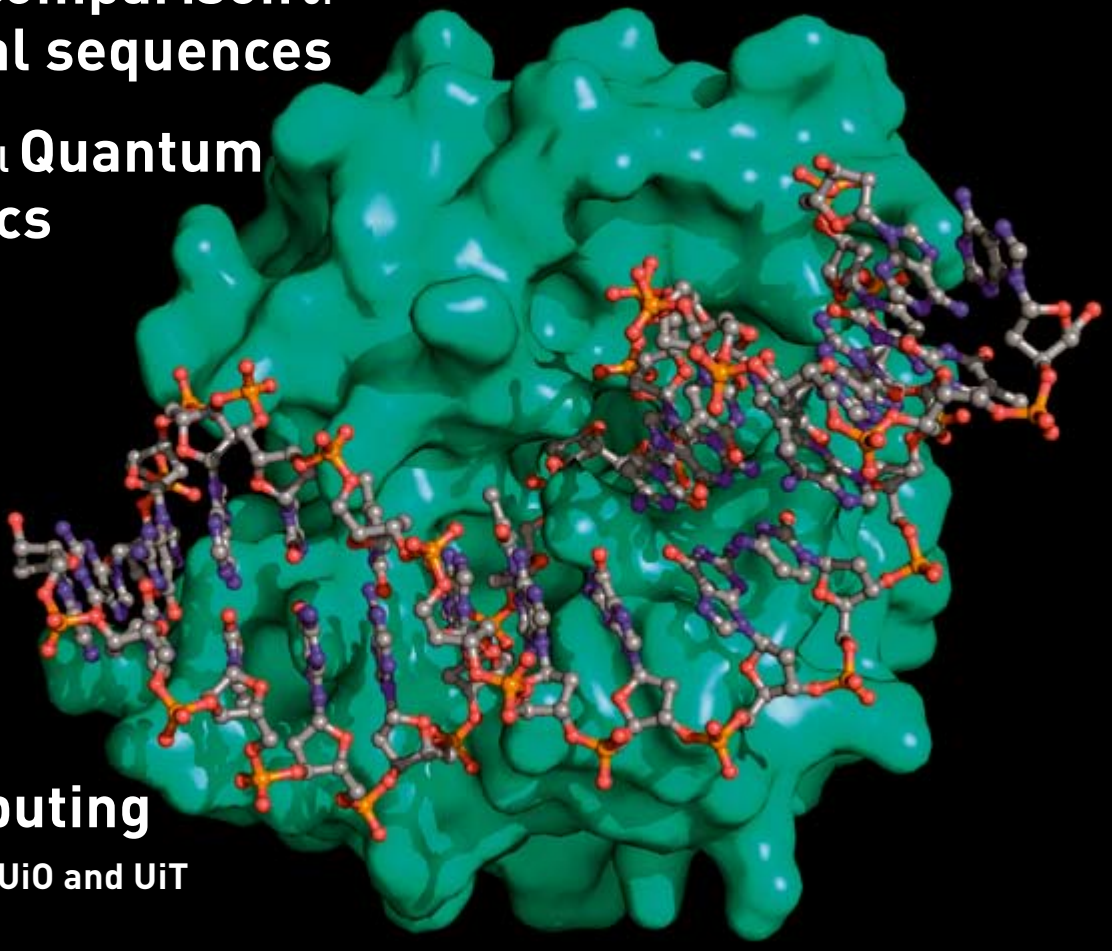
Number 2 : 2007

meta

A magazine published by the Notur II project

Large-scale comparison of biological sequences

Computational Quantum Mechanics



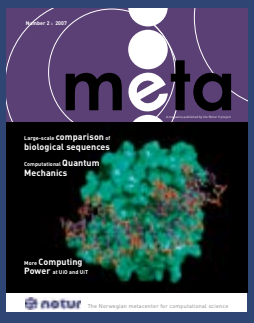
More Computing Power at UiO and UiT

notur

The Norwegian metacenter for computational science

CONTENTS

Cover picture: *Thermus thermophilus* UDG protein bound to DNA containing an abasic site. The illustration has been generated with PyMOL (http://www.pymol.org) from the Protein Databank structure 2DDG of the RIKEN Structural Genomics/Proteomics Initiative. Copyright:Jon K. Lærdahl, CMBN, Rikshospitalet HF.



Number 2 : 2007
A magazine published by the Notur II project – The Norwegian metacenter for computational science

Editorial Director: Jacko Koster
Contributing Editor: Eva Irene Haugen
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

e-mail address: sigma@uninett.no. Phone: +47 73 55 79 00. Fax: +47 73 55 79 01.
Publisher: Innrykk. Print: Skipnes

EDITORIAL

In August, the compute custer titan at the University of Oslo was expanded with a total of 448 dual-core processors. The dual-core processors will be replaced by quad-core processors before the end of the year. Together with the upgrade of the already existing part of the cluster, a total of 2432 processor cores will be added. The University of Tromsø installed a new HP cluster early November that contains 5632 processor cores. The University of Bergen installs a new Cray XT4 early 2008 with 5552 processor cores. The total number of processing units in the Notur project increases in a few months by almost a factor of ten.

Within a couple of years, the newest HPC resources in Norway will contain tens of thousands of processor cores. Technology developments are such that processor clock frequencies no longer increase at the same rate as they did in the last fifteen years. Manufacturers of computer chips are now producing chip architectures that contain multiple processing cores. In 2007, quad-core processors are being shipped in large quantities by the major vendors. In the near future one can expect the release of processors with tens of cores.

The scalability of several widely-used scientific applications is however inherently limited and a considerable part of the research community does not automatically benefit of new systems that contain much larger numbers of processing cores. The rapidly increasing gap between the growing numbers of processing units in modern compute resources and the limited scalability of the application software running on it, must be bridged. Significant effort is required to investigate and improve the scalability of relevant scientific applications, also in the Notur project.

In addition, multi-core architectures contain complex features like multi-threading and multi-level cache/memory hierarchies. It is necessary to build competence on such architectures to be able to design new (and redesign old) applications that can exploit the full potential of multi-core architectures. Redesign may require structural changes in e.g. the underlying algorithms and in memory-processor data flow. Especially if one wants to enable grand challenge applications, serious effort may be needed.

The effort that is needed is not only related to enhancing application performance and scalability. A researcher cannot always scale up his/her activity immediately in a sensible manner on a larger system, even if the application software would allow it. Larger computer simulations to study a physical phenomenon in more detail often requires more detailed input data (with finer resolutions, better accuracy, etc.). These data sets may not be readily available. Also, more refined mathematical models may be required that take into account physics that could be neglected in coarser models.

We can expect to encounter such problems regularly in the Notur project in the coming years.

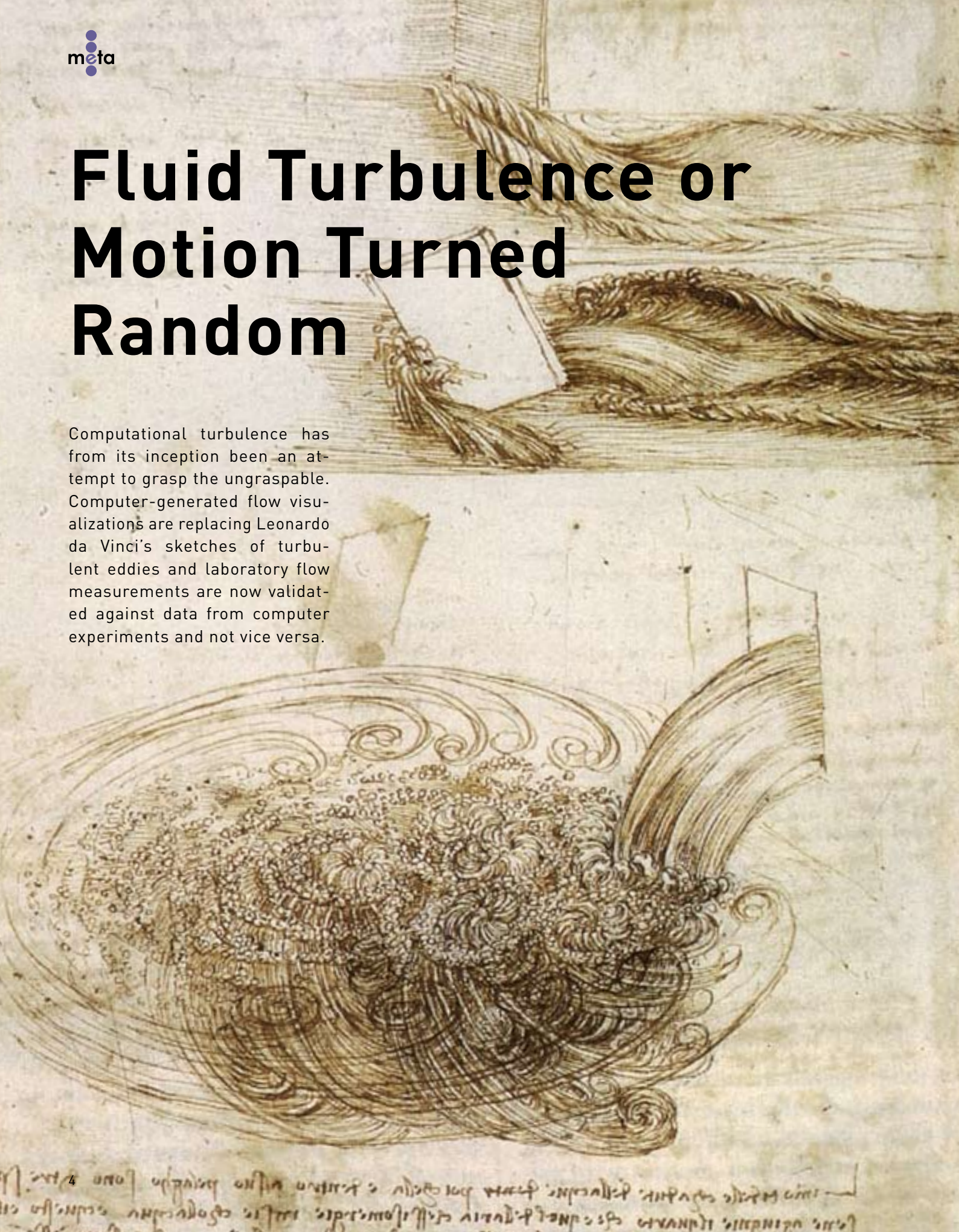
Jacko Koster, Project Coordinator Notur II, Managing Director UNINETT Sigma AS

Autumn berries

FEATURES

Fluid Turbulence or Motion Turned Random

Computational turbulence has from its inception been an attempt to grasp the ungraspable. Computer-generated flow visualizations are replacing Leonardo da Vinci's sketches of turbulent eddies and laboratory flow measurements are now validated against data from computer experiments and not vice versa.



