapping hydrates for hydrocarbons is of big interest, because the amount of gas is large. But this trapping cannot occur without destroying the network. An idea is then to exchange methane with carbon dioxide, and use the hydrate to store the greenhouse gas.

To explore this idea, means to acquire extensive thermodynamic data of types of gas hydrates. This can be done favorably by computer simulations. Such simulations can give knowledge of typical thermodynamic conditions, when this procedure would be possible and the most efficient way to actually do it.

Data from molecular dynamics simulations

Until very recently the only source for thermodynamic data for gas hydrate structures were obtained in the laboratory. Such experiments are however very costly and require a lot of time, in particular because it is difficult to predict how and when hydrate forms. By doing molecular dynamics simulations data can be performed in a relatively cheap way. Following the equations of motion for the molecules (Molecular Dynamics) or sampling the phase space according to a statistical distribution (Monte Carlo) one resembles the real thermodynamic properties. Our aim has been to develop these techniques for hydrates, by calculating

- the cage occupancy (the loading)
- the Helmholtz energy

The loading is the number of guest gas molecules per unit cell. The maximum value is 8, while the hydrate could contain less than 8. The adsorption isotherm presents the loading curve versus the external pressure. The stability is determined by the Helmholtz energy in the case of hydrates. The lower this energy is, the more stable is the compound. The task becomes therefore to calculate the Helmholtz energies for different hydrate loadings and compare them.

Cage occupancy

Figure 2 shows the cage occupancy as a function of pressure in the case of methane, ethane of carbon dioxide.

Figure 2 shows how ethane, carbon dioxide and methane molecule are filling up a hydrate cage structure. The hydrate has two types of cages: 6 large ones and 2 small ones per unit cell. The total number of cages is the same as the maximum number of the guest molecules per unit cell. Only one of these molecules can be trapped per cage. As the pressure increases, the
amount of trapped molecules in the whole hydrate increases as well.

Consider first methane and carbon dioxide. The methane molecule is small and symmetric, while the carbon dioxide molecule is quite large and has a dipole form. The size of the molecules and the cages are such that both methane and carbon dioxide molecules can easily fit into the large cage, but the large carbon dioxide molecule does not prefer to go into the small cages. This means that, while at small pressures the methane molecules distribute evenly among all the cages, the carbon dioxide molecules prefer to go only to the large cages (compare the two lower parts of Fig. 3). The small cages start to be filled in by the carbon dioxide only after all the large cages are already filled.

**Helmholtz energy variations**

There are several ways to compute the Helmholtz energy of a hydrate, and in our case the simplest one is by using the adsorption isotherm [2]. It is known, that the Helmholtz energy depends on the number of the molecules and the chemical potential. The amount of molecules present in the hydrate is found for a specified chemical potential. In Monte-Carlo simulations, this is done with the help of imaginary gas, which is being in equilibrium with the hydrate. The chemical potentials of the phases in equilibrium are the same, and one can therefore control the properties via the imaginary gas pressure and temperature. After the adsorption isotherm is determined, the Helmholtz energy is found by integration of this curve.

This method predicts that the Helmholtz energy of a carbon dioxide hydrate is always less than the one of the methane hydrate at the same loading. This is illustrated in Fig. 3. The figure shows that the ethane hydrate is more stable than both the others. But the carbon dioxide hydrate is more stable than the methane hydrate, which gives the hope for the original idea. Indeed, if the methane hydrate already exists at specified condition at the sea bottom, it means that the carbon dioxide hydrate will a fortiori exist at these conditions.

**A process for methane - carbon dioxide exchange**

Figure 3 does not tell us anything about how to convert the methane hydrate into the carbon dioxide hydrate, or which thermodynamic path [process] to follow [3]. To answer this question we need to study mixtures. In studying mixtures, in addition to the question of stability, the question of selectivity becomes important. Namely, it is important to know, how many molecules of each of the components we have at the specified conditions.

The cage-filling properties of Fig. 2 are then central. At low pressure, the amount of methane and carbon dioxide present in the hydrate is approximately the same: the large carbon dioxide molecules go to the large cages, while small methane molecules fill both small and large cages. In contrast, at large pressures, when all the large cages are filled in, it is mostly methane, which likes to get into the hydrate. It is hard to force the carbon dioxide molecules to get into hydrate at these conditions.

This suggests a thermodynamic path to convert a methane hydrate into a carbon dioxide hydrate. As at large pressures carbon dioxide does not like to go to small cages, so one needs to decrease the pressure to the level, where some of the large cages are available. Then, one has to gradually exchange the molecules, and at the end, when no methane is present in the mixture anymore, increase the pressure such that carbon dioxide fills the rest of the cages. In the absence of the other easy ways to decrease the Helmholtz energy (there is no methane any more), all the carbon dioxide molecules will fill the small cages as well. More data are needed to substantiate this idea, however. In particular, the reference that is now used (an empty network) must be replaced by properties of real water and gas.

**Acknowledgement**

This work has been carried out with the support of the Statoil-Vista program on enhanced oil recovery.

**References**

When you move a muscle, trillions of myosin motors are moving actin filaments in the muscle fibre. The myosin motors must be able to attach to the actin and this is only possible as long as Ca\(^{2+}\)-ions are present and bound to troponin proteins on the actin filaments. The Ca\(^{2+}\)-ions are stored in the lumen of sarcoplasmic reticulum (SR) and released during movement of the muscle. In order for the muscle to relax when the movement is completed, Ca\(^{2+}\)-ions must be moved back into the sarcoplasmic reticulum. This is accomplished by the Calcium-pump (see Figure 1), which is P-Type ATPase (it uses ATP as fuel for moving the ions) and it is essential for you being able to move.

Modelling of the protein
Since its discovery in the 1960’s, the pump has been extensively studied experimentally. Today, the vast kinetic and mutagenic data available has positioned the pump as a prototype for the P-Type ATPases. Conventionally, the Calcium-pump is described by kinetic models but experimental studies have shown that the pump releases heat during operation [2] and that the fraction of the energy from the ATP-hydrolysis released as heat can be regulated by the pump. This is ignored in the kinetic models of the pump. However, this shortcoming can be lifted by applying mesoscopic non-equilibrium thermodynamics (MNET) which places the description of the pump on a sound thermodynamic foundation [1].

The MNET theory is phenomenological in nature, and transport coefficients are needed for the description.

