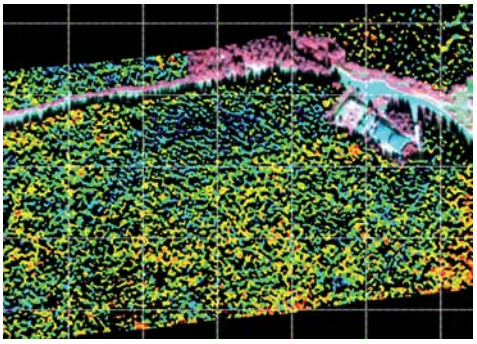
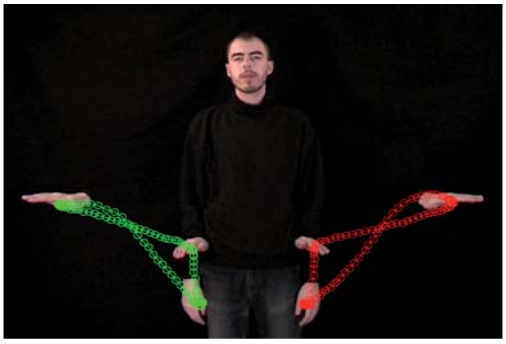
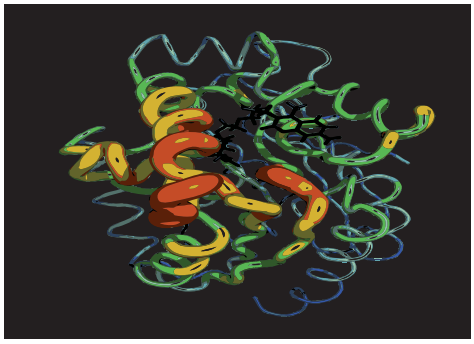
MetaCentrum







Yearbook 2009

MetaCentrum Yearbook 2009

Ivana Křenková, David Antoš,
and Luděk Matyska (Eds.)

MetaCentrum Yearbook 2009

Editors

Ivana Křenková, David Antoš, and Luděk Matyska
CESNET, z. s. p. o.
Zikova 4
Prague, Czech Republic

ISBN 978-80-904173-7-3

Preface

MetaCentrum is a collaborative activity of several Czech public universities and CESNET association to provide the distributed computing environment—Grid—to the research community in the Czech Republic. With its more than 10 year history—the MetaCentrum was conceived by Masaryk University in 1996—it has a long track of successful support of computing needs of diverse scientific areas. This 2009 MetaCentrum Annual Report provides an insight to the scientific work done and results achieved by users of this infrastructure. This is preceded with a brief overview of the computing and storage facilities MetaCentrum manages, as well as activities MetaCentrum is involved in.

The 2009 was a very important year for grid activities in Europe. The European Grid Initiative (EGI) Collaboration finally took a more formal shape, with the support from the EGI Design Study project funded by the European Commission within the 7th Framework Programme and coordinated by CESNET (MetaCentrum Director served as the EGI DS project Director). The EGI Collaboration is composed from National Grid Initiatives (NGI) complemented by international organizations (EIROForum members). In January 2009, during their Prague meeting, the NGI representatives endorsed the EGI Blueprint as the document defining the structure and functions of the European Grid Infrastructure. In March, during the large EGEE conference in Barcelona, the NGI representatives agreed to Amsterdam as a location of the EGI.eu, a new organization expected to coordinate the European Grid Infrastructure. The organization was formally established on 8th February 2010.

Based on the findings described in the EGI Blueprint and accompanying deliverables of the EGI DS project, European Commission shaped the call for the Distributed Computing Infrastructure (DCI), opened in July 2009. The EGI Collaboration reacted on this call through a preparation of a series of project proposals. MetaCentrum actively participated in several most important ones, namely the EGI InSPIRE and EMI projects, defining the operational aspects of EGI and the middleware development for this infrastructure, respectively. Both projects were accepted for funding and will define the European grid landscape for the next three to four years, with MetaCentrum as an active contributor and partner. MetaCentrum will be also a member of the accepted CHAIN project that focuses on the global (i.e., outside Europe) links of the EGI Collaboration.

MetaCentrum is naturally reacting to the changes in the European landscape, and already started a transformation into a National Grid Initiative during 2009. Users were encouraged to test and start use the EGEE operational environment (the predecessor of the EGI infrastructure) within the Virtual Organization for Central Europe (VOCE), managed by MetaCentrum and providing combined resources of the Central European partners. MetaCentrum continued in the fabrics virtualization efforts as the best means to support different environments and fulfill different user's requirements on a shared infrastructure. This will also

allow to fully support the EGI infrastructure in 2010 without compromising the services already available to MetaCentrum users at the national level.

Scientific results achieved using the MetaCentrum infrastructure, as presented in this Annual, clearly demonstrate the value and utility of the MetaCentrum, its infrastructure and activities. New resources, either directly provisioned by MetaCentrum or by collaborating institutions, are not able to keep pace with the increased demand on computing power. As part of the overall CESNET strategy to be one of the most important national large infrastructures for the research, development and innovations, MetaCentrum transformation into an NGI is a necessary step to guarantee a future for distributed computing infrastructure in the Czech Republic. As the Czech NGI, MetaCentrum will strengthen its role as the coordinator of the computing and directly associated storage resources, continuing also in the future with its support of research in different areas at the national and international levels, as also confirmed by this Annual.

Brno, June 2010

Luděk Matyska
MetaCentrum Director

Table of Contents

Part I

The MetaCentrum Infrastructure

The MetaCentrum Infrastructure in 2009

1 Introduction

The first part of this Yearbook describes the infrastructure of the National Grid Initiative MetaCentrum. We give an overview of available hardware resources and software, developments in cluster operation, resulting in significant resource utilisation increase in 2009, and an overview of information resources about MetaCentrum. Statistical information about resource utilisation is also provided.

Long-term goal of MetaCentrum, grid activity supported by CESNET, is to build and operate the National Grid Infrastructure in the Czech Republic. MetaCentrum includes operation of distributed production grid, development of grid middleware and involvement in international grid infrastructure through participation in corresponding international projects. In recent years, this activity covers not only distributed computing resources, but also operation of storage capacity, serving as a repository for both computational and general data.

During 2009, MetaCentrum activity was significantly influenced by involvement in preparation of the national e-infrastructure projects and in transition to dependable, long term grid infrastructure; both on national and international level.

Focus on user support and improved monitoring of traffic and user activities were identified as key topics, supplementing our continued push to creation of user-adaptive grid environment built around consistent virtualisation of computing environment. This strategy is fully compatible with current trends in the so-called Cloud computing, providing elastic computing and storage capacities that users can customize according to their needs.

Strong and continuously growing user interest in MetaCentrum is demonstrated by significant increase in usage of computational resources managed by MetaCentrum. During the year 2009, almost half million of jobs were computed, consuming more then 4.5 millions of CPU-hours.

The purpose of the infrastructure is to allow users to do research, therefore, scientific papers referring to MetaCentrum have been collected since 2008 and users with registered scientific results are prioritised. Total of 37 papers published in international journals and prestigious international conferences, with acknowledge to MetaCentrum, were registered in 2009.

Not only the users produce research and development results, the infrastructure must evolve, too. MetaCentrum was involved in preparation of six new EU projects, including the following three that were accepted and are either already running or in the last stages of the preparation—EGI Inspire (the European grid project, responsible for the operation and further development of pan-European grid infrastructure), EMI (further development of the grid middleware and contribution to the UMD), and CHAIN (linking EGI infrastructure

with regions outside Europe). Besides preparation of new projects, during 2009 MetaCentrum has been participating in the whole range of national and international projects in the area of grid infrastructure development. A list of the most significant projects follows:

- EPIKH (March 2009–March 2013). Exchange Programme to advance e-Infrastructure Know-How. Its aim is to reinforce the impact of e-infrastructure in scientific research defining and delivering stimulating programme of educational events, including Grid School and High Performance Computing courses. `http://www.epikh.eu/`
- EUAsiaGrid (April 2008–June 2010) Towards a common e-Science Infrastructure with the European and Asian Grids. It aims to promote awareness of the EGEE infrastructures, middleware and services in the Asian countries, based on results of the EGI_DS projects. `http://www.euasiagrid.org/`
- EGEE III (May 2008–April 2010) Enabling Grids for E-science in Europe. EGEE III is the third in a series of grid infrastructure projects. Its goal is to extend and maintain pan-European grid infrastructure available to European scientific community. MetaCentrum representative served also as the representative of the Central Europe region at the Project Management Board. `http://www.eu-egee.org/`
- EGI_DS (September 2007–December 2009) European Grid Initiative Design Study. It represents an effort to establish a sustainable grid infrastructure in Europe. In this project, CESNET (MetaCentrum) served as the project coordinator. `http://web.eu-egi.eu/`
- MediGrid (January 2005–December 2009). Development of methods and tools for GRID application in biomedicine. `http://www.medigrid.cz/`

2 Hardware and Software Resources

This section describes hardware and software portfolio available in MetaCentrum.

2.1 Computations

MetaCentrum offers a wide range of hardware resources. The portfolio includes multiprocessor machines with shared memory (`mat` with 48 CPUs), the middle class is represented by popular SMP servers with 16 or 32 cores and 64 to 256 GB memory (`eru`, `aule`, and `manwe`). Typical "high density" clusters consist of many 2- or 4-CPU servers connected with fast networks: 1 Gb/s (Gigabit Ethernet), 2.5 Gb/s (Myrinet) and 20 Gb/s (Infiniband).

Procurement activity was concentrated primarily on the purchase of a new cluster, which serves as a replacement for the old cluster "skurut" running 32-bit processors located at the CESNET headquarters. The new cluster, equipped with Intel Nehalem processors (28 nodes, with two quad-core processors each, 224 cores in total), together with QDR Infiniband network, was installed in the last days of 2009.

Furthermore, two Sun X4600 machines were purchased, each with 32 cores, as continuation of the successful and popular series of "manwe" mid-range machines.

Several service nodes were upgraded, notably new backup servers for tape library management (as direct implication of acquiring additional tape capacity) and a dual-node high-availability server, together with shared disk-storage, used for running various services in virtualised environment.

List of clusters and their parameters, as of the end of 2009, is shown in Table 1.

Machine	**CPU cores**	**description**	**memory per node**	**owner**
Acharon	16	SGI Altix 350	48 GB	UK
Ajax	8	SGI Altix 350	72 GB	ZČU
Alela	64	PC Cluster	8 to 32 GB	VUT
Dali	10	SGI Onyx 350	16 GB	UK
Eru	64	Cluster SUN	12 GB	CESNET
Hermes	48	PC Cluster Xeon	16 GB	JČU
Hydra	12	PC Cluster Xeon	1 GB	KIV/ZČU
Konos	120	PC Cluster AMD	2 to 8 GB	KMA/ZČU
Manwe, Aule	128	Cluster SUN X4600	32 to 128 GB	MU, CESNET
Mat	48	SGI Origin 2000	12 GB	UK
Nympha	160	PC Cluster HP	16 GB	CESNET
Orca	72	PC Cluster AMD	8 GB	NCBR/MU
Perian	190	PC Cluster SGI/HP	2 to 8 GB	NCBR/MU
Skirit	216	PC Cluster Xeon	1 to 8 GB	CESNET, FI/MU
Skurut	124	PC Cluster	n/a	CESNET
Dedicated clusters				
Loslab	24	PC Cluster AMD	4 GB	LL/MU
Quark	62	PC Cluster Xeon	2 to 18 GB	MU
Wood	16	PC Cluster	4 GB	MZLU
Other	86	dedicated clusters	n/a	various

Table 1. Machines and clusters connected in MetaCentrum, 1468 CPU cores in total.

MetaCentrum resources are physically hosted in the following locations: Masaryk University in Brno (Supercomputing Center Brno), Charles University in Prague (Supercomputing Center UK), University of West Bohemia in Pilsen (Westbohemian Supercomputing Center ZČU), University of South Bohemia in České Budějovice (Faculty of Science's cluster JČU), University of Technology in Brno (Faculty of Electrical Engineering and Communication VUT), Insti-

tute of Wood Science of the Mendel University in Brno (Supercomputing Center MZLU), and CESNET headquarters in Prague.

The individual clusters are connected with a network based on the CESNET2 backbone. The clusters accept jobs submitted through the common PBS scheduling system and share data with distributed file systems. The infrastructure formed a virtual cluster with 1468 CPU cores by the end of 2009.

Up-to-date number of machines, CPUs and their utilisation is displayed on the MetaCentrum portal `http://meta.cesnet.cz`.

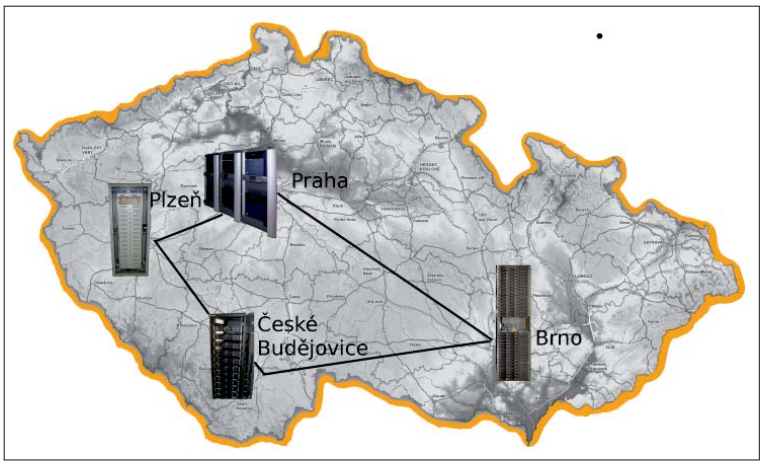


Fig. 1. MetaCentrum major sites

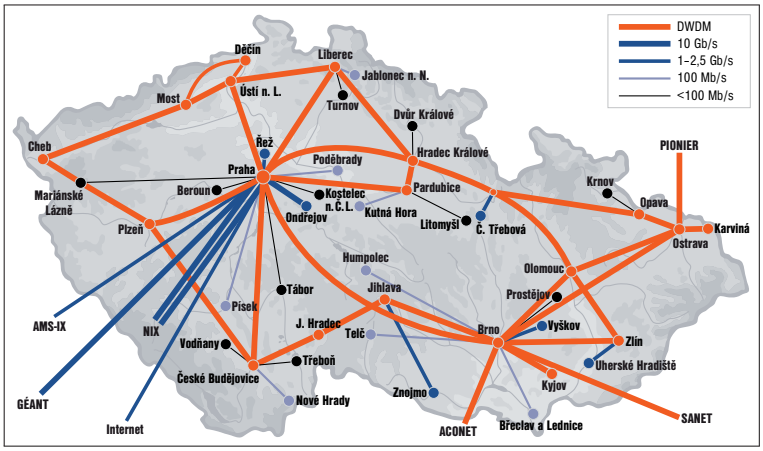


Fig. 2. CESNET2 network topology

2.2 Storage

MetaCentrum runs several data storage systems with both local and global access. Each cluster node has a /scratch space (sized in tens to hundreds GBs) to store intermediate results of currently running computations. Home directories are usually local for a cluster and realised by a local NFSv3 or NFSv4 server.

Main storage capacities are available through the NFSv4 protocol. All clusters except Loslab have the /storage volume attached, currently having access to 44 TB total usable capacity.

Measurements show that the NFSv4 central server is not suitable to centrally provide cluster home directories; data access from Pilsen to Brno is about 10-times slower than local access which is caused by latencies of protocol control messages. Despite this drawback, the NFSv4 volume is suitable to keep long-term data that is staged in computation nodes.

MetaCentrum still offers a global filesystem based on AFS. Software distributions are accessed via AFS, but it can be deployed directly by users to benefit from fine-grained access control possibilities.

For backup purposes, MetaCentrum uses tape libraries NEO8000 from Overland Storage with 500 tape positions based on LTO-3 technology. Two identical libraries are installed with total on-line capacity of 400 TB (uncompressed), one at the University of West Bohemia and one at the Masaryk University. Supercomputing Center Brno. The libraries are managed with a combination of EMC Legato NetWorker suite and in-house developed software. The distributed backup environment provides a robust and fault-tolerant solution that could survive even catastrophic events such as a complete destruction of one of the hosting sites.

2.3 Application Programs

Software portfolio of MetaCentrum covers wide spectrum of application programs. The following list of installed software is not complete, it serves merely as an overview for both current and potential users. It is roughly divided into several problem areas.

Software packages in MetaCentrum can be installed on user request. More details about the application programs and licence policies can be found at the MetaCentrum web site `http://meta.cesnet.cz/en/resources/`.

Computational Chemistry & Molecular Modelling

- *Amber*—a suite of programs for molecular dynamics simulations, particularly on biomolecules.
- *Babel*—a converter of a number of file formats currently used in molecular modelling. It is capable of assigning hybridization, bond order, and connectivity when these elements are not present in the input file.

- *Gaussian/GaussView*—a package of programs for quantum mechanics computations, used by chemists, chemical engineers, biochemists, physicists and other scientists. Gaussian predicts the energies, molecular structures, vibrational frequencies and molecular properties of molecules and reactions in a wide variety of chemical environments.
- *Molden*—a general molecular and electronic structure processing program. It can read output from the ab initio packages GAMESS (US), Gaussian, MOLPRO, and from semi-empirical packages such as MOPAC; display molecular orbitals or electron density as contour plots or 3D grid plots and output to a number of graphical formats; animate reaction paths and molecular vibrations. Molden has a powerful Z-matrix editor for full control over the geometry.
- *Gromacs*—an engine to perform molecular dynamics simulations and energy minimization. It is primarily designed for biochemical molecules like proteins, lipids and nucleic acids that have a lot of complicated bonded interactions, but since GROMACS is extremely fast at calculating the nonbonded interactions (that usually dominate simulations) many groups are also using it for research on non-biological systems, e.g., polymers.
- *VMD*—a molecular visualization program for displaying, animating, and analyzing large biomolecular systems (proteins, nucleic acids, lipid bilayer assemblies, etc.) using 3-D graphics and built-in scripting. VMD can act as a graphical frontend for an external MD program by displaying and animating a molecule undergoing simulation on a remote computer.
- *Gamess*—a program for ab initio molecular quantum chemistry. GAMESS (US) can perform a number of general computational chemistry calculations, including Hartree–Fock, density functional theory (DFT), generalized valence bond (GVB), and Multi-configurational self-consistent field (MCSCF).
- *deMon*—a software package for density functional theory (DFT) calculations (density of Montréal). It uses the linear combination of Gaussian-type orbital (LCGTO) approach for the self-consistent solution of the Kohn-Sham (KS) DFT equations.
- *PC GAMESS/Firefly*—a freely available ab initio and DFT computational chemistry program developed to offer high performance on Intel-compatible x86, AMD64, and EM64T processors. It shares lots of functionality with GAMESS(US).
- *MolPro*—a complete system of ab initio programs for molecular electronic structure calculations. As distinct from other commonly used quantum chemistry packages, the emphasis is on highly accurate computations, with extensive treatment of the electron correlation problem through the multiconfiguration-reference CI, coupled cluster and associated methods.
- *Tinker*—a flexible system of programs and routines for molecular mechanics and dynamics as well as other energy-based and structural manipulation calculations.
- *NAMD*—a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems. Based on Charm++ parallel ob-

jects, NAMD scales to hundreds of processors on high-end parallel platforms and tens of processors on commodity clusters using gigabit ethernet.
– *VASP*—a package for performing ab-initio quantum-mechanical molecular dynamics (MD) using pseudopotentials and a plane wave basis set.
– *Autodock Vina*—a program for drug discovery, molecular docking and virtual screening, offering multi-core capability, high performance and enhanced accuracy and ease of use.

Technical and Material Simulations

– *ANSYS (including module LS-DYNA)*—a general purpose finite element modeling package for numerical solution of a wide variety of mechanical problems, including static/dynamic structural analysis (both linear and nonlinear, in geometry as well as physical properties used), heat transfer and fluid dynamics, acoustic, electromagnetics, and other problems.
– *Fluent*—a powerful and flexible general-purpose computational fluid dynamics (CFD) package used for engineering simulations of all levels of complexity. It offers a comprehensive range of physical models that can be applied to a broad range of industries and applications. The wealth of physical models in FLUENT allows to accurately predict laminar and turbulent flows, various modes of heat transfer, chemical reactions, multiphase flows, and other phenomena with complete mesh flexibility and solution-based mesh adaption.
– *MSC.Marc*—based on finite element method serves for solving large linear and non linear (physical and geometrically), structural, thermal, thermal-mechanical, electromagmetic and acoustic jobs. Program is extensively used for application in impulse phenomenon area and large deformation problem.
– *Open Foam*—Open Field Operation and Manipulation CFD Toolbox can simulate anything from complex fluid flows involving chemical reactions, turbulence and heat transfer, to solid dynamics, electromagnetics and the pricing of financial options.

Mathematical and Statistical Modelling

– *Maple*—an essential technical computing software for mathematics and modelling. Maple can to do quick calculations, develop design sheets, teach fundamental concepts, or produce sophisticated high-fidelity simulation models.
– *Matlab*—a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation in a wide range of applications, including signal and image processing, communications, control design, test and measurement, financial modeling and analysis, and computational biology.
– *SNNS*—Stuttgart Neural Network Simulator is a software simulator for neural networks on Unix workstations. It consists of two main components: simulator kernel and graphical user interface.
– *R*—a language and environment for statistical computing and graphics. R provides a wide variety of statistical and graphical techniques, and is highly extensible.

Development Tools and Environments

- *SGI Development environment*
- *PGI CDK*
- *TotalView*
- *Vampir*
- *Paradise*
- *SICStus Prolog*
- *Wine*
- *GCC*
- *Grace*
- *Numpy*
- *Python*
- *Scientific Python*
- *Scipy*
- *JDK*
- *Lisp-CMUCL*

Structural Biology and Bioinformatics

- *MrBayes*—a program for the Bayesian estimation of phylogeny.
- *QUEEN*—QUantitative Evaluation of Experimental Nmr restraints.
- *X-PLOR*—X-ray crystallography and NMR.
- *CS-Rosetta*—chemical shifts based protein structure prediction.

3 Operation of the Infrastructure

The operations group, in addition to routine administration, maintenance and hardware configuration issues, was focused on fine-tuning, stabilization and improvement of services deployed in the previous year.

Scheduling of parallel jobs and jobs with large memory requirements was substantially improved. Before the new scheduling scheme was deployed, aggressive scheduling of single-processor jobs, despite of their lower priorities, resulted in starvation of large parallel jobs. In current configuration, the scheduler can automatically reserve resources (nodes) for multi-processor parallel jobs, avoiding queue-jumping of small jobs.

Similar technique is used for scheduling jobs submitted by privileged users on clusters owned by cooperating organisations and managed by MetaCentrum. It guarantees the users priority access to their resources without unnecessary restricting other users. Modified setup of the queuing system also ensures fair-share access, when job priority is modified based on number of jobs and CPU time consumed by the user during last month. We achieved reduced waiting time for high-priority jobs, while preserving efficient handling of short jobs.

Job management systems in MetaCentrum has been based on centralized instance of PBSPro for almost ten years. This approach allows implementation of new planning policies relatively easily, as proved during 2009, as it provides global status of all resources in MetaCentrum and enables implementation of fair-share scheduler with utilisation of global knowledge of all running jobs, as well as history of already finished jobs. Despite of undoubted advantages, with growing number of sites and processors, the central scheduler is becoming the most significant single point of failure in MetaCentrum and performance bottleneck. In 2009, we started work on a new concept of job scheduling, in which

weaknesses indicated above should be eliminated, while desired global properties are preserved. We have designed an architecture of mutually cooperating planners, based on a peer to peer network of semi-independent schedulers installed at each site. Each scheduler accepts jobs for the whole infrastructure, cooperating with other schedulers on implementation of global policies like central job accounting, fair-share, or submission of jobs across several sites. The scheduling system is integrated with the Magrathea system to support scheduling of virtual clusters, including setup of their internal networks, eventually spanning several sites. Implementation based on local schedulers and user gateways guarantees functionality (though limited to the site) even in case of network outages between centres. Details can be found in the technical report [5].

In the virtualisation area, year 2009 was devoted to integration of virtualisation tools into a new service, representing a major step towards the concept of cloud computing, providing resources as a service and leaving decisions on provided environment on user preferences. The concept of virtual clusters provides users the possibility to request sets of virtual nodes according to their specifications. Cluster setup may include dedicated virtual network, isolating the clusters on network level. Requests on node instantiation can be submitted using the same tools used currently for standard job submission and virtual clusters are handled as ordinary jobs from scheduling point of view. Our technical solution is fully integrated into the existing grid infrastructure, and it allows provisioning of both services efficiently on one set of physical resources. The main benefit for the users will be greater flexibility, the ability to use MetaCentrum resources in accordance with their habits and needs, allowing groups of users to run their own cluster installations fine-tuned to their needs. More information is available in papers [6, 2, 1].

The most significant result achieved by the security group is the MetaCentrum authorisation service introduced at the end of 2009 to improve access control mechanisms. The authorisation service makes it possible to specify access policies at a single place and distribute them to the end services. The distribution is triggered on each change in the policies and ensures translation of the policy to a format acceptable by the end service (e.g., a snippet of a configuration to load). End services make access control decision solely on the information stored locally and do not to contact any remote service. A working prototype of the service was finished in 2009 and put into operation at the end of the year. More detailed description of the authorisation service can be found in [3].

We also went on with developing of the Pakiti system to monitor deployment of security patches. Pakiti identifies machines where security patches have not been applied properly. The development was performed mainly during our participation in the EU EGEE project and the result is in use by several sites within the project and also for monitoring of the MetaCentrum infrastructure.

For a long time we have been focusing on utilisation of identity federations. During last year, we provided technical background for the atlases of histological images `https://atlases.muni.cz/` collected at the Faculty of Medicine, Masaryk University. The service is integrated with multiple identity federations

world-wide, witch helped significantly in exploring mechanisms and procedures needed for multi-federation arrangement. Details of federation approach to user authentication can be found in [4].