AUTHOR

Helge Andersson
Professor in Fluid Mechanics at Department of Energy and Process Engineering, Norwegian University of Science and Technology

Turbulence – a flow phenomenon

Leonardo da Vinci was probably the first to use the term la turbulenza in his recognition of two distinct states of fluid motion. The pioneering scientific studies of turbulence date back to Osborne Reynolds' (1883) classical pipe flow experiments where he systematically distinguished between direct and sinuous states of flow and showed that the birth of eddies depended on the value of a particular combination of parameters to which his name was later attached. Subsequently, Lord Kelvin introduced turbulence into the English vocabulary as a derivative of the Latin 'turbare' (disorderly). Turbulence thus refers to the irregular and violent motion of liquids and gases which occurs when a certain parameter that characterizes the flow, the so-called Reynolds number, exceeds a certain limit beyond which the fluid viscosity is no longer able to stabilize the motion.

Turbulence, i.e. the state of being turbulent, is a generic phenomenon which may arise in any liquid or gas provided that the flow velocity is sufficiently high. Turbulent flows are the dominating flow regime both in nature and engineering. The structure of a turbulent flow is extremely complex: the flow is unsteady and intrinsically three-dimensional and apparently random and chaotic. The complexity arises from the fact that fluid flow is a non-linear phenomenon. The inherent non-linearities of fluid motion are essential for the creation and maintenance of turbulence and results in a wide range of the size of the whirling eddies which are the vital constituents of all turbulent flows. Fortunately, the tiniest eddies, which still are significantly larger than the intermolecular distances, are viscously dissipated into heat. Fluid turbulence is therefore a continuum phenomenon which can be described by the classical laws of fluid motion.

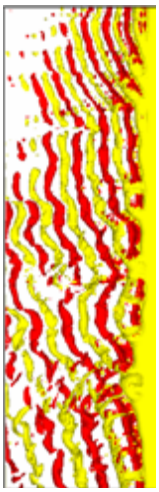


Figure 1: Coexistence of different flow regimes in the wake behind a linearly tapered circular cylinder. The cylinder diameter varies linearly along the span and makes the local Reynolds number higher near the lower part of cylinder than at the top part. The flow is from right to left and the yellow and red colour signifies vortices of opposite sense of rotation. The snapshot reveals a laminar wake behind the upper part of the cylinder, followed by a transitional flow and a turbulent wake near the lower part. The shedding frequency is observed to be significantly higher near the top. This research is part of a strategic university programme (SUP) devoted to Marine CFD; see Narasimhamurthy et al. (2007).

The mathematical model of turbulence

The flow of gases and liquids is governed by a seemingly simple partial differential equation discovered independently by the Irish mathematician George S. Stokes and the French engineer Claude L.M.H. Navier more than 150 years ago. This celebrated equation, which is the backbone of modern fluid dynamics, is nothing but Newton's law of motion formulated for the flow of a fluid rather than for the motion of a solid body. The Navier-Stokes equation is the ultimate model for both laminar and turbulent flows of liquids and gases. The complexities of fluid turbulence are all embodied in this equation: unsteadiness, non-linear inertia, non-locality, viscous diffusion and dissipation.

The Navier-Stokes equation was of limited use for more than a century until the computer processing rates became sufficiently fast such that numerical solutions could be obtained for some simple two-dimensional flows in the 1960s. Even today when computer simulations of three-dimensional fluid flows are common, fluid turbulence is not yet fully exploited and is sometimes referred to as the major unresolved problem in classical physics.

Computational turbulence research

Technological and industrial flow analysis, i.e. CFD (computational fluid dynamics), is performed with computer software based on phenomenological turbulence models that often suffer from severe shortcomings. The only viable route to computational turbulence research is to model the turbulence from first-principles. To this end a turbulent flow can be realized numerically by solving the complete Navier-Stokes model in time and three spatial coordinates on a computational mesh sufficiently fine to resolve even the tiniest scales of the turbulent motions. This approach is known as DNS (direct numerical simulation) and serves as computer experiments; see for instance Moin & Manesh (1998) and Sagaut (2004).

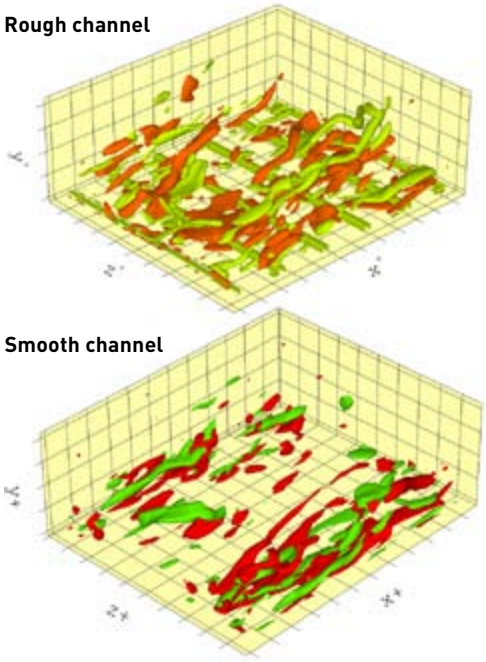


Figure 2: Turbulent flow along a rough surface is known to be conceptually different from the flow along a smooth surface. By means of realistic computer-based flow realisations, we were able to investigate the distinguishing features of turbulence in the vicinity of a surface roughened by transverse rods. The figure shows a perspective view on the whirling eddies near the structured surface (upper) and a smooth (lower) surface. The flow is from left to right and the snapshots show whorls rotating in opposite sense (distinguished by either brown or green). Ashrafian & Andersson (2006).

The benefits of DNS are plentiful from a scientific point of view: extremely detailed information on the coherent flow structures is embodied in the numerically generated flow field. The spatial resolution is substantially better than in any laboratory study. Correlations between fluctuating quantities (e.g. pressure and velocity) that cannot be obtained experimentally with today's measurement techniques can readily be deduced from the simulations. Computer experiments are particularly advantageous if a laboratory experiment is either too difficult, too expensive, or too dangerous to conduct. A striking example is our pioneering investigation (Kristoffersen & Andersson 1993) of turbulence in a long rotating duct. This

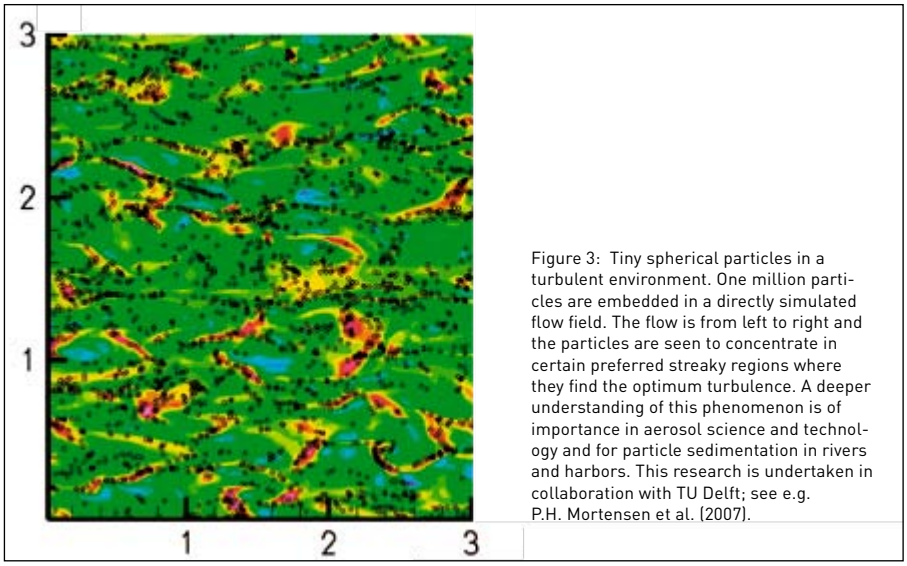


Figure 3: Tiny spherical particles in a turbulent environment. One million particles are embedded in a directly simulated flow field. The flow is from left to right and the particles are seen to concentrate in certain preferred streaky regions where they find the optimum turbulence. A deeper understanding of this phenomenon is of importance in aerosol science and technology and for particle sedimentation in rivers and harbors. This research is undertaken in collaboration with TU Delft; see e.g. P.H. Mortensen et al. (2007).

was also the first ever DNS study undertaken in Scandinavia.

Even today computer experiments (DNS) can only be performed for fluid flow at relatively modest speeds. This limitation is due to the multitude of scales present in fluid turbulence. The smallest dissipative eddies are smaller than the geometrical length scales by a factor N which varies from less than 100 at low flow speeds up to about 100 000 for the turbulence in air flowing over the fuselage of a cruising aircraft or in a pressurized gas pipeline. The numerical integration of the Navier-Stokes equations must therefore be made on a discrete three-dimensional mesh with roughly N grid points in each coordinate direction in order to capture the tiniest whirling eddies. The outcome of a direct numerical simulation will accordingly be discrete values of the four flow variables (three components of the instantaneous velocity vector together with the scalar pressure) in N^3 grid points.