Recently, several 3D-structures of the pump have been obtained, and this opens the possibility of using atomistic computer simulations (e.g. molecular dynamics simulations) in order to obtain the transport coefficients. We have performed simulations of the protein in a water solution and we have obtained thermal properties [e.g. thermal conductivity] [3].

Atomistic simulations can also be used to study the structural coupling between

**Authors**

Signe Kjelstrup  
Department of chemistry,  
Norwegian University of Technology and Science, Trondheim, Norway

Anders Lervik  
Department of chemistry,  
Norwegian University of Technology and Science, Trondheim, Norway

Fernando Bresme  
Department of Chemistry,  
Imperial College, London, United Kingdom
the protein and the membrane. There is experimental evidence that the activity of the Calcium-pump strongly depends on the nature of the phospholipid membrane. The pump reaches maximum activity (measured in terms of ATP hydrolysis) when the protein is immersed in bilayers consisting of lipid chains containing between 16 and 20 carbon atoms. The activity decreases above and below this number of carbon atoms. In order to advance in the explanation of the factors influencing the pump activity as well as the associated non-equilibrium effects, further work is needed to characterize the structure of the protein-membrane system.

We have analysed the protein immersed in a POPC-model membrane and quantified the structure. A snapshot of the simulated system is shown in Figure 2.

The POPC lipid chain length is such that the CaATPase can operate in the range of maximum activity. We find that the POPC bilayer is thinner near the protein surface and that the POPC molecules close to the protein are tilted. This suggests that membrane deformation and high activity are not mutually exclusive. We also find that very long simulations are necessary in order for the pump to relax in the bilayer. More than 25 ns is required before a stable configuration is reached. This is illustrated by the variation in potential energy of the system as a function of time in fig. 3. We see that the potential energy reaches a stable level after about 25 ns. Still we are far away from the characteristic time for pump operation, which is 0.1 ms.

**Supercomputer time is essential**
A simulation of such a large molecular system on the time scale of 100 ns is not possible without supercomputers. In our case they were performed in parallel, using the GROMACS software package and 96 Xeon 2.66 GHz quad-core processors with infiniband connectivity on the Stallo computer cluster at the University of Tromsø.

**Conclusion and perspectives**
We have studied the Calcium-pump and obtained thermal properties in water solution, and the structure of the membrane. We will now study the protein under non-equilibrium conditions (i.e. a steady temperature gradient) and the protein immersed in bilayers consisting of lipid molecules with chain lengths corresponding to different activities of the pump.

**References**


Antimicrobial peptides represent a versatile class of promising molecules to combat the growing problem of antibiotic resistance. Recent decades have seen a boom in the cases of antibiotic resistance among different bacteria, which unfortunately has not been followed by a similar rise in new classes of drugs to combat them.

Antimicrobial peptides

After the first discovery of cationic (positively charged) antimicrobial peptides (CAPs) in the early 1980s through the identification of cecropin from silk moths and magainin from the claw footed frog, workers in the field have been aware of the unique properties as potential antimicrobial agents [1]. Antimicrobial peptides are diverse molecules typically ranging from 12 to 100 amino acid residues. Antimicrobial peptides are evolutionary ancient weapons, and their widespread distribution throughout the animal and plant kingdom suggests that they have served a fundamental role in the successful evolution of complex multicellular organisms. Antimicrobial peptides are part of the immune system as the first line of defense versus a variety of organisms. The diversity of antimicrobial peptides discovered makes it difficult to categorize them, except broadly based on their secondary structure. The fundamental structural principle found in all classes is the ability of the molecule to segregate hydrophobic (nonpolar) and positive charges (polar) regions in discrete sectors of the molecule (“amphipathic” design). Some peptides only adopt this organization when they enter a membrane, whereupon they assume an amphipathic secondary structure.

The first attempts to bring CAPs into clinical use started more than 20 years ago, but despite significant efforts, to date, no CAP has yet reached the clinic [2]. This reflects the difficulty in making peptide drug, as they are liable to degradation by proteolytic enzymes in our body. In addition, the production cost for making large peptides becomes commercially unacceptable. Lactoferrin is a globular protein found at high concentrations in “first milk”. It is one of the components of the immune system of the body, and has antibacterial activity. It is part of the innate defense. Lactoferrin provide human infants with antibacterial activity. In the early 1990s, the antimicrobial activity of lactoferrin was identified as the N-terminal part, corresponding to an antimicrobial peptide (lactoferricin) resulting from cleavage by pepsin. Based on this, a systematic study was initiated to identify the active components responsible for the antibacterial activity, and several years of study has led to the identification of the minimum antibacterial motif in cationic (positively charged) antibacterial peptides [CAPs] [3-4]. These peptides can be surprisingly small, and pronounced antibacterial activity has been shown in a series of di- and tripeptides [5]. Generally, two units of bulk, each similar to a phenyl group in size and polarity, along with two cationic charges are sufficient to obtain an active peptide [Fig. 1]. Our research group has in collaboration with Lytix Biopharma [11] tried to bring these molecules closer to clinical use by incorporation of non-genetically coded amino acids. One of the compounds has entered clinical phase II trials for topical treatment of infections of multiresistant bacteria.

The key questions we have tried to answer through the use of various molecular modeling techniques are the proteolytic stability of short CAPs and their mechanism of action.

Proteolytic stability of antimicrobial peptides

One of the key challenges antimicrobial peptides face when administered orally is the digestive enzymes in the gastrointestinal tract. These enzymes normally break down peptides and proteins into their smaller building blocks (amino acids) to facilitate absorption by the body. Obviously, oral activity requires stability towards degradation by these enzymes. In order to examine the proteolytic stability, and thereafter potentially redesign CAPs, we have examined how CAPs interact with the key enzymes found in the gastrointestinal tract.
tract using computational docking and scoring algorithms. Docking refers to a technique in the field of molecular modeling that tries to predict the best orientation of one molecule relative to another in the bound state to form a stable complex. This knowledge is then used to predict the strength of binding using scoring functions. Most docking programs are inherently parallel in the sense that one can submit a subset of molecules to different processors for docking and scoring to one target molecule. Many different conformations can be generated for each molecule if we allow fully flexible docking, and the calculations can require evaluation of millions of complexes. All our calculations are done using the Schrödinger program package at the supercomputer facility in Tromsø (Stallo). We have carried out such experiments to examine the proteolytic stability towards degradation by chymotrypsin and trypsin [6], which are two enzymes with specificity towards peptides and proteins that are hydrophobic and cationic respectively. The docking experiments tell us not only which peptide is likely to be decomposed into its building blocks (amino acids), but also the most likely cleavage site and the products formed (Figure 2). The in silico data must of course be validated with experimental data, typically obtained form kinetic assays where the peptide is preincubated with the enzyme before a substrate is added. Based on the docking experiments we can suggest how modifications of the CAPs can make the resistant to enzymatic degradation, and thereby reduce the cost, time and effort in the experimental synthesis.