We also started a close collaboration with the CESNET CERTS team, which is responsible for coordination and handling of security incidents. The goal of the collaboration is to use the procedures and mechanisms established to cover also domain of grid computing in CESNET. We also actively participated to the Operational Security Coordination Team of EGEE, where we were responsible for coordination of security monitoring development.

4 Interaction with Users

User support encompasses wide range of related tasks: informing users about the current state of the centre, addressing user problems, maintaining collaboration with user communities, monitoring resource availability, providing evaluated feedback to the operation activities, and many others.

4.1 Information Services

The portal `http://meta.cesnet.cz/` is the basic information gateway of MetaCentrum. The portal provides both general information for casual visitors and specific information for registered users (authentication may be required).

In addition to general documentation adding and improving, two new web applications were developed and/or installed in 2009. A new version of *Pbsmon*[1] has been deployed. Pbsmon displays current status of virtualised infrastructure, taking into account relationship of physical and virtual machines. To make user orientation in growing number of computing resource properties easier, a "qsub command builder" has been developed. The tool, accessible from the web portal, helps users to configure job submission configurations, based on knowledge of option relationships, also allowing to display machines satisfying the requested options and their current utilisation.

4.2 User Courses and Workshops

A Grid Computing Workshop was held in Brno in November 2009. It was a one day meeting organised by MetaCentrum in cooperation with SGI. We are very proud that a speech of Dr. Eng Lim Goh, the SGI Chief Technology Officer, was part of the programme. The aim of the event was to inform current as well as new potential users about high-performance computing, storage, and other MetaCentrum services available for solving of numerous research problems and challenges at national as well as international level. More than 50 users participated in the workshop.

[1] `http://meta.cesnet.cz/pbsmon2/`

MetaCentrum was also disseminated on various events. To name a few:

- 7th Discussions in Structural Molecular Biology in November 2009 (Nové Hrady)—user support staff presented MetaCentrum activities here (a poster was presented and a lecture was given)
- Conference Internet and Technology 09 (organized by CZ.NIC) where our staff presented EGI as a base of the European Research Area (ERA) and Virtualising Network for Grid Environments.
- Presentation "Logging&Bookkeeping User Interaction Interfaces" was given at CyGrid Training the Trainers, Nicosia, Cyprus

Many presentations were also given to the grid development communities at various international and national conferences.

5 Infrastructure Usage—Operational Statistics

Data collected from monitoring and accounting tools is available at the MetaCentrum portal `http://meta.cesnet.cz`. The statistical data collecting must take infrastructure complexity into account. It is necessary to keep records on clusters, machine membership in clusters, physical and virtual machine outages, user reservation of resources, coexistence of virtual machines on a single physical computer etc.

Total of 345 physical machines with 1468 CPU cores were available for the 371 registered MetaCentrum users by the end of 2009. The increase of available CPU resources is shown in Table 2.

Computation jobs were run by 219 users. Tens of users did not actively run computation jobs, they used the infrastructure just to access storage capacities. It is often caused by the fact that they are members of groups where their colleagues submit jobs and they work with the results only. Direct MetaCentrum access allows them to study the results without the need of transferring data.

5.1 Cluster Usage

Significant increase in usage of computing resources was observed in 2009. Users consumed about 4.5 million CPU-hours in nearly half a million jobs (only 1.5 million of them and 125 thousands jobs of them in the first half of the year). We attribute the increase to two basic changes:

- clusters Perian and Orca, dedicated to owner exclusive use previously, have been added to the pool of shared resources,

year	2004	2006	2008	2009
CPU cores	262	600	1200	1468

Table 2. CPU core number increase in MetaCentrum during last years.

– number of jobs per queue and user have been set to unlimited, causing more efficient resource usage.

During the second half of 2009, utilisation of MetaCentrum clusters reached 90%, which is an excellent result for an infrastructure running heterogeneous jobs, meaning practically a complete saturation, indicating strong user interest in the infrastructure. This also corresponds to job waiting times which generally increased. In order to prevent job waiting times to grow into unacceptable levels, it is obvious that further investment into the infrastructure is necessary.

Fig. 3 displays numbers of jobs according to waiting times from submitting to the start of execution, dividing the halves of the year for comparison. While majority of the jobs waits no more than a couple of minutes, it can be seen that the infrastructure utilisation increase in the second half of the year caused several thousands of jobs wait days.

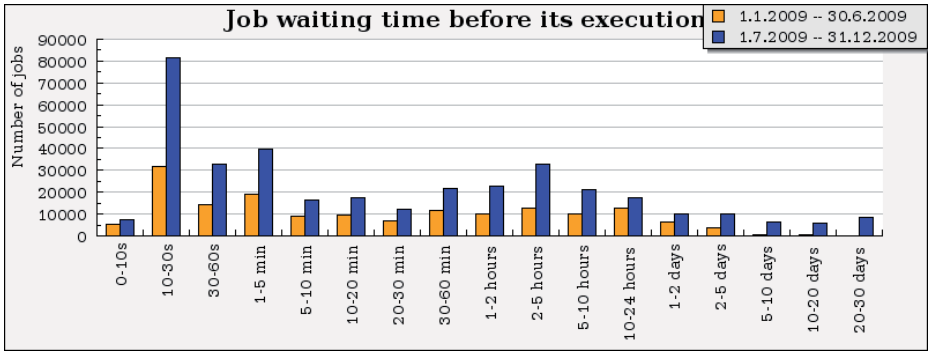


Fig. 3. Job average waiting times

Running times of jobs (Fig. 4) offer an interesting comparison, too. Extremely short jobs usually indicate a configuration and/or job submission error (users with excessive amount of such jobs are offered help by the User Support). The graph doesn't distinguish sequential and parallel jobs, but in general, jobs running more than several hours tend to be parallel (4–32 CPUs). Despite seemingly long job waiting times caused by high resource utilisation, MetaCentrum still makes user (real time) waiting for results shorter: a five-hour 16-CPU job would last more than 3 days on a single CPU system and nearly a day on a 4-CPU system.

Following overview displays graphs of hardware utilisation in 2009. The base (100%) is the total number of available CPU-core-seconds minus the CPU-core-seconds of machines under maintenance. The values are CPU-core-seconds of running jobs and reservations. Complete statistics can be found at the MetaCentrum portal[2].

[2] `http://meta.cesnet.cz/cs/about/results/stats/index.html`

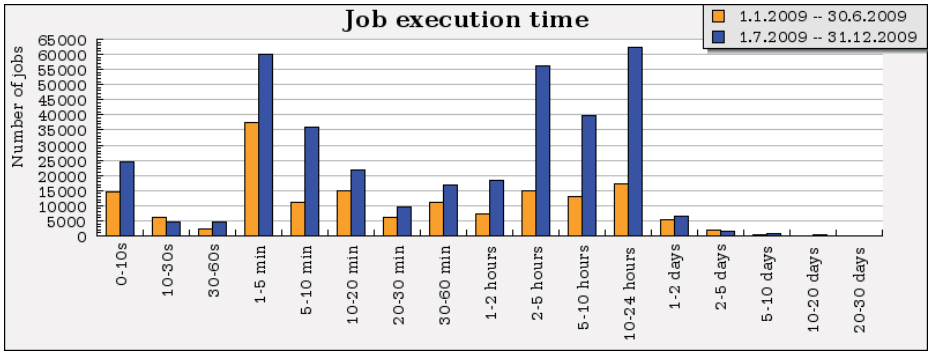


Fig. 4. Job running time

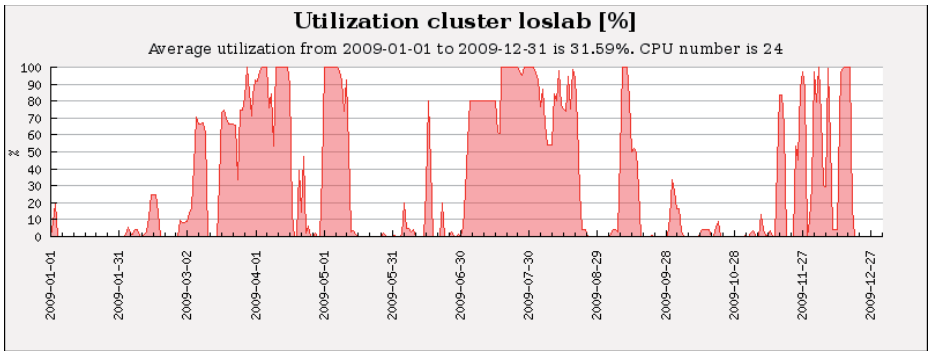


Fig. 5. Loslab cluster utilisation

Dedicated clusters (e.g., Loslab) belong to their owners and are accessible by the owners only. MetaCentrum has no chance to influence the workload of that clusters, which is far below standard utilisation of the open access clusters.

Clusters with open access are available to all MetaCentrum users, their utilisation is much higher than utilisation of the dedicated clusters. Owners of the machines have priority access to their resources, users outside the community can utilise the remaining capacities.

The effect of job queue policy is best seen on the Hermes cluster, having 8 CPU cores per node. It was put into operation in January, reaching 50% utilisation in the first half of the year when dedicated to the owner, but 80% in the second.

The advantage of open access can also be demonstrated by utilisation of clusters Orca and Perian. Both of them belong to the NCBR research centre. Originally, they were dedicated to their owners only, since summer 2009 they have been accessible to all MetaCentrum users. Jobs from priority queues accessible only by the owners can preempt task from the other queues. Utilisation of Orca increased from 18 % to 76 % in the second half of the year and owners

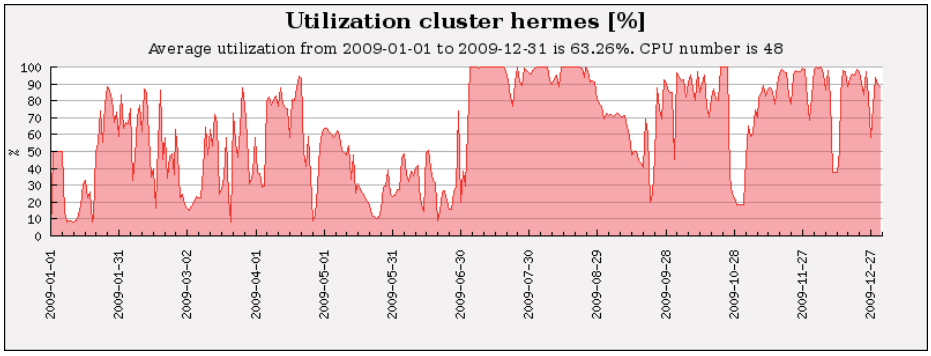


Fig. 6. Hermes cluster utilisation

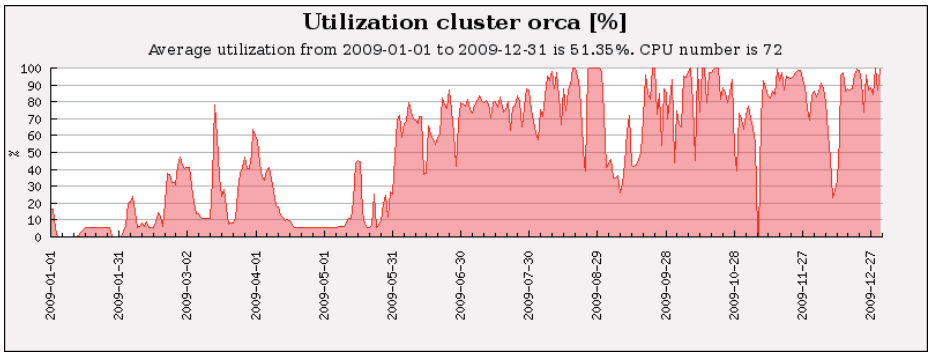


Fig. 7. Orca cluster utilisation

still have privileged access to their resources without any loss of perceived access quality.

5.2 Storage Usage

The NFSv4 */storage* volume, which is the main high-capacity online storage facility for MetaCentrum, has total capacity of 44 TB. User data occupied 31 TB in more than 30 milions files by the end of 2009.

The number of users keeping some data on the /storage volume—289—exceeds the number of users that run jobs. Top three users have more than 5 TB of data each, 85 users have at least 1 GB of data on the /storage volume.

5.3 Top Users

Graphs in Figs. 8 and 9 illustrate quantitative differences of number of submitted jobs and consumed CPU time among users. The top user run more than 147 thousand jobs and consumed more than 1 million CPU-hours (about 120

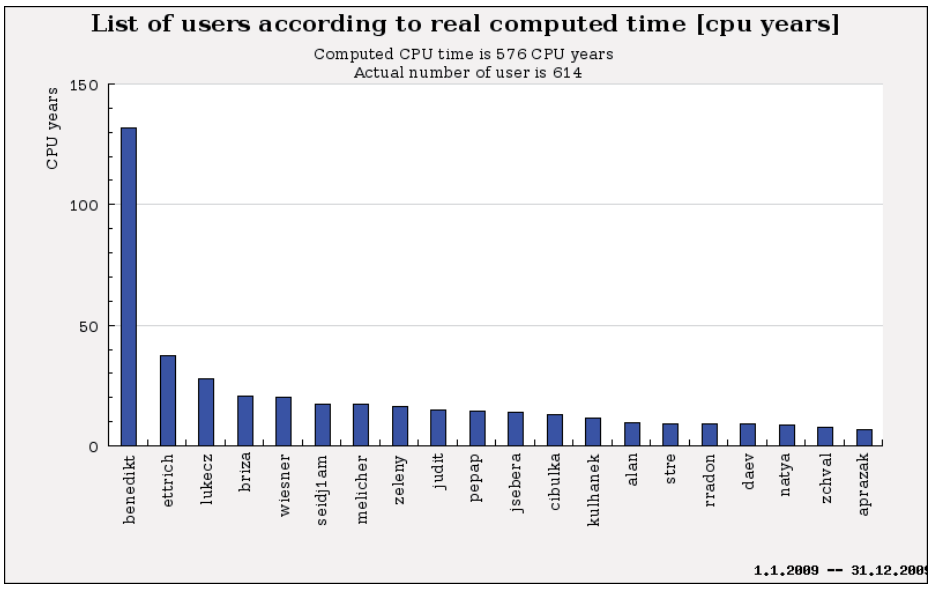


Fig. 8. CPU time by top users

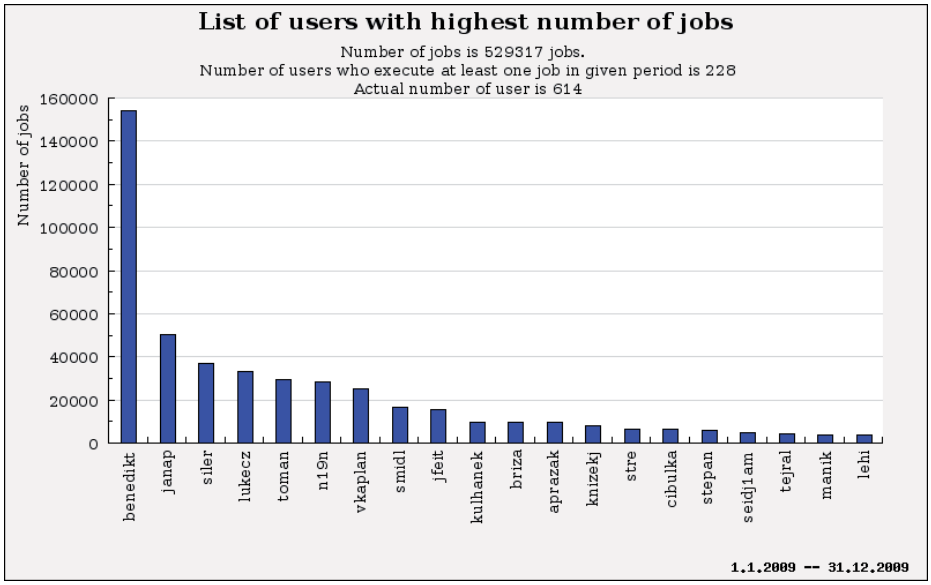


Fig. 9. No. of jobs by top users

CPU-years). In the second half of 2009, this user consumed more than one third of machine time of the whole MetaCentrum.

The applications consuming most of CPU time are shown in Fig. 10. The "Others" column represents sum of commercial applications with smaller usage, undistinguished home-grown user scripts and applications.

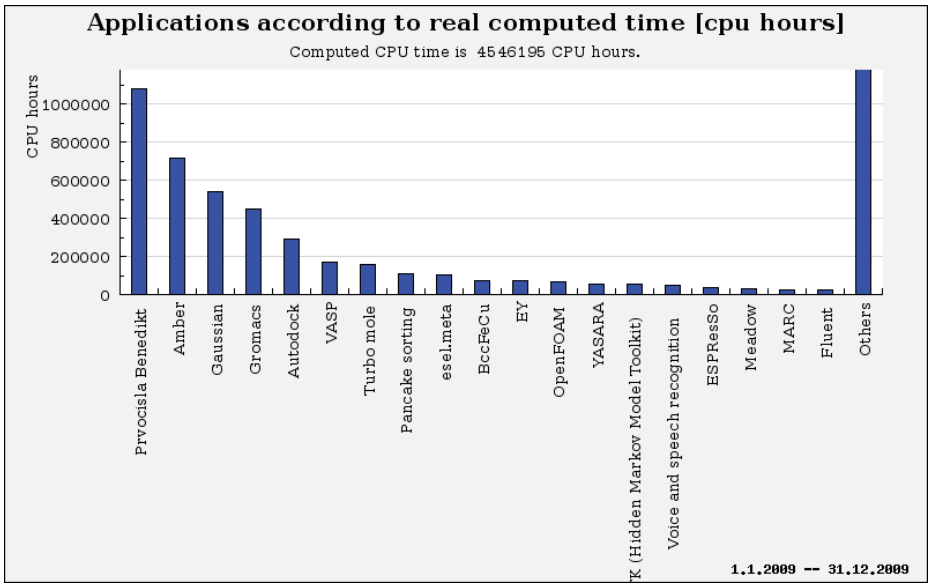


Fig. 10. CPU time by top applications

5.4 User Interaction: Request Tracking System

User interaction and solving user's problems is realised with a request tracking system (RT). The users submitted 924 new tickets in 2009, either by mail and/or MetaCentrum web portal (new account requests, quota changes, ...). Additional 751 tickets were generated by infrastructure administrators (hardware issues, software errors, internal tasks, etc.). More than 200 tickets had remained unsolved from 2008 and 1508 tickets have been solved and closed.

6 MetaCentrum Personnel

As is implied by its mission, most of the people contributing to the Metacentrum success in 2009 work only part time as CESNET employees or are members of the international projects teams. Here we list all the people who in some way worked for CESNET (usually in a part time job) in 2009 and contributed to either the MetaCentrum activities or were involved in the EGEE III project.

David Antoš
Jiří Chudoba
František Dvořák
Jiří Filipovič
Ladislav Havlík
Lukáš Hejtmánek
Petr Holub
Jana Hrdličková
Martin Klubal
Jan Kmuníček
Tomáš Kouba
Daniel Kouřil
Gabriela Krčmářová
Iva Krejčí
Aleš Křenek
Ivana Křenková
Igor Krnáč
Martin Kuba
Jan Kundrát
Slávek Licehammer
Miloš Lokajíček
Luděk Matyska
Miloš Mulač
Jan Pospíšil
Michal Procházka
Miroslav Ruda
Zdeněk Salvet
Jaroslava Schovancová
Zora Sebestianová
Jiří Sitera
Zdeněk Šustr
Jan Švec
Šimon Tóth
Michal Voců
Vlasta Žáková

7 End-User Papers

All the infrastructure activities do not have much sense without users that unleash the potential of the shared computing and storage infrastructure. The statistics provided earlier confirm that the CPUs are not staying idle, the actual scientific value of the computing power consumption is demonstrated on the following pages that are the bulk of this Yearbook.

This report includes 22 contributions written by individual scientists as well as large scientific teams covering practically all the scientific areas that use the distributed computing infrastructure. The papers were selected to include all the major infrastructure users together with contributions that did not consume enormous amount of computing power but were of interest due to the methods they deployed, tools they used or results they achieved.

Each paper references one or more publications, usually from high quality journals or international conferences, as a confirmation of the scientific achievements done by the MetaCentrum user community with the direct use of provided resources.

References

1. David Antoš, Luděk Matyska, Petr Holub, and Jiří Sitera. VirtCloud: Virtual Network for User-controlled Virtual Clusters. Technical Report 1/2009, CESNET, z. s. p. o., 2009.
2. David Antoš, Luděk Matyska, Petr Holub, and Jiří Sitera. VirtCloud: Virtualising Network for Grid Environments–First Experiences. In Irfan Awan, Muhammad Younas, Takahiro Hara, and Arjan Durresi, editors, *The 23rd IEEE International Conference on Advanced Information Networking and Applications AINA*

2009, pages 876–883, Bradford, UK, May 2009. IEEE Comp. Soc. ISBN 978-0-7695-3638-5, ISSN 1550-445X.

3. Daniel Kouřil and Michal Procházka. Fault-tolerant Access Control in Distributed Environment – the MetaCentrum Authorization Infrastructure. Technical Report 4/2009, CESNET, z. s. p. o., 2009.
4. Michal Procházka, Daniel Kouřil, and Luděk Matyska. User Centric Authentication for Web Applications. In *Proceedings of the 2010 International Symposium on Collaborative Technologies and Systems*. IEEE, 2010. ISBN 978-1-4244-6620-7.
5. Miroslav Ruda and Šimon Tóth. Transition to Inter-Cluster Scheduling Architecture in MetaCentrum. Technical Report 21/2009, CESNET, z. s. p. o., 2009.
6. Miroslav Ruda, Zdeněk Šustr, Jiří Sitera, David Antoš, Lukáš Hejtmánek, and Petr Holub. Virtual Clusters as a New Service of MetaCentrum, the Czech NGI. In *Cracow Grid Workshop '09*, pages 64–71. Academic Computer Centre CYFRONET AGH, 2010. ISBN 978-83-61433-01-9.

Part II

MetaCentrum User Reports

Fredholm Alternative for the p-Laplacian at Higher Eigenvalues

Jiří Benedikt*

Department of Mathematics, Faculty of Applied Sciences
University of West Bohemia
Univerzitní 22, 306 14 Plzeň, Czech Republic
benedikt@kma.zcu.cz

1 Introduction

We are interested in the existence of a weak solution u of the quasilinear boundary value problem

$$\begin{cases} -\left(|u'|^{p-2}u'\right)' = \lambda_k |u|^{p-2}u + f(x), \quad x \in (0, \pi_p) \\ u(0) = u(\pi_p) = 0 \end{cases} \tag{1}$$

where $p > 1$, $\pi_p \stackrel{\text{def}}{=} \frac{2\pi}{p \sin(\pi/p)}$, $f \in L^\infty(0, \pi_p)$ is a given function, and λ_k stands for the k-th eigenvalue of the one-dimensional p-Laplacian. Let $\sin_p$ denote the odd $2\pi_p$-periodic extension of the eigenfunction associated with λ_1. Then $\sin_p kx$ is the eigenfunction associated with λ_k. Notice that for $p = 2$ the equation in (1) becomes a linear one, i.e. $-u'' = \lambda_k u + f$, and $\pi_2 \equiv \pi$, $\sin_2 \equiv \sin$.

In the linear case $p = 2$, the classical Fredholm alternative yields that the orthogonality condition $L(\pi_p) = 0$ where

$$L(x) \stackrel{\text{def}}{=} \int_0^x f(s) \sin_p ks \, \mathrm{d}s$$

is both sufficient and necessary for solvability of (1). Del Pino, Drábek and Manásevich [2] proved that in the quasilinear case $p > 1$, $p \neq 2$, and at the first eigenvalue ($k = 1$), $L(\pi_p) = 0$ is a sufficient condition for solvability of (1) (it is also known that it is not necessary). At the higher eigenvalues ($k \geq 2$), neither sufficiency nor necessity of $L(\pi_p) = 0$ is proved at present. Manásevich and Takáč [3] proved that for $k \geq 2$ and $p > 1$, the condition $L(\pi_p) = 0$ together with $(p-2)Q(\pi_p) \neq 0$ where

$$M(x) \stackrel{\text{def}}{=} \int_0^x f(s) \sin_p' ks \, \mathrm{d}s,$$
$$Q(x) \stackrel{\text{def}}{=} \int_0^x L(s) M'(s) \, \mathrm{d}s,$$

* The author has been supported by the Ministry of Education, Youth, and Sports (MŠMT) of the Czech Republic, Research Plan no. MSM 4977751301 and by grant no. ME 10093.

is sufficient for existence of a solution of (1).

We focus on the case $k \geq 2$, assuming $p \in (1,2)$. The approach used in [1] allows us to handle more degenerate sufficient conditions on f then any previous work. Our main result is that (1) has at least one weak solution provided $L(\pi_p) = Q(\pi_p) = (2p-3)R(\pi_p) = 0$ and $(3p-4)S(\pi_p) \neq 0$ where

$$\begin{aligned}
H(x) &\stackrel{\text{def}}{=} \int_0^x f(s) \sin_p'' ks \,\mathrm{d}s, \\
R(x) &\stackrel{\text{def}}{=} \int_0^x L(s)^2 H'(s) \,\mathrm{d}s - (p-2) \int_0^x M(s)^2 L'(s) \,\mathrm{d}s, \\
U(x) &\stackrel{\text{def}}{=} \int_0^x f(s) \sin_p''' ks \,\mathrm{d}s, \\
S(x) &\stackrel{\text{def}}{=} \int_0^x L(s)^3 U'(s) \,\mathrm{d}s + 3 \int_0^x L(s)^2 M(s) H'(s) \,\mathrm{d}s \\
&\quad - 6(p-2) \int_0^x M(s) L'(s) \left(\int_0^s L(t) H'(t) \,\mathrm{d}t \right) \mathrm{d}s \\
&\quad - (2p-3)(p-2) \int_0^x M(s)^3 L'(s) \,\mathrm{d}s.
\end{aligned}$$

We remark that at the first eigenvalue ($k = 1$), our conditions are unsatisfiable. In fact, it follows from Del Pino, Drábek and Manásevich [2] that $L(\pi_p) = 0$ implies $Q(\pi_p) \neq 0$ at the first eigenvalue. Hence, it is very important to find an example of f such that our sufficient conditions are satisfied at a higher eigenvalue. In other words, to show that our conditions are satisfiable at all. This example is presented in the next section and it yields that our result indeed significantly generalize known sufficient conditions that guarantee existence of a weak solution of (1) at higher eigenvalues.