Our first DNSs were run with N = 128 (i.e. $N^3 \approx 2.10^6$ grid points) on the CRAY X-MP/28 in Trondheim and a CRAY-2 at EPFL in Lausanne nearly two decades ago. Today, with access to the NOTUR II infrastructure, we have performed DNS with N typically in the range from 200 to 400. This is way below the DNS with N up to 4096 performed by Kaneda et al. (2003) on the Earth Simulator with peak performance 16.4 Tflops. This is an extreme simulation, but the move towards gradually larger simulations is inevitable, first of all to allow higher-Reynolds-number flows, in which the scale separation is larger, to be explored numerically. In order to stay abreast with international research simulations with N > 500 (more than 10^8 grid points) are being planned.

Multiphysics

In its infancy, DNS was confined to studies of turbulent flows, i.e. pure fluid dynamics. Now, a move towards inclusion of other phenomena can be observed, for instance to include chemical reactions or aeroacoustics in a DNS. That may bring additional time and length scales into the coupled problem and therefore challenge the numerical solution scheme to be used. Our recent simulations of tiny particles (e.g. aerosols) in a DNS is a noteworhty example. Here, the particles were governed by Newton's law of motion and their motion in the turbulent flow field were obtained in a Lagrangian manner by solving coupled sets of ordinary differential equations for each and every particle. The snapshot in the above figure shows that the particles tend to cluster in certain regions near the wall. This preferential concentration of particles has been observed before in laboratory measurements, but the computer simulation enables an indepth investigation of this phenomenon.

Outlook

In most areas of computational science and engineering, the physical realism of a computer simulation improves when a simple mathematical model is replaced by a more complex model. This is not the situation in computational turbulence where the classical Navier-Stokes model is generally accepted as a complete mathematical representation of the physics of turbulence.

Computational turbulence or DNS has evolved into the third direction in turbulence research and complements theoretical deductions and laboratory experiments. The multiscale nature of fluid turbulence makes DNS computationally demanding, both with respect to computer processor speed and storage capacity. The international research trend is a gradual move in either of three directions: towards higher Reynolds numbers, more complex flow configurations and coupled (multiphysics) problems. Easy accessibility to high-performance computers is therefore of uttermost importance in this field of research. The need for efficient techniques for extraction and visualisation of information from the huge data sets generated is becoming increasingly important.

Computational turbulence research is a viable route to traverse for physicists and fluid dynamicists in order to tackle many of the challenges that remain in order to fully comprehend the variety of turbulent flow phenomena.

Bibliography

A. Ashrafian & H.I. Andersson (2006) "The structure of turbulence in a rod-roughened channel", International Journal of Heat and Fluid Flow, Vol. 27, pp. 65-79.

P.A. Durbin & B.A. Pettersson Reif (2001) "Statistical Theory and Modeling for Turbulent Flows", John Wiley & Sons.

Y. Kaneda, T. Ishihara, M. Yukokawa, K. Itakura & A. Uno (2003) "Energy dissipation rate and energy spectrum in high resolution direct numerical simulations of turbulence", Physics of Fluids, Vol. 15, pp. L21-L24.

R. Kristoffersen & H.I. Andersson (1993) "Direct simulations of low-Reynolds-number turbulent flow in a rotating channel", Journal of Fluid Mechanics, Vol. 256, pp. 163-197.

P. Moin & K. Manesh (1998) "Direct numerical simulation: a tool in turbulence research", Annual Reviews of Fluid Mechanics, Vol. 30, pp. 539-578.

P.H. Mortensen, H.I. Andersson, J.J.J. Gillissen & B.J. Boersma (2007) "Particle spin in a turbulent shear flow", Physics of Fluids, Vol. 19, 078109.

S.B. Pope (2000) "Turbulent Flows", Cambridge University Press.

O. Reynolds (1883) "An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous and of the law of resistance in parallel channels", Philosophical Transactions of the Royal Society, Part III, pp. 935-982.

V.D Narasimhamurthy, H.I. Andersson & B. Pettersen (2007) "Direct numerical simulation of vortex shedding behind a linearly tapered circular cylinder", IUTAM Symposium on Unsteady Separated Flows and Their Control (CD rom).

P. Sagaut (2004) "Turbulence direct numerical simulation and large-eddy simulation", Chapter 9 in Volume 3 "Encyclopedia of Computational Mechanics", John Wiley & Sons.

Large-scale comparison of biological sequences

The complete genome sequences of many hundred organisms are now known, and the number is growing rapidly. Within each genome, there are many genes. Humans seem to have about 25 000 genes. It is a huge challenge to determine the precise function of all these genes. Knowing their functions will help understanding what happens if the genes are changed, and help fight associated diseases. As a result of evolution, many of the genes of the different organisms are related to each other. Because related genes, known as homologs, often have a similar function in the cells, it makes life easier for biologists knowing which genes are related before they start out with laboratory experiments. Many homologs can be identified due to similar DNA or protein sequences. However, the similarities may be subtle, making it difficult or impossible to detect. To identify the relationships between all known genes we need to compare the sequences of all the genes with each other. This is only possible in reasonable time using large computational resources.

AUTHOR
Torbjørn Rognes
Department of Informatics, University of Oslo
Centre for Molecular Biology and Neuroscience (CMBN), Institute of Medical Microbiology, Rikshospitalet
PubGene Inc., Oslo

Genes are evolving. New genes appear and old genes disappear. By duplicating an existing gene, a new gene is created. Initially, the two gene copies will be very similar both in sequence and in function, but during the course of evolution they might evolve to fill different roles in the cells and their sequences will also diverge. As new organisms appear, the sets of genes in the different organisms will also evolve differently. Genes corresponding to each other in different organisms are called orthologs. One example of a set of orthologs is the gene called uracil DNA glycosylase, UDG, which is present both in humans and bacteria, as well as in almost all other organisms. The UDG genes have very similar functions across all these organisms, namely to remove certain illegal bases from the genomic DNA. Figure 1 shows structural models of the protein encoded by the UDG gene in man and in the bacteria Escherichia coli. As can be seen from the figure, these proteins look almost identical. Even if this protein had been studied only in the bacterium, it would be possible to make a good guess about what kind of function it would have in humans. In this case, the genes are exceptionally well conserved. In most cases the similarity will be less obvious. This kind of functional inference is very powerful. But it is not foolproof: sometimes small changes can have a large impact, and the function of the gene can be quite different.

Because one cannot do experiments on humans, molecular biologists have for a long time been using a range of so-called model organisms to study the effect of mutations and deletions of genes, in order to better understand the mechanisms of human genetic disorders. Model organisms include several bacteria, flies, yeasts, plants, as well as rats and mice. In order to relate experiments in the model organisms to their assumed effect in humans it is essential to know which genes in these organisms correspond to which genes in humans.

Unless two related genes have diverged too far, there will still be significant similarity between their sequences. Figure 2 shows a sequence alignment of the amino acid sequences corresponding to the two proteins shown in figure 1. This so-called optimal local alignment represents the best fit between parts of the two sequences, maximising the alignment score. The score is computed by adding up scores for each pair of amino acids as well as deducting penalties for insertion of gaps into the aligned sequences. The probability that the two sequences are evolutionary related is then calculated based on the resulting alignment score.

To identify all genes that are significantly similar to a given gene, database search

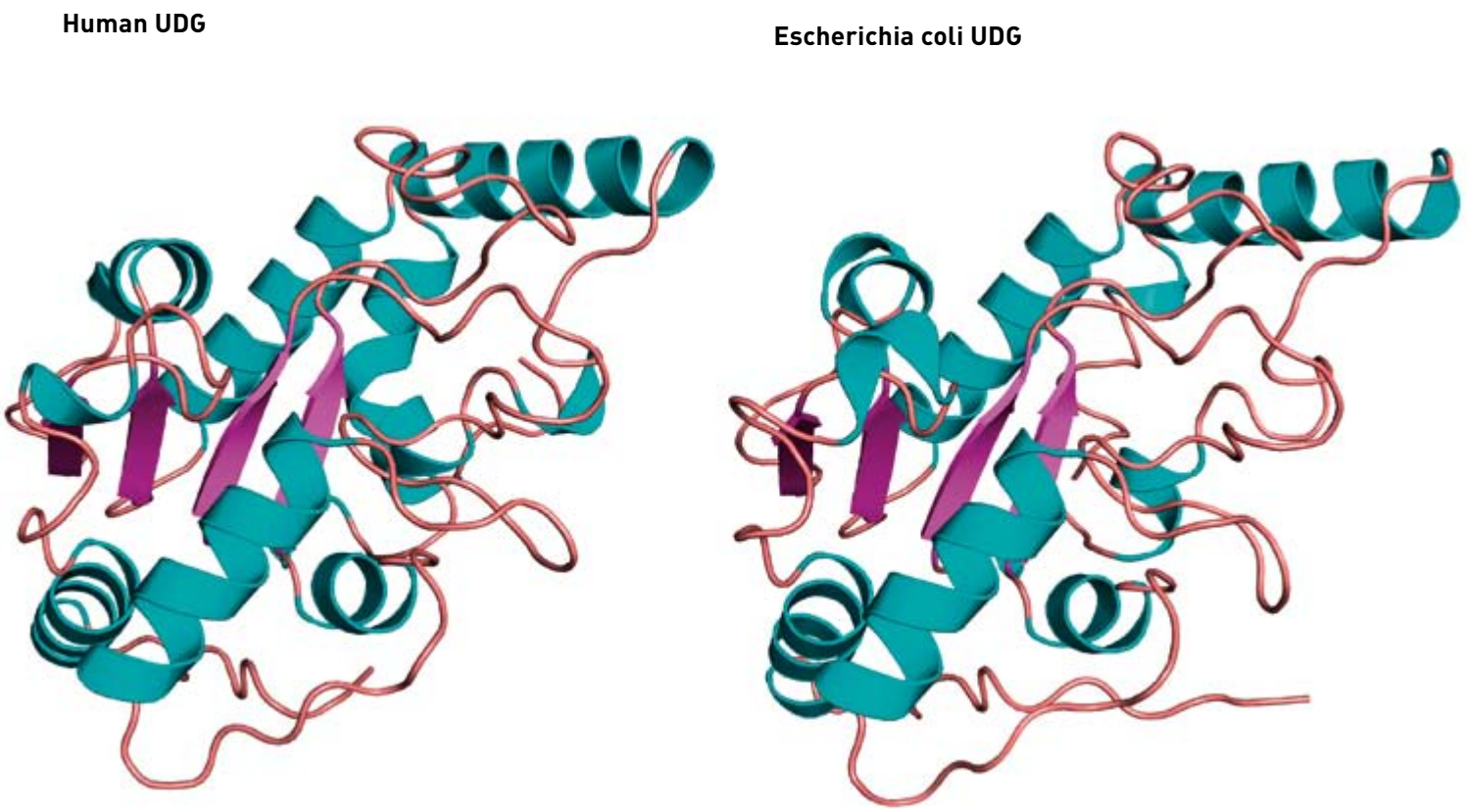


Figure 1. Structural models of the protein encoded by the UDG gene from humans (PDB ID: 1AKZ) and the bacterium *Escherichia coli* (PDB ID: 1EUI). These genes are corresponding genes, or orthologs, in the two organisms. There is a striking similarity between the structures and the functions of the genes.

programs are used. These programs work by comparing a given query sequence to each of the database sequences and determining the alignment score and significance for each database sequence. The database sequences with the highest scores and significant similarity are then identified. Methods for carrying out searches have been developed since around 1980 and are known by names like BLAST, FASTA and Smith-Waterman. In order to make sure that the optimal alignment is found, a large number of possible alignments must be examined. Some programs are guaranteed to find the mathematical optimal alignment, while others take shortcuts to save time, at the expense of a little sensitivity.