**Mechanism of action**

Nature has designed antimicrobial peptides to target the fundamental difference in the cellular membranes of microbes and multicellular animals. The outermost leaflet of the bilayer in bacterial membranes is heavily populated by lipids with negatively charged phospholipid headgroups. In contrast, lipids with no charge are found in the outer leaflet of plant and animals. The conventionally accepted Shai-Matsuzaki-Huang model [7-9] explains that antimicrobial peptides exert their action by binding to the cell membrane, causing cell death by direct membrane disruption or by translocation into the cytosol where potential targets are affected. This model was derived based on the properties of longer CAPs [12-100 amino acids], but, interestingly, CAPs with three to four residues, follow many of the established observations in terms of hydrophobicity, charge and amphipathicity. To examine this in more detail, molecular dynamics (MD) simulations were carried out.

An atomistic model of the system is first built, where each atom is treated as a sphere and bonds between atoms are modeled as springs. The potential energy of the system is described using a force field, which typically contain terms and parameters describing the energy associated with bonds, angle, rotation around bonds and nonbonded interactions (van der Waals and electrostatics). Random velocities are then assigned to each atom in the system, according to a Boltzmann distribution. In order to obtain information on how the atomic positions change as a function of time we need to solve Newton’s equations of motion. From the force that acts on each particle we can calculate the acceleration working on it using Newton’s second law (F=ma), and change in the velocities can be calculated from the acceleration. Now the new positions of the system can be calculated from the velocities. The force calculation is then recalculated with the new coordinates of the system. This is in essence what we call molecular dynamics simulations, and integration of the equations of motion is typically done in millions of steps. The final result is what we call a trajectory of the system, which specifies how the positions and velocities of the system change as a function of time. MD simulations are not particularly memory demanding, but are highly parallel in the sense that the energy and forces that act on different particles can be calculated simultaneously on different processors. Many MD programs now scale well to more than thousand processors, but for most applications 64-512 processor cores are used.

Amphipathicity is a property required for the antimicrobial peptides to interact with and penetrate into cellular membranes. MD simulations of a series of CAPs consisting of three amino acids revealed that
data obtained from the MD simulations is in very good agreement with experimental solution data from NMR (Nuclear Magnetic Resonance) spectroscopy.

Knowledge of the interaction between small molecules and lipid bilayers are obviously important, but still rather limited. Experiments on lipid bilayers are complicated due to the “soft” nature of the system. Perhaps the most important for us has been the ability to gain atomic insight into how these small peptides interact with models of the bacterial cell membrane. We have carried out MD simulations where one peptide is placed outside a lipid bilayer consisting of phospholipids typically found in bacterial cell membranes. The supercomputing facility in Tromsø, Stallo, has been used for this, along with the Schrödinger program package. Our MD simulations revealed that the cationic charges in the peptides rapidly begin to interact with the negatively charged phosphate groups. As a consequence, the molecules become amphipathic, and their hydrophobic groups move from the solvent to the more favorable membrane environment. In contrast, inactive peptide stays in the solvent and is not observed to enter the membrane during the simulations, which probably reflects its poorer antibacterial properties. An example what we can observe in our simulations is illustrated in Figure 3. Figure 3 A shows that despite that the peptide initially interacts with the membrane, it moves into solvent [water] where it cannot “see” the membrane. However, after some more simulation time, the peptide begins to interact with the lipid surface through polar interactions, but now the nonpolar groups also enter the membrane. Again, the mechanism of action emerging from theoretical MD simulations finds support from NMR experiments of peptides titrated into liposome (spherical lipid bilayer) dispersions [10].

References


Acknowledgments

We thank the HPC group in Tromsø for assistance and Notur and the University of Tromsø for allocations of CPU time on the supercomputers. Geir Isaksen and Annfrid Sivertsen have been the main people involved in this project through their PhD work. John Sigurd Svendsen is gratefully acknowledged for a long and inspiring collaboration.
The forecasted tidal wave of scientific data [1] has reached the Norwegian Life Sciences community as DNA sequencing technology and other high throughput experimental technologies produce drastically more data, at a lower cost per data amount and with a higher service availability than ever before. NorStore is the national high-capacity storage resource for scientific data in Norway. Through the StoreBioinfo project, NorStore and the FUGE Bioinformatics platform has joined forces to develop solutions for the Life Sciences community to manage and utilise their data.

Introduction

The new DNA sequencing technologies are referred to as a game changer in many respects [2], and are causing large changes in the scientific protocols of many fields in Norwegian science. Scientists that traditionally are not used to amount of data in the terabyte scale, are now in need of tools and infrastructure to manage and process their data sets of biological sequences, and to interpret them in the context of available information in the public knowledge space in Life Sciences. Publicly funded research projects in Norway are obliged to make all necessary data that supports scientific findings available for minimum 10 years. Consequently, an infrastructure that supports proper documentation of datasets as well as features for governing the life-span of data is important to ensure a proper return of investment of Norwegian research funding.

As the user bases of Notur and NorStore are steadily growing, and new user communities appear on the scientific computing arena, new models are needed to efficiently interact with user communities instead of maintaining a direct communication with all individual research
skills traditionally required for the usage of the national computing and storage resources. Furthermore, many actors across different roles are often contributing to the same project involving high-throughput data: actors from the range of lab-bench technicians, instrument specialists, bioinformaticians, statisticians, biological/biomedicine research group staff such as technicians, PhD, PostDocs and professors. They all need access to the project data. The Bioinformatics platform has provided tailored data processing and analysis pipelines towards the Life Science community as part of our service activity the past 10 years. As specific data types are utilized by a large user mass, an infrastructure for performing standard analysis pipelines integrated with national storage resources supporting project oriented data management, is a natural aim for the Bioinformatics platform.