2 Example

To provide an example of f such that the solvability of (1) is guaranteed by our sufficient condition, but not by any previously known one, we consider $p = 3/2$, $k = 2$ and define

$$f_{1,2}(x) \stackrel{\text{def}}{=} 1 + \frac{40\pi}{9\sqrt{3}} \left(-1 \pm \sqrt{1 + \frac{243}{80\pi^2}} \right) \sin_p' 2x.$$

It is not difficult to prove analytically that $L(\pi_p) = Q(\pi_p) = (2p-3)R(\pi_p) = 0$ for both f_1 and f_2 (see [1]). However, it is apparently impossible to calculate $S(\pi_p)$ analytically, and so we proceed with a computer assisted proof of $S(\pi_p) \neq 0$. We compute $\sin_p$, its derivatives $\sin_p^{(i)}$, $i = 1,2,3$, and all the above functions L, M, Q, H, R, U and S in interval arithmetic. It means that we represent them as interval valued functions for which the exact function lies in the corresponding

interval. Of course, integration in interval arithmetic is very time consuming, so we performed parallel computations in the METACentrum. It turns out that

$$S(\pi_p) \in [-1.862637604294, -1.860015068451] \quad \text{for } f = f_1 \quad \text{and}$$
$$S(\pi_p) \in [41.790518158498, 42.150573506827] \quad \text{for } f = f_2.$$

Consequently, $S(\pi_p) \neq 0$ in both cases and our main result guarantees existence of a weak solution of (1) with $f = f_{1,2}$.

References

1. Benedikt, J., Girg, P., Takáč, P., On the Fredholm alternative for the p-Laplacian at higher eigenvalues (in one dimension), Nonlinear Anal. 72 (2010), 3091–3107.
2. Del Pino, M. A., Drábek, P., Manásevich, R. F., The Fredholm alternative at the first eigenvalue for the one-dimensional p-Laplacian, J. Differential Equations 151 (1999), 386–419.
3. Manásevich, R. F., Takáč, P., On the Fredholm alternative for the p-Laplacian in one dimension, Proc. London Math. Soc. 84 (2002), 324–342.

Research Projects of Loschmidt Laboratories

Jan Brezovský, Eva Chovancová, and Jiří Damborský

Loschmidt Laboratories, Institute of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic
briza@chemi.muni.cz, akllupe@chemi.muni.cz, jiri@chemi.muni.cz

1 People Involved in the Projects

Lada Biedermannová (lada)
Jan Brezovský(briza)
Eva Chovancová (akllupe)
Antonín Pavelka (tonda)
Lukáš Daniel (lukecz)
Martin Klvana (martink)
Jiří Damborský (jiri)

2 Engineering of the Pathways for Product Release in the Haloalkane Dehalogenases

The main objective of this project was to study the effect of mutations in the tunnel of haloalkane dehalogenases on product release. We have mutated several amino acid residues positioned in the two tunnels connecting active site of haloalkane dehalogenase DhaA with the bulk solvent. Product release in the wild type and mutant enzymes were studied using classical and random accelerated

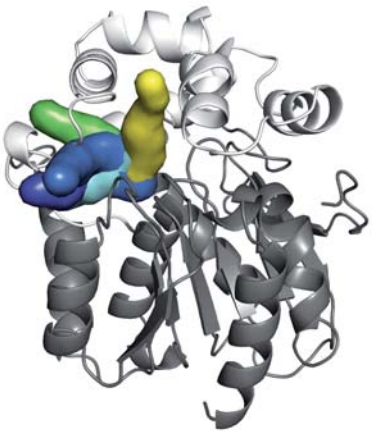


Fig. 1. Pathways identified in the wild-type and the mutants of DhaA enzyme. Main tunnel (p1) is in yellow, slot tunnel and its derivatives (p2a, p2b and p2c) are in blue and the auxiliary pathway (p3) is in green.

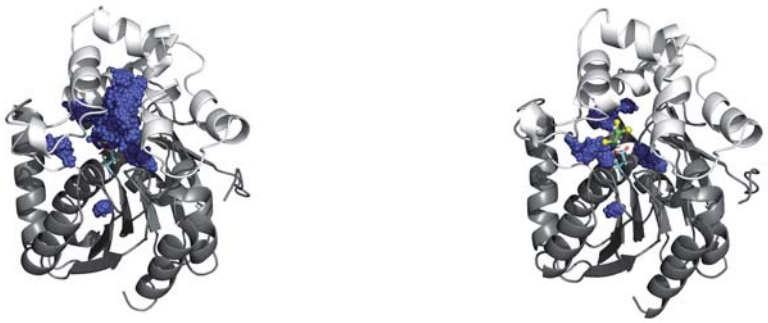


Fig. 2. Solvatation of the wild-type DhaA (left) and the mutant DhaA31 (right). The accessibility of active sites for water molecules is indicated by their cumulative position during the molecular dynamics simulation (blue spheres). The nucleophile in the active site is in cyan stick and the substrate TCP is in green/yellow ball and sticks.

molecular dynamic (RAMD) simulations. The released products passed through five different pathways (p1, p2a, p2b, p2c and p3) of which only p1 (main tunnel) and p2a (slot tunnel) were observed by X-ray crystallography (Figure 1). The mutations of smaller residues forming pathway p1 for bulky aromatic residues led to reduced accessibility of this pathway to ligands as well as change in the nature of this tunnel from permanent to ligand-induced. Moreover, these mutants exhibited increased activity toward anthropogenic compound 1,2,3-trichloropropane (TCP). Therefore, we experimentally randomized the key residues identified by RAMD simulations and obtained the mutant DhaA31, which was 32× more active with TCP than the wild-type DhaA. The mechanism of the activity enhancement was consequently revealed by molecular dynamic simulation. The mutant DhaA31 has all tunnels very narrow, preventing the access of water molecules to its active site. The desolvated active site of the mutant DhaA31 allowed more frequent formation of the reactive complex, since no water molecules could compete with TCP for the nucleophile (Figure 2). The subsequent kinetic analyses confirmed that mutations improved the rate of nucleophilic attack on TCP, while making the product-release rate-limiting. Identification and engineering of functionally important tunnels in the enzyme with buried active site was found to be successful strategy for modification of its catalytic properties. The importance of tunnels and their dynamics for protein function led us develop, in collaboration with the group of Professor Jiří Sochor from the Faculty of Informatics of Masaryk University, the software CAVER, which is applicable for calculation of tunnels from molecular dynamic simulations.[1]

[1] http://loschmidt.chemi.muni.cz/caver/

3 HotSpot Wizard: A Web Server for Identification of Hot Spots in Protein Engineering

We have developed a web server HotSpot Wizard for identification of "hot spots" for rational engineering of substrate specificity, activity and enantioselectivity of enzymes and for annotation of protein structures. The "hot spots" for mutagenesis are selected through the integration of structural, functional and evolutionary information obtained from selected bioinformatics databases and tools (Figure 3). HotSpot Wizard requires a structure of the query protein an e-mail address as the only obligatory inputs. In the first step, the Catalytic Site Atlas and UniProt databases are used to determine the residues indispensable for protein function. HotSpot Wizard then searches for the "functional" residues that potentially come into contact with substrates or products by CASTp, identifying the active site pocket, and by CAVER, calculating tunnels connecting buried cavities with the outside solvent. Mutagenesis targeting these residues has a better chance to produce enzyme variants with novel catalytic properties than mutations localized in other parts of the protein structure. On the other hand, some of the "functional" residues may be indispensable for catalysis and their replacement lead to complete loss of enzyme function. To avoid this, HotSpot Wizard uses the information about evolutionary conservation to estimate the mutability of individual amino acid positions of the query protein. Only "functional" residues with high mutability, i.e., high tolerance to amino acid replacements, are assigned as "hot spots". Calculation of evolutionary conservation is a multistep process involving the collection of related sequences by the BLAST search against the nr database of NCBI, sequence clustering by CD-HIT and alignment by MUSCLE. Constructed multiple sequence alignment serves as an input for the Rate4Site program for estimation of evolutionary conservation of individual amino acid positions and assignment of mutability. HotSpot Wizard results are mapped on the protein structure and can be visualized directly in the web browser using the Jmol applet or downloaded as the input for the PyMOL visualization software. For individual residues, information about their mutability, structural location, functional role and annotations are provided.

For validation purposes, the "hot spot" residues identified by HotSpot Wizard were compared with the mutations extracted from the Protein Mutant Database and the primary literature for several enzymes, e.g. haloalkane dehalogenase, phosphotriesterase, 1,3-1,4-β-D-glucan 4-glucanohydrolase and β-lactamase. This comparison confirmed that mutagenesis targeting the "hot spots" positions leads to changes in catalytic properties of enzymes. The effectivity of mutagenesis targeting highly variable positions was assessed by comparing HotSpot Wizard results with the data from the published systematic mutagenesis studies: (i) 3315 mutants of lactose repressor, (ii) 1930 mutants of T4 lysozyme, (iii) 676 mutants of barnase and (iv) 364 mutants of HIV reverse transcriptase. The proportion of deleterious mutations in the positions assigned as highly mutable by HotSpot Wizard were compared with deleterious mutations in the entire protein. Calculated ratio of deleterious mutations in the mutable positions versus in the entire structure were: (i) 2.7 % versus 25.3 % for lac-

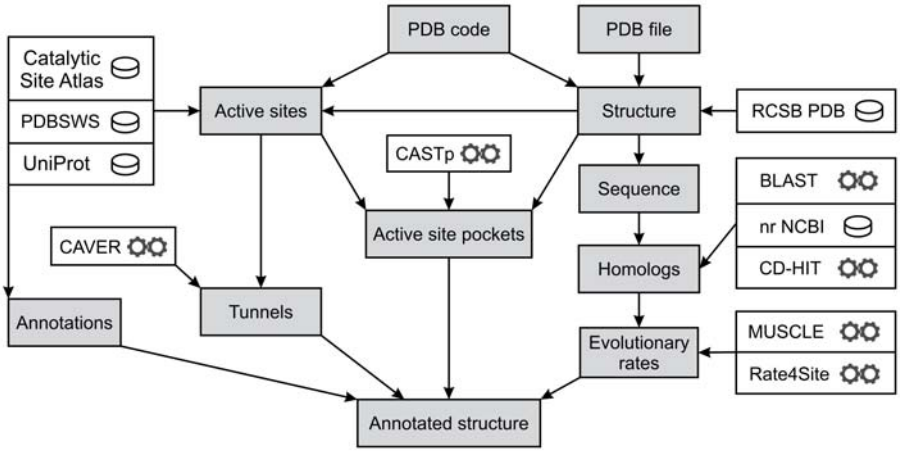


Fig. 3. Flow chart of the HotSpot Wizard calculation. The "hot spots" for mutagenesis are selected through the integration of structural, functional and evolutionary information from the bioinformatics databases and the computational tools.

tose repressor, (ii) 1 % versus 9.1 % for T4 lysozyme, (iii) 0.4 % versus 4.8 % for barnase and (iv) 3.9 % versus 30.7 % for HIV reverse transcriptase. These data confirm that mutagenesis targeting the "hot spot" positions identified by HotSpot Wizard provides significantly higher proportion of viable variants, than random mutagenesis. HotSpot Wizard is available to the scientific community at http://loschmidt.chemi.muni.cz/hotspotwizard/.

4 Dynamics and Solvatation of Tunnel Mouth of Haloalkane Dehalogenases

The dynamics, solvatation and structure of tunnel mouth of two haloalkane dehalogenases, DbjA and DhaA, were studied in collaboration with the group of Professor Martin Hof from the Heyrovsky Institute of the Academy of Scicences of the Czech Republic. The tunnel mouth of DbjA and DhaA were specifically labeled by the fluorescence dye—coumarine. The coumarine dye was covalently attached to the enzyme active site via its nucleophile. The time-resolved fluorescence spectroscopy allowed us to investigate properties of coumarine probe and its microenvironment such as polarity, accessibility and mobility. These results differed significantly between DbjA and DhaA enzymes. By combining the experimental results with molecular dynamic simulations, the differences in the dye microenviroment could be related to the anatomy (Figure 4), dynamics (Figure 5) and solvatation of the tunnel mouths of investigated enzymes. Moreover, the dynamics of waters around this region is extremely slowed down in comparison to that of bulk solvent suggesting functional importance of this region, which is evolutionally the most variable in the haloalkane dehalogenases. The

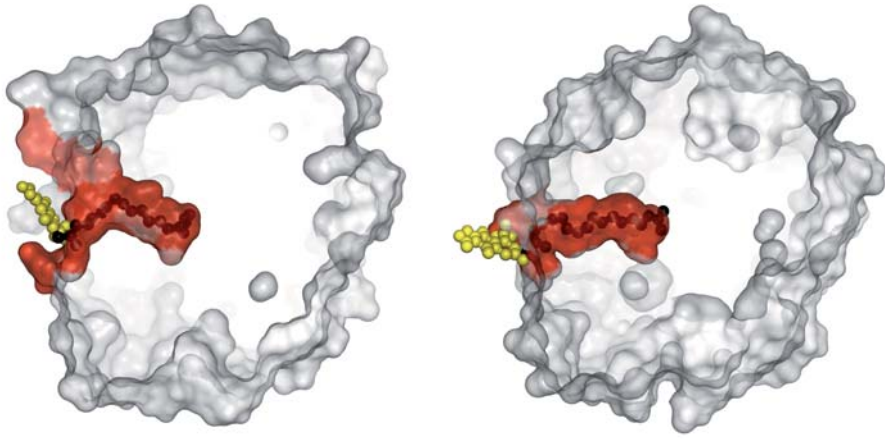


Fig. 4. Anatomy of the tunnel mouth of DbjA (left) and DhaA (right) enzymes specifically labeled by coumarin dye. Surface of enzyme is depicted in gray, main tunnel in red, fluorescent dye in yellow balls and covalent linker connecting enzyme with dye in black balls.

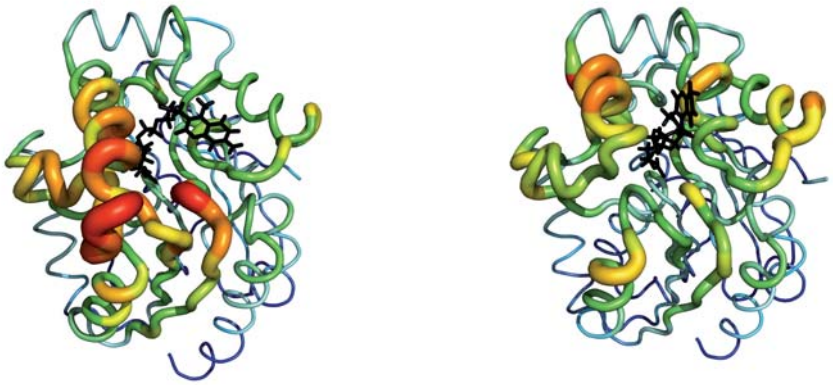


Fig. 5. Flexibility of the protein backbone in the proteins with covalently attached coumarine probe (black stick) in DbjA (left) and DhaA (right). The color scale and thickness of the ribbon indicate flexibility of the residues quantified by B-factors as follows: dark blue < light blue < green < yellow < orange < red.

relevance of tunnel mouth to enzyme activity and specificity was confirmed by recent mutagenesis study (see the project of Engineering of the Pathways for Product Release in the Haloalkane Dehalogenases).

Financing Sources

- the Grant Agency of the Czech Republic(203/08/0114 and 201/07/0927)
- the Ministry of Education, Youth and Sports of the Czech Republic (LC06010, MSM0021622412 and MSM0021622413)
- the Grant Agency of the Czech Academy of Sciences (IAA401630901)

Used Programs and Applications

AMBER 7 – 10
AUTODOCK 3, 4
BLAST 2.2.20
CD-HIT 07.01.31
CHARMM C33
FOLDX 3.0 beta4
GAUSSIAN 94, 03
MUSCLE 3.6
PYMOL 0.99 – 1.1
RATE4SITE 2.01
VMD 1.8.5 – 1.8.7

List of Publications Dedicated to MetaCentrum

1. Klvaňa, M., Pavlová, M., Koudeláková, T., Chaloupková, R., Dvořák, P., Prokop, Z., Stsiapanava, A., Kutý, M., Kutá-Smatanová, I., Dohnálek, J., Kulhánek, P., Wade, R.C., Damborský, J., 2009: Pathways and Mechanisms for Product Release in the Engineered Haloalkane Dehalogenases Explored using Classical and Random Acceleration Molecular Dynamics Simulations. *Journal of Molecular Biology* 392: 1339-1356. **IF = 4.146**
2. Pavlová, M., Klvaňa, M., Chaloupková, R., Banáš, P., Otyepka, M., Wade, R., Nagata, Y., Damborský, J., 2009: Redesigning Dehalogenase Access Tunnels as a Strategy for Degrading an Anthropogenic Substrate. *Nature Chemical Biology* 5: 727–733. **IF= 14.612**
3. Pavelka, A., Chovancová, E., Damborský, J., 2009: HotSpot Wizard: A Web Server for Identification of Hot Spots in Protein Engineering. *Nucleic Acids Research* 37: W376–W383. **IF = 6.878**
4. Jesenská, A., Sýkora, J., Olżyńska, A., Brezovský, J., Zdráhal, Z., Damborský, J., Hof, M., 2009: Nanosecond Time-Dependent Stokes Shift at the Tunnel Mouth of Haloalkane Dehalogenases. *Journal of the American Chemical Society* 131: 494-501. **IF = 8.091**

Presentations and Poster Prepared Using MetaCentrum Resources

1. Fořtová, A., Sýkora, J., Olzyńska, A., Brezovský, J., Zdráhal, Z., Hof, M. and Damborský, J.: Study of Water Molecule Dynamics at the Tunnel Openings of Haloalkane Dehalogenase, March 12-14, 2009, Nové Hrady, Czech Republic (presented by AF).

2. Damborský J.: Engineering of Tunnels in Proteins with Buried Active Sites, March 29- April 3, 2009, Sant Feliu, Spain, (presented by JD).
3. Brezovský, J., Prokop, Z., Sato, Y., Florián, J., Mozga, T., Chaloupková, R., Koudelaková, T. , Jeřábek, P., Natsume, R., Nagata, Y., Senda, T. and Damborský J.: Two Bases of Enantioselectivity in Haloalkan Dehalogenase DbjA, April 14-15, 2009, Brno, Czech Republic, (presented by JB).
4. Pavlová, M., Klvaňa, M., Prokop, Z., Chaloupková, R., Banáš, P., Otyepka, M., Wade, R.C., Nagata, Y. and Damborský J.: Construction of Improved Biocatalyst by Engineering of the Access Tunnels: A New Concept in Enzyme Engineering, July 5-9, 2009, Bern, Switzerland, (presented by MP).
5. Damborský, J., Prokop, Z., Fořtová A., Chaloupková, R., Pavlová, M., Chovancová E, Brezovský, J., Klvaňa, M., Koudeláková, T., Mozga, T., Pavelka, A., Dvořák P.and Straková, M.: Engineering of Biocatalysts with Modified Activity, Specificity and Stability by Modification of Access Tunnels, September 20-24, 2009, Groningen, The Netherlands, (presented by JD).
6. Pavelka, A., Chovancová, E., Damborský, J.: HotSpot Wizard: a Web Server for the Identification of Mutagenesis Hot Spots in Enzyme Structures, March 12-14, 2009, Nové Hrady, Czech Republic, (presented by AP).
7. Brezovský, J., Prokop, Z., Sato, Y., Florián, J., Mozga, T. , Chaloupková, R., Koudeláková, T., Jeřábek, P. , Natsume, R., Nagata, Y., Senda, T., and Damborský, J.: Modulation of Enantioselectivity in Haloalkane Dehalogenase DbjA by Engineering of a Surface Loop, March 12-14, 2009, Nové Hrady, Czech Republic, (presented by JB).
8. Dvořák, P., Pavlová, M., Klvaňa, M., Brezovský, J., Chaloupková, R., Prokop, Z. and Damborský, J.: Increasing the Activity of Haloalkane Dehalogenase DhaA with Non-natural Anthropogenic Substrate 1,2-dichloroethane using Methods of Focused Directed Evolution, September 20-24, 2009, Groningen, The Netherlands, (presented by PD).

Metal Ions and Their Interactions with Nucleic Acids

Zdeněk Chval[1] and Ingrid Romancová[2]

[1] Faculty of Health and Social Studies, University of South Bohemia,
J. Boreckého 27, 370 11 České Budějovice
[2] Faculty of Science, University of South Bohemia,
Branišovská 31, 370 05 České Budějovice
chval@jcu.cz

During last two years the computational resources of Metacenter have been used for the following projects:

1. Influence of Environmental Effects on Specificity of Metal Binding to Uracil
   Uracil $\cdots$ Mg^{2+} ion was chosen as a model system to evaluate influence of the environmental effects, namely of the metal ion hydration shell, the presence of the phosphate group, the implicit bulk water on the specificity of the binding. This work was motivated by the fact that most DFT and ab initio studies about interactions of metal ions with biomolecules consider a bare metal ion in the gas phase calculations which in case of nucleic acid bases lead to a clear preference of the bidentate binding sites leading to a stabilization of the rare tautomers.
   In case of the uracil-Mg^{2+} system we show that after already the inclusion of the first hydration shell of the Mg^{2+} ion no clear preference of the bidentate binding can be found. Further inclusion of the implicit "bulk" water molecules and the presence of the negatively charged phosphate group clearly supports monodentate Mg^{2+} binding to the keto group of the canonical uracil tautomer.
2. Role of the Mg^{2+} ion for 2'OH Activation in RNA Catalysis.
   The mechanism of the 2'OH activation of catalytically active self-cleaving ribozymes was studied theoretically by DFT and MP2 methods. Influence of the outer/inner sphere coordination of the Mg^{2+} ion on the pKa value change of 2'OH group was evaluated in the presence of either a hard OH^- or a soft nucleobase nucleophile. Kinetics of 2'OH activation by the inner-sphere Mg^{2+} coordination is determined by the coordination number change (from 6 to 5) step on the Mg^{2+} center since outer-sphere coordination of 2'OH to Mg^{2+} ion was considered in the starting structure. Factors influencing feasibility of the pentacoordinated Mg^{2+} ion with respect to the hexacoordinated Mg^{2+} ion are discussed in detail. This project is done in cooperation with dr. Leclerc, MAEM UHP Nancy, France.
3. The mechanism of the hydrolysis of the RAPTA-B complex.
   The mechanism of the two step hydrolysis of the RAPTA-B complex was studied using a density functional theory (B3LYP). All relevant structures (minima and transition states) were optimized in the gas phase and in the

implicit solvent (COSMO and PCM models). A detailed (and very time consuming) analysis of the B3LYP/PCM reaction paths has been doing using IRC method with the subsequent CCST(T) single point calculations. This project is done in cooperation with prof. J.V. Burda, MFF UK, Praha.

Used Programs

Gaussian03, Amber10

Financing Sources

Czech Science Foundation: grant No. 204/09/J010, Ministry of Education of the Czech Republic: grant No. ME09062

List of publications dedicated to MetaCentrum

1. I. Romancová: Theoretical Study of Tautomeric Forms of Uracil and Their Interactions with Mg^{2+}, Bachelor Thesis, Faculty of Science, University of South Bohemia, České Budějovice, 2009
2. Z. Chval, F. Leclerc, D. Chvalová: Modeling Reaction Pathways for 2'OH Activation in RNA Catalysis: Metal-ion/Nucleobase Mechanisms in Self-Cleaving Ribozymes, to be submitted
3. I. Romancová and Z. Chval: Influence of Environmental Effects on Specificity of Metal Binding to Uracil, in preparation

Sorting Permutations by Prefix Reversals

Josef Cibulka

Department of Applied Mathematics, Charles University, Faculty of Mathematics and Physics, Malostranské nám. 25, 118 00 Prague, Czech Republic;
cibulka@kam.mff.cuni.cz

1 Introduction

In 1975, Jacob Goodman posed the pancake problem in American Mathematical Monthly. We are given a stack of pancakes each two of which have different sizes and our aim is to sort them in as few operations as possible to obtain a stack of pancakes with sizes increasing from top to bottom. The only allowed sorting operation is a "spatula flip", in which a spatula is inserted beneath an arbitrary pancake, all pancakes above the spatula are lifted and replaced in reverse order.

We can see the stack as a permutation π. A flip is then a prefix reversal of the permutation. The set of all permutations on n elements is denoted by S_n, $f(\pi)$ is the minimum number of flips needed to obtain $(1, 2, 3, \ldots, n)$ from π and

$$f(n) := \max_{\pi \in S_n} f(\pi).$$

In general $15\lfloor n/14 \rfloor \leq f(n) \leq 18n/11 + O(1)$. The exact values of $f(n)$ were previously known for $n \leq 17$, see Table 1. Using the computers of the METACentrum grid, it was possible to compute the values of $f(18)$ and $f(19)$.

A related problem in which the reversals are not restricted to intervals containing the first element received considerable attention in computational biology; see, e.g., Brian Hayes, Sorting out the genome, Amer. Scientist, 95:386–391, 2007.

Gates and Papadimitriou proposed a variation on the problem in which pancakes are burnt on one of their sides. In this burnt pancake problem, the aim is not only to sort them by their sizes, but we also require that at the end, they all have their burnt sides down. Let I_n be the stack of n sorted pancakes with all burnt side down and $-I_n$ the sorted stack with all burnt sides up.