The latest release of the GenBank database that contains most of the publicly available sequence data contains more than 100 million nucleotide sequences with a total of more than 180 billion nucleotides. These sequences encode more than 5 million proteins with nearly 2 billion amino acids in total. Storing the entire GenBank including both the actual sequences and their associated information currently requires about 300GB. Steady progress in the technology for sequencing DNA has made the amount of biological sequence data growing exponentially. The size of GenBank has been doubling in about every 18 months for more than two decades. It actually seems to be growing faster than the improvements in computer technology as predicted by Moore's law, thus increasing the time or cost for linear searches in the entire database, as well as increasing the cost of storing it. This trend will continue for sure. Recent advances in sequencing technology might even increase the rate in the future. Both storing and searching these data therefore poses great challenges.

We would like to know which genes are related to other genes. We would also like to know which genes correspond to each other in different species. Based on this we can build networks of related genes, and tables of corresponding genes. We need to compare all genes with each other in order to achieve this.

Comparing every gene in a large database with every other gene in the same database using an accurate comparison method will take a long time. Parallel processing is therefore employed at different levels. The commonly used BLAST program works well on several microprocessor cores in a single compute node. We have developed the PARALIGN software to exploit parallelism within a single core (single instruction multiple data technology), as well as between nodes in a cluster. We are now employing the Titan II cluster at the University of Oslo to carry out such large scale searches. Comparing all the 5 million protein sequences with each other using BLAST is expected to require on the order of 10 cpu years of computing time. And since the databases are growing every day, that's just the beginning.

We are grateful to NOTUR and USIT for supporting this project.

Centre for Molecular Biology and Neuroscience

The Centre for Molecular Biology and Neuroscience is a Norwegian Centre of Excellence at the University of Oslo and Rikshospitalet-Radiumhospitalet Medical Centre. CMBN shall take on a leading role in elucidating the role of DNA repair and genome maintenance mechanisms in preventing neurological disease and brain ageing.
The Centre will develop and apply stem cell technology and targeted repair to broaden the range of therapeutic strategies in neurological disease.

Read more about CMBN at:
http://www.cmbn.no/

Thermus thermophilus UDG protein bound to DNA containing an abasic site. The illustration has been generated with PyMOL (http://www.pymol.org) from the Protein Databank structure 2DDG of the RIKEN Structural Genomics/Proteomics Initiative.
Copyright:Jon K. Lærdahl, CMBN, Rikshospitalet HF.

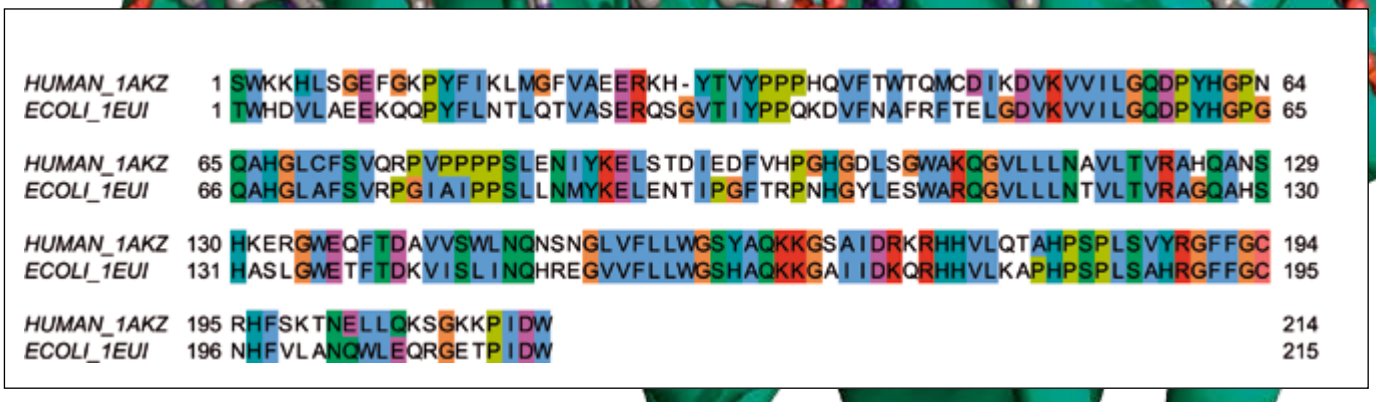


Figure 2. Sequence alignment of the amino acid sequences of the two proteins shown in figure 1, indicating the high level of similarity at the sequence level. Identical residues are indicated with a vertical bar (|), while similar residues are indicated by a plus sign (+).

Computational Quantum Mechanics

A theoretical understanding of the behavior of quantum mechanical systems with many interacting particles, normally called many-body systems, is a great challenge and provides fundamental insights into systems governed by quantum mechanics, as well as offering potential areas of industrial applications, from semi-conductor physics to the construction of quantum gates. The capability to simulate quantum mechanical systems with many interacting particles is crucial for advances in such rapidly developing fields like materials science.

AUTHOR

Morten Hjorth-Jensen
Professor in Physics
Department of Physics and
Center of Mathematics for
Applications
University of Oslo

Most quantum mechanical systems of interest in physics consist of a large number of interacting particles. The total number of particles N is usually sufficiently large that an exact solution (viz., in closed form) cannot be found. One needs therefore reliable numerical methods for studying quantum mechanical systems with many particles.

Computational quantum mechanics is thus a field of research which deals with the development of stable algorithms and numerical methods for solving Schrödinger's or Dirac's equations for many interacting particles, in order to gain information about a given system. Typical examples of popular manybody methods are coupled-cluster methods, various types of Monte Carlo methods, perturbative expansions, Green's function methods, the density-matrix renormalization group, ab initio density functional theory and largescale diagonalization methods. The numerical algorithms cover a broad range of mathematical methods, from linear algebra problems to Monte Carlo simulations.

Studies of many-body systems span from our understanding of the strong force with quarks and gluons as degrees of freedom, the spectacular macroscopic manifestations of quantal phenomena such as Bose-Einstein condensation with millions of atoms forming a coherent state (see Fig. 1), to properties of new materials, with electrons as effective degrees of freedom. The length scales range from few micrometers and nanometers, typical scales met in materials science, to 10^{-15}–10^{-18} m, a relevant length scale for the strong interaction. Energies can span from few meV to GeV or even TeV. In some cases the basic interaction between the interacting particles is well-known. A good example is the Coulomb force, familiar from studies of atoms, molecules and condensed matter physics. In other cases, such as for the strong interaction between neutrons and protons (commonly dubbed as nucleons) or dense quantum liquids one has to resort to parameterizations of the underlying interparticle interactions. But the system can also span over much larger dimensions as well, with neutron stars as one of the classical objects. This star is the endpoint of massive stars which have used up their fuel. A neutron star, as its name suggests, is composed mainly of neutrons, with a small fraction of protons and probably quarks in its inner parts. The star is extremely dense and compact, with a radius of approximately 10 km and a mass which is roughly 1.5 times that of our sun. The quantum mechanical pressure which is set up by the interacting particles counteracts the gravitational forces, hindering thus a gravitational collapse. To describe a neutron star one needs to solve Schrödinger's equation for approximately 10^{54} interacting particles! A possible composition of matter for a neutron star is shown in Fig. 2.