System design and architecture
We have adhered to the design paradigm of Service Oriented Architecture (SOA) that is often utilized in larger software systems either coordinating existing solutions for different sub tasks or being designed from scratch to divide work
between independent modules of software. A system based on SOA will implement a set of interoperable services, that can be used by different parts of the system to coordinate a series of events to complete a task. A convenient way to integrate external clients to a system, is to provide programmatic access to a set of these services. XML is commonly used for interfacing with SOA services, but is not a requirement from the design paradigm.

The eSysbio project that started in Bergen in 2008 is developing a SOA based online software system for interdisciplinary collaborative life science research, and was the starting point in StoreBioinfo for developing the SOA based infrastructure for accessing the NorStore resources from autumn 2010. The independent services are implemented as W3C SOAP Web Services, and these services together then constitute the interface both towards a generic storage portal and towards specialised analysis programs as illustrated in Figure 1.

The authentication and security service is a central service that has the responsibility of authenticating users based on credentials, and issuing a security token in the SAML format [3], that is needed for accessing all other services in the infrastructure. All services govern authorization of their own resources based on the security token given to them in any operation a client code is calling. The dedicated authentication service allows for easy integration with federated user databases such as Feide [4]. Currently the system maintains a local StoreBioinfo user database, while there is work in progress to explore user authentication through the Feide system.

As illustrated in Figure 1, the system consists of three conceptual layers, the persistance, service and front end layers. In the persistance layer at the very backend, the data files uploaded by users are stored in an iRods instance on the NorStore high-capacity disk systems, and meta data about files, users, quotas, projects etc are maintained independently by the different services in relational databases. The StoreBioinfo portal accesses the different operations in the service layer to authenticate a user, query for projects the user participates in, and managing the datasets according to the user’s interaction events with the web pages of the portal.

The operation of transferring files to and from NorStore storage is of course central to the system and paid particular attention to due to the size requirements of Life Science data sets and the network based access interface. The main method of transfer is for the user to initiate a transfer process from a server accessible over the internet to the NorStore resources using the data storage service as illustrated in Figure 2. The data storage service then handles the transfer of the content and update of all meta-data without the front-end having to be connected during the whole transfer. For the portal this means users can stage several uploads and keep on their daily work on their workstations without being affected by the transfer, and even log out for the day and return the next day without interrupting the file transfer. StoreBioinfo currently implements both internal and e-mail based status notifications of file transfer operations.

In addition to the challenges of data transfer, the system has to cater for the wide size range of individual files in a Life Science dataset, that can vary in the range of a few kilobytes to hundreds of gigabytes. Combined with a relative large minimum block size of file systems that need to span large-capacity installations like NorStore, it is clear that individual files can not be stored directly on to the NorStore disks. Instead the data storage service repackage all files of a datasets into archives that are then transferred to the iRods backend, while all meta-data about the files are stored in the data storage persistance database. The package are based on sub types of data that are natural to download as a set, described further below. The iRods backend also provides direct programmatic access to extract any individual file from an archive, and this is used together with the meta-data to provide access to individual files stored in the backend, although they are stored more optimally in terms of consumed space/footprint on disk.

**StoreBioinfo web portal**

The StoreBioinfo web portal is the main interface to interact with the NorStore system for Life Science users, and the first mean of access made publicly available as an alternative to the traditional command-line tools requiring linux/unix skills. Key features are project centric sharing of data, quota management,
A typical usage scenario for a researcher not registered with the StoreBioinfo resources before is as follows:

1. The researcher applies for a quota for her project via email to contact@bioinfo.no.
2. She gets invited to join the system by email, with a confirmation ID attached.
3. Visiting https://storebioinfo.norstore.no she creates her account herself using the confirmation ID.
4. The project and quota are made by an admin user when the application has been evaluated.

5. The user stages an upload with data types and server settings and starts the upload.
6. Upload status is visible in the portal.
7. Notification is sent by email or internal message (at the choice of the user) when upload operation has completed.

Figure 3 displays the portal view of the user at step 5, where a server and the locations of folders or archive files of data to be uploaded are specified. A Java Web Start file upload client is also supported, as a mean of access for users to upload files only available at their own workstation and not accessible on any server.

Since the web portal simply orchestrates the back-end services to provide all available functionality, the architecture lends itself for tight integration to any software system that is given access to the service interface level.
Integration of StoreBioinfo services in the external Genomic HyperBrowser application

The Genomic HyperBrowser application [5] is developed by the UiO node of the Bioinformatics platform, and the main type of data it is operating on are genomic tracks. These are essentially annotations of biological features linked to a position in a genomic sequence. As an example, the current definition of the official human genome with the placement, extent, different components etc of all known human genes are encoded as one genome track. HyperBrowser enables scientists to query for relationships between several genome tracks and test for statistical significance of such relationships, for instance between experimentally obtained markers in one or several tracks and the genes in the human genome reference track.

HyperBrowser is the first application with active development work to interface the StoreBioinfo storage services. The exact same Web Services as the web portal are interfacing are utilized to achieve the integration, performing the same steps to login the user and handle the security token to access the other services. Figure 4 displays at the top a user navigating her datasets in StoreBioinfo through Hyper-Browser, and at the bottom a screenshot of how such a loaded track can be used as an input to a standard HyperBrowser analysis.

Conclusion and future plans

The first version of the StoreBioinfo web portal is now available for use, and we are currently enhancing several aspects of the portal to make the user experience even better and more productive. Another main milestone for the project will be the release of the full Web Service layer to the public, expected to happen medio 2012. In terms of scientific utilisation of data, storage only covers half the picture. An just as important aspect for the user is how to get the data processed and analysed. Some of these computing tasks will require large scale facilities provided by Notur, and other tasks will only require access to relatively standard scale processing power. StoreBioinfo will contribute actively to the development of the needed interface between the storage resources and computational resources available in NorStore and Notur, as this is an urgently needed feature both from application developers and users in the Life Sciences community.

A special thanks goes to the programmers of the StoreBioinfo, eSysbio and HyperBrowser teams for their hard work towards realising this infrastructure.

References

Emerging Needs for eInfrastructure in Climate Science

The need for climate datasets has become ubiquitous nowadays. Most countries need to better prepare and cope with changes in climate - which requires the best state-of-the-art projections available. The cost of not planning or adapting can be devastating. More and more nations are producing and gathering climate information for impact studies - and there has been an emphasis on higher and higher spatial resolutions. The demands for faster and greener computers have also become inherent to this process.
The fast-paced computer world under Moore’s law is difficult to keep up with. As soon as a new computer system is purchased, it becomes almost obsolete in a few months time. The same can be said of storage systems, which in many cases can become bottlenecks down the line (i.e.: slow and unstable storage systems). While the development of fast computers progresses so does the development of climate models. Today, climate models are expanded to take full advantage of multi-core processing. Model development goes almost hand-in-hand with advances in computer science. In fact, even one of the first applications of computers was for weather forecasting.