Let $g(C)$ be the minimum number of flips needed to obtain I_n from C and let

$$g(n) := \max_{\pi \in S_n, v \in \{0,1\}^n} g((\pi, v)).$$

Exact values of $g(n)$ were previously known for all $n \leq 12$ and our computations extended this to all $n \leq 17$, see Table 1.

In 1979 Gates and Papadimitriou provided the bounds $3n/2 - 1 \leq g(n) \leq 2n + 3$. Since then these were improved only slightly by Cohen and Blum to $3n/2 \leq g(n) \leq 2n - 2$, where the upper bound holds for $n \geq 10$. The result $g(16) = 26$ further improves the upper bound to $2n - 6$ for $n \geq 16$. Cohen and Blum also conjectured that the maximum number of flips is always achieved for the stack $-I_n$. But our computations found two counterexamples with $n = 15$.

n	$f(n)$	$g(n)$	$g(-I_n)$
2	1	4	4
3	3	6	6
4	4	8	8
5	5	10	10
6	7	12	12
7	8	14	14
8	9	15	15
9	10	17	17
10	11	18	18
11	13	19	19
12	14	21	21
13	15	**22**	22
14	16	**23**	23
15	17	**25**	24
16	18	**26**	26
17	19	**28**	28
18	**20**		29
19	**22**		**30**
20			**32**

Table 1. known values of $f(n)$, $g(n)$ and $g(-I_n)$

2 Results obtained using the METACentrum grid

The values $f(18) = 20$ and $f(19) = 22$ were computed by the method of Kounoike et al. and Asai et al. . It is an improvement of the method of Heydari and Sudborough and decreased the number of permutations to be tested from $19! \simeq 10^{17}$ to roughly 10^{13}. Let $\mathbb{U}_n^m$ be the set of stacks of n unburnt pancakes requiring m flips to sort. For every stack $U \in \mathbb{U}_n^m$, 2 flips always suffice to move the largest pancake to the bottom of the stack, obtaining stack U'. Since then, it never helps to move the largest pancake. Therefore U' requires exactly the same number of flips as U'' obtained from U' by removing the largest pancake and thus U'' requires at least $m - 2$ flips.

To determine $\mathbb{U}_n^i$ for all $i \in \{m, m+1, \ldots, f(n)\}$, it is thus enough to consider the set $\cup_{m'=m-2}^{f(n-1)} \mathbb{U}_{n-1}^{m'}$. In each stack from this set, we try adding the pancake number n to the bottom, flipping the whole stack and trying every possible flip. The candidate set composed of the resulting and the intermediate stacks contains all the stacks from $\cup_{i=m}^{f(n)} \mathbb{U}_n^i$.

What remains is to determine the value of $f(U)$ for each stack U in the candidate set, which is done using the A* search. During the A* search, we need to compute a lower bound on the number of flips needed to sort a stack. It is done differently then in the previous implementations, which leads to considerable speedup: We try all possible sequences of flips that create an adjacency in every flip. If some such sequence sorts the stack, it is optimal and we are done. Otherwise, we obtain a lower bound equal to the number of adjacencies that are

needed to be made plus 1 (here we count pancake n at the bottom of the stack as an adjacency).

In addition, we also use a heuristic to compute an upper bound. If the upper bound is equal to the lower bound they give the exact number of flips.

It was previously known, that $f(18) \geq 20$ and $f(19) \geq 22$. No candidate stack of 18 pancakes needed 21 flips thus $f(18) = 20$. Then $f(19) = 22$ because $f(19) \leq f(18) + 2 = 22$.

A modification of this method was also used to compute the values of $g(n)$ up to $n = 17$.

The stack obtained from $-I_n$ by flipping the topmost pancake is known as J_n. Let Y_n be the stack obtained from $-I_n$ by changing the orientation of the second pancake from the bottom. The two found stacks of 15 pancakes requiring 25 flips are J_{15} and Y_{15} and they are the first known counterexamples to the Cohen-Blum conjecture which claimed that for every n, $-I_n$ requires the largest number of flips among all stacks of n pancakes. However, no other J_n or Y_n with $n \leq 20$ is a counterexample to the conjecture.

Data and source codes of programs mentioned above can be downloaded from the following webpage: http://kam.mff.cuni.cz/ cibulka/pancakes.

Grant support:

Work on this paper was supported by the project 1M0545 of the Ministry of Education of the Czech Republic and by the Czech Science Foundation under the contract no. 201/09/H057.

Publications and presentations:

1. Josef Cibulka, Average number of flips in pancake sorting, submitted, preprint available at http://arxiv.org/abs/0901.3119.
2. Josef Cibulka, Pancake sorting, CanaDAM 2009, Montréal, Canada
3. Josef Cibulka, Palačinkové třídění, STTI 2009, Praha, Czech Republic

Hypertext Atlases of Pathology

Josef Feit[1], Luděk Matyska[2], Lukáš Hejtmánek[2], Michal Procházka[2], Vladimír Ulman[3], Věra Feitová[4], Hana Jedličková[5], Marta Ježová[1], and Mojmír Moulis[1]

[1] Institute of Pathology, Medical Faculty, Masaryk University, Brno
University Hospital Brno, Jihlavská 20, 66263 Brno, Czech Republic
jfeit@ics.muni.cz, martazej@centrum.cz, mmoulis@fnbrno.cz
[2] Institute of Computer Science, Masaryk University
Botanická 68a, 60200 Brno, Czech Republic
ludek@ics.muni.cz, xhejtman@ics.muni.cz, michalp@ics.muni.cz
[3] Faculty of Informatics, Masaryk University
Botanická 68a, 60200 Brno, Czech Republic
xulman@fi.muni.cz
[4] Department of Radiology, Medical Faculty, Masaryk University
St. Anna Hospital Brno, Pekařská 53
65691 Brno, Czech Republic
vera.feitova@fnusa.cz
[5] Department of Dermatology, Medical Faculty, Masaryk University
St. Anna Hospital Brno, Pekařská 53, 65691 Brno, Czech Republic
hana.jedlickova@fnusa.cz

Abstract. Hypertext Atlases of Pathology consist of short texts and image collections. The first of them, the Atlas of Dermatopathology, is available since 1998. Later the Atlas of Fetopathology, Neonatal Pathology and Atlas of Pathology of Bone Marrow were added. These atlases are available in Czech and English and are meant as reference and teaching resources for pregraduate as well as postgraduate students. The Atlas of Pathology for pregraduate students is available only in Czech.
The image collections contain various types of images: macroscopic images (clinical pictures, pictures from the autopsies and other), endoscopical pictures, images from CT, MRI and angiography and of course histological slides.

1 Introduction

Evaluating images (especially histological) is essencial for pathologists. In this way diagnoses of many diseases are made. Without histological evaluation any reliable oncological diagnosis is impossible.

Pathologists evaluate typical diagnostic signs using the microscope. Not only presence of certain features, but their absence is important as well. When making the diagnosis, pathologists evaluate the tissues using various stainings and magnifications. Typical structures on organoid level are evaluated at low magnifications by looking at large areas of tissues. Cytological details are evaluated using high magnifications on cellular level. Techniques using virtual slides (scanned

piece after piece using high resolution objective, where low magnifications are simulated by diminishing the images) are excellent tools for teaching, reference and (in near future) for diagnostic work as well.

We wanted publicly available image collection, accessible over the Internet and independent of the operating system. No additional software should be necessary to full use of the Atlas.

JavaScript is used to enhance the possibilites to work with the images. Users can activate arrows pointing to the important diagnostic parts of images. It is possible to go through the time series or set of planes of angiographic or CT/MRI slides. From the earliest versions of the atlas the virtual microscope to access the histological slides is available.

Images are annotated: arrows can be activated, pointing to important diagnostic signs of each image. This feature is especially useful for students, who use the atlases as studying meterial.

2 Histological Images

The goal was to create image collection of histological images of highest quality and resolution possible. At early years we used special scanning digital camera Leica S1 mounted to the microscope, giving images 5012×5012 px. These images were processed and virtual microscope, based on JavaScript driven synchronization of two windows of the internet browser were used to show overview as well as the detail of histological pictures. The maximmum size of image accessible in this way was 7200 px.

Since the performance of microscopic lenses does not allow to create images more than 2500×2500 px, we have started using robotic microscope with movable scanning table and digital camera. The system was equipped with autofocus and enabled creating very large images by gradual taking small images and stitching them later together.

By the end of 2008 by participating in MediGrid project (CESNET) we have obtained a new hardware, based on Nikon 90i motorized microscope, driven by NIS AR software (LIM Prague). It became possible to obtain very large images in several focusing planes. Another project was started, which used the Metacenter computing facilities to process these images.

2.1 Step by Step Image Acquisition

Histological images (parts of future images) are now taken using 2560×1920 px 3×8 bit digital camera. The system performs autofocus on every other (or third) tile and then can take images in several focusing planes (usually 5). The system performs the necessary shading correction of each image taken. These images are stored on a disk for future processing. The system can take usually two tile stacks in one minute (including autofocus).

The tiles are later processed (color correction, sharpness and contrast, saturation, size change) and stitched together into large images.

2.2 Stitching the Image Parts Together

Special algorithm for image stitching and processing for the virtual microscope was developed through cooperation with the Faculty of Informatics, MU (dr. V. Ulman).

Individual tiles are taken with some overlap (9%) and the information in overlapping areas is used to exactly align the tiles. Because of unavoidable change of the focusing plane these areas are not exactly equal. Therefore the images are stitched along randomly generated curve to provide the best results. Because the images contain non-contrast parts (or parts without any image), the system calculates the position of the tiles starting with the most contrast overlapping areas.

2.3 Manipulation with Large Images

We are able to create images of more than 100×100 tiles, enabling us to cover relatively large areas of the histological slide. The actual area depends on the lens used (20×, 40× or 60×oil immersion). After decreasing the size of the tiles and cropping the resulting image we obtain images over 50000×50000 px. Our largest image so far is of 118252×133200 px in its final size. And of course, each such image is produced 5 or 7× for each focusing plane. Raw images may have over 40 GB for each focusing plane, over 200 GB for one slide.

Working with images of that size presents another challenge. The only image processing software we have found, capable of manipulating such images, is `nip2/vips`. This software is especially suitable for batch processing, is open-source and can run in various operating systems, namely linux. We use this software for cropping, rotating, color and contrast improvement and size changes of our images. After processing these images are converted to PNG and archived.

2.4 Virtual Microscope

Large images are cut into squares of increasing size (256, 512, 1024, ...). These squares are downsized into 256×256 px jpg images and organized into hierarchic directories, understood by JavaScript programme, running on user's computer. The program reacts to image drag, focusing, window resize and other actions by requesting proper image parts, putting them on the proper place and releasing the parts which are not necessary. The system is similar to Google Maps.

The window of the virtual microscope can remain small, showing only detail of the area of the main window, where the user clicked in the window with basic image. Or can be stretched (even fullscreen) to show as much of each picture as possible (picture 1).

The number of partial images in the directory of one large multifocal image is well over 1 million.

The virtual microscope (and similar CT/MRI/angiography browsers) are written in JavaScript. No additional software is necessary. The microscope requires no server and can run directly from hard disk, flash drive or other media.

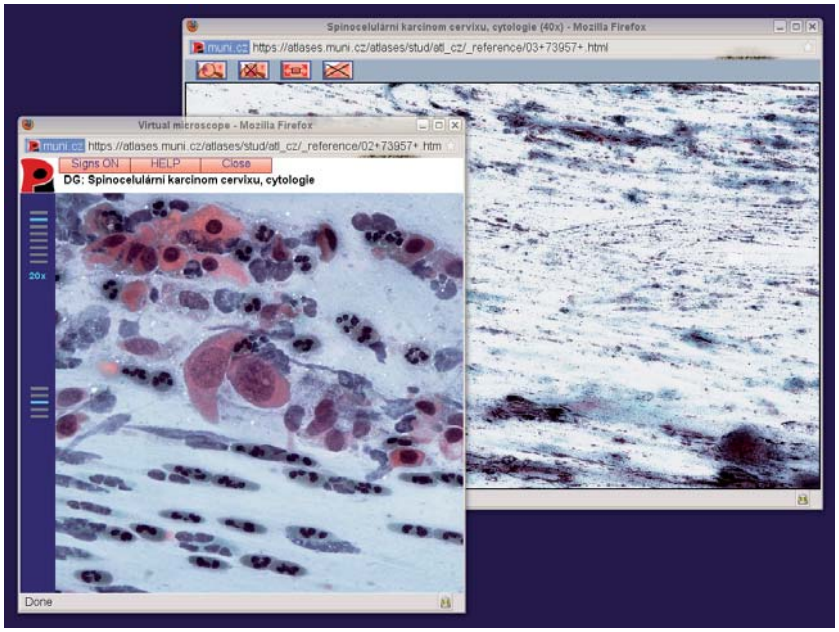


Fig. 1. The window with the image in basic size (900 px) is in the background, in the left foreground is the virtual microscope window. The top left scale changes the magnification, bottom left scale is for focusing.

3 Metacenter

Distributed computing facilities of the Metacenter are used from individual tile processing, over the large image manipulation to virtual slide preparation.

3.1 Cluster Optimizations

Distributed computing facilities of the Metacenter contributed mainly by processing power. However, running scanned images transformation in Metacenter environment required substantial changes of processing scripts and programs.

First of all, all scripts had to be changed so that they run as jobs through a batch system. To utilize large distributed computing facilities, we had split tasks that processed particular steps, into many small jobs that run in parallel.

The second problem has been with processing of many small files used both as input and output data. Distributed computing facilities do not perform well on small files stored on a kind of shared storage. Creating and opening of a file is expensive operation that is not amortized in reading/writing of the file in such conditions. Therefore, we changed the programs that manipulated with those small files so that they use `tar` archives which avoids the expensive operations and improves overall running time. We were able to cut running time from several hours to tens of minutes.

The third problem was in suboptimal program chaining. We had one program that cuts images into tiles and the second program that scale tiles. It needs the intermediate tiles to be stored somewhere with highly utilize system resources. It also means that we had to start image scaler for every small tile which is costly as starting the scaler takes comparable time to scaling itself. We changed cutting program so that it cuts and scales tiles at once and no storage area for intermediate tiles is needed and no image scaler needs to be run. This optimization also significantly cut down overall running time.

The already mentioned `nip2` program is able to manipulate really large images (many gigapixels) but it is not very fast mainly if you are using low memory (e.g., 4, 8 GB RAM) computing node. We measured that running time significantly improves if the computing node is equipped with more memory (at least half of the image size, i.e., for 32 GB image you need 16 GB RAM) or equipped with fast disk storage such as solid state disks (SSD). Metacenter provided three such computing nodes equipped with SSDs. Running time of the `nip2` program was cut from about four hours to less than one hour.

4 Results

All our atlases are available over the Internet for free, but registration is necessary to access the images. Recently all the atlases combined contain about 7000 images, most of them are annotated. Of those over 3000 images were obtained by image stitching. Video snippets are being added as well.

About 300 very large multifocal images were computed using distributed computational capacity of the Metacenter. The number of registered users is over 10 000 (March 2010) and access rate is about 1000 visits/day.

Hypertext Atlases are supported by the following grant projects:

- MediGrid: From this project resources was acquired new automatic microscope Nikon 90i, camera, software, antivibration table and computers. MediGrid servers are used to process the images from tiles to virtual microscope. The MediGrid project was covered by the grant of the Academy of Sciences of the Czech Republic "MediGrid—methods and tools for using GRID network in biomedicine" (T202090537).
- MEFANET: Building the educational network of Czech medical faculties for creation and sharing of teaching materials on common platform. (1. 1. 2009 – 31. 1. 2009), Code: RP MŠMT C15/2009.

Modeling of Vibrational Molecular Properties

Jan Horníček, Jakub Kaminský, Valery Andrushchenko, Martin Dračínský, and Petr Bouř

Institute of Organic Chemistry and Biochemistry, Academy of Sciences,
166 10 Prague, Czech Republic
hornicek@uochb.cas.cz, kaminskj@gmail.com, vandrush@uochb.cas.cz,
dracinsky@uochb.cas.cz, bour@uochb.cas.cz

Prediction and interpretation of vibrational properties of molecules require the knowledge about the dependence of energy and other variables on atomic positions. In general, this can be obtained by quantum-chemical electronic methods, which do not need any additional input parameters. The price for this convenience is a high computational cost resulting in many practical restrictions including limited accuracy and size of systems than can be investigated. The computer restrictions need to be overcome both by faster computers and more clever computational methods.

Proper description of molecular vibrations is essential, for example, for structural spectroscopic studies of small molecules [1], proteins [2], or nucleic acids [3]. From the comparison of calculated and experimental spectral curves we get useful information about the geometry, dynamics, and interactions between molecules. Also, accurate simulations of nuclear magnetic resonance (NMR) and optical rotatory dispersion (ORD) parameters should involve a vibrational contribution [4]. In particular, the vibrational part is important for a correct description of temperature dependencies and isotopic effects.

Lately, we performed a detailed theoretical and experimental study of the N-methylacetamide molecule that contains the amide group, which is the principle chromophore monitored in peptide and protein vibrational studies [5]. High-resolution infrared spectra were obtained within collaboration with the University of Wyoming. The computational analysis not only explained finer features in the spectra (Figure 1), but revealed an interesting property of the out of plane (oop) nitrogen deviation potential. The group thus becomes more non-planar and flexible in hydrophobic environments, which can have implications in molecular biology.

Another application of molecular computing can be documented on simulations of the alanine hydration shell and its influence on NMR chemical shifts and indirect spin-spin coupling constants [6]. Different theoretical models provide different distribution of the water molecules around the solvent (Figure 2) and their validity could be estimated by comparison of the calculated and experimental spectroscopic variables.

Intermolecular hydrogen motion can be encountered in many biological processes, including photosynthesis and respiratory cycles. We have studied this phenomenon in calixphyrin derivatives (Figure 3) that are also important complexation compounds [7]. The calculated energy profiles could be related to experimental activation energies obtained from coalescent NMR spectral lines at

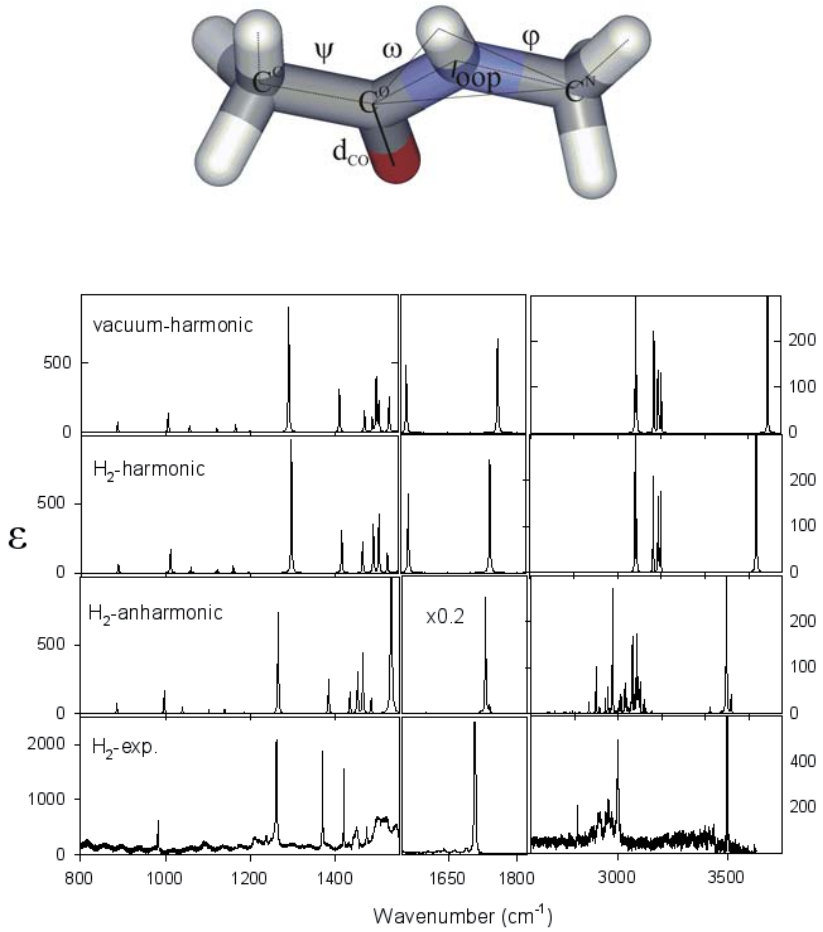


Fig. 1. N-methylacetamide molecule with principle coordinates and absorption spectra calculated (MP2/6-311++G**/CPCM) in vacuum, in a hydrogen-like continuum at the harmonic and anharmonic levels, and the experimental spectrum recorded in solid-hydrogen matrix at 2 K.

low temperatures, and revealed the transfer mechanism. Interestingly, the hydrogen motion is significantly influenced by the distant substituents.

Acknowledgement

We appreciate the support of the GAČR, GAAV, and MŠMT grant agencies (grant numbers 203/06/0420, 202/07/0732, M200550902, A4005507020, and MSM6383917201), and the Metacentrum staff.

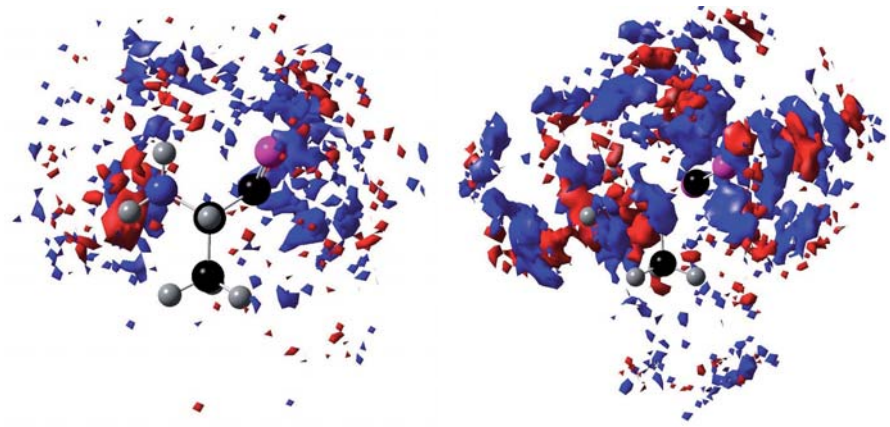


Fig. 2. Water oxygen (red) and hydrogen (blue) probability distributions as obtained from the classical (left) and quantum CPMD (right) dynamics. The quantum distribution provides better NMR spectroscopic parameters.

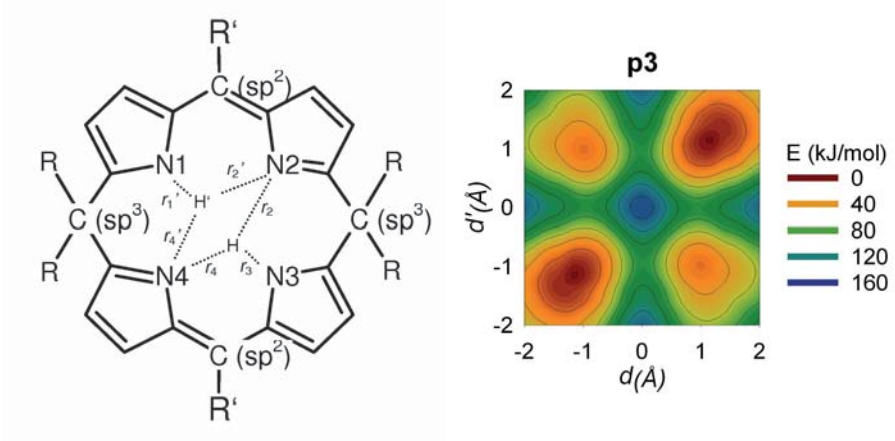


Fig. 3. Calixphyrin cycle (left) and example of calculated energy profile in generalized hydrogen coordinates ($d = r4 - r3$ and $d' = r2' - r1'$).

Used Programs

Gaussian and special Fortran codes.

References

[1] M. Buděšínský, P. Daněček, L. Bednárová, J. Kapitán, V. Baumruk, and P. Bouř, J. Phys. Chem. A 112, 8633 (2008).

[2] P. Bouř, J. Kim, J. Kapitán, R. P. Hamer, R. Juany, L. Wu, and T. A. Keiderling, Chirality 20, 1104 (2008).
[3] V. Andrushchenko and P. Bouř, J. Comput. Chem. 29, 2693 (2008).
[4] M. Dračínský, J. Kaminský, and P. Bouř, J. Phys. Chem. 130, 094106 (2009), J. Kaminský, I. Raich, K. Tomčáková, P. Bouř, J. Comput. Chem., in print, 2010.
[5] V. Andrushchenko, P. Matějka, D. T. Anderson, J. Kaminský, J. Horníček, L. O. Paulson, and P. Bouř, J. Phys. Chem. A 113, 9727 (2009).
[6] M. Dračínský, J. Kaminský, and P. Bouř, J. Phys. Chem. B (113), 14698 (2009).
[7] J. Horníček, H. Dvořáková, and P. Bouř, J. Phys. Chem. A, 114, 3649 (2010).

Experiments with Different Language Models in the Automatic Speech Recognition Task

Pavel Ircing

University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics
Univerzitní 8, 306 14 Pilsen, Czech Republic
ircing@kky.zcu.cz

1 The Task at Hand

The aim of the automatic speech recognition (ASR) is to transcribe human speech into text. The state-of-the-art ASR methods are based on the so-called noisy channel model which assumes that the speech generation process and the consequent speech recognition can be described by the model depicted on Figure 1.

The human speaker is shown as consisting of two parts—the source of the communication is his mind, which specifies the word sequence W that will be pronounced by his vocal apparatus. Resulting acoustic waveforms are then processed by the acoustic processor, yielding the sequence of the acoustic features O. The aim of the last block of the model, the linguistic decoder, is to find the most likely sequence of words $\hat{W}$ given the observed acoustic data O.