With a given interparticle potential and the kinetic energy of the system, one can in turn define the so-called many-particle Hamiltonian H which enters the solution of Schrödinger's equation or Dirac's equation in case relativistic effects need to be included. For many particles, Schrödinger's equation is an integro-differential equation whose complexity increases exponentially with increasing numbers of particles and states that the system can access. Unfortunately, apart from some few analytically solvable problems and one and two-particle systems that can be treated numerically exactly via the solution of sets of partial

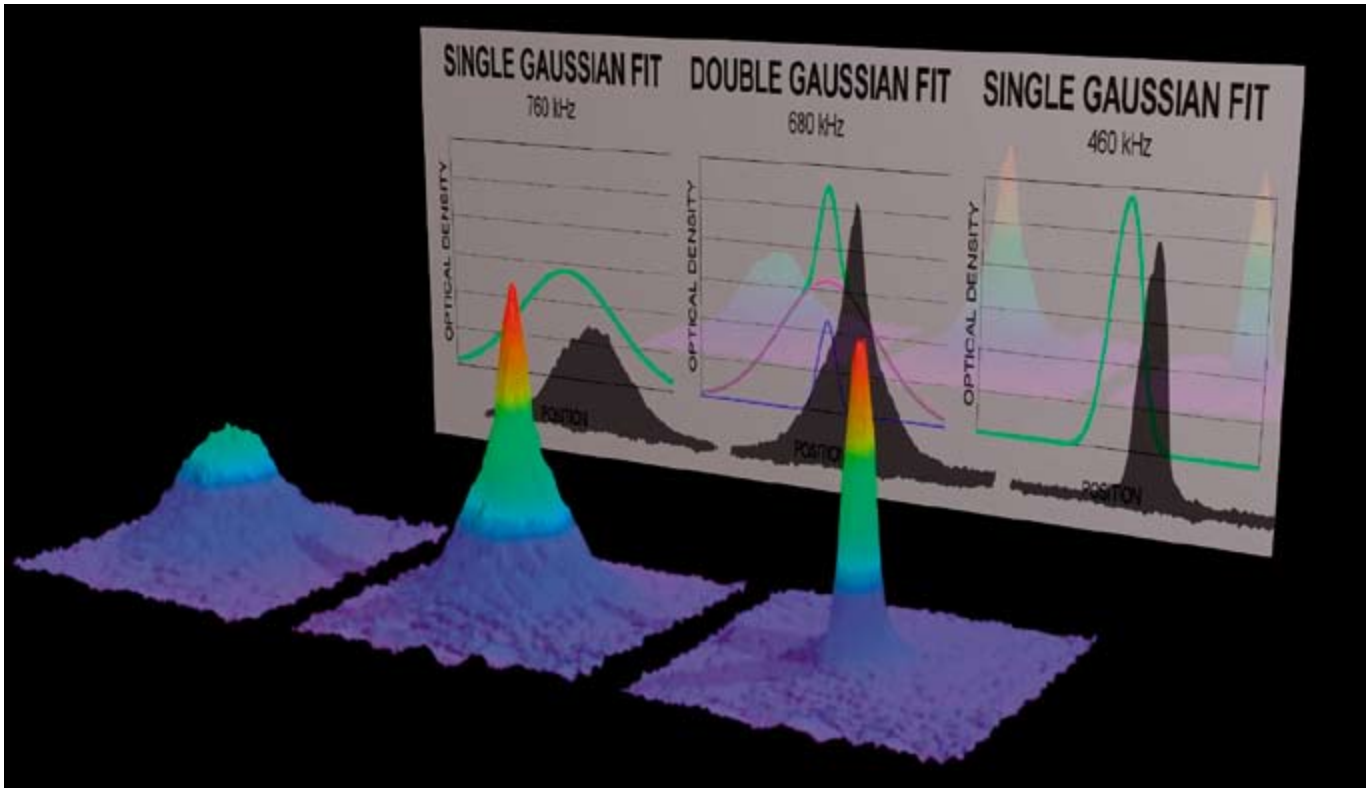


Figure 1: Observation of Bose-Einstein condensation by absorption imaging. Shown is absorption vs. two spatial dimensions. The top row shows shadow pictures, which, in the lower row, are rendered in a three-dimensional plot where the blackness of the shadow is represented by height. The "sharp peak" is the Bose-Einstein condensate, characterized by its slow expansion observed after 6 msec time of flight. The left picture shows an expanding cloud cooled to just above the transition point; middle:just after the condensate appeared; right: after further evaporative cooling has left an almost pure condensate. The width of the images is 1.0 mm. The total number of atoms at the phase transition is about 700,000, the temperature at the transition point is 2 microkelvin. This figure and its corresponding text are taken from http://cua.mit.edu/ketterle_group/Popular_papers/Ultralow_temperatures.htm.

differential equations, the typical absence of an exactly solvable contribution to the many-particle Hamiltonian means that we need reliable numerical many-body methods. These methods should allow for controlled approximations and provide a computational scheme which accounts for successive many-body corrections in a systematic way.

To understand the large dimensionalities involved, consider the following two examples. As an example from the nuclear many-body problem, we have Schrödinger's equation as a differential equation

$$\hat{H}\Psi(\mathbf{r}_1, .., \mathbf{r}_A, \alpha_1, .., \alpha_A) = E\Psi(\mathbf{r}_1, .., \mathbf{r}_A, \alpha_1, .., \alpha_A)$$

where $\mathbf{r}_1, .., \mathbf{r}_A$ are the coordinates and $\alpha_1, .., \alpha_A$ are sets of relevant quantum numbers such as spin and isospin for a system of A nucleons with $A = \nu+\pi$, ν being the number of neutrons and π the number of protons. The Hamiltonian is $\hat{H}$. It acts on the many-body wave function $\Psi(\mathbf{r}_1, .., \mathbf{r}_A, \alpha 1; .., \alpha_A)$ and yields an eigenvalue problem with energies E and the wave functions Ψ as the set of eigenvectors. We can rephrase this equation as a set of coupled second-order ordinary differential equations. There are

$$2^A \times \binom{A}{\pi}$$

coupled second-order differential equations in 3A dimensions. For a nucleus like ^{10}Be, with four protons and six neutrons, this number is **215040**. Methods like partial differential equations can at most be used for 2-3 particles and are of a limited value in computational quantum mechanics, unless one can reduce the problem to an effective one-particle problem. The latter has an applicability only for systems with weak interactions and/or where single-particle degrees of freedom are predominant. Strongly interacting cases such as those found in nuclear physics or quantum liquids have to be treated in full glory.

The other example is taken from variational Monte Carlo studies, where one can approximate the expectation value of a given Hamiltonian, resulting in the energy of the system, by a multi-dimensional integral over the various coordinates. For N particles, we can write this expectation value as

$$\langle H \rangle = \frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r} \Psi^*(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) H(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)}{\int d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N \psi^*(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)},$$

an in general intractable problem. The function ψ is a guess for the exact solution to Schrödinger's equation. Via the variational principle, one can then improve the guess until one obtains the best possible energy minimum. Hopefully then the variance of the above expectation value is zero. Then one has the exact solution. This

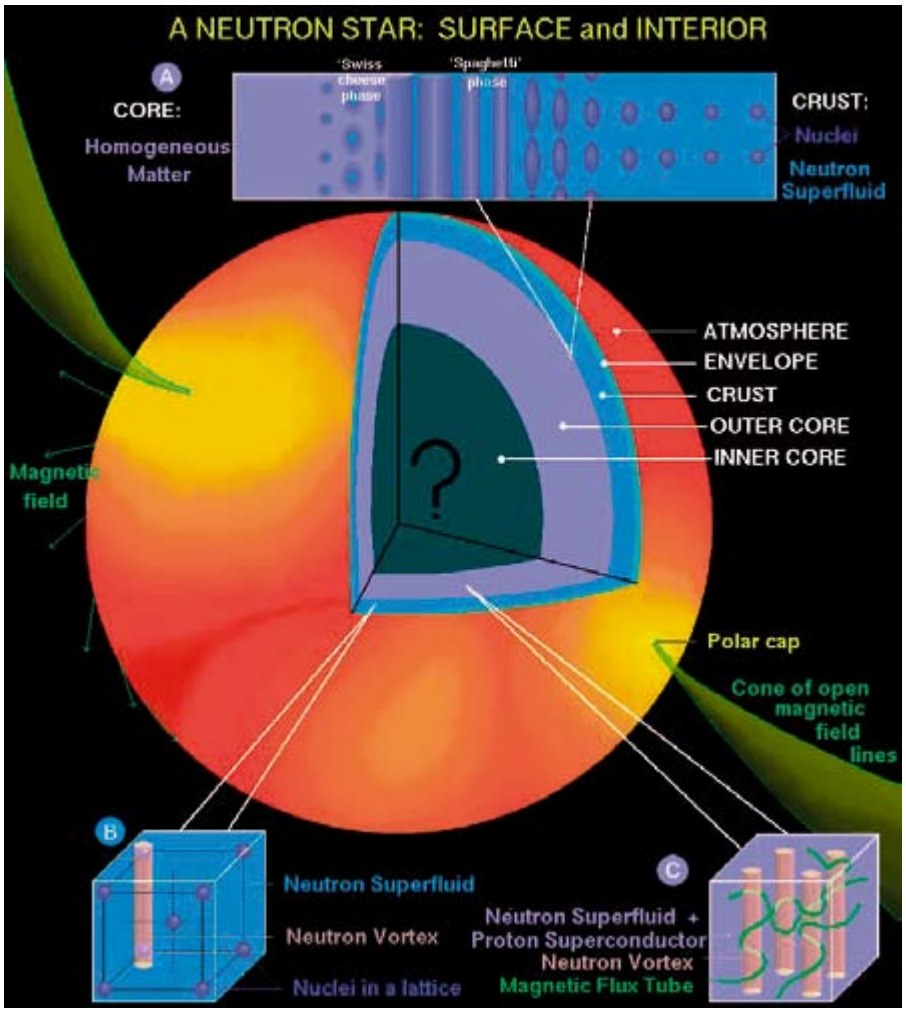


Figure 2: Tentative composition of matter inside a neutron star.

integral is actually the starting point in a Variational Monte Carlo calculation, since for more than five to six dimensions, we can forget methods based on Gaussian quadrature. With 10 particles and 10 mesh points (a small and insufficient number in most applications) for each degree of freedom and an ideal 1 Tflops machine (all operations take the same time), the reader can easily convince himself/herself that the time time needed for any quadrature method quickly exceeds the lifetime of the universe, $T \approx 4.7 \times 10^{17}$s. The integral has to be evaluated by stochastic methods such as Monte Carlo methods.

Both examples demonstrate the dimensionality problem one has to face when studying quantum mechanical systems. In the rest of this article I will focus on many-body methods applied to problems in nuclear physics, a field where comparison between experiment and exact numerical results, is still lagging behind the precision obtained in fields like quantum chemistry. One major reason for this is the lack of a proper knowledge of the underlying interaction. In nuclear physics, the number of degrees of freedom normally exceed those encountered in quantum chemistry problems. In addition, the interaction itself is not known analytically and three-body forces and correlations play an important role. The latter provide up to 10 % of the total binding energy.

Fig. 3 illustrates schematically the challenges one finds in nuclear physics, where the aim is to understand the structure of a nucleus (at the end of the road) in terms of the underlying interactions. Theory and experiment go hand in hand in this endavour, with ramifications to many other fields, with astrophysical applications as a growing field. The various acronyms stand for various few and many-body methods which are used in nuclear physics.