Norway is a privileged country when it comes to computer resources. The Research Council of Norway funding through the NOTUR project provides computational hours for Research centers and universities throughout the country.

The Makeings of a Climate Model

Climate models are created based on mathematical equations, such as, the momentum equations for a spherical Earth; the thermodynamic energy equation; the continuity equation for total mass and the ideal gas law. These are called primitive equations and the operational and research models are based on some version of these equations (see Warner, 2011 - Chapter 2, for further details). These equations form the so-called “dynamical core”, which is the component that treats the resolvable scales in the system. They are then solved numerically at “...points defined by a quasi-regular, three-dimensional spatial grid” (Warner, 2011) and integrated in time.

It may sound simple to create a model, but in fact, developing a full model system requires the effort of a team of researchers and many years of work: “It is estimated that a fully coupled ocean-atmosphere general circulation model (OAGCM) takes about 25-30 person-years to code, and the code requires continuous updating as new ideas are implemented and as advances in computer science are accommodated” (McGuffie and Henderson-Sellers, 2005).

**The GREEN group**

The Global and Regional climatE projEctioNs group at the Uni Bjerknes Centre is committed to develop and apply climate models to regions throughout the world as well as to provide capacity building. The capacity building involves training provided to developing countries, where GREEN has projects with, for example: Bangladesh, India, Nepal, and Vietnam. One such training activity is the TERI-BC-CR Research School² in October, which will give a hands-on experience and training on using a Global Climate Model (GCM) as well as classes on Atmospheric Sciences and numerical methods.

The GREEN group also works on the NorClim and EarthClim projects, led by Professor Helge Drange. This project focuses on the development of the Norwegian Earth System Model (NorESM). In Bergen, NorESM is the successor model to the Bergen Climate Model (BCM), which was used to create data to the fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) in 2007. Together with collaborating partners in Norway, Dr. Mats Bentsen and PhD student Ingo Bethke have recently made NorESM model data available to the Coupled Model Intercomparison Project 5 (CMIP5). These NorESM climate change experiments, together with contributions from modeling centers around the world, will be assessed in the upcoming IPCC Report due 2013.

The Norwegian Earth System model or NorESM is comprised of a combination of several models made up of different components of the climate - forming a coupled system. NorESM is based on the Community Climate System Model 4 (CCSM4) from the National Center for Atmospheric Research (NCAR), Boulder, USA. The NorESM model components are shown in Table 1.

NorESM data needs to be stored for decades. The total raw data for the planned NorESM contribution to CMIP5 takes up 120 TB in addition to 60 TB for post-processed

<table>
<thead>
<tr>
<th>Component</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmosphere</td>
<td>Community Atmosphere Model (CAM 4.1.08)</td>
</tr>
<tr>
<td>Land</td>
<td>Community Land Model (CLM 3.7.10)</td>
</tr>
<tr>
<td>Sea-ice</td>
<td>CICE (CICE 4.0.20100323)</td>
</tr>
<tr>
<td>Ocean</td>
<td>Miami Isopycnic Coordinate Ocean Model (MICOM) extensively modified at the Nansen Center</td>
</tr>
<tr>
<td>Atmospheric chemistry</td>
<td>Chemistry-aerosol-cloud package in CAM by University of Oslo and met.no</td>
</tr>
<tr>
<td>Ocean Carbon Cycle</td>
<td>Hamburg Model of Ocean Carbon Cycle (HAMOCC) adopted for use with an isopycnic ocean model</td>
</tr>
<tr>
<td>Coupler</td>
<td>CPL 7</td>
</tr>
<tr>
<td>Ice sheet</td>
<td>Currently none, but NCAR and LANL are working on this</td>
</tr>
</tbody>
</table>

1 The ENIAC, in 1950, created the first weather forecast.
2 See [http://terii.org/TERI-BCCR-CRS](http://terii.org/TERI-BCCR-CRS) for further information
The regional high-resolution simulations also need long-term storage for about 10+ years. For the tropical channel domain, each model timestep raw file has a size of ~800 MB. This gives around 1 PB of raw data considering ~900 simulation years:
• 4 timesteps per day * 365 days * 800 MB = 1.14 TB per model year
• ~900 simulation years = 1 PB (raw data)

For the European domain, less than 1 PB is needed. Due to the large sizes of raw files in high-resolution models, data “stripping” is thus necessary: one needs to remove model levels, variables and other elements in order to reduce the data size - which can lead to loss of valuable data. This also points to the fact that there is a need for faster computer servers to be able to post-process and analyze these datasets.

The regional part of the GREEN group works with the Weather Research and Forecasting model (WRF), developed at the National Center for Atmospheric Research (NCAR) in the United States. Several projects involve the downscaling of atmospheric data for: Norway, Europe, Southeast Asia and the tropical belt. The tropical belt downscaling is made to allow waves in the tropics to travel more naturally. The domain used is cyclic and it extends from 45S to 45N. A simulation using the tropical-channel setup at 36 km resolution is shown in an animation (see link in figure caption). This animation was produced using the Weather Research and Forecasting (WRF) model as part of a sensitivity study related to different microphysics and cumulus parameterization schemes.

The regional high-resolution simulations also need long-term storage for about 10+ years. For the tropical channel domain, each model timestep raw file has a size of ~800 MB. This gives around 1 PB of raw data considering ~900 simulation years:
• 4 timesteps per day * 365 days * 800 MB = 1.14 TB per model year
• ~900 simulation years = 1 PB (raw data)

For the European domain, less than 1 PB is needed. Due to the large sizes of raw files in high-resolution models, data “stripping” is thus necessary: one needs to remove model levels, variables and other elements in order to reduce the data size - which can lead to loss of valuable data. This also points to the fact that there is a need for faster computer servers to be able to post-process and analyze these datasets.

The Need for Computer Power
Climate research requires state-of-the-art computer power. This is necessary so as the climate runs can be made at a reasonable amount of time. The fast development in climate research in the past decades has been made possible also because of the development of fast computers, as well described by McGuffie and Henderson-Sellers (2005): “Without the recent growth in computational power and the reduction in computing costs, most of the developments in climate modelling that have taken place over the last four decades could not have happened”.