Described formally, the task of the linguistic decoder is to find the word sequence $\hat{W}$ satisfying the equation

$$\hat{W} = \arg\max_{W} P(W|O). \tag{1}$$

This can be transformed (using the Bayes formula and the fact that $P(O)$ is constant for a given acoustic signal) into

$$\hat{W} = \arg\max_{W} P(O|W)P(W). \tag{2}$$

The ASR system could therefore be decomposed into two components—the acoustic model determining $P(O|W)$ (i.e., modeling the way in which words get converted into the acoustic features) and the language model computing the prior probabilities of all the word sequences $P(W)$. The optimal ratio between the contribution of the acoustic and the language model needs to be determined experimentally, as well as number of parameters of the language model itself. The MetaCentrum computing facilities were used exactly for this purposes.

Financing Sources

– The Ministry of Education, Youth and Sports of the Czech Republic: Project LC536

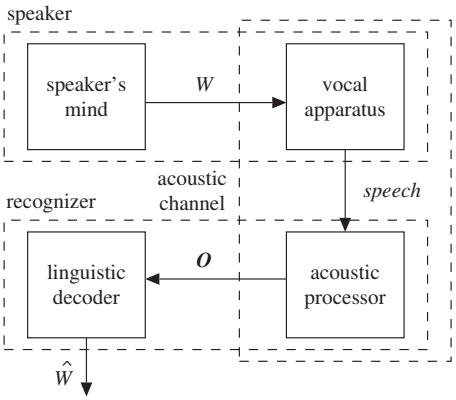


Fig. 1. The noisy channel model of speech recognition

– The Ministry of Education, Youth and Sports of the Czech Republic: Project 2C06020
– European Commission: Project IST-FP6-034434

Involvement in International Research Projects

– European Commission Sixth Framework Programme Information Society Technologies Integrated Project IST-34434

Used Programs and Applications

– AT&T Finite State Machine Speech Recognition Decoder

Publications dedicated to MetaCentrum

1. Pražák, A., Ircing, P., Müller, L.: Language model adaptation using different class-based models. In: Proceedings of SPECOM 2007, Moscow (2007) 449–454
2. Ircing, P., Psutka, J.V., Psutka, J.: Using Morphological Information for Robust Language Modeling in Czech ASR System. IEEE Transactions on Audio, Speech, and Language Processing **17**(4) (2009) 840–847

Integrated Parametric Study of Hybrid-stabilized Argon-water Arc Under Subsonic and Supersonic Regimes

J. Jeništa[1], H. Takana[2], H. Nishiyama[2], M. Bartlová[3], V. Aubrecht[3], P. Křenek[1], M. Hrabovský[1], T. Kavka[1], V. Sember[1], and A. Mašláni[1]

[1] Institute of Plasma Physics, AS CR, v.v.i., Thermal Plasma Department, Za Slovankou 3, 182 00 Praha 8, Czech Republic
jenista@ipp.cas.cz
[2] Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi, 980-8577, Japan
[3] Brno University of Technology, Technická 8, 616 00 Brno, Czech Republic

Abstract. This research presents numerical simulation of temperature and flow fields in the discharge and near outlet regions of the hybrid argon-water stabilized electric arc. Calculations for 300–600 A and high argon mass flow rates reveal transition from a transonic plasma flow for 400 A to a supersonic one for 600 A with the maximum Mach number of 1.6. Comparison with available experimental temperature and velocity profiles shows very good agreement.
Keywords: hybrid plasma torch, mass flow rate, partial characteristics, shock diamonds.

1 Introduction

The so-called hybrid stabilized electric arc, which was developed a few years ago at IPP AS CR in Prague, utilizes a combination of gas and vortex stabilization. In the hybrid argon–water plasma torch, the arc chamber is divided into the short cathode part, where the arc is stabilized by tangential argon flow, and the longer part, which is stabilized by water vortex. This arrangement provides not only the additional stabilization of the cathode region and the protection of the cathode tip but also offers the possibility of controlling plasma jet characteristics in wider range than that of pure gas- or liquid-stabilized arcs [1]. The arc is attached to the external water-cooled rotating disk anode at a few millimeters downstream of the torch orifice. The experiments made on this type of torch [1] showed that the plasma mass flow rate, velocity, and momentum flux in the jet can be controlled by changing the mass flow rate in the gas-stabilized section, whereas thermal characteristics are determined by the processes in the water-stabilized section. At present, this arc has been used for plasma spraying using metallic or ceramic powders injected into the plasma jet, as well as for the pyrolysis of waste (biomass) and production of syngas [2], which seems to be a promising environmentally friendly application of thermal plasma jets.

Numerical simulation provides an efficient tool for the optimization of operating conditions (arc current and mass flow rate) and prediction of temperature and velocity structures for these applications.

There are two basic aims of this research: 1) a numerical study of characteristics and processes in the hybrid arc under subsonic, transonic and supersonic plasma flow regimes especially at high currents and high argon mass flow rates; 2) a detailed comparison of the calculated results with experiments.

2 Physical Model and Numerical Approach

In the numerical simulation, we assume one-fluid two-dimensional axisymmetric compressible and turbulent plasma flow in local thermodynamic equilibrium with homogeneous mixing of water and argon species. The resulting set of conservative governing equations for density, velocity and energy (continuity, momentum and energy equations) includes both temperature- [3] and pressure-dependent transport and thermodynamic properties for Ar + H_2O plasma mixture.

Radiation loss from the argon-water arc is calculated in two ways: by the net emission coefficient for the required arc radius of 3.3 mm and by the partial characteristics method for different molar fractions of argon and water plasma species [4, 5] as a function of temperature and pressure. Continuous radiation, discrete radiation consisting of thousands of spectral lines, molecular bands of O_2 (Schuman-Runge system), H_2 (Lyman and Verner systems), H_2O and OH have been included in the calculation of partial characteristics. Broadening mechanisms of atomic and ionic spectral lines due to Doppler, resonance and Stark effects have been considered.

For time integration, LU-SGS method [6] is used, which is coupled with Newtonian iterative method. To resolve compressible phenomena, convective term is calculated by using a third-order MUSCL-type TVD scheme. For electric potential, we applied TDMA algorithm enforced with the block correction method. Large eddy simulation (Smagorinsky subgrid-scale model) is applied to capture possible turbulent behavior. The task has been solved on an oblique structured grid with nonequidistant spacing. The total number of grid points was 38 553, with 543 and 71 points in the axial and radial directions respectively.

3 Results

Calculations have been carried out for the currents 300, 400, 500 and 600 A. Mass flow rate for water-stabilized section of the discharge was taken for each current between 300 and 600 A from our previously published work [7], where it was determined iteratively as a minimum difference between the numerical and experimental values of the outlet quantities. The resulting values are 0.228 $g \cdot s^{-1}$(300 A), 0.315 $g \cdot s^{-1}$(400 A), 0.329 $g \cdot s^{-1}$(500 A), 0.363 $g \cdot s^{-1}$(600 A). Argon mass flow rate was varied in agreement with experiments in the interval from 22.5 slm (standard liters per minute) to 40 slm, namely 22.5, 27.5, 32.5 and 40 slm. It was proved in experiments [8] that part of argon is taken away before

it reaches the torch exit because argon is mixed with vapor steam and removed to the water system of the torch. The amount of argon transferred in such a way from the discharge is at least 50 % for currents studied here. Since the present model does not treat argon and water as separate gases and the mechanism of argon removal is not included in the model, we consider in the calculations that argon mass flow rate present in the discharge equals one-half of argon mass flow rate at the torch inlet.

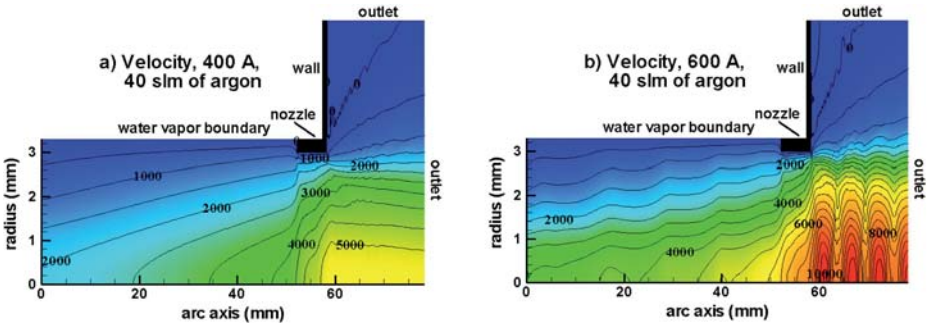


Fig. 1. Velocity contours ($m \cdot s^{-1}$) for (a) 400 A and (b) 600 A discharges, partial characteristics model. Water mass flow rates are 0.315 $g \cdot s^{-1}$(400 A) and 0.363 $g \cdot s^{-1}$(600 A); argon mass flow rate is 0.554 $g \cdot s^{-1}$(50 % of 40 slm) for both currents. The transonic flow field for 400 A converts to a supersonic flow structure with clearly distinguished shock diamonds at the outlet for 600 A. Contour increment is 500 $m \cdot s^{-1}$.

Fig. 1(a) and (b) shows velocity fields in the discharge region and the near outlet for 400 and 600 A for the partial characteristics radiation model with the argon mass flow rate of 40 slm. Argon flows axially into the domain from the left side, whereas water evaporates in the radial direction from the "water vapor boundary". A relatively high value of argon mass flow rate, used also in experiment, was chosen here to demonstrate compressible phenomena. The result for 400 A shows a transonic case with a Mach number of around 1.1 at the axial outlet region, a maximum temperature of 17 200 K and a velocity of 5 400 $m \cdot s^{-1}$. A qualitatively different picture is obvious for 600 A with the formation of shock diamonds in the downstream of the torch exit. The velocity maxima overlap with the temperature minima and vice versa. Since the pressure decreases at the torch exit, the computed contours of axial velocity and temperature correspond to an under-expanded atmospheric-pressure plasma jet. The maximum velocity 10 100 $m \cdot s^{-1}$occurs near the axial position of 61.1 mm with the Mach number 1.6, further downstream the velocity amplitudes decrease. The maximum and minimum temperatures in the shock diamonds range between 21 500–23 500 K. The net emission model provides qualitatively the same results but with slightly higher temperatures, slightly lower velocities and nearly the same Mach numbers.

Fig. 2 presents the radial profiles of the Mach number 2 mm downstream of the nozzle exit with the argon mass flow rate of 40 slm. It is clearly demonstrated that the partial characteristics model gives higher values of the Mach number and the difference regarding the net emission coefficients model is below 0.1 at the arc axis. For currents higher than 400 A, a supersonic rare plasma in the central parts of the discharge is surrounded by a subsonic, much denser but still hot plasma.

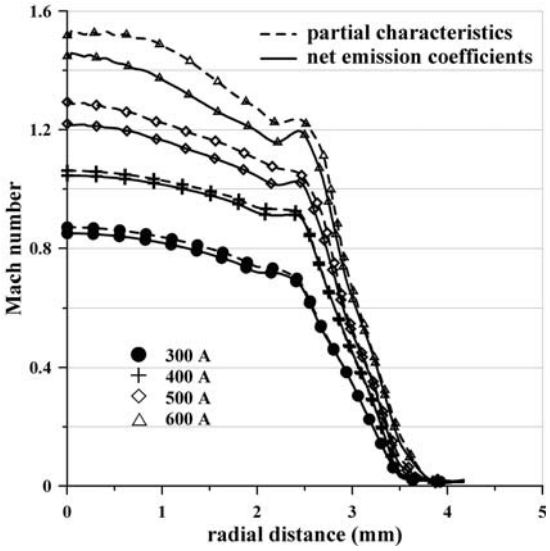


Fig. 2. Radial profiles of the Mach number 2 mm downstream of the nozzle exit with argon mass flow rate of 40 slm. The partial characteristics model gives higher values of the Mach number; the difference is less than 0.1 at the arc axis.

A number of experiments have been carried out on the hybrid stabilized electric arc in recent past for the currents 300–600 A with 22–40 slm of argon. Temperature is one of the fundamental plasma parameters, needed also for evaluation of the other quantities.

In experiment, the radial profiles of temperature at the nozzle exit were calculated from optical emission spectroscopy measurements. The procedure is based on the ratio of emission coefficients of hydrogen H_β line and four argon ionic lines using calculated LTE composition of the plasma for various argon mole fractions as a function of temperature [3]. From the calculated molar fractions of hydrogen and argon it is easy to obtain emission coefficients of H_β and argon lines. The

temperature corresponding to an experimental ratio of emission coefficients is then found by cubic spline interpolation on the theoretical data.

Fig. 3 compares measured and calculated temperature profiles 2 mm downstream of the nozzle exit for two cases—500 A, 32.5 slm of argon; 600 A, 40 slm of argon. Excellent agreement is demonstrated for 600 A where the measured profile nearly coincides with the two profiles calculated using the net emission and partial characteristics radiation methods. In the case of 500 A, agreement is better for the profile calculated by the net emission model (black color). Calculation for broad range of currents and argon mass flow rates proved that the maximum relative difference between calculated and experimental temperature profiles is lower than 10 % for the partial characteristics and 5 % for the net emission radiation model used in the present calculation.

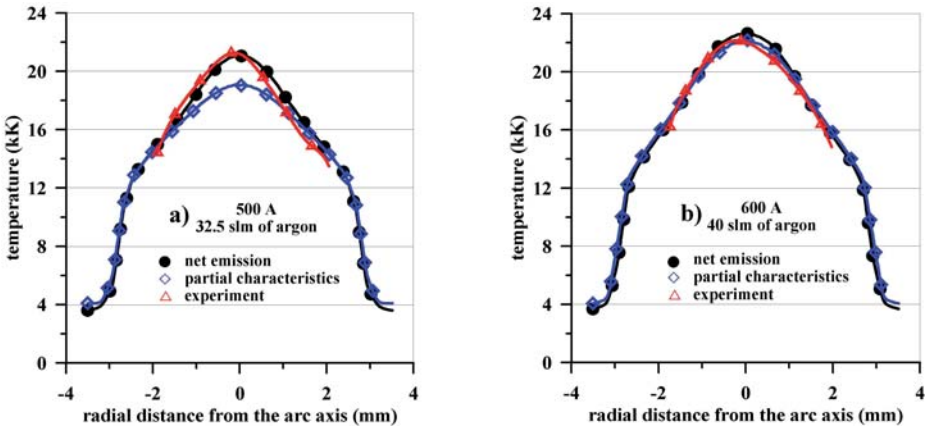


Fig. 3. Experimental and calculated temperature profiles 2 mm downstream of the nozzle exit for a) 500 A, 32.5 slm of argon; b) 600 A, 40 slm of argon.

Besides temperature profiles, velocity profiles at the nozzle exit and mass and momentum fluxes through the torch nozzle are important to characterize the plasma torch performance. Velocity at the nozzle exit was determined in the experiment from the temperature profile and power balance assuming LTE [9]. First, the Mach number M is obtained from the simplified energy equation integrated through the discharge volume; second, the velocity profile is derived from the measured temperature profile using the definition of the Mach number

$$u(r) = M \cdot c\{T(r)\},$$

where $c\{T(r)\}$ is the sonic velocity for the experimental temperature profile. The drawback of this method is the assumption of the constant Mach number over the nozzle radius. The existence of supersonic regime (i.e., the mean value of the Mach number over the nozzle exit is higher than 1) using this method

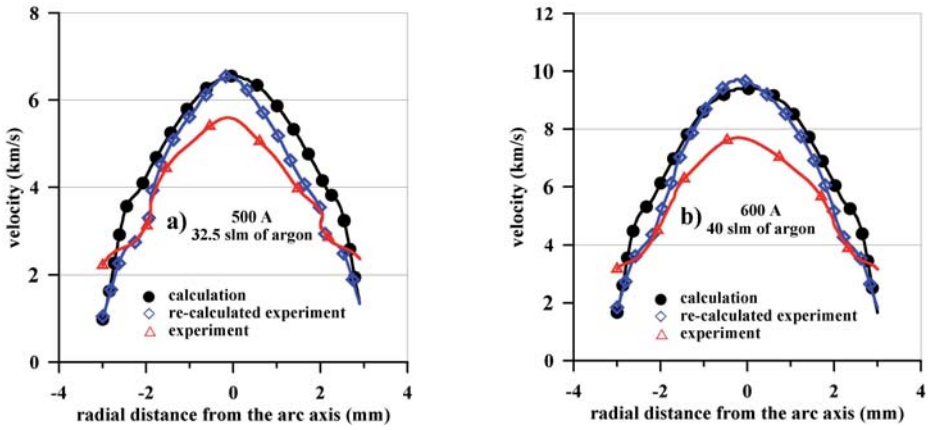


Fig. 4. Velocity profiles 2 mm downstream of the nozzle exit for a) 500 A, 32.5 slm of argon; b) 600 A, 40 slm of argon. Calculation—partial characteristics model, re-calculated experimental profile is based on the calculated Mach number and the experimental temperature profile.

was proved for 500 A and 40 slm of argon, as well as for 600 A for argon mass flow rates higher than 27.5 slm. Similar results have been also reported in our previous work [10].

Comparison between calculated and experimental radial velocity profiles 2 mm downstream of the nozzle exit is shown in Fig. 4, for the same operating conditions as in Fig. 3. The blue curves represent "re-calculated" velocity profiles, i.e. velocity profiles obtained as a product of the Mach number profiles obtained from the present numerical simulation and the sonic velocity based on the experimental temperature profile (the so-called "integrated approach"). It is clearly visible that agreement of such re-evaluated experimental velocity profiles with the numerical ones is much better than between original experiments and calculation. More detailed analysis of the re-evaluated experimental velocity profiles for the all operating conditions studied in this research proved that the relative difference between the calculated and re-evaluated profiles is less than 26 %, while for the original experimental profiles the difference reaches up 50 %.

4 Conclusions

The numerical results carried out for 300–600 A and for argon mass flow rates of 22.5–40 slm proved that the plasma flow in the hybrid torch is supersonic for 500 A and 600 A if argon mass flow rate exceeds 27.5 slm. The supersonic structure with shock diamonds occurs in the central parts of the discharge at the outlet region. It was also evidenced that laminar and turbulent regimes give nearly the same results, so that the plasma flow can be considered to be laminar for the operating conditions and simplified discharge geometry studied in this

research. The partial characteristics radiation model gives slightly lower temperatures but higher outlet velocities compared to the net emission model. The reabsorption of radiation in the discharge ranges between 31–45 %, it decreases with current and it also slightly decreases with argon mass flow rate. Calculated and experimental radial temperature and velocity profiles at the nozzle outlet exhibit very good agreement.

Acknowledgments

J. Jeništa is grateful for financial support under the Fluid Science International COE Program from the Institute of Fluid Science, Tohoku University, Sendai, Japan, and their computer facilities. Financial support from the projects GA CR 202/08/1084, GA AV B100430701, and MSM 0021630503 from the Czech Ministry of Education is gratefully acknowledged.

Our appreciation goes also to the following Institutions for granting their computational resources:

- Institute of Plasma Physics AS CR, v.v.i.,
- Luna project of the Institute of Physics ASCR, v.v.i.,
- MetaCentrum project.

The computer code is being developed by the author and is written in the FORTRAN77 language.

References

[1] Březina V, Hrabovský M, Konrád M, Kopecký V and Sember V 2001 *New plasma spraying torch with combined gas–liquid stabilization of arc*, ed A Bouchoule et al (Proc. ISPC 15) (Orleans, France) pp 1021–26.
[2] Van Oost G, Hrabovský M, Kopecký V, Konrád M, Hlína M, Kavka T, Chumak O, Beckman E and Verstraeten J 2006 *Vacuum* **80** 1132–37.
[3] Křenek P 2008 *Plasma Chem. Plasma Process.* **28** 107-22.
[4] Aubrecht V and Bartlová M 2004 *Czech. J. Phys.* **54** C759-C765.
[5] Bartlova M and Aubrecht V 2006 *Czech. J. Phys.* **56** B632-B637.
[6] Jameson A and Yoon S 1987 *AIAA Journal* **25** 929-35.
[7] Jeništa J 2003 *J. High Temp. Mat. Processes* **7** 11-16.
[8] Kavka T, Chumak O, Sember V and Hrabovský M 2007 *Processes in Gerdien arc generated by hybrid gas-water torch* ed J Schmidt et al (Proc. ICPIG 2007) (Prague, Czech Republic) pp 1819-22.
[9] Kavka T, Maslani A, Chumak O and Hrabovsky M 2008 *Character of plasma flow at the exit of DC arc gas-water torch* ed S Maruyama (Proc. ICDF 2008) (Sendai, Japan) OS8-11.
[10] Jeništa J, Takana H, Hrabovský M and Nishiyama H 2008 *IEEE Trans. Plasma Sci.* **36** 1060-61.

Publications Related to the MetaCentrum Computer Facilities in 2009

1) J. Jeništa, H. Takana, H. Nishiyama, M. Bartlova, V. Aubrecht, M. Hrabovsky, *Parametric study of hybrid argon-water stabilized arc under subsonic and supersonic regimes*, Journal of High Temperature Material Processes, vol. 14(issue 1), pp. 55-70, 2010.
2) J. Jeništa, H. Takana, H. Nishiyama, M. Bartlová, V. Aubrecht, P. Křenek, M. Hrabovský, T. Kavka, V. Sember, A. Mašláni, *Integrated parametric study of hybrid-stabilized argon-water arc under, subsonic and supersonic regimes*, ISPC-proceedings on CD-ROM, editors: A. von Keudell, J. Winter, M. Böke, V. Schulz-von der Gathen, 19th International Symposium on Plasma Chemistry, Bochum, Germany, 26.-31. 7. 2009.
3) J. Jenista, H. Takana, H. Nishiyama. M. Hrabovsky, *Modelling of Supersonic and Turbulent Hybrid Arc for Biomass Gasification*, Proceedings of the ninth international symposium on advanced fluid information and transdisciplinary fluid integration, Sendai, Japan, (2009), pp. 40-41. (ISSN 1344-2236, IFS-TIM019).

Raditive Transfer Simulations Using DART Model

Věroslav Kaplan, Zbyněk Malenovský, Jan Hanuš, and Petr Lukeš

Institute of Systems Biology and Ecology AS CR, v.v.i.,
Poříčí 3b, Brno, 60300
http://www.usbe.cas.cz/
Remote sensing of vegetation processes workgroup
kaplan@usbe.cas.cz, zbynek.malenovsky@gmail.com, jan.hanus@usbe.cas.cz,
petr.lukes@usbe.cas.cz

The main focus of our research activities is the physically based retrieval of biophysical and biochemical parameters of vegetation from hyperspectral remote sensing image data. The quantitative parameters of interest, representing also bio-indicators of the actual vegetation state, are currently mainly leaf area index (LAI) and total concentration of green photosynthetically active pigments (chlorophyll a+b). We are capable to estimate the indicated bio-parameters from remote sensing data of high spatial resolutions by means of radiative transfer modelling at the leaf (PROSPECT model) and crown/canopy (DART model) level. The pilot ecosystems are montane forest stands of Norway spruce (*Picea abies (L.) Karst.*).

Discrete Anistrophic Radiative Transfer model (DART[1]) was developed in CESBIO laboratory (Center for the Study of the BIOsphère) for simulation of radiative transfer in the "Earth-Atmosphere" system. The input of the model is a 3D model of landscape including optical, geometric and biophysical parameters of model elements—i.e., optical properties of the ground, positions and dimensions of trees on the landscape, distribution of the biomass in the trees, type of leaves in trees, and many other properties. The DART model results are simulated images which would be (theoretically) acquired by airborne/satellite sensors flight over the simulated scene.

Although the DART model is able to simulate in any wavelength in the optical domain (visible, thermal and infra-red), we currently perform our simulations only in visible and near-infrared domains, which corresponds to airborne image sensor AISA/EAGLE also operated by Remote sensing group.

In addition of forward run (radiative transfer) of the model, it is possible to run model in reverse mode, which makes possible retrieval of several characteristics from the acquired image, e.g., chlorophyll content in the tree leaves.

Software Used in MetaCentrum

- Discrete Anisotropic Radiative Transfer model (DART)—licensed for use in METACentrum by CESBIO

[1] http://www.cesbio.ups-tlse.fr/us/dart.htm

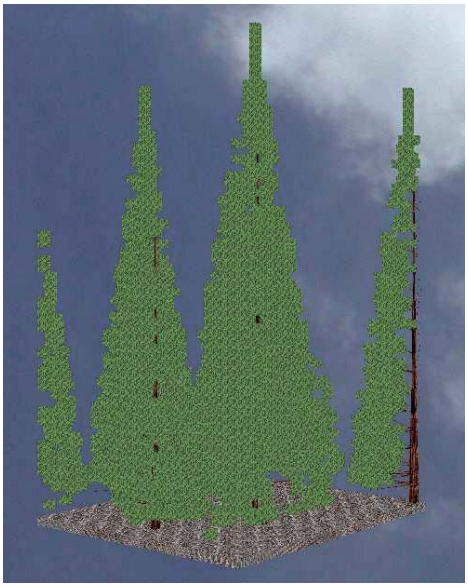


Fig. 1. Input data: DART 3D scene example

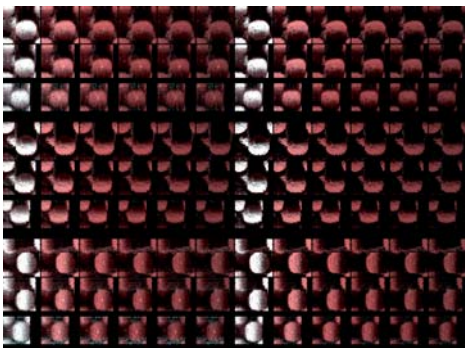


Fig. 2. Output data: Results of simulations—one tile represents one full simulation. Simulations vary in optical and structural properties, all simulations are merged together for further processing.