The nuclear many-body is perhaps the worst case scenario when it comes to the applicability of many-body methods. To complicate life even more, current experimental programs in low-energy nuclear physics focus on nuclei close to the stability line. This means that one needs to account for the fact that many of the nuclei close to the stability line can be weakly bound and therefore the nuclear interactions will couple bound, continuum, and resonant states. This leads to a further increase in dimensionality compared with stable nuclei. The current worldwide experimental programs address topics which are of great relevance for our understanding of the stability of matter itself, with applications and relevance to many other fields, such a the synthesis of the elements, how stars end their lives, electroweak interactions with matter and so forth. Fig. 4 indicates some of this relevance along the nuclear chart.

In order to tackle these challenges, we have chosen to focus on coupled cluster methods in our discussion of systems involving many single-particle degrees of freedom. The ab initio coupled-cluster theory is a particularly promising candidate for such endeavors due to its enormous success in quantum chemistry. Our national quantum chemistry groups, recently awarded grants for a center of excellence on computational chemistry, have been working on coupled cluster methods for more than a decade. Coupled cluster methods results in amazingly precise estimates for systems governed by electrostatic interactions with approximately 100-200 electrons. For larger systems one resorts often to density functional theory, which seldomly goes beyond mean-field approaches (independent particle approximations). The price is often a loss of a quantitative predicability.

Coupled cluster methods allow to study ground- and excited-state properties of systems with dimensionalities beyond the capability of present large scale diagonalization approaches, with a much smaller numerical effort when compared to diagonalization methods aiming at similar accuracies. Our hope is then that coupled cluster methods, which actually originated from nuclear physics but have found large areas of applications and theoretical developlements in quantum chemistry, can be used to shed light on different many-body correlations in nuclear physics. With a given approach to a many-body Hamiltonian, we hope then to be able to extract enough information to claim that an eventual disagreement with experiment is due to missing physics in the starting Hamiltonian.

The aim of our recent many-body studies is therefore to delineate a many-body scheme for nuclear physics problems based on coupled-cluster theory that incorporates as many as possible of the following features. The theory

- should be fully microscopic and start with present two- and threebody interactions derived from, for example, effective field theory or eventually lattice quantum-chromodynamics. The latter is most likely possible by 2013-2015 with the advent of peta-scale computing. One may then have enough statistics to reliably compute various partial wave contributions within the framework of lattice QCD. From lattice QCD calculations, one can constrain the off-shell character of the nucleon-nucleon interaction from the fundamental QCD Lagrangian, removing thereby several ambiguities in the nuclear many-body problem. Presently, we have several methods which allow almost exact solutions of the nuclear many-body problem. The large uncertainty resides in the derivation of the underlying interactions.

- The theory can be improved upon systematically by the inclusion of three-body interactions and more complicated correlations.

- It allows for the description of both closed-shell systems and valence systems.

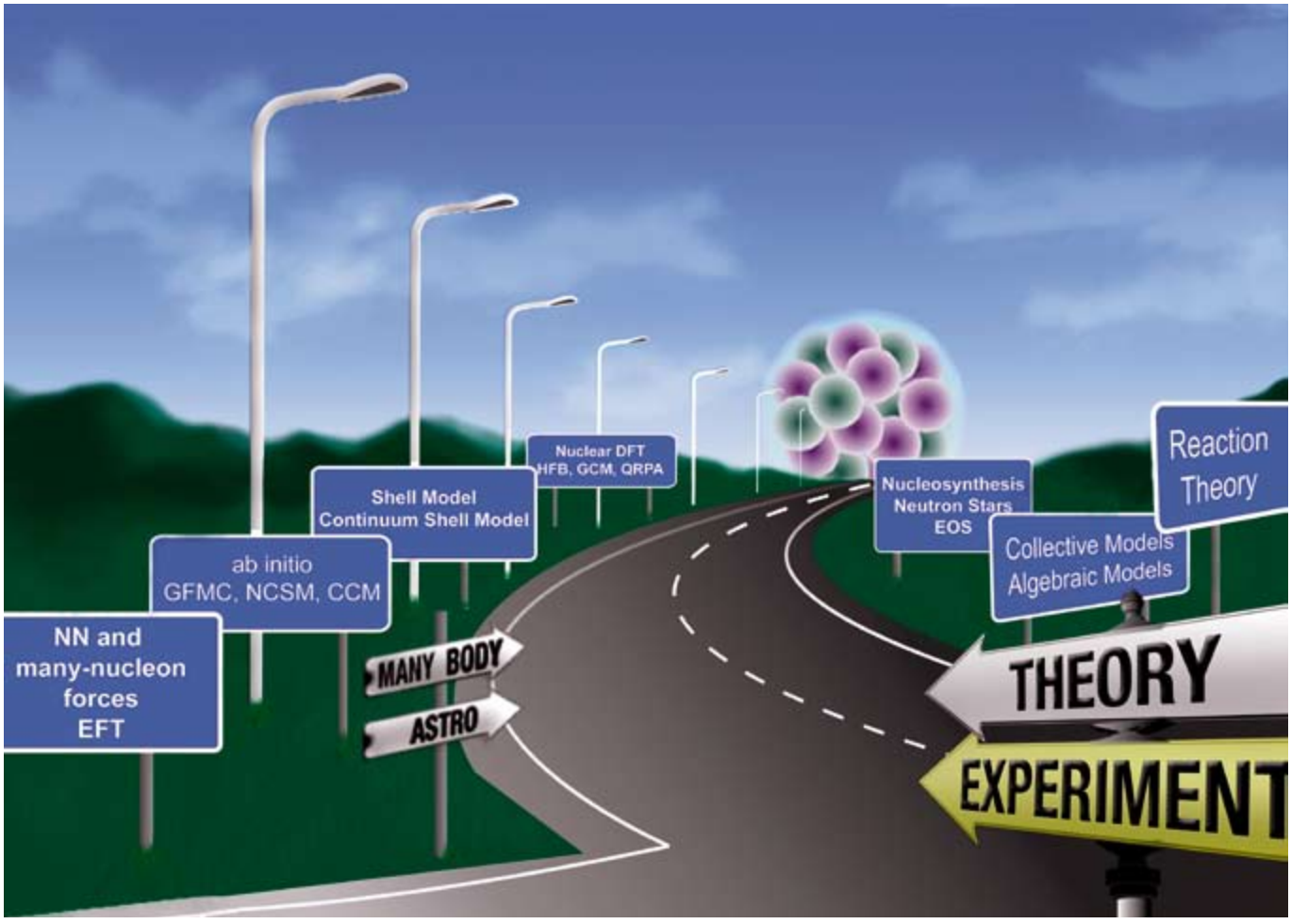


Figure 3: The plethora of many-body methods and applications, with emphasis on the nuclear many-body problem. This figure portrays a kind of roadmap for nuclear physics studies.

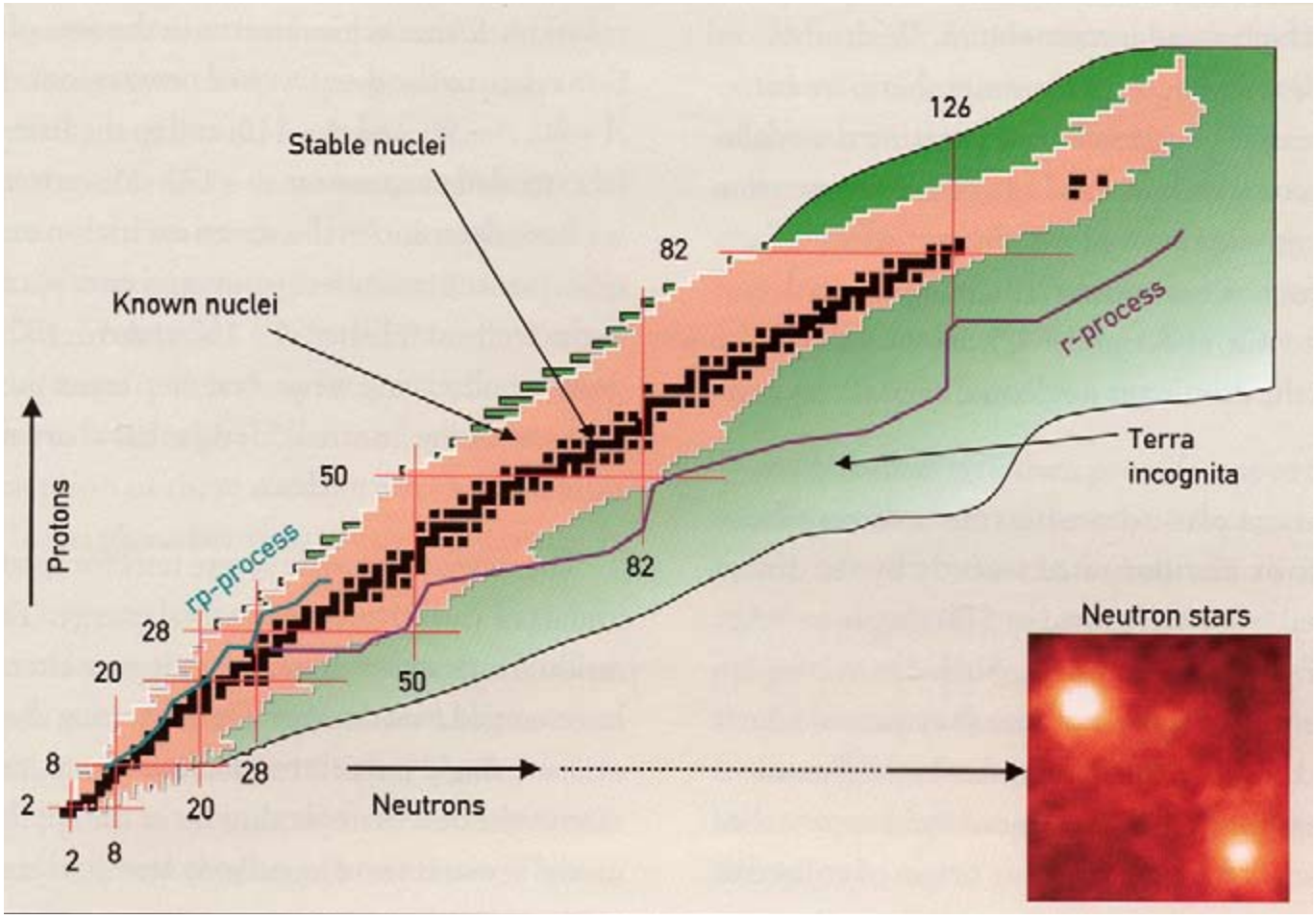


Figure 4: Chart of stable and unstable nuclei, with links to important applications.