There are, however, challenges. For example, there is a need for higher resolution datasets for climate adaptation and adaptation research. Hexagon has many users and the queue time can take up to a week for a simulation to start. In addition, there are challenges for some of the simulations to be completed and safely stored.

NOTUR Support
Thanks to the support by NOTUR and NorStore and the collaboration with the System Engineers at Parallab, the GREEN group is able to provide state-of-the-art datasets and climate projections to the different projects at Uni Bjerknes Centre.

In order to improve and speed up our work, our group has recently purchased a computer server, that will provide greater capacity for the group to post-process and analyze the climate datasets from Hexagon. These are local additions to alleviate some of the more specialized processing of the data, i.e. the downscaled computations that are still executed on Hexagon and data that still need to be archived (long-term) on NorStore.

References

© BCCS
Some words from three climate researchers and their experience with the present supercomputer Hexagon

Ingo, Mats and Helge are among the researchers who this summer delivered the model data for the next IPCC Assessment Report 5 (AR5), the first of several data batches that will be delivered to the AR5. The model output is produced on the supercomputer Hexagon, using the national resources via Notur, and then moved to the NorStore storage facilities.

Ingo Bethke
PhD Student at the University of Bergen and Bjerknes Centre for Climate Research

When setting up the current model system NorESM we used a considerable amount of time to tune the model in order to represent a realistic climate without too much climatic drift. This tuning process delayed our data delivery by several months. Yet, the computational resources were adequate for our model setup and more resources would not necessarily have led to a much earlier delivery. It should be noted, however, that the choice of using an intermediate resolution setup was made because the computational resources were not sufficient to use the [more desirable] high resolution setup of the NorESM.

Our old model system, the Bergen Climate Model, is still used for paleo and coupled carbon cycle research. The reason is that the computational costs of the NorESM are too high for their applications and the coupled carbon cycle is not implemented yet. Currently it is only possible to run the BCM on one computer system in Norway, and this computer is soon to be phased out.

A continuation of the research will depend critically on support from Notur to port the model to a new computational platform. An application for such support has been submitted to Notur.

We have received good support from NOTUR. In particular, when porting a model system to a new computational platform, the advanced-user-support is indispensable (i.e. dedicated support from trained IT-engineers that work with our technical problems full-time over a longer period). All in all, I’m very satisfied with the arrangement with Notur and NorStore.

Mats Bentsen
Research scientist, Bjerknes Centre for Climate Research, Uni Research

I have been using Hexagon extensively for development of NorESM and production of climate model data. Like Ingo mentioned, making our models run correctly and efficiently on new computers are challenging. Although Hexagon was new in 2008, it will be replaced in 2012, and considering the time to make the model systems work properly the actual production period on a computer can be quite short. We also have to adjust the models to new technology. After several generations of computers using rather comparable technology, we are now facing a shift with a transition towards computations being done on GPUs instead of CPUs. This likely means extensive modification of our models, something we will start working on this autumn.

We are planning our CPU-usage one year ahead but there is always the possibility for delays due to unforeseen problems. Fortunately there is some flexibility in the allocation system so in most cases we have been able to meet our project milestones.

I really like this arrangement with distributed computing resources organized by Notur, and believe it is a very efficient way to use the computers. If every research unit should have their own computer, the idle time would often be extensive and one probably could not afford a computer large enough for desired peak usage. Also, the Notur machines are taken care of by competent people with supercomputer experience, and we have good support.

Our models also produce an enormous amount of data, and we have to find good ways to transfer, store, and analyse them. Compared to producing the data, analysing the data requires very different tools and work flow.

Helge Drange
Professor, climate modelling, University of Bergen and Bjerknes Centre for Climate Research

Generally speaking, the need for computing capacity in climate research is large. This is particularly the case for societal needs related to address possible changes in local climate and extreme weather. In Bergen, the supercomputer Hexagon only partly fills this need. When Hexagon first came it was quite powerful, but now it is far too small.

Like Mats mentioned, we need large and stable space for data storage in addition to computational power. Today we have difficulty storing the model output. Storage is a general problem in Norway. When connection between science and politics can be close, as is the case for climate science, it is of paramount importance to be able to go back and reanalyse the data, and to make the data available for any interested parties.

I would also like to add that we get, in general, good knowledge support from Notur. We need people who understand the different machine architectures. What is lacking in Norway is a more holistic view on the full chain of HPC resources and infrastructure.
Cloud computing is a new model for the delivery of information and communication technology (ICT) as a service rather than a product, whereby shared resources and information are provided as a utility, similar to the water distribution or electricity grid. Cloud computing can bring significant benefits to the services and efficiency provided to a researcher’s daily working environment.
Clouds typically involve the provisioning of dynamically scalable services and include computation, storage, data access and software [1]. The services are almost always virtualized, e.g., they do not require end-user knowledge of the physical location and configuration of the resource that delivers the services. Cloud computing services are delivered based on Internet protocols, e.g., via web applications/web services, while the user/customer software and data are stored on servers at a remote location. Most cloud computing infrastructures consist of services delivered through [shared] data centers and through a single point of access.

One can distinguish between several deployment models of clouds. We only describe two here. In a public cloud, resources are dynamically provisioned to the general public on a self-service basis by an off-site third-party provider (commercial actor) who bills on a pay-as-you-use basis. The business case for public clouds includes lower up-front capital costs and less hands-on management. At the other side of the spectrum, there are private clouds. A private cloud is infrastructure that is usually operated for a single organization, whether managed internally or by a third-party and hosted internally or externally. For private clouds, it is possible that the underlying systems often still have to be bought, built and managed by the organization itself and this makes the business case less obvious.

In all considerations concerning the utilization of clouds, one has to keep in mind that data is the new gold. The ability to move, access, change, annotate and archive data will be a determining factor in one’s research output and how it will be used by others, now and in the future.

The result of adopting cloud services is that one’s working environment becomes ever more networked and dynamic. Research communities and service providers are increasingly intertwined. Especially now that we see more classes of cloud services emerging. With companies like Google, Amazon, Microsoft (and the next Gorilla we don’t know yet), that are providing a wide range of cloud services, there is a case to be made to outsource part of the infrastructure to these big players: these companies offer services as a commodity what national infrastructures like Notur and NorStore have to establish and operate themselves. In addition, there is “green computing”. More and more research requires extensive computation and the management of data sets. Computers consume energy and dissipate heat. Energy efficient, cost efficient and scalable solutions for hosting compute and storage resources are increasingly important. (Green computing was addressed in more detail in the previous issue of META.)