- custom scripts for managements of mock-ups and simulations—bash, Python, Numeric Python, Python MPI
- custom scripts for processing and evaluation of DART simulation results

We simulated mainly coniferous forest stands scenes during the year 2009. Simulated scenes were used for evaluation and sensitivity analysis of reflectance compared to miscellaneous properties of the forest stand (influence of LAI, TAI and understory to the reflectance of crown canopy). We also simulated forest stand scenes in configuration compatible with AISA/Eagle airborne scanner and

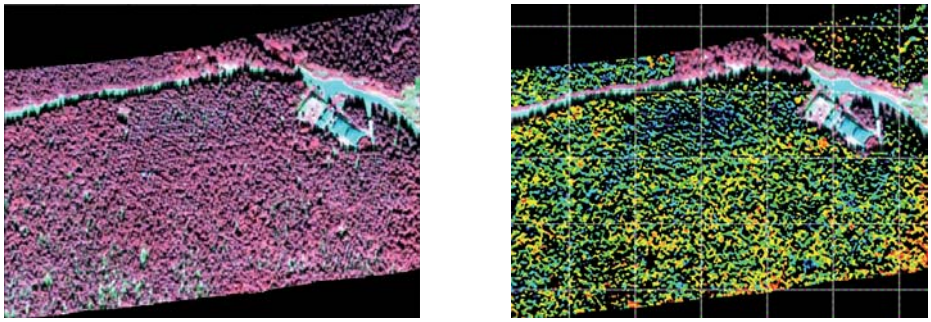


Fig. 3. Example of application. Left: AISA/EAGLE acquired image of test site (Bílý Kříž, Beskydy Mts.). Right: retrieval of chlorophyll content in the same area.

CHRIS/PROBA satellite scanner during the ESA PECS grant. Simulated scenes were used later to train artificial neural networks in order to perform retrieval on images acquired by sensors afterwards.

Grants

Sentinel 2: SPECTRA—Spectral-spatial scaling from leaf to forest level by using spectral and directional approach as a support for exploitation of hyperspectral scanner on-board Sentinel 2. ESA. http://www.czechspace.cz/earth-observations/sentinel-2-spectra. Defended in March 2009.

Presented Results

[1] Lukeš, P.; Malenovský, Z.; Hanuš, J.; Kaplan, V.; Homolová, L.; Pokorný, R.: Challenges In Accuracy Assessment of Norway Spruce Leaf Chlorophyll Content Estimated from Airborne and Satellite Imaging Spectroscopy Data, *Workshop on the Retrieval of Geophysical Variables Using High Spatial Resolution Optical Imagery*, ESA/ESTEC workshop, October 14th-16th October 2009
[2] Kaplan, V.; Homolová, L.; Hanuš, J.; Lukeš, P.; Malenovský, Z.: *Effective Parametrization of the DART Model for Retrieval of Vegetation Properties of Coniferous Forest*. Workshop on the Retrieval of Geophysical Variables Using High Spatial Resolution Optical Imagery, ESA/ESTEC workshop, October 14th-16th October 2009

Experiments with Job Scheduling in MetaCentrum

Dalibor Klusáček, Hana Rudová, and Miroslava Plachá

Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno
Czech Republic
{xklusac,hanka}@fi.muni.cz

1 Introduction

Large computing clusters and Grids have become common and widely used platforms for the scientific and the commercial community. Efficient job scheduling in these large, dynamic and heterogeneous systems is often a very difficult task [14]. Therefore, a lot of testing and evaluation is needed before some scheduling algorithm is applied in the production system such as PBSpro [6] or LSF [15]. Due to several reasons, such as the cost of resources, reliability, varying background load or the dynamic behavior of the components, experimental evaluation is not usually performed in the real systems. Many simulations with various setups that simulate different real-life scenarios must be performed using the same and controllable conditions to obtain reliable results. This is hardly achievable in the production environment.

2 Available Data Sets

When performing simulations and testing, usually the workload traces from the Parallel Workloads Archive (PWA) [3] or Grid Workloads Archive (GWA) [1] are used to represent users' jobs. However, these data do not contain several parameters that are important for realistic simulations. Typically, very limited information is available about the Grid or cluster resources. The number of machines in particular clusters, their architecture, the CPU speed, the RAM size or the resource specific policies are not usually known. However, these parameters often significantly influence the decisions and performance of the scheduler [10]. Moreover, no information concerning background load, resource failures, or specific users' requirements are available. In heterogeneous environments, users often specify some subset of machines or clusters that can process their jobs. This subset is usually defined either by the resource owners' policy (user is allowed to use such cluster), or by the user who requests some properties (cluster location, library, software license, execution time limit, etc.) offered by some clusters or machines only. Also, the combination of both owners' and users' restrictions is possible. When one tries to create a new scheduling algorithm all such information and constraints are crucial, since they make the algorithm design much more

complex. If omitted, resulting simulation may provide misleading or unrealistic results as we show in Section 4.

3 MetaCentrum Data Set

Since the current archives miss to provide truly complete data sets, we were very happy that we were kindly allowed by the MetaCentrum team to create the data set covering many previously mentioned issues. Namely, this data set contains trace of 103,620 jobs executed during the first five months of 2009 as well as detailed description of computational nodes. Job description is very complex, including e.g., the maximum runtime limits for jobs or their specific requirements concerning target platform (CPU architecture, location, network interface, etc.). Also the description of clusters contains detailed information involving CPU speed, RAM size, CPU architecture, operating system and the list of supported properties (allowed queue(s), cluster location, network interface, etc.). Together, these detailed information about jobs and machines allow to use so called *specific job requirements* representing the "job-to-machine" suitability. Moreover, information about machines that were not available has been collected, covering the time periods when machines were either in maintenance (failure/restart) or dedicated for specific purposes. Finally, the list of queues including their time limits and priorities is provided. The MetaCentrum data set is publicly available at `http://www.fi.muni.cz/~xklusac/workload`. Certainly all information in the data set containing private information such as user, machine, queue or job names or names of specific parameters were anonymized.

4 Evaluation

Once the complex data set from the MetaCentrum was available, we could answer the question whether the additional information and constraints such as machine failures or specific job requirements influence the quality of the solution generated by the scheduling algorithms. For this purpose, two different problems have been considered and then simulated using the Alea job scheduling simulator [8]. The first BASIC problem involved the use of MetaCentrum data set where both machine failures and specific job requirements were ignored. This setup is quite similar to the typical amount of information available in the GWA or PWA archives that do not provide information about machine failures or specific job requirements. The second EXTENDED problem used all information available in the MetaCentrum data set, therefore both machine failures and specific job requirements have been used during the simulation. Five different scheduling algorithms have been used in this evaluation. The algorithms FCFS, EASY backfilling (EASY) [12], and conservative backfilling (CONS) [4, 13] represent standard queue-based algorithms. The other two algorithms were developed as a part of our work. The CP algorithm is based on the ideas of constraint programming [11] and the local search (LS) [9] is an optimization procedure using

conservative backfilling for construction of the initial solution. The average slowdown [5] and average wait time [2] have been used as the evaluation criteria here (additional results are available in [10]).

The Figure 1 shows the results for all algorithms applied in this study. Clearly,

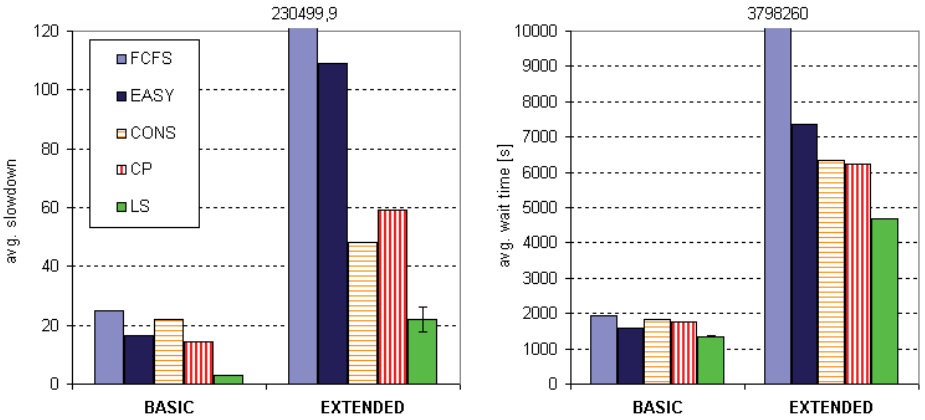


Fig. 1. The average slowdown (left) and average wait time (right) for BASIC and EXTENDED problem.

when the BASIC problem is applied, the differences between algorithms are not very large, while the differences start to grow as soon as the EXTENDED problem is used. The solution produced by a given algorithm for the EXTENDED problem, is always worse than for the BASIC problem. Moreover, the relative differences between algorithms are much higher which can be seen especially in the extreme case of FCFS, which totally failed to generate acceptable results. On the other hand, LS optimization of CONS is very successful, significantly decreasing both slowdown and response time. Clearly, additional features such as machine failures or specific job requirements add nontrivial constraints into the decision making process of selected algorithms. Experimental results showed that these constraints should not be ignored otherwise the simulation results are very unrealistic.

5 Conclusion

The use of complete and "rich" data set may significantly influence the quality of generated solution as we have published in [10, 7]. In addition, we have shown that similar observations can be made also for other publicly available data sets [10]. If possible, complete data sets should be collected and used to evaluate scheduling algorithms under harder conditions. Their use may narrow the gap between the "ideal world" of simulations and the real-life experience,

producing more reliable and realistic experiments. Realistic simulations help to quickly identify possible weaknesses in the algorithm design, allowing to make them more robust and scalable. From this point of view, complex data set from MetaCentrum represents an important source of valuable information for the scientific community.

Acknowledgment

We appreciate the gracious support of the Ministry of Education, Youth and Sports of the Czech Republic under the research intent No. 0021622419. We also highly appreciate the help of MetaCentrum team who provided us the workload data and help us to interpret them correctly, as well as the access to the MetaCentrum computing facilities provided under the research intent MSM6383917201.

References

1. Dick Epema, Shanny Anoep, Catalin Dumitrescu, Alexandru Iosup, Mathieu Jan, Hui Li, and Lex Wolters. Grid workloads archive (GWA). Available at: `http://gwa.ewi.tudelft.nl/pmwiki/`.
2. Carsten Ernemann, Volker Hamscher, and Ramin Yahyapour. Benefits of global Grid computing for job scheduling. In *GRID '04: Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing*, pages 374–379. IEEE, 2004.
3. Dror G. Feitelson. Parallel workloads archive (PWA). Available at: `http://www.cs.huji.ac.il/labs/parallel/workload/`.
4. Dror G. Feitelson. Experimental analysis of the root causes of performance evaluation results: A backfilling case study. *IEEE Transactions on Parallel and Distributed Systems*, 16(2):175–182, 2005.
5. Dror G. Feitelson, Larry Rudolph, Uwe Schwiegelshohn, Kenneth C. Sevcik, and Parkson Wong. Theory and practice in parallel job scheduling. In Dror G. Feitelson and Larry Rudolph, editors, *Job Scheduling Strategies for Parallel Processing*, volume 1291 of *LNCS*, pages 1–34. Springer Verlag, 1997.
6. James Patton Jones. *PBS Professional 7, administrator guide*. Altair, April 2005.
7. Dalibor Klusáček and Hana Rudová. Complex real-life data sets in Grid simulations. In *Cracow Grid Workshop 2009 Abstracts (CGW'09)*, 2009.
8. Dalibor Klusáček and Hana Rudová. Alea 2 – job scheduling simulator. In *Proceedings of the 3rd International Conference on Simulation Tools and Techniques (SIMUTools 2010)*, Torremolinos, Malaga, Spain, 2010.
9. Dalibor Klusáček and Hana Rudová. Efficient grid scheduling through the incremental schedule-based approach. *Computational Intelligence: An International Journal*, 2010. To appear.
10. Dalibor Klusáček and Hana Rudová. The importance of complete data sets for job scheduling simulations. In *Proceedings of the 15th Workshop on Job Scheduling Strategies for Parallel Processing*, Atlanta, USA, 2010.
11. Miroslava Plachá. Dynamické rozvrhování úloh na výpočetní zdroje, 2010. Submitted as a Master Thesis at Faculty of Informatics, Masaryk University, Brno, Czech Republic.

12. Joseph Skovira, Waiman Chan, Honbo Zhou, and David Lifka. The EASY - LoadLeveler API project. In Dror G. Feitelson and Larry Rudolph, editors, *Job Scheduling Strategies for Parallel Processing*, volume 1162 of *LNCS*, pages 41–47. Springer, 1996.
13. Srividya Srinivasan, Rajkumar Kettimuthu, Vijay Subramani, and P. Sadayappan. Selective reservation strategies for backfill job scheduling. In Dror G. Feitelson, Larry Rudolph, and Uwe Schwiegelshohn, editors, *Job Scheduling Strategies for Parallel Processing*, volume 2537 of *LNCS*, pages 55–71. Springer Verlag, 2002.
14. Fatos Xhafa and Ajith Abraham. Computational models and heuristic methods for grid scheduling problems. *Future Generation Computer Systems*, 26(4):608–621, 2010.
15. Ming Q. Xu. Effective metacomputing using LSF multicluster. In *CCGRID '01: Proceedings of the 1st International Symposium on Cluster Computing and the Grid*, pages 100–105. IEEE, 2001.

Utilization of MetaCentrum Computational Resources by National Centre for Biomolecular Research

Petr Kulhánek, Jiří Wiesner, Judit Šponer, Josef Pasulka, Jan Alán, Zora Střelcová, Radek Matuška, Tomasz Pawlak, Zdeněk Kříž, Stanislav Kozmon, Lucie Novosadová, Jakub Štěpán, Jiří Fukal, Crina-Maria Ionescu, Richard Štefl, Arnošt Mládek, Pavel Kadeřávek, Jan Vícha, Jan Slavík, Stanislav Standara, Jana Přecechtělová, Radka Svobodová Vařeková, Petr Novák, Alexej Kulaš, Monika Pěntáková, Stanislav Geidl, Sushil Kumar Mishra, Josef Chmelík, Leona Šerá, Michal Ďurech, Radka Kolínková, Zuzana Novotná Jiroušková, Jan Novotný, Barbora Benešová, Martin Babinský, Jan Adam, Martin Prokop, Radek Marek, and Jaroslav Koča

National Centre for Biomolecular Research, Faculty of Science,
Masaryk University, Kotlarska 2,
611 37 Brno, Czech Republic
{kulhanek, jkoca}@chemi.muni.cz
http://ncbr.chemi.muni.cz

1 Introduction

National Centre for Biomolecular Research (NCBR) is an independent institute of the Faculty of Science, Masaryk University, Brno. The main activities of NCBR involve research in specific areas of chemistry and biology as well as teaching, especially at advanced and Ph.D. levels. The research activities of NCBR include projects in the following areas: Computational chemistry and molecular modeling, Computational studies on nucleic acids, NMR spectroscopy, Glycobiochemistry, RNA/Protein interactions, RNA processing and degradation, DNA reparation, and Nanobiotechnology. The NCBR research strategy combines and balances both theoretical and experimental approaches.

While the experimental research requires good experimental facilities, the theoretical research requires strong computational resources to obtain results precise enough to complement experimental observations. Such resources are available in MetaCentrum supercomputing centre and this article describes how they are utilized by the computational chemists from NCBR in order to predict the behavior and properties of a wide range of molecular systems. Achievements in our own software development will be outlined. Briefly, they either accelerate calculations or make them more comfortable in the supercomputing environment. Last but not least, the hardware provided by NCBR to common MetaCentrum computational infrastructure will be overviewed shortly.

2 Computational Chemistry

The research effort in computational chemistry branch focuses on several areas. First research field represents applications of quantum chemistry to small molecules. These studies include searching for chemical reaction pathways and transition states in particular, properties of supramolecules, and calculations of NMR parameters. Second area comprises studies of conformational behavior and dynamics of larger molecules of biological importance. The most widely used tools in this field are molecular dynamics and docking. Results obtained by theoretical models are validated against available experimental data, especially from NMR and X-ray. Tuned models and methods are then able to provide information that is not available via conventional experimental tools and thus very helpful for experimentalists. Selected projects are shortly overviewed in the following subsections.

2.1 NMR Properties Calculations

Nucleic acid structure determination by NMR typically suffers from the lack of restraints for the phosphodiester backbone. Especially, due to the inaccessibility of $^3J_{OPOC}$ spin-spin couplings via experiment, no Karplus equations are available for backbone torsion angles α and ξ. The relationships between ^{31}P chemical shifts, $\delta_{iso}(^{31}\mathrm{P})$, and the two P–O torsion angles are therefore of particular interest as they could be used to derive restraints unattainable from other sources. In our study, we calculate phosphorus chemical shifts in B-DNA by combining Molecular Dynamics (MD) and Density Functional Theory (DFT). Snapshots extracted from the MD simulation of $[\mathrm{d(CGCGAATTCGCG)}]_2$ are used to construct cluster models of backbone phosphate groups consisting of dimethyl phosphate and water molecules of the first solvation shell. The cluster models are then employed for DFT chemical shift calculations. Via analyzing the results obtained, we 1) reveal the trends in ^{31}P chemical shifts as influenced by α and ξ, 2) examine the effects of the continual breaking and reforming of hydrogen bonds on $\delta_{iso}(^{31}\mathrm{P})$, and 3) determine the difference in ^{31}P chemical shifts between BI and BII conformational states of B-DNA.

Another project is aimed to find the optimal approach for calculating the NMR shielding constants in transition metal complexes, for example, platinum complexes with nitrogen basis known as cytostatics, important drugs for cancer treatment. Various pure and hybrid DFT functionals are tested and the results are compared to the experimental NMR data. This project involves geometry optimizations and subsequent calculations of the nuclear shielding at nonrelativistic and scalar relativistic levels of theory due to presence of heavy transition-metals. The selected methodology will be applied to a range of structural problems in chemistry of complexes and organometallics, e.g., determination of the coordination geometry, solvent effects.

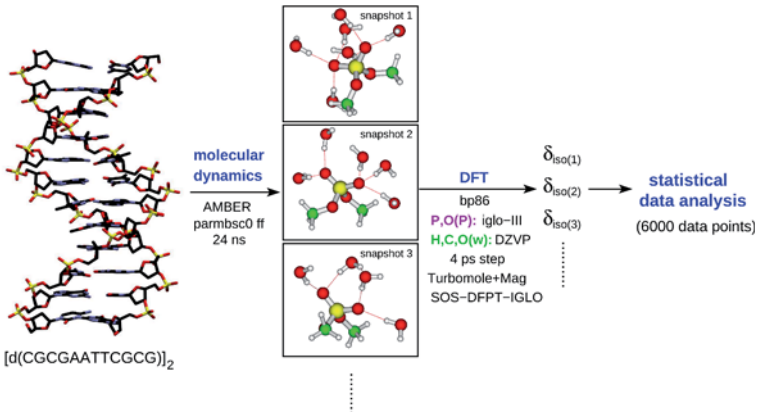


Fig. 1. Procedure describing the calculations of NMR chemical shifts in a DNA backbone

2.2 Interaction Energy Calculations

Recent experimental studies on the Watson-Crick type base pairing of triazine and aminopyrimidine derivatives suggest that acid/base properties of the constituent bases might be related to the duplex stabilities measured in solution. High level quantum chemical calculations and molecular dynamics simulations were used to evaluate the base pairing and stacking interactions of seven selected base pairs, which were common in that they were stabilized by two N-HO hydrogen bonds separated by one N-HN hydrogen bond. It was shown that neither the base pairing nor the base stacking interaction energies correlated with the reported pK_a data of the bases and the melting temperatures of the duplexes. This suggested that the experimentally observed correlation between the melting temperature data of the duplexes and the $\mathsf{pK}_a s$ of the constituent bases is not rooted in the intrinsic base pairing and stacking properties. In addition, since our calculations were carried out with extrapolation to the complete basis set (CBS) of atomic orbitals and with inclusion of higher electron correlation effects, they provide reference data for stacking and base pairing energies of non-natural bases.

World wide distributed bacterium Ralstonia solanacearum causes lethal wilt in many agricultural crops. It was shown that it produces a highly selective and strong fucose-binding lectin called RSL. It seems that dispersive interactions play a key role in molecular recognition of carbohydrates by this lectin. Thus several models of binding site were studied by high-level *ab initio* and DFT methods. Based on the comparison of the optimized structures and the X-ray structures, the best method for the further optimizations has been chosen. An extrapolation to CBS has been computed to compare the DFT-D (DFT with dispersion correction) and MP2 interaction energy behavior. The coupled cluster corrected MP2 interaction energy in the CBS has been used as the reference value. Basis Set Superposition Error (BSSE) correction has been used in the

computation of all interaction energies. On the basis of these benchmarks, a method with the best accuracy/CPU performance ratio has been chosen for the further computations. The benchmark results show that DFT-D method with the BP functional and Grimme's empirical dispersion correction is the best to match these criteria.

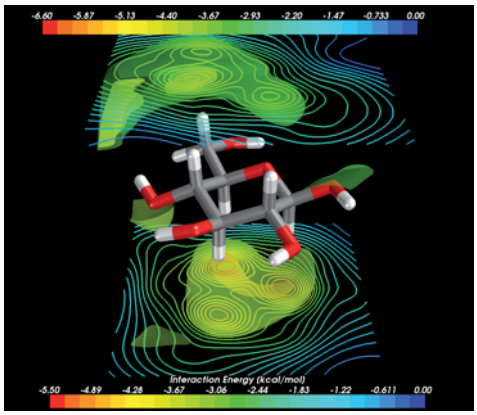


Fig. 2. Visualization of the regions with high values of the dispersion interaction energy in the benzene/glucose model complexes

2.3 Partial Atomic Charge Calculations

The subject of this research topic is the calculation of partial charges on atoms in molecules via Electronegativity Equalization Method (EEM), which is based on DFT. The main advantage of EEM is its accuracy, which is comparable to that of quantum mechanical methods, with markedly shorter execution times. Since EEM is an empirical approach that needs parameterizations, our current effort is put into development of a new and broader parameter set, which will cover larger biomolecular systems including metalloproteins. The *ab initio* charges used for training were calculated for about 30 000 molecules. These molecules were small organic molecules obtained from the Cambridge Structural Database. Another training set containing about 100 proteins was extracted from the PDB database. To decrease computational expenses, proteins from the protein set were modified with the TRITON program and unnecessary side chains were removed. The resulting protein-like structures contained the maximum of 300 atoms per molecule.

2.4 Dynamical Behaviour of Biomolecules

Molecular dynamics simulations on Pseudomonas aeruginosa lectin-II and its complexes with four different monosaccharides were performed. The saccharide-

free, saccharide-occupied, and saccharide- and ion-free forms of the lectin were compared. The water density pattern around the binding site in the free lectin molecular dynamics was fitted with that in the X-ray structure and with the hydroxyl groups of the monosaccharide within the lectin/monosaccharide complexes. The best ligand was predicted based on the best fit. Interestingly, the water density pattern around the binding site in the uncomplexed lectin exactly fitted the O2, O3, and O4 hydroxyl groups of the lectin-fucose complex. This observation could lead to a hypothesis that the replacement of these three water molecules from the binding site by the monosaccharide decreases entropy of the complex and increases entropy of the water molecules in favor of the binding. It suggests that the high density solvent peaks around the binding site in the free protein could be a tool to predict hydroxyl group orientation of the sugar in the protein/sugar complexes.

Another of our projects deals with extensive molecular dynamics studies of a RNA molecule derived from Transactivation Response Element of HIV-1. These molecules serve often as a binding site and also play role in HIV replication. The wild type molecule has two binding sites, UCU bulge and apical loop of CUGGGA base composition. The apical loop was replaced by UUCG tetraloop in our case. Four calcium ions were found by X-ray crystallography positioned in the proximity of the UCU bulge. The effect of replacement of the calcium ions by magnesium ions is studied as well as the case when they are omitted completely. Additionally the effect of different ionic strength and the type of counter-ions used to neutralize backbone charge will be studied. The molecular dynamics for all mentioned cases are performed using two different force fields to compare them. Especially, we focus on two objectives. First, we are interested how the mentioned differences in the conditions and used force fled influence the structure of non-canonical parts of the molecule. Second, we want to investigate the changes of molecular motion. The motional parameters will be extracted from the trajectory using Pine program. The obtained dynamical information will be compared to the published experimental data based on NMR experiments.

2.5 Free Energy Calculations

One of the challenging applications of the free energy calculations is a pK_a computation for ionizable residues. One of the suitable methods for this task is a thermodynamic integration. This method was used in benchmark calculations of pK_a shifts in a small protein, thioredoxin. The preliminary results indicate that the choice of the charge model is less important and that the proper sampling of the ionized species is the most vital factor. We will transfer this knowledge to acetylcholinesterase and apply it to compute the pK_a shifts of ionizable residues in the active site. The knowledge of pK_a shifts will have an impact on the design of efficient reactivators.

Thermodynamic integration was also used for assessment of binding of various carbohydrate ligands into PA-IIL and CV-IIL lectins and their mutants. The obtained results were in a very good agreement with experimental values. For crude estimations of affinities towards CV-IIL lectin, MM/PBSA and

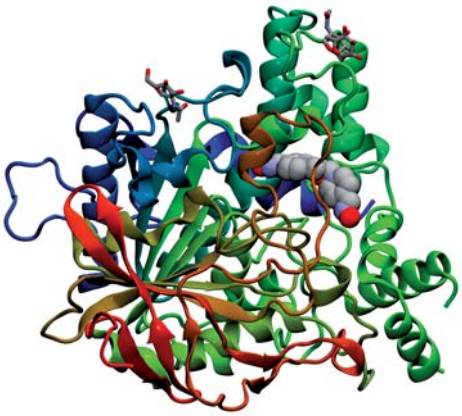


Fig. 3. Structure of acetylcholinesterase with bound reactivator

MM/GBSA methods were used to compute absolute binding free energy from 20 ns long molecular dynamics trajectories. Entropic contributions to the binding free energies were computed by a normal mode analysis and by analysis of a covariation matrix. Obtained results confirmed hypothesis that some of the designed mutants bind carbohydrates stronger then the native protein.