- It is amenable to parallel computing. It can be used to generate excited spectra for nuclei where many shells are involved and describe weakly bound systems with or without resonances and couplings to the continuum.

- Enables the derivation of effective interactions to be used in reduced space appropriate for large-scale diagonalization techniques with matrices of dimensionalities 10^{10} x 10^{10}.

- Enables microscopic nuclear structure results to be married with microscopic reaction studies.

We present here recent results from a large-scale coupled-cluster calculation for Helium isotopes, for more details see Refs. [1, 2, 3, 4]. The calculations were mainly carried out at the Jaguar supercomputing facility at Oak Ridge National Laboratory. The newly upgraded Cray XT4/XT3 supercomputer has a computing power of 101.7 Teraflops, resulting in second place on the top500 computers list of June 2007. In our largest runs we used between 1000 to 2000 nodes.

Fig. 5 presents coupled-cluster results with to two-particle-two-hole correlations (so-called singles and doubles) for ground state energies of the $^{3-10}$He isotopes for increasing number of partial waves. In our largest calculation we include 5*s*5*p*5*d*4*f*4*g*4*h*4*i* proton orbitals and 20*s*20*p*5*d*4*f*4*g*4*h*4*i* neutron orbitals, with a complex basis in order to reproduce eventual resonances. Our calculations show excellent convergence with respect to the single-particle basis size. We obtain a convergence within 10 keV for the real part and within 0.1 keV for the imaginary part of the ground state energy.

Our largest calculation of ^{10}He with ~850 active single particle orbitals, would correspond to a dimension of ~10^{22} basis states. In a recent work [5] we computed the ground state energy of ^{40}Ca with a single-basis set which results in a dimensionality of 10^{62} basis states. The computed decay widths of the helium isotopes are in semi-quantitative agreement with experiment. The comparison of binding energies shows that ^{5}He and ^{7}He are unstable with respect to one-neutron emission, while ^{8}He is stable with respect to the emission of up to three neutrons. The nucleus ^{6}He is sta-

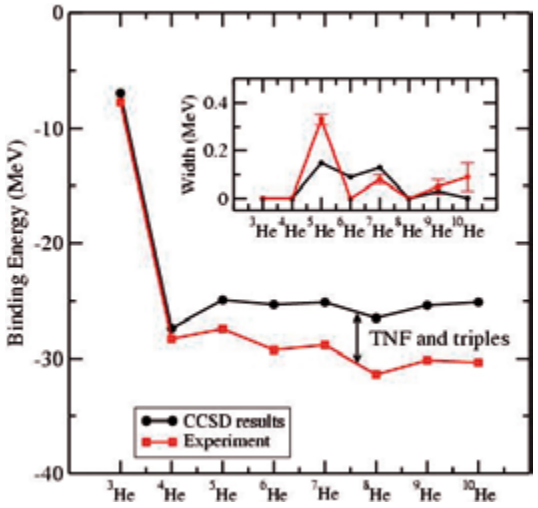


Figure 5: Real (large picture) and imaginary (insert) energies for the chain of Helium isotopes up tp ^{10}He. Experimental values are included. See text for further details.

ble with respect to one-neutron emission but unstable with respect to two-neutron emission. It has a nonzero decay width. All helium isotopes are unstable with respect to ^{4}He plus residual neutrons in the continuum.

Our results here represent the first time that decay widths have been computed in an *ab-initio* way for an isotopic chain. The decay widths of unbound nuclei are in semiquantitative agreement with experimental data, and the binding energies meet expectations for *ab-initio* calculations based on two-body Hamiltonians. The calculated masses follow the experimental pattern in that ^{5,7,9}He are unstable with respect to one-neutron emission and ^{6,8}He stable with respect to one-neutron emission. The missing agreement with experiment is probably due to the lack of the inclusion of three-nucleon clusters and three-nucleon forces. With the inclusion of the latter we may hopefully be able to tell how much of the spectrum is driven by a coupling to resonances and the non-resonant continuum and how much is due to possible three-nucleon forces. The latter would aid us in explaining one of the major unresolved problems in low-energy nuclear physics, namely how nuclei evolve towards the line of stability.

Finally, the methods we have developed for studies of weakly bound systems can easily be applied to other quantum mechanical systems where resonances play important roles.

Acknowledgments

I am much indebted to my collaborators Torgeir Engeland (UiO), David Dean, Gaute Hagen and Thomas Papenbrock, all Oak Ridge National Laboratory. In addition, I ackowledge many discussions with Mihai Horoi, Takahura Otsuka and Piotr Piecuch. The reported results were supported in part by the U.S. Department of Energy under Contract Nos. DE-AC05-00OR22725 (Oak Ridge National Laboratory, managed by UT (Battelle, LLC), DE-FG02-96ER40963 (University of Tennessee), DE-FG05-87ER40361 (Joint Institute for Heavy Ion Research). Computational resources were provided by the Oak Ridge Leadership Class Computing Facility, the National Energy Research Scientific Computing Facility, and the Research Council of Norway via the Notur project (Supercomputing grant NN2977K).

References

[1] D. J. Dean and M. Hjorth-Jensen. *Phys. Rev. C*, 69:54320, 2004.
[2] K. Kowalski, D. J. Dean, M. Hjorth-Jensen, T. Papenbrock, and P. Piecuch. *Phys. Rev. Lett.*, 92:132501, 2004.
[3] M. Wloch, D. J. Dean, J. R. Gour, M. Hjorth-Jensen, K. Kowalski, T. Papenbrock, and P. Piecuch. *Phys. Rev. Lett.*, 94:132501, 2005.
[4] G. Hagen, D. J Dean, M. Hjorth-Jensen, and T. Papenbrock. *Phys. Lett. B*, in press, 2007.
[5] G. Hagen, D. J Dean, M. Hjorth-Jensen, T. Papenbrock, and A. Schwenk *Phys. Rev. C*, 76:44305, 2007.

Competing in the International Science Race

– Why the University of Tromsø installs a 60 Teraflop supercomputer in November 2007

When the University of Tromsø installs its "Stallo" during November 2007 it will be the largest supercomputer in the Nordic and increase the available compute power in Norway by a factor of 3. This is the result of a strategic focus on modelling and simulation as the "third way" of doing science and is also acknowledging the fact that scientists that make use of supercomputers are able to get their results published faster and in journals with higher impact. In addition to being a workhorse for parallel computations in all disciplines the new system will be aimed specifically towards important fields in theoretical and computational chemistry and for some time give Norwegian researchers an advantage in this highly competitive arena.

"Stallo"

The new supercomputer at the University of Tromsø will be 15 meters long and weigh 16 tons. To reflect the grand dimensions we had to find a suitable name for the installation. The term "Stallo" is known from Sami folklore to be the biggest, fiercest "troll" or "monster" you can imagine. The saga claims that a Stallo can not be outdone by using force or power, only by cleverness and cunning tactics. So users should take notice and be aware of what awaits when the new system is installed. It is also said that a Stallo will always act with rage if a bigger opponent emerges.

The Sea Troll by Theodor Kittelsen (1887)

AUTHOR

Svenn A. Hanssen
Head of Systems
Operations
University of Tromsø

To boldly go

Scientific research is a pillar of society and one of the biggest contributions coming from the universities. From the day men and women sat foot on earth we have had the curiosity to explore, investigate and challenge ideas about different aspects of the world: "To boldly go where no one has gone before..." To followers of Star Trek this is a familiar quote. If you are aiming for outer space, or looking into the behaviour of molecular structures, scientific methods and paradigms are needed to generate information, theories and applications. Advancement of knowledge and theoretical understanding are fundamental to the development of scientific research, and computers are perhaps the most important tools in this process.

The international science race

Research is a highly competitive arena. Of course scientists collaborate and are dependent on a broad "social network" to become successful. To be included in what other outstanding investigators are doing is important if you want to become dominant in your field. One important thing to notice is that this competitive game is not performed only at a national or even European level. It is a global race to get published in top rated journals and to reach a high level of citations. To get to this level you have to recruit the best brains and have the best tools available, including supercomputers.

In 2006 the University of Tromsø published the first annual "Research Report". In this document every published article by all the universities in Norway were counted and rated depending on the level of the publication. In short there are two levels: Level one which gives 1 publication point, and level two which gives 3 publication points per article. Because each publication point generate income to the university the higher impact level 2 publications become very important.

University	Points	Level 1	Level 2
NTNU	1842,3	81,9%	18,1%
UMB	381,0	84,1%	15,9%
UIB	1600,8	79,7%	20,3%
UIO	2868,7	78,6%	21,4%
UIS	232,7	92,7%	7,3%
UIT	665,6	83,4%	16,6%
UIT HPC	32,1	51,9%	48,1%

Table 1: Publication points and levels 2005 (Copyright: UiT)

In 2005 the University of Tromsø on average produced only 16,6% level 2 publications [Table 1], but when we looked isolated at the production related to the users of supercomputers in Tromsø (UIT HPC) the results were significantly better. The percentage of level 2 publications achieved with the use of the supercomputer in Tromsø during 2005 was 48,1. A majority of the publications were made in the area of chemistry.

Building locomotives

To the University of Tromsø this fact did not go unnoticed. For a relatively small university that covers every classical discipline it is quite easy to get too sparse when it comes to research. It is therefore important to create locomotives that motivate others and pull the organisation in the right direction. In December 2006 decades of hard work from the Chemistry department, in close collaboration with influential research groups at the University of Oslo, resulted in such a locomotive: A Centre of Excellence in Theoretical and Computational Chemistry (CTCC).