Cloud computing

The use of cloud services in a national infrastructure for computing like Notur can be illustrated by a number of cases. The first use case is related to the correct usage of the HPC infrastructure. A significant fraction of the jobs running on the HPC infrastructure are single-threaded or fit in a single compute node. These are candidates to be executed on a non-HPC infrastructure. As the number of such jobs will vary over time, a flexible solution is needed that is available on demand.

Cloud computing can bring significant benefits to the services and efficiency provided to a researcher’s daily working environment.”

The second use case consists of serving scientific communities that are not (easily) served by the current infrastructure, but that need access to qualitatively reliable, on-demand computing. The present infrastructure is largely Linux-based which can be difficult to adopt by communities that have no expertise in this. Cloud computing can provide convenient and cost-efficient access to new scientific end-user groups.

The third use case would be “temporary capacity” and sometimes is a subset of the second use case. Examples are:

- Studying software architectures in an educational setting, where the “computing lab” is on during the course period only.
- Qualitatively software profiling or testing deployment scenarios for scientific software packages.
- As infected machines can be destroyed and cleanly launched on demand, be it in an educational or research setting, security research on computer architectures can be done on virtual machines that are paid for by the hour and then destroyed.

A public cloud adds capacity in a wide variety of communities in these use cases. This allows to establish services for new communities without massive upfront investments in time or money. As new uses of the cloud and their respective usage communities grow, they might or might not be moved to a private cloud environment depending on the community’s size and needs.

Today, the market leader for public cloud computing is Amazon. Since the beginning of 2009, the prices per compute hour (one pays for usage only) have been slowly declining. Recently other price components have changed as well, mostly related to data traffic: no incoming traffic charges and lowered prices per tier for outgoing bandwidth charges. The steady decline in prices will make cloud computing increasingly attractive. Private clouds need close collaboration between network administrators and system administrators; it is not likely they will reach the same maturity level at operational cost before mid 2012, if at all.

The utilization of cloud computing may not always pay off, e.g., in the case of special...
applications with a small user community. In addition, cloud computing requires that end-user applications are portable. This can be a challenge for applications that depend on technologies that are not supported by the cloud environment. For such cases, software and data management will have to be revised.

Cloud storage
We address here the challenge to share data between groups of users, both in space (e.g., different universities) or in time (e.g., accessed now and in the future). Direct usage of cloud storage as part of a computation as described is not addressed here. In 2010-2011, we have developed a prototype tool that enables system administrators as well as end-users to integrate cloud backed storage into their daily operations and applications. This activity started under the 1-year NEoN project from the Nordic Data Grid Facility in 2010. The findings of that project were carried further in 2011 by Notur and NorStore, in collaboration with the Swedish National Infrastructure for Computing (SNIC).

Cloud storage such as Amazon’s S3, Rackspace’s CloudFiles or OpenStack’s object storage all look similar: you have a name space, often called a bucket, in which you can store an unlimited number of “objects” of various sizes. Depending on the cloud used, the maximum object size is presently between 5 GB and 5 TB. The object space itself is flat and does not have a hierarchy like a traditional file system.

Points that are relevant in the design of a cloud backed storage system include the following:

- Encryption. Data that is stored in the cloud should be encrypted.
- Metadata. Many research communities have metadata schemes that can be attached to their data sets and that will preserve the meaning of the data, especially in the long term.
- Support for multiple storage back ends. You might want to migrate your data to a different provider for a myriad of reasons, e.g., cost. A solution should be easily extendable to add multiple storage back ends.
- Easy access. Most cloud storage provides specialized interfaces which require considerable programming time or specialized knowledge in order to use them. A solution should work across a lot of different devices without having to install or maintain dedicated clients (of course, one always might do so for special needs). In other words: a storage service should be accessible without special tools a priori. Easy access is also important because users will often chose convenience over privacy. An example is the success of Dropbox, which is based in the US. Unfortunately, there’s been a number of security leaks outside any university or company’s control that has plagued this and similar services in 2011.

Support for multiple storage back ends is relatively easily achieved: the familiar “plug-in” model can be easily designed (and has been) in the eventual cloud backed storage solution. Making it easy to use and secure is a harder task. And then we also have to add metadata to make research results verifiable and available in the long term.

As it turned out, easy access went hand in hand with metadata. We chose WebDAV as primary protocol to interface with cloud backed storage. Though an old protocol in Internet lifetime (since 1988), it is widespread and goes via HTTP. HTTP is so widespread that a whole slew of network connectivity problems disappear for the end user (e.g., firewall issues). But there are more reasons to use WebDAV:

1. WebDAV is available from iPods and smart phones to Linux servers to Windows 7 and OS X.
2. WebDAV allows you to attach arbitrary XML properties to files and folders. And the XML standard is used by many communities to define their metadata. This makes it possible to create a storage front end where data and metadata are kept together.

Figure 1. Schematic overview of cloud backed storage.
If we consider storing the file system structure and all its metadata (such as file size, date created, date modified), the user-specified metadata per community and additional metadata (such as encryption keys per file), then the solution becomes obvious: separate the storage of the metadata from the actual data. Once that separation is made, one can add different “front ends” for access (like WebDAV) and back ends for storage.

Figure 1 gives a schematic overview of how the current solution works. There is also an interface for management of data and convenience. It is easy to add multiple ways to access and store your data, should the need arise (e.g., web services).

The metadata contains all the information - from filename to encryption key of a file - and can be maintained and stored locally by the user. The actual data that is stored in the public cloud, possibly in a different country, is nothing more than a set of encrypted set of objects. As encryption is handled transparently by the website and WebDAV layer, users need not do something special as long as the metadata is located on a trusted site.

The fact that the metadata holds all the information allows any layer on top of it [e.g., the WebDAV disk interface] to be linearly scalable. It also allows the storage back-end to be just a “dumb” plug-in that stores encrypted objects. The metadata store itself might become the single point of failure then. However, it has replication built in (even across data centers if needed).

The described cloud backed storage design and implementation are available as an open source project [2]. This allows institutions to deploy instances locally and tune them to specific needs, while at the same time providing a basic cloud service.