Other studied systems are pseudorotaxane complexes. Pseudorotaxanes are interlocked molecules composed of a linear molecule (so called axle) threaded into a macrocyclic molecule (so called wheel). The axle is held in the wheel cavity by non-covalent interactions, which prevent the complex from dissociation but also allow the motion of both components. Possible molecular shuttle in such complexes was studied by the potential of mean force method. To obtain converged results, Adaptive Biasing Method was used together with Multiple Walkers Approach. The supramolecular interactions between series of 1-alkyl-3-methylimidazolium guests with variable alkyl substituent lengths and cucurbit[6]uril were investigated in the solution and the conformation behavior of such structures was predicted and compared using methods of molecular dynamics and quantum mechanics.

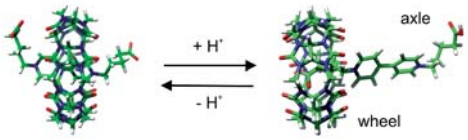


Fig. 4. Molecular shuttle in the studied pseudorotaxane complex

2.6 Docking

Interactions of a bacterial lectin PA-IIL and its mutants with saccharide ligands were studied using various types of docking software. The effect of different partial charges assigned to the calcium ions was tested and evaluated in terms of the best agreement with crystal structure. The docking results were further optimized by molecular dynamics and rescored using AMBER. The agreement of the docked structures and the provided binding energies were evaluated in terms of prediction accuracy. The molecular docking methods proved to be efficient in identifying the correct binding modes in terms of geometry and partially also in predicting the mutation-caused changes in sugar preference.

3 Software

Computational chemistry calculations described in the previous chapter were performed in the following software. Gaussian, Turbomole, and Molpro were used in quantum chemical studies. The benefit of MetaCentrum lies in the possibility of their parallel executions, which significantly decrease the time required for these very time consuming calculations. Molecular dynamics studies are solely performed using the AMBER software package. Docking and related studies are performed using Dock, Autodock, and Autdock Vina. Besides the commonly used software packages, several applications have been developed at NCBR. They simplify job preparation, management and analysis of results as described below.

3.1 Infinity

Infinity provides uniform and simple tools for an easy job submission and management in various computer environments such as computational clusters or grid environments. Infinity is targeted on users using a command line interface, which provides greater extent of flexibility than web based interfaces. This is especially important for fast and innovative scientific research requiring use of computational resources. Thus Infinity also provides an innovative Advanced Module System, which simplifies selection of software according to available architecture and parallel environments.

3.2 TRITON

TRITON is a graphical tool for modeling protein structures and their properties. It offers preparation of input files and visualizations and analyses of output data. TRITON can be used for modeling protein mutants and their properties. Moreover, it provides modeling of chemical reactions at semi-empirical levels with various tools assessing mutant activities. Another feature of TRITON is an interface to AutoDock, which simplifies a docking procedure and result analyses. It offers visualizations of affinity maps and calculations of electrostatic potential interactions between a ligand and individual residues of binding site.

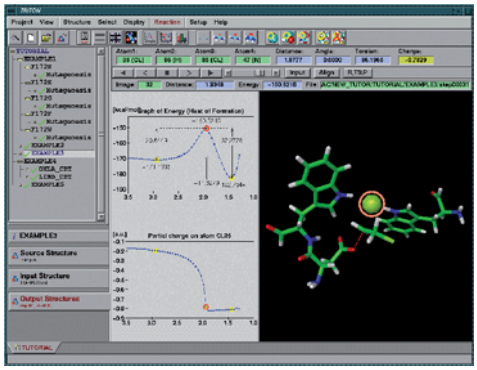


Fig. 5. TRITON screenshot

3.3 PMFLib

PMFLib is a set of libraries and programs supporting the calculations of free energies by the means of Potential of Mean Force (PMF) methods in the framework of molecular dynamics simulations. Currently implemented methods are as follows: Adaptive Biasing Force method, Blue Moon method, Metadynamics, and Umbrella Sampling. Due to a sampling problem, free energy calculations are time consuming. Thus we implemented Multiple Walkers extension to Adaptive Biasing Force method and Metadynamics. This extension employs N independent molecular dynamic simulations, the so-called walkers, that feel the estimated PMF accumulated from all walkers. The interaction among walkers and estimated PMF was achieved by simple client-server architecture and successfully tested in heterogeneous MetaCentrum and VOCE grid environments.

3.4 Other Software

Several other programs and utilities have been developed, namely dynutil, cats, and qmutil. For example, dynutil precycle utility together with Infinity enables an effective automation of job resubmissions. This allows employing MetaCentrum resources limited to short and normal queues for performing very long molecular dynamics simulations. In addition, tools for an effective parallel analysis of molecular dynamics trajectories have been developed and used in the calculations of binding free energies using MM/PBSA method.

4 Hardware

NCBR provided to common MetaCentrum infrastructure about 340 CPUs in year 2009. These resources were organized in two clusters: orca (18 nodes, 2× dual-core Opterons @2.6 GHz) and perian (48 nodes, 2× Intel Xeon @2.4 GHz; 8 nodes, 2× AMD Opteron @1.8 GHz; 10 nodes, 2× quad-core Intel Xeon @3 GHz;

10 nodes, 2× quad-core Intel Xeon @2.5 GHz). Temporarily unused resources were available to other MetaCentrum users via special queues (short, normal and preemptible). This setup enabled preferred access to these resources by NCBR members.

5 Conclusions

Research conducted at NCBR in the field of computational chemistry uses state-of-the art computational approaches, which require up-to-date computational resources. There are diverse requirements for job executions ranging from a huge number of small single CPU jobs to a smaller number of highly parallelized jobs. These different requirements are fulfilled in MetaCentrum. We can therefore conclude that computational resources and services provided by MetaCentrum are helpful for effective research in the field of computational chemistry held at National Centre for Biomolecular Research.

Acknowledgements. The access to the MetaCentrum supercomputing facilities provided under the research intent MSM6383917201 is highly appreciated. This work has been supported by Ministry of Education of the Czech Republic (MSM0021622413, LC06030) and Czech Science Foundation (301/09/H004). The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement no. 205872.

6 Selected Publications Prepared with the Use of MetaCentrum

6.1 Articles

1. Kolman, V., Marek, R., Strelcova, Z., Kulhanek, P., Necas, M., Svec, J., Sindelar, V.: Electron Density Shift in Imidazolium Derivatives upon Complexation with Cucurbit[6]uril. Chemistry-A European Journal. 15, 6926-6931 (2009).
2. Jirouskova, Z., Varekova, R., Vanek, J., Koca, J.: Electronegativity Equalization Method: Parameterization and Validation for Organic Molecules using the Merz-Kollman-Singh Charge Distribution Scheme. Journal of Computational Chemistry. 30, 1174-1178 (2009).
3. Grycova, L., Dommisse, R., Pieters, L., Marek, R.: NMR determination of pK(a) values of indoloquinoline alkaloids. Magnetic Resonance in Chemistry. 47, 977-981 (2009).
4. Bartl, T., Zacharova, Z., Seckarova, P., Kolehmainen, E., Marek, R.: NMR Quantification of Tautomeric Populations in Biogenic Purine Bases. European Journal of Organic Chemistry. 1377-1383 (2009).
5. Macek, P., Chmelik, J., Krizova, I., Kaderavek, P., Padrta, P., Zidek, L., Wildova, M., Hadravova, R., Chaloupkova, R., Pichova, I., Ruml, T., Rumlova,

M., Sklenar, V.: NMR Structure of the N-Terminal Domain of Capsid Protein from the Mason-Pfizer Monkey Virus. Journal of Molecular Biology. 392, 100-114 (2009).
6. Fadrna, E., Spackova, N., Sarzynska, J., Koca, J., Orozco, M., Cheatham, T., Kulinski, T., Sponer, J.: Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. Journal of Chemical Theory and Computation. 5, 2514-2530 (2009).
7. Mladek, A., Sharma, P., Mitra, A., Bhattacharyya, D., Sponer, J., Sponer, J.: Trans Hoogsteen/Sugar Edge Base Pairing in RNA. Structures, Energies, and Stabilities from Quantum Chemical Calculations. Journal of Physical Chemistry B. 113, 1743-1755 (2009).

6.2 Bachelor, Diploma, and Doctoral Theses

1. Chmelik, J. Three dimensional structure determination of a biopolymer using molecular dynamics and experimental restraints measured in isotropic media and oriented phases. Ph.D. Thesis, MU Brno. (2009).
2. Babinsky, M. Binding mode of quaternary isoquinoline alkaloids to d(AAGAATTCTT)2. Diploma Thesis, MU Brno. (2009).
3. Pentakova, M. Studies on OB fold and its dynamical features. Diploma Thesis, MU Brno. (2009).
4. Durech, M. Study of interactions of RS20L lectin with saccharides using molecular modeling methods. Bachelor Thesis, MU Brno. (2009).
5. Kolinkova, R. Interaction of chelerythrine with d(GGAAGCTTCC)2. Bachelor Thesis, MU Brno. (2009).
6. Matuska, R. Studies on protein/carbohydrate interactions by computational tools. Bachelor Thesis, MU Brno. (2009).
7. Sera, L. Study of the interaction of monosacharides with selected lectins using molecular modeling. Bachelor Thesis, MU Brno. (2009).

Mechanisms of the Different DNA Adduct Forming Potentials of the Urban Air Pollutants 2-nitrobenzanthrone and Carcinogenic 3-nitrobenzanthrone

Václav Martínek and Marie Stiborová

Department of Biochemistry, Faculty of Science, Charles University, Albertov 2030, 128 40 Prague 2, Czech Republic
vacmar@natur.cuni.cz

Numerous epidemiological studies have shown increased mortality and morbidity from respiratory and cardiovascular diseases due to ambient air pollution. Nitro-polycyclic aromatic hydrocarbons (nitro-PAHs) are widely distributed environmental pollutants found in extracts of exhausts from diesel and gasoline engines and on the surface of ambient air particulate matter. Although environmental levels of nitro-PAHs are lower compared to unsubstituted PAHs, certain nitro-PAHs exhibit high direct-acting mutagenic potency in bacterial and mammalian bioassays. They have become of enormous concern also because of their carcinogenicity. The increased lung cancer risk after exposure to these environmental pollutants and their detection in the lungs of non-smokers with lung cancer has led to considerable interest in assessing their potential cancer risk.

The nitroaromatic 3-nitrobenzanthrone (3-nitro-7*H*-benz[*de*]anthracen-7-one, 3-NBA, Fig. 1) occurs in diesel exhaust and in airborne particulate matter. In recent years this aromatic nitroketone has received much attention due to its extremely high mutagenic activity in the Ames Salmonella assay. 3-NBA is also carcinogenic in rats, causing lung tumours after intratracheal instillation. The uptake of 3-NBA in humans has been demonstrated and it is a suspected human carcinogen. 3-NBA is a genotoxic carcinogen forming DNA adducts after metabolic activation through reduction of the nitro group. The predominant DNA adducts formed are 2-(2'-deoxyguanosin-N^2-yl)-3-aminobenzanthrone and *N*-(2'-deoxyguanosin-8-yl)-3-aminobenzanthrone and these are most probably responsible for the G to T transversion mutations induced by 3-NBA in Muta Mouse and in the human *TP53* gene in TP53 knock-in (Hupki) murine embryonic fibroblasts. The metabolic activation of 3-NBA is mediated primarily by human and rat cytosolic NAD(P)H:quinone oxidoreductase 1 (NQO1), while *N,O*-acetyltransferases (NATs) and sulfotransferases (SULTs) are the major activating phase II enzymes. Hepatic microsomal NADPH:cytochrome P450 (CYP) reductase also activates 3-NBA effectively *in vitro*. 3-NBA is activated *in vivo* predominantly by cytosolic NQO1 rather than microsomal NADPH:CYP reductase (Fig 1).

The predominant DNA adducts formed are 2-(2'-deoxyguansin-N^2-yl)-3-aminobenzanthrone and *N*-(2'-deoxyguanosin-8-yl)-3-aminobenzanthrone, and these

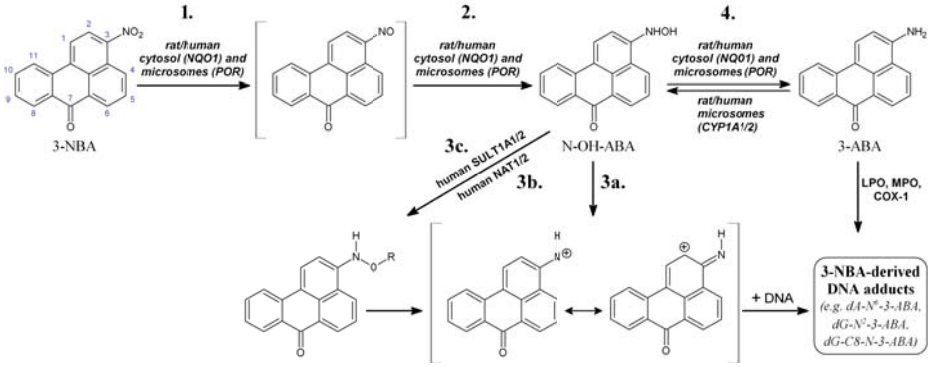


Fig. 1. Pathways of metabolic activation and DNA adduct formation of 3-nitrobenzanthrone and 3-aminobenzanthrone. See text for details. Individual reaction steps are assigned by numbers 1–4 (see also Table 1).

are most probably responsible for the induction of GC to TA transversion mutations induced by 3-NBA.

2-Nitrobenzanthrone (2-NBA), an isomer of 3-NBA, has been detected in ambient air particulate matter. It has been suggested that 2-NBA might be formed more specifically by atmospheric processes while 3-NBA seems to be formed preferentially by combustion processes, such as in a diesel engine. Using a semiempirical quantum mechanical approach recent studies indicated small differences between 2-NBA and 3-NBA in the reduction potential and the geometry of the nitro group attached to a benzanthrone skeleton. Only small differences in hydrophobicity were observed between these NBA isomers, indicating that penetration through cell membranes is expected to be similar for both compounds. In contrast, the mutagenic and genotoxic potential of 2-NBA is much lower than that of 3-NBA. In the standard strain TA98 2-NBA is over 1000-fold less mutagenic than 3-NBA; in strain YG1024 even as much as 2000 times. Although 2-NBA has been shown to be genotoxic *in vitro*, lack of genotoxicity *in vivo* in rats (e.g., DNA adduct formation) was reported previously. Nevertheless, its higher abundance than 3-NBA in ambient air urges further investigation to assess its potential hazard to human health.

The aim of the present study was to investigate the reasons for the large differences in DNA adduct formation by 2-NBA and 3-NBA found previously in several cells in culture and in rats exposed to these compounds. Previously, we have characterized the first two steps of the metabolic activation cascade to species forming DNA adducts for 3-NBA, which comprise reduction of the nitro group forming the corresponding hydroxylamine (Fig. 1).

All evaluated nitrobenzanthrones are almost chemically inert at the physiological conditions; therefore their enzymatic transformation is necessary step toward their activation and/or detoxication. Evaluation of interactions of NBAs

with the active centers of major human enzymes metabolizing NBAs (NQO1) is required for further elucidation of their activation mechanism.

In order to examine the molecular basis of the differences in the reductive activation of 2-NBA and 3-NBA by NQO1, their binding to the active centre of NQO1 was modeled. The 3D model of the human NQO1 enzyme protein was based on revised human NQO1 crystal structure. NQO1 3D coordinates in 2.50 Å resolution are available from RCSB Protein Data Bank under PDB ID 1DXO. This NQO1 structure was used because its active site adopts a small planar ligand (Duroquinone) that sterically resembles the planar molecules of 2-NBA and 3-NBA. There are also other NQO1 structures with co-crystallized ligands (available, but their ligands, unlike the 2- or 3-NBA, also contain an extra-planar group. 2-NBA and 3-NBA ligand geometries were obtained via *ab initio* calculations on the Hartree-Fock level of theory in conjunction with the 6-31G(d) basis set.

We employed a hybrid global-local Lamarckian genetic algorithm implemented in Autodock v4 program suit to evaluate binding free energies and preferred binding modes for 2-NBA and 3-NBA. In order to allow enzyme to adapt new ligand we ran soft-soft docking calculations. All rotatable bonds of ligands and Tyr126 (chain C), Tyr128 (chain C) and Trp105 (chain A) and His161 (chain A) residue side chains were allowed to rotate freely.

The calculated model structures for the NQO1 2-NBA and NQO1 3-NBA complexes indicated that both compounds fit well into the active site of NQO1, having similar binding affinities. Docking calculations (using a dimeric molecule of NQO1, which is biologically active) predict negligible differences between 2-NBA and 3-NBA binding affinities (-5.8 and -5.7 kcal/mol). The best orientations (facilitating the reduction) of the two NBA isomers are different, however, resulting in a greater distance of the nitro group of 2-NBA to the hydrogen on N5 of the isoalloxazine skeleton of FAD, the NQO1 flavin prosthetic group (4.1–4.4 Å), than that of 3-NBA, which is located close to hydrogens of this part of molecule of FAD (2.4–3.5 Å) (Fig. 2). These spatial arrangements allow an electron transfer to facilitate reduction in the case of 3-NBA, but not of 2-NBA.

In order to evaluate the difference in reaction free energies during individual reaction steps from Figure 1, the geometry optimizations of all reactants and products were done using *ab initio* approach implemented in Gaussian03 program suite. All calculations were performed on the Hartree-Fock (HF) level of theory in conjunction with 6-31+G(d) basis set. Initially, the solvent effect was estimated by performing energy optimizations using the polarizable conductor calculation model (CPCM) with default atomic radii. The thermal corrected Gibbs free energies in CPCM solvent environment were obtained from electronic calculations and harmonic vibration frequencies of these optimized structures. The reaction Gibbs free energies of individual reaction steps evaluated here were calculated as the total energies of products minus the total energies of reactants.

Several reasons might be responsible for the different potency of 3-NBA and 2-NBA to form DNA adducts in several enzymatic systems *in vitro* and in hu-

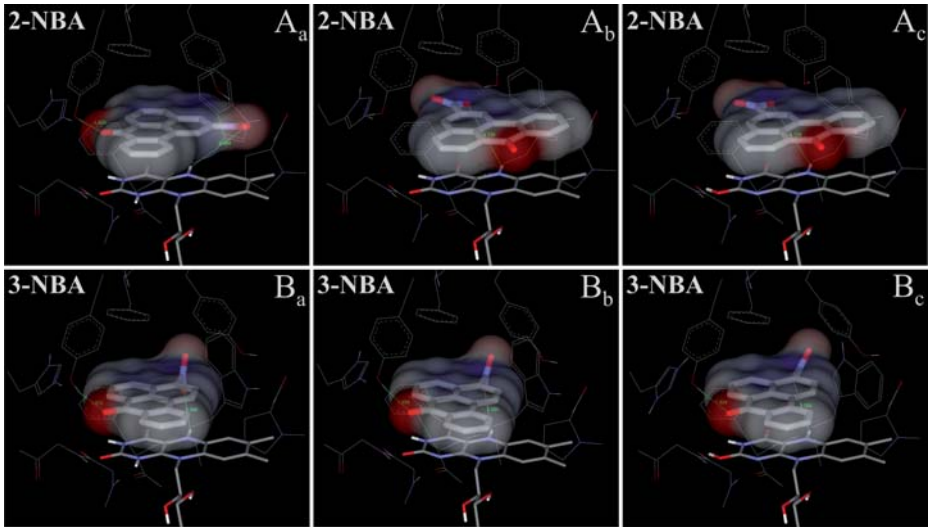


Fig. 2. The binding modes facilitating the hydrogen transfer form N5 of isoalloxazine ring to NO_2 group of 2-NBA (A_a-A_c) or 3-NBA (B_a-B_c) are shown docked to the active site of human NQO1. Results for three possible forms of reduced isoalloxazine ring—ketoform, ionized enolate and protonated enol form bound to the NQO1 are labeled with subscript a, b or c, respectively. Both ligands are positioned parallel to the flavin prosthetic group. 2-NBA or 3-NBA, FAD cofactor and amino acids residues within 5.5 Å from ligand are rendered as bold sticks, sticks and lines, respectively.

man hepatocytes. Amenability of 3-NBA and 2-NBA to chemical and enzymatic reduction might be one of the reasons.

As shown in Table 1, the absolute values of reaction free energies of individual reaction steps (1, 2 and 4 in Fig. 1) calculated using CPCM approach are lower than -25 kcal/mol. Therefore, these reaction steps should not, in presence of suitable catalyst, be controlling the overall rate of the whole reaction sequence. The dissociation of *N*-hydroxylamines (the step 3a in Fig. 1) in water solution is not thermodynamically favored. Slightly negative reaction energies were obtained for hydrolysis of conjugates with acetic or sulfuric acid, the step 3b or 3c in Fig. 1). Therefore, the dissociation of *N*-hydroxylamines is the suitable candidate for the rate limiting step determining the overall reaction rate of the nitrenium ion formation from 2-NBA and 3-NBA.

Interestingly, the dissociation of *N*-hydroxy-3-ABA is thermodynamically disfavored by 6.3 kcal/mol over dissociation of *N*-hydroxy-2-ABA (Table 1). This fact results in more than four orders of magnitude-fold difference in reaction equilibrium constants of *N*-hydroxy-3-ABA *versus* *N*-hydroxy-2-ABA. Such significant differences in the reaction free energy ΔG of nitrenium ion formation could partially contribute to the large disparity in DNA adduct forming potential of 2-NBA and 3-NBA. Very similar tendency was detected also for hydrolysis

Reaction step (Scheme 1)	1. $NO_2 \rightarrow NO$	2. $NO \rightarrow NHOH$	3a.[a] $NHOH \rightarrow NH^+ + OH^-$	3b. $NHOH \rightarrow NH^+ + Ac^-$	3c. $NHOSO_3^- \rightarrow NH^+ + SO_4^{2-}$	4. $NHOH \rightarrow NH_2$
$\Delta G^{0' wat(PCM)}_{\text{rea2NBA}}$ [kcal/mol]	-44.4	-25.6	28.6	0.2	-1.6	-58.7
$\Delta G^{0' wat(PCM)}_{\text{rea3NBA}}$ [kcal/mol]	-46.9	-28.3	22.3	-6.1	-8.3	-64.5
$\Delta\Delta G^{0' wat(PCM)}_{\text{rea2NBA-3NBA}}$ [kcal/mol]	-2.5	-2.7	**-6.3**	**-6.3**	-6.7	-5.8
$\Delta pK'_{\text{eq2NBA-3NBA}}$[b]	-1.8	-2.0	-4.6	-4.6	-4.9	-4.3

[a] the standard (biochemical) free energy change of this reaction step was corrected to pH=7

[b] difference in equilibrium constant (2-NBA minus 3-NBA)

Table 1. The comparison among the standard reaction free energies of individual reaction steps of 2-NBA and 3-NBA conversion calculated by quantum chemical approach in combination with CPCM solvation model.

of corresponding conjugates with acetic acid, and slightly more pronounced differences were predicted for conjugates with sulfuric acid (Table 1).

Publications Prepared with the Use of METACenter

[1] Stiborová M, Martínek V, Svobodová M, Šístková J, Dvořák Z, Ulrichová J, Šimánek V, Frei E, Schmeiser HH, Phillips DH, Arlt VM.: Mechanisms of the Different DNA Adduct Forming Potentials of the Urban Air Pollutants 2-nitrobenzanthrone and Carcinogenic 3-nitrobenzanthrone. Submitted to *Chemical Research in Toxicology.*

Used Programs

Gaussian 03, Autodock 4

Simulation of Ion Scattering on Solid State Surfaces

T. Matlocha

Institute of Physical Engineering, Brno University of Technology,
Brno, Czech Republic.
tomasmatlocha@seznam.cz

Surfaces and interfaces of materials finding applications in semiconductor industry or in nanotechnology applications in general have been commonly studied by particle scattering. Information about surface atom composition or structure can be obtained by an analysis of ions scattered from these surfaces. Ion spectroscopic methods LEIS and TOF-LEIS[1] utilize primary energies of incident ions (e.g. of inert gases) in the order of units of keV. Complex interpretation of corresponding energy or time spectra is a main disadvantage of these methods. The interpretation problems of experimental results can be resolved by ion scattering simulations.

For these purposes the CDLS computer code has been developed in the IPE FME BUT Brno. The physical model of ion scattering implemented into the CDLS[2] code is based on the simple principles of classical dynamics. The mutual binary interaction between the projectile and target atoms is described by the screened potential [1]

$$V(R) = \frac{1}{4\pi\varepsilon_o}\frac{Z_2 Z_2 e^2}{R}\phi(R), \tag{1}$$

where R is the interatomic distance between interacting particles with atomic numbers Z_1 and Z_2, respectively, and $\phi(R)$ is the screening function of the potential $V(R)$. The mutual interactions between target atoms is neglected. In addition to collisions with nuclei of target atoms the primary ions interact with electrons and this interaction is described by the Lindhard-Scharf formula [1]. Also, the thermal vibrations of target atom structures have been implemented into the CDLS code based on the Debye-Waller theory [2].

During the year 2009 we studied a Si surface covered by Ga droplets and Cu(100) surface by He^+ ion scattering simulations and presented our results at the IBA International Conference in Cambridge. As well, a manuscript of an article entitled "A study of a LEIS azimuthal scan behavior: Classical dynamics simulation" was sent to the journal Surface Science. Here, the results of He^+ ion scattering from the Cu(100) surface were summarized.

[1] Time of flight low energy ion spectroscopy.
[2] LEIS Classical Dynamics Simulator.

He^+ Ion Scattering from the Cu(100) Surface

In the mentioned manuscript we built upon already published experimental results [3, 4] of the group of Peter Bauer from Johannes Kepler University in Linz. The group has been involved in the study of surfaces by low energy ion scattering for many years. First, we tested a validity of our simulation results by comparing the simulated azimuthal scans $N(\phi)$ and experimental results from Linz (Fig. 1). We extended our simulated angular scans $N(\phi)$ to an ion scattering map $N(\phi, \beta)$ (Fig. 2) which includes information on all He^+ ions scattered above the probed Cu(100) surface. This extension brought more light into ion scattering phenomenas at the Cu(100) surface and even allowed us to identify positions of Cu surface atoms (Fig. 2b) causing a rapid intensity decrease of detected ions in $N(\phi, \beta)$ scattering maps.