CTCC covers a broad range of research. Theory and methods in quantum chemistry as well as application development are equally important to the success of the new Centre. The research includes linear scaling and periodic boundary condition methods, multiwavelet basis functions, and subsystem optimization. A number of projects are also closely coupled with ongoing experimental work at the two Universities.

Supercomputers are of vital importance to conduct accurate simulations on realistic molecules at CTCC. In 2005 nearly 40% of the total computer time allocated by the research council was used for chemistry related simulations. As the size of the supercomputers increase, so does the level of complexity on the programming level. For CTCC the approximation of the Schrödinger equation is one of the key areas. The development of methods with linear scaling to replace traditional routines with exponential cost will make more efficient use of large parallel supercomputers as well as leapfrog scientific research in Norway to become a driving force in this area.

Creating a competitive advantage

In November 2007 a new 60 teraflop supercomputer aimed directly towards chemistry related simulations will be installed at the University of Tromsø. The new 5632 core system will have 12 Terabyte of distributed memory and more than 200 Terabyte of storage. Our goal is to contribute to make Norwegian scientists the most efficient users of high performance computing. Of vital importance is to create the best support environment for the users and present them to the most efficient solutions to support their highest demands in HPC. Our vision is that Norwegian researchers shall have a competitive advantage in the international science race.

About the new Cluster

"Stallo" will have a theoretical peak performance of 60 teraflop/s (trillion floating point operations per second). It will consist of 704 HP BL460c blade computers each with two 2,66 GHz Intel Xeon X5355 quad core processors, giving a total of 5632 cores. 50 of the compute nodes will have 32 GB memory each, while the rest will have 16 GB. 384 compute nodes will be interconnected by Infiniband. Centralised storage is based on HP Scalable File Share with a total volume of 128 TB and a total I/O throughput of 6 GB/s.

The user environment is based on ROCKS Linux and will be more or less the same as the existing clusters in Norway with regards to operating system, compilers and batch system.

UIT is building a new 6 million NOK machine room to host the system. Because of the extreme power density water cooled racks had to be installed instead of traditional room cooling. The location of the system enables reuse of some of the 370 kW excess heat to warm up the surrounding building complex.

TECHNICAL SPECIFICATION

System name	Stallo
Machine type (node)	HP BL 460c
Processors (node)	2
Processor type	Intel Xeon 2,66 Ghz
Core technology	Quad core
Network	GigE + Infiniband
Compute racks	11
Infrastructure racks	2
Storage racks	4
Compute nodes	704
Compute cores	5632
Internal memory	12 Terabyte
Internal storage	84 Terabyte
Centralized storage	128 Terabyte
Peak performance	60 Teraflop/s
Dimensions (wxdxh)	15x1,3x2 m
Weight	16 tons

Link: http://www.hpc.uit.no/

Titan – a general-purpose high-throughput PC cluster

On October 1st 2007, the compute cluster Titan at the University of Oslo was included in the national HPC project Notur. Titan is an easy-to-use, high-throughput resource for individuals and groups in research and education in Norway.

AUTHOR

Jostein Kandal Sundet,
PhD and Group leader,
University of Oslo

The setup and operation of the PC cluster Titan at the University of Oslo (UiO) has been done in conjunction with an increase in human resources at UiO with focus on scientific computing. The main goal of the Scientific Computing group (SCG) at UiO's Center for Information Technology Services (USIT) is to provide researchers at UiO and collaborating organizations with easy access to computing, storage, and related relevant services. SCG is a competent support group that has experience with a broad range of scientific applications, from social science and economy to medicine and climate simulations. The group currently counts 15 people of which seven hold a PhD degree.

USIT has operated the PC cluster Titan since June 2005. In the first 18 months, it was a system set up for research groups at UiO and USIT that financed the cluster. About one third of the cluster was originally set up with a high speed interconnect (Infiniband) for relatively tightly coupled applications (climate models, cosmic simulations, etc.). In addition to the high speed interconnect, all the 200 nodes (400 CPUs) in the cluster were connected by a 32 Gbit capacity Ethernet ring. For the central disk system, USIT decided to use a parallel file system (GPFS) that enables each researcher to access their files from all compute nodes in the cluster. The initial ten research groups (medicine, economy, astrophysics, bio-informatics, atmospheric modeling, chemistry, and legislative politics) used a wide variety of applications and each with their own usage pattern. It soon became apparent that the utilization of the PC cluster could be improved by using a back-fill technique that allows the execution of tasks on otherwise unused CPU resources and, an "overflow queue" that consisted of all the nodes not currently in use were set up.

> THE PURPOSE OF TITAN IS TO HAVE AN EASY TO USE, HIGH THROUGHPUT RESOURCE FOR THE INDIVIDUAL NOTUR AND UIO RESEARCHER.

Titan2 back view
(Copyright: UiO)

Sun Grid Engine (SGE)

Executing applications on Titan is, as on all the Notur facilities, done through a batch system that allows multiple applications to use the system simultaneously using disjoint sets of processors. A batch system uses advanced algorithms to map (parallel) applications onto the available processors thereby aiming to optimize job throughput while also allowing the prioritization of jobs. Titan uses the open-source Sun Grid Engine (SGE). The user has only few job classes (or queues) to select from. The default development and execution environment on Titan uses the Pathscale compiler suite and Scali's native implementation of the Message Passing Interface (ScaMPI). Experience shows that this combination produces in general the most efficient executables on Titan.

As a result, a user on Titan can execute an application on whatever resources there are unused at any time, but with the risk that the owners of nodes can claim the nodes before the application of the user has been completed. In practice, this setup has been very successful and usage of Titan soared from 30% to more than 80% of the theoretical capacity within a couple of months after it was implemented.

After nine months of operation, the demand exhausted the capacity and the need to expand Titan was evident. Late December 2006, 225 new AMD-based SUN X2200 nodes were delivered and put in operation in February 2007 as Titan2. Around the same time, Notur decided to partake in Titan and it was decided that Titan would be further expanded and included in the national HPC project Notur such that also researchers outside UiO and its collaborating partners would get access to the resource. Titan is currently a 450 node dual-CPU dual-core AMD cluster (in total 1800 cores) in addition to the original nodes (350 cores). Titan has currently 20 different owners of which Notur has the largest share. In addition to being a general computational resource, Titan is also an important resource for grid computing, with 90 TB storage and 128 cores available for the Nordic Tier-1 center that is part of the CERN WLCG collaboration.

A cluster is a "moving target" that undergoes continuous changes and upgrades which eventually makes it an inhomogeneous resource. This is also true for Titan. Currently, there is 4 Gigabyte per core. Early December, the AMD dual-core CPUs will be replaced by the new quad-core CPUs (2.2 GHz Barcelona). Each node will then have 8 cores with 2 Gigabyte of memory per core. Since many of the other owners (besides Notur) of Titan steadily need to expand their computational resource, the CPU pool will continue to grow. We expect that by the end of 2007 Titan will consist of about 4500 cores , with more than 7 Terabyte of memory and about 400 Terabyte of local disk. Most of the nodes will be connected with a high speed interconnect. Central storage will be about 200 Terabyte, which includes the 150 TB Tier-1 storage.

TECHNICAL SPECIFICATION, CURRENT (UPGRADE)	
System name	Titan
Machine type (compute node)	SUN Fire X2200m2
Machine type (server node)	SUN Fire X4600m2
Processors (node)	2
Processor type	AMD 2.6GHz (AMD 2.2GHz)
Core technology	dual core (quad-core)
Network	Gigabit Ethernet, Infiniband
Compute racks	14
Infrastructure racks	3
Storage racks	3
Compute nodes	450
Compute cores	1900 (3700)
Internal memory	7 Terabytes
Internal storage	340 Terabytes
Centralized storage	160 Terabytes
Peak performance	13 Teraflop/s (35 Teraflops)
Dimensions (wxdxh)	13x1x2 m

Relevant links

Technical support pages for Titan can be found at **http://docs.notur.no**.
Queries related to Titan can be sent to help-desk support at **support-uio@notur.no**.
Statistics about status and current usage of Titan can be found on **http://titan.uio.no/**

See http:**//www.notur.no/quotas** on how to apply for access.

Titan2 consists of 1U SUN Fire X2200m2 nodes, 32 nodes per rack - note the infiniband switch on top (Copyright: UiO)

notur www.notur.no

Return address: UNINETT Sigma AS, 7465 Trondheim, NORWAY

Advanced user support: call for applications

Applications for advanced user support by the Notur project can be submitted continuously until **February 28, 2008**. Applicants will be notified of the result of the evaluation within four weeks after submission of the application.

User survey 2007

A User survey was conducted among the Notur user community earlier this year. The results were extracted from the 41 questionnaires that were returned by projects that had access to the Notur supercomputer facilities in 2006. The 41 answers represent 65% of the 63 projects that were contacted.

93% of the respondents consider their applications time demanding, while 68% of the respondents consider their applications also memory demanding. High machine availability and reliability and high overall processing speed are important for the applications. Large overall memory size, processor interconnect and large temporary storage capacity at run-time are also considered important for the applications.

Overall satisfaction with the hardware resources is high. In general, satisfaction with the provided software resources is also high. Programming tools and libraries and performance and debugging tools score somewhat lower than the other.

Regarding user support, the projects reported high satisfaction with technical advice, response times (including those for urgent requests), follow-up to queries and reported problems, and overall time to resolve problems.
The project managers expressed satisfaction to high satisfaction with most aspects of the Notur project and the overall service provided. Users commented positively on Notur as an important provider for large computational resources. From the aspects addressed, the available information and documentation and website were considered the weakest points. The technical support pages will be improved further by the project in the coming year.

The full report of the results will be published on the Notur web pages: **www.notur.no**

The Notur II project provides the 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), the University of Tromsø (UiT), and the Meteorological Institute (met.no). The project has a 10-year duration (2005-2014). The project is funded in part by the Research Council of Norway (through the eVITA programme) and in part by the consortium partners.