References


To subscribe to the paper version of the META magazine, please visit: www.notur.no/publications/magazine
e-Infrastructure
Use Roadmap

The eInfrastructure Scientific Opportunities Panel, appointed by the Research Council of Norway, is responsible for monitoring the development of scientific use of eInfrastructure. Last year, the panel produced a report describing the Scientific Case for the eInfrastructure that can best serve Norwegian research groups and operational forecasting from 2015.

That document is a prelude to the second report, the eInfrastructure Use Roadmap, that has now been drafted by the panel. This roadmap document covers current and new scientific areas for eInfrastructure use, taking into account the opportunities offered by existing and emerging large-scale international collaborations. The roadmap should also make some international comparisons and matches users and application areas against services. The document is intended for all stakeholders in Norwegian eInfrastructure including users and support staff, while the recommendations we make are intended for decision makers at universities, the Research Council, and the Ministry of Education and Research.

A draft version for the eInfrastructure Use Roadmap can be downloaded here: www.metacentre.no/esop/roadmap

e-IRG White
Paper 2011

The 7th White Paper of the e-Infrastructure Reflection Group was released in July 2011. The White Paper addresses some of the most interesting questions related to new and on-going e-Infrastructure challenges, such as:

- How do we deal with the increasing energy demands of computing?
- What software is needed to fully harness the power of future HPC systems?
- What are the appropriate governance models for e-Infrastructures?
- How can we facilitate access, discovery and sharing of large and diverse sources of scientific data?
- How can we further advance research networks, and adopt and implement new e-Infrastructure services

Download the White Paper here: www.e-irg.eu

PRACE Preparatory Access

Preparatory Access allows researchers to apply for code scalability testing and also support for code development and optimisation from PRACE software experts. Preparatory access allows researchers to optimise their codes before responding to PRACE calls for project access.

There are three types of Preparatory Access:

A. Code scalability testing to obtain scalability data which can be used as supporting information when responding to future PRACE project calls. This route provides an opportunity to ensure the scalability of the codes within the set of parameters to be used for PRACE project calls, and document this scalability. Assessment of applications is undertaken using a light-weight application procedure. The maximum allocation time is two months which includes submission of the final report.

B. Code development and optimisation by the applicant using their own personnel resources (i.e. without PRACE support). Applicants need to describe the planning for development in detail together with the expert resources that are available to execute the project. The maximum allocation time is six months which includes submission of the final report.

C. Code development with support from experts from PRACE. The maximum allocation time is six months which includes submission of the final report.

The maximum allocation that can be requested depends on the type of access that is applied for (and choice of target system), but varies from 50 000 to 250 000 core hours. Preparatory access calls are rolling calls, researchers can apply for resources all year. There are no closing dates. Assessment of the applications received will be carried out at least every two months.

For more information: www.notur.no/news/archive/20110201.html
EUDAT is a three-year EC-funded project that will deliver a federated data infrastructure with the capacity and capability for meeting future researchers needs in a sustainable way. Its design will reflect a comprehensive picture of the data service requirements of the research communities in Europe and beyond. This will become increasingly important over the next decade as we face the challenges of massive expansion in the volume of data being generated and preserved (the so-called ‘data tsunami’) and in the complexity of that data and the systems required to provide access to it.

The EUDAT consortium includes representatives from each stage of the value chain that has evolved to deliver scientific knowledge to researchers, citizens, industry and society as a whole. It includes funding agencies that invest in research infrastructures and programmes of research, infrastructure operators and research communities who rely on the availability of data-management services, national data centres and providers of connectivity and, of course, the users who rely on the availability of data and services. With the inclusion of disciplines from across the spectrum of scientific endeavour sharing a common infrastructure, EUDAT also provides the opportunity for data-sharing between disciplines and cross-fertilisation of ideas.

The EUDAT consortium includes 25 partners and third-parties. The EC-contribution is 9.3 MEuro. UNINETT Sigma and the University of Oslo participate for Norway. The Finnish partner CSC is the project coordinator. The kick-off meeting for EUDAT will be held in Helsinki, October 18-19.

More information will be available soon on www.eudat.eu

New supercomputer at NTNU

NTNU and Silicon Graphics International Corp (SGI) signed a contract for a new supercomputer June 24, 2011. The system will be installed at NTNU by the end of 2011.

The new system will include the new generation Xeon Intel processors, and will have theoretical peak performance of 300 Teraflops. This is almost thirteen times the capacity of today’s njord.

The system will be named “Vilje” after the Norwegian god Vili. According to Nordic mythology, Vilje gave humans wit (intelligence) and the ability to move.

More information about Vilje in the next issue of META.
Upcoming Events

SC11 - International Conference for High Performance Computing, Networking, Storage and Analysis
November 12-18, Seattle, WA, USA
sc11.supercomputing.org

7th IEEE International Conference on e-Science
December 5-8, Stockholm, Sweden
www.escience2011.org

EGI Community Forum 2012
March 26-30, 2012, Munich, Germany
cf2012.egi.eu

IEEE CloudCom 2011,
3rd IEEE International Conference on Cloud Computing, Technology and Science
November 29 – December 1, Athens, Greece
http://2011.cloudcom.org

HPC-Europa2

The EU-funded project HPC-Europa2 calls for proposals from researchers working in Europe to visit one of the seven HPC centres in its Transnational Access programme. The programme offers visiting researchers access to some of the most powerful HPC facilities in Europe, consultancy from experienced staff, and opportunities to collaborate with scientists working in related fields at a relevant local research institute. The programme covers travel costs, subsistence expenses and accommodation. Visits can last between 3 weeks and 3 months. The next closing date for proposals is November 15, 2011

More information at www.hpc-europa.org

The Notur II project provides the national e-infrastructure for computational science in Norway

The infrastructure serves individuals and groups involved in education and research at Norwegian universities, colleges and research institutes, operational forecasting and research at the Meteorological Institute, and other groups who contribute to the funding of the project. Consortium partners are UNINETT Sigma AS, the Norwegian University of Science and Technology (NTNU), the University of Bergen (UiB), the University of Oslo (UiO) and the University of Tromsø (UiT). The project is funded in part by the Research Council of Norway and in part by the consortium partners. The Notur project is complemented by two other projects that are financed in part by the Research Council of Norway. NorStore is a national infrastructure for scientific data. The NorGrid project deploys and operates non-trivial services for work load management and aggregation of the resources provided by the Notur and Norstore resources.

Subscription: If you would like to subscribe to the paper version of the META magazine, please go to www.notur.no/publications/magazine/