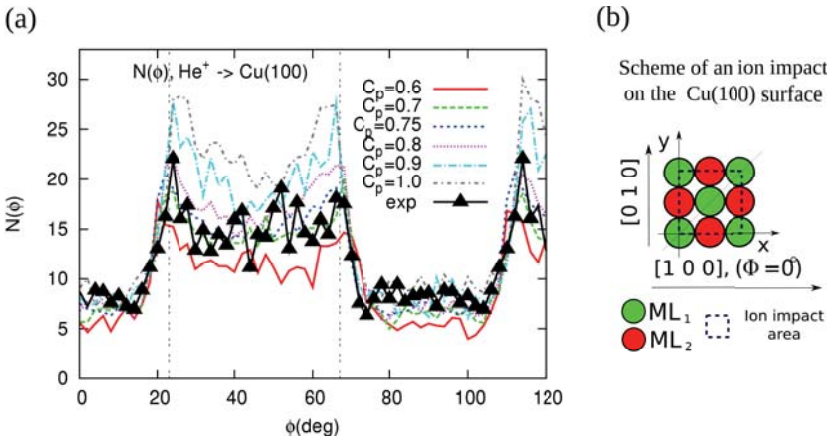


Fig. 1. a) Simulated azimuthal scans $N(\phi)$ of e He^+ ions scattered from the Cu(100) surface for different correction factors of the screening function in the Thomas-Fermi-Moliere potential (C_p = 0.6,.., 1.0), together with the experimental azimuthal scan. b) A scheme of an ion impact area on the Cu(100) surface used in the CDLS simulations.

He^+ Scattering on a Ga(Pt)/Si Surface

The group of Surfaces and Thin Films at the Institute of Physical Engineering at BUT mainly focuses on the study of objects of small dimensions, such as nanostructures. A significant part of the scientific capacity of this group has been dedicated to a study of physical properties of such nanoobjects by using many analytical methods such as AFM, XPS, SIMS, etc. It opens a question whether the TOF-LEIS method, which is primarily intended for structural resp. element analysis of surfaces, is capable and sensitive enough to recognize

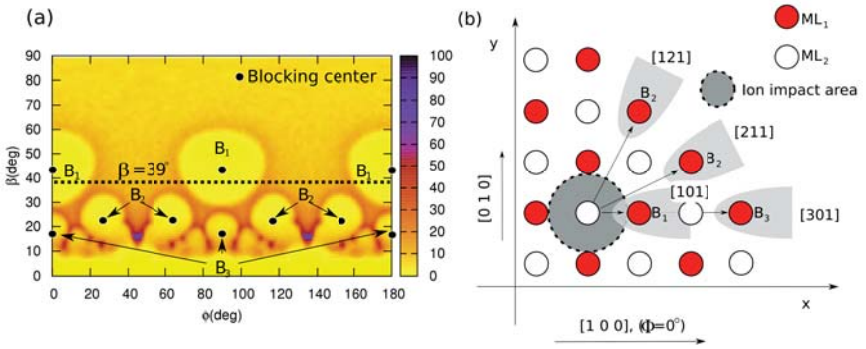


Fig. 2. a) Simulated ion scattering maps $N(\phi, \beta)$ of He^+ ion scattering from the 2^{nd} atomic monolayer ML_2 of the Cu(100) surface with depicted blocking centers B_{1-3} . b) Schematic view of the ion blocking centers (1^{st} monolayer of Cu(100)) affecting He^+ ion trajectories.

– the shape and size of these nanostructures or to measure their ordering on surfaces.

The sensitivity of TOF-LEIS method to the shape and dimensions of nanoobjects have already been experimentally confirmed, and published in our article [5]. The question of LEIS sensitivity to ordering nanoobjects on surfaces remains unanswered for now, lacking experimental evidence. Therefore, in this work, we investigate an effect of ordered resp. non ordered systems of surface nanoobjects on the resulting angular dependence of $N(\phi)$ and $N(\phi, \beta)$, theoretically. We expect that in the future we will succeed in comparing experimental and simulated angular dependences, and thus in extending the application of TOF-LEIS methods.

The real surface (Ga system droplets deposited on a Si surface) was modelled by a system of rotational ellipsoids periodically distributed over the modeled surface (Fig. 3). Using this model we could examine the effects on the resulting azimuthal scans $N(\phi)$ (Fig. 4) and $N(\phi, \beta)$ (Fig. 5b). The influence of a droplet structure on the intensities in $N(\phi)$ and $N(\phi, \beta)$ scans is discussed in Fig. 3b. Alternatively, we also tested the effect of a heavier element than Ga (like Pt) in the model. Pt has a much greater influence on the attenuation of the signal in comparison with Ga.

The resulting simulated angular dependences $N(\phi)$ and $N(\phi, \beta)$ show that the ability to detect the periodicity of ordered nano-objects by ion beam scattering is in principle possible. This is especially true for small scattering angles and the h/r ratios of Ga droplets bigger than one.

In the year 2010 it is planned to continue the experiment of He^+ ion scattering on a Ga/Si surface and to confirm or disprove the idea of the application of TOF-LEIS in the field of nanotechnologies. For this purpose, and in particular due

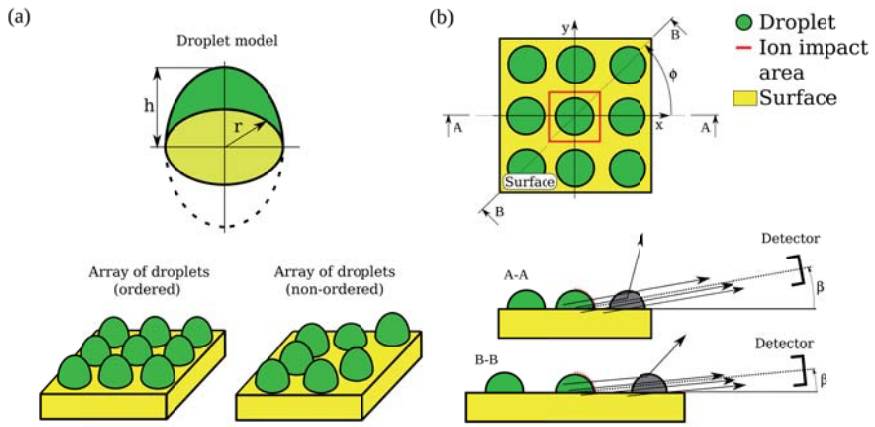


Fig. 3. The studied model of ordered and non ordered arrays of Ga(Pt) droplets on a Si surface, one droplet is modelled by an rotational ellipsoid with the radius r and height h. b) the principle of ion scattering attenuation resulting in an intensity decrease of scattered ions in the azimuthal dependencies $N(\phi)$ and $N(\phi, \beta)$ in directions with the greatest density of nanoobjects ($\phi = 0°, 45°, 90°, \ldots$).

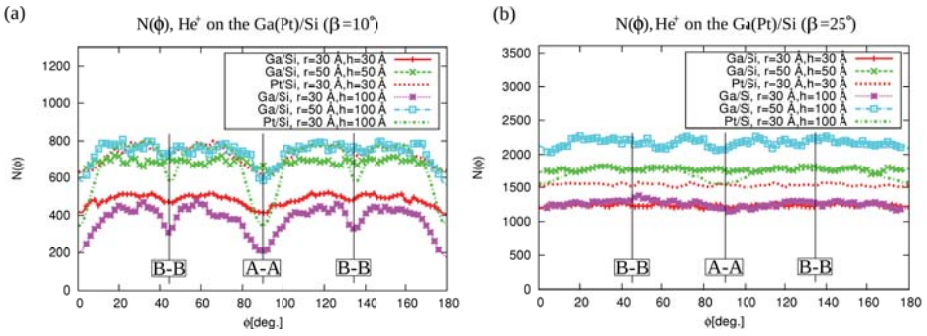


Fig. 4. Azimuthal scans $N(\phi)$ of the He^+ ion scattering from a Ga(Pt)/Si surface (ordered array of droplets). a) $N(\phi)$ of ions scattered from the surface under the exit angle $\beta = 10°$ (the exit angle β is measured from the surface). b) $N(\phi)$ for $\beta = 25°$. The areas corresponding to the largest intensity drops of detected ions are marked by A-A resp. B-B.

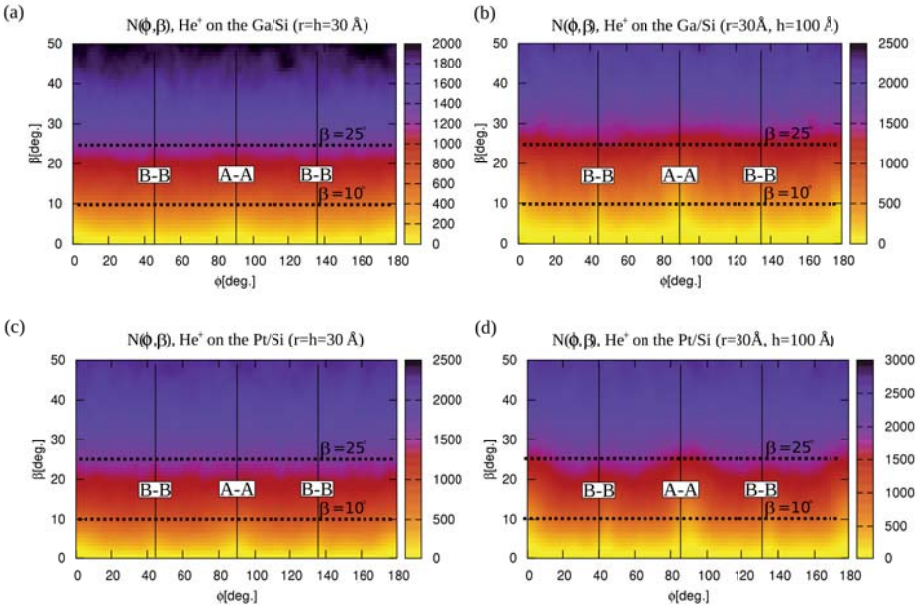


Fig. 5. Simulated scattering maps $N(\phi, \beta)$ of the He^+ ion scattering from a Ga(Pt)/Si surface with droplet dimension a) $r = 30°$, $h = 30°$, b) $r = 30°$, $h = 100°$, c) $r = 30°$, $h = 30°$, d) $r = 30°$, $h = 100°$. The areas corresponding to the largest intensity drops Simulated scattering maps $N(\phi, \beta)$ of the He^+ ion scattering from a Ga(Pt)/Si surface with droplet dimension a) $r = 30$ Å, $h = 30$ Å, b) $r = 30$ Å, $h = 100$ Å, c) $r = 30$ Å, $h = 30$ Å, d) $r = 30$ Å, $h = 100$ Å. The areas corresponding to the largest intensity drops of detected ions are marked by A-A resp. B-B of detected ions are marked by A-A resp. B-B.

to corresponding time consuming calculations it is essential to use the computer networks of the METACentrum.

Grant Support

– The research grant programmes of Ministry of Education CR (Projects No. MSM0021630508 and LC06040)
– GAAV (Projects No. IAA1010413 and KAN400100701)
– EUROCORES GACR project (FON/06/E001)

Used Programs and Applications

– CDLS—home-built code

List of Publications Dedicated to MetaCentrum

– T. Matlocha, S. Průša, M. Kolíbal, P. Bábor, T. Šikola, D. Primetzhofer, N. Markin, P. Bauer: *A study of a LEIS azimuthal scan behavior: Classical dynamics simulation.* Sent manuscript.

References

1. Eckstein W.: *Computer Simulation of Ion-Solid Interactions.* Springer-Verlag Berlin Heidelberg 1991. ISBN 0-387-19057-0.
2. Niehus H., Heiland W., Taglauer E.: *Low-energy ion scattering at surfaces.* Surf. Sci. Rep. 17(1993) 213-303.
3. D. Primetzhofer, S.N. Markin, R. Kolarova, M. Draxler, R. Beikler, E. Taglauer, P. Bauer, Nucl. Instr. Meth. B 258 (2007), p. 36.
4. D. Primetzhofer, S.N. Markin, M. Draxler, R. Beikler, E. Taglauer, P. Bauer, Surf. Sci. 602 (2008), p. 2921.
5. M. Kolíbal, O. Tomanec, S. Průša, M. Plojhar, S.N. Markin, L. Dittrichová, J. Spousta, P. Bauer, T. Šikola: *ToF-LEIS spectra of Ga/Si: Peak shape analysis*, Nuclear Instruments and Methods in Physics Research Section B, Vol.265,(2007), No. 2.

Molecular Dynamics Simulations of Multimeric Protein Complexes

Milan Melicherčík, Žofie Sovová, Morteza Khabiri, Natalia Kulik, and Rüdiger Ettrich

Institute of Systems Biology and Ecology,
Academy of Sciences CR, Zámek 136, 37333 Nové Hrady
and
Institute of Physical Biology,
University of South Bohemia, Zámek 136, 37333 Nové Hrady
melichercik@nh.usbe.cas.cz, sovova@nh.usbe.cas.cz,
khabiri@nh.usbe.cas.cz, kulik@nh.usbe.cas.cz, ettrich@nh.usbe.cas.cz

The MD (molecular dynamics) computational method is based on each atoms' interactions. Interactions are described using classical Newton equations with empirically derived parameters from experiments and quantum calculations. MD simulations let us observe motions and movements on an atomic level (advantage over experimental methods) and it allows simulations of much larger systems (comparing to quantum methods). Presently, it is possible to simulate systems containing multimeric systems of proteins, nucleic acids, lipid membranes, etc. (hundreds of thousands of atoms, including the solvent) up to several hundreds of ns. Taken these limitations, we can observe complex quarternaly changes in protein systems which enables us to propose hypothesises e.g. about mechanism of the protein functionality. Of course, experimental feedback is appreciated to verify the computational predictions.

1 Project 1: Functionality of Arginine Repressor

Arginine repressor (ArgR) is the master regulator of the arginine regulon in a wide variety of bacteria [1], acting as direct sensor and transcriptional transducer of intracellular L-arginine (L-arg) concentrations to provide feedback control over biosynthesis and catabolism of L-arg. The co-effector L-arg binds to a central hexamerization domain, altering DNA affinity and specificity [2] of peripheral domains. The structural organization of ArgR into N- (ArgRN) and C-terminal (ArgRC) domains, and the functional division of labour between them, are conserved even among distant orthologs that display an unexpected diversity of reported biochemical properties, notably the L-arg dependence of hexamerization and DNA-binding equilibria [3-7]. An allosteric mechanism was inferred by comparison of crystallized intact unliganded apoprotein from the thermophile *Bacillus stearothermophilus* with its liganded C-terminal domain fragment, which differ by ≈ 15 degrees rotation about the trimer-trimer interface that was ascribed to L-arg binding and presumed to be transmitted to the DNA-binding domains [8]. A similar degree of rotation was reported recently between apo- and holoArgRC of *Mycobacterium tuberculosis* [9]. Unlike the *Bacillus*

and *Mycobacterium* proteins, crystal structures of the *E. coli* ArgR C-terminal domain with (holoEcArgRC) and without (apoEcArgRC) bound L-arg are essentially identical [10]. Crystalline apo- and holoEcArgRC hexamers are also entirely symmetric, a finding that is seemingly incongruent with the complex thermodynamics of L-arg binding. Isothermal titration calorimetry (ITC) confirms that EcArgR and EcArgRC hexamers bind six equivalents of L-arg, but with a multiphasic binding mechanism in which the first binding event has ≈ 100-fold greater affinity than the subsequent five events [11]; thus L-arg binding is itself allosteric.

Molecular dynamics simulations with ArgRC, the hexameric domain that binds L-arginine with negative cooperativity, reveal that conserved arginine and aspartate sidechains in each ligand-binding pocket promote rotational oscillation of apoArgRC trimers by engagement and release of hydrogen-bonded salt bridges. Binding of exogenous L-arginine displaces resident arginine residues and arrests oscillation, shifting the equilibrium quaternary ensemble and promoting motions that enable formation of a cooperative hydrogen-bond network among subunits while maintaining the configurational entropy of the system. A single L-arg ligand is necessary and sufficient to arrest oscillation, and the symmetry of the hexamer is maintained despite the conceptual asymmetry of the partially-liganded state. The results also offer the first opportunity to describe in structural and thermodynamic terms the symmetric relaxed state predicted by the concerted allostery model of Monod, Wyman, and Changeux [12], revealing that this state is achieved by exploiting the dynamics of the assembly and the distributed nature of its cohesive energy. This example reveals that symmetry can be maintained even when binding sites fill sequentially due to negative cooperativity, which was not anticipated by the Monod, Wyman, and Changeux model. The molecular mechanism identified here neither specifies nor requires a pathway for transmission of the allosteric signal through the protein, and it suggests the possibility that binding of free amino acids was an early innovation in the evolution of allostery.

Bibliography

1. W. K. Maas, The arginine repressor of *Escherichia coli*, Microbiol. Mol. Biol. Rev. 58 (1994) 631-640.
2. D. Szwajkajzer, L. Dai, J.W. Fukayama, B. Abramczyk, R. Fairman, J. Carey, Quantitative analysis of DNA binding by the *Escherichia coli* arginine repressor, J. Mol. Biol. 312 (2001) 949-962.
3. D. Lim, J.D. Oppenheim, T. Eckhardt, W.K. Maas, Nucleotide sequence of the argR gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 6697-6701.
4. L.G. Czaplewski, A.K. North, M.C. Smith, S. Baumberg, P.G. Stockley, Purification and initial characterization of AhrC: The regulator of arginine metabolism genes in *Bacillus subtilis*, Mol. Microbiol. 6 (1992) 267–275.

5. C.-D. Lu, J.E. Houghton, A.T. Abdelal, Characterization of the arginine repressor from *Salmonella typhimurium* and its interactions with the carAB operator, J. Mol. Biol. 225 (1992) 11-24.
6. M. Dion, D. Charlier, H. Wang, D. Gigot, A. Savchenko, J.-N. Hallet, N. Glansdorff, V. Sakanyan, The highly thermostable arginine repressor of *Bacillus stearothermophilus*: Gene cloning and repressor-operator interactions, Mol. Microbiol. 25 (1997) 385-398.
7. A. Morin, N. Huysveld, F. Braun, D. Dimova, V. Sakanyan, D. Charlier, Hyperthermophilic *Thermotoga* arginine repressor binding to full-length cognate and heterologous arginine operators and to half-site targets, J. Mol. Biol. 332 (2003) 537-553.
8. J. Ni, V. Sakanyan, D. Charlier, N. Glansdorff, G.D. Van Duyne, Structure of the arginine repressor from *Bacillus stearothermophilus*, Nat. Struct. Biol. 6 (1999) 427-432.
9. L.T. Cherney, M.M. Cherney, C.R. Garen, G.J. Lu, M.N. James, Structure of the C-domain of the arginine repressor protein from *Mycobacterium tuberculosis*, Acta Cryst. D64 (2008) 950-956.
10. G.D. Van Duyne, G. Ghosh, W.K. Maas, P.B. Sigler, Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*, J. Mol. Biol. 256 (1996) 377-391.
11. L. Jin, W.F. Xue, J.W. Fukayama, J. Yetter, M. Pickering, J. Carey, Asymmetric allosteric activation of the symmetric ArgR hexamer, J. Mol. Biol. 346 (2005) 43-56.
12. J. Monod, J.-P. Changeux, F. Jabcob, Allosteric proteins and cellular contorol systems, J. Mol. Biol. 6 (1963) 306-329.

2 Project 2: Interactions of Mouse Activatory Natural Killer Cell NKR-P1 Receptors

Natural killer cells (NK cells) are the third major group of lymphocytes derived in blood morrow. They do not express the T-cell receptors or B-cell receptor and mediate natural killing against prototype NK-cell-sensitive targets. NK cells are involved in rejecting tumorous, infected or differently harmed cells by the innate immune system. They distinguish these broken cells from normal cells by interacting with the receptors on the cellular surface. This mechanism is described by so-called missing-self hypothesis [1].

One group of receptors on their surface is NKR-P1 receptors, which are homodimeric type II transmembrane proteins each containing extracellular C-type lectin-like domains (CTLD). CLTDs [2, 3] are alpha and beta proteins having N- and C-terminal close together and forming antiparallel β-sheet. The first β-strand is followed by helix–sheet–helix motif and than by so-called long loop region, that has its own mainly β-structure. This loop is connected to the C-terminal β-strand. Four cysteines are fully conserved among all CLTDs. They form two disulfide bonds that stabilize the structure of the domain. First pair of cysteines can be found between N-terminal α-helix and the last β-strand, the

second one on the N-terminus of long loop, where it stabilizes an antiparellel β-sheet. Long type CLTDs [2], like NKR-P1s, have additional disulfide bond stabilizing N-terminal β-hairpin, that is the characteristic feature for long type CLTDs.

Functional NKR-P1s can be found in rat and mice as well as in human and chicken. Extensive searches of genome databases [4, 5] revealed, that orthologues of these receptors are also in other mammals such as cows, dogs, chimpanzees... However functionality of these have not yet been proofed.

Two types of ligands were reported to interact with NKR-P1 receptor: saccharides [6] and Clr proteins [7, 8]. As no experimental structural information is available to date, homology models of rat NKR-P1A, NKR-P1B receptors, mouse NKR-P1A, NKR-P1C, NKR-P1F, NKR-P1G and human NKR-P1 receptors were generated and analyzed with respect to understand the structural features determining the function. All models were refined by at least 10 ns of molecular dynamics simulations at room temperature in explicit solvent. Experimentally, rat NKR-P1A/B and mouse NKR-P1A/C receptors were structurally analyzed by Raman and infrared spectroscopy. Thermal dynamics measurements from 5 to 90 °C were performed and Raman and infrared spectra were subsequently analyzed by two-dimensional correlation analysis. The analysis showed that prior to a standard sequence of denaturation process a part with high β-turn content, and probably containing Trp, is rearranged. This is in agreement with resultant homology models. Two regions, that are conserved in all analyzed NKR-P1s and under evolutionary pressure, do not contain core secondary structure elements, as one would expect, but represent loop regions, one of them the long loop region with only a minor portion of secondary structure, that even is ambiguously in the different models. We conclude that this region plays a key role in the ligand-specifity and that random coil structure is involved in ligand binding. Also the high flexibility of this region, with conserved anchoring sequences at the beginning and end, indicates that its relative position to the rest of the protein might be variable. Theoretical results will are confirmed experimentally by the collaborating group of prof. Bezouska from Charles University.

Bibliography

1. Ljunggren HG, Karre K (1990) In search of 'missing self': MHC molecules and NK cell recognition. Immunol. Today 11:237–244 homodimeric type II transmembrane proteins each containing extracellular C-type lectin-like domains
2. Zelensky AN, Gready JE (2005) The C-type lectin-like domain superfamily. FEBS J. 272:6179–6217
3. Weis WI, Taylor ME, Drickamer K (1998) The C-type lectin like superfamily in the immune system. Immunol. Rev. 163:19–34
4. Zelensky AN, Gready JE (2003) Comparative analysis of structural properties of the C-type-lectin-like Domain (CTLD). Proteins 52: 466–477
5. Hao L., Klein J., Nei M (2006) Heterogenous but conserved natural killer receptor gene complexes in four major orders of mammals. PNAS 103:3192–3197

6. Bezouška K, Yuen CT, O'Brien J, Childs RA, Chai W, Lawson AM, Drbal K, Fišerová A, Pospíšil M, Fiezi T (1994) Oligosaccharide ligands for NKR-P1 protein activate NK cells and cytotoxicity. Nature 372:150–157
7. Kveberg L, Dai KZ, Westgaard IH, Daws MR, Fossum S, Naper C, Vaage JT (2009) Two major groups of rat NKR-P1 receptors can be distinguished based on chromosomal localization, phylogenetic analysis and Clr ligand binding. Eur. J. Immunol. 39: 541–551
8. Tian W, Nunez R, Cheng S, Ding Y, Tumang J, Lyddane C, Roman C, Liou HC (2005) C-type lectin OCILRP2/Clr-g and its ligand NKR-P1f costimulate T cell proliferation and IL-2 production. Cell. Immunol. 234:39–53

3 Project 3: Influence of Organic Solvents on the Structure and Enzymatic Activity of Haloalkane Dehalogenases

Organic solvents can have different effect of the catalytic reactions such as lowering the energy of activated complex or stabilization of enzyme in solution. Moreover using solution of organic solvents can improve the solubility of some reactants. The effect of some non-ionizing water-soluble organic solvents on the enzyme activity was studied both by experiments and molecular dynamics simulations. As many mono or divalent cations and anions can enhance the activity of enzymes in aqueous solutions of organic solvents, we examined the case of haloalkane dehalogenase in mixtures of organic solvents with water. Organic solvents can increase the solubility of haloalkanes which have low solubility in water. Solvation and dynamics of some organic solvents such as DMSO, isopropanol and formamide in water have been studied by means of molecular dynamics simulations. Different concentrations of four organic solutions are used: ACETONE 20 %, Formamide 5 %, Isopropanol 10 %, DMSO 42 %. Simulations for these solutions are run till full homogenity up to 20 ns. 4 different proteins, dehalogenases (DhaA, DhaA57, LINB, DBJA), were solvated in these solutions and simulated for 35 ns. The effect of these organic solvents on the structure and dynamics of dehalogenases have been studied. Additionally, dynamics, solvation and structure of different ions in different organic solvents have been studied. Ion pair formation and propensity to the solution/air interface have been observed for some of organics solvents.

4 Project 4: 4-Deoxy-substrates for β-N-acetylhexosaminidases: How to Make Use of "Wobbling Specificity"

METAcentrum was used for MD calculations of substrate—enzyme complexes of hexosaminidases from *A. oryzae* and *P. oxalicum* with modified carbohydrates. Optimization of MD parameters for calculation of these complexes revealed that

despite of the high activity of *P. oxalicum* at pH 3—better reproducibility of 'wet' experiment results is reached at pH 7.

The results of MD with oxidized substrates a the active site of *A. oryzae* were used to explain substrate specificity and revealed strong changes at active site with docked dehydrated substrates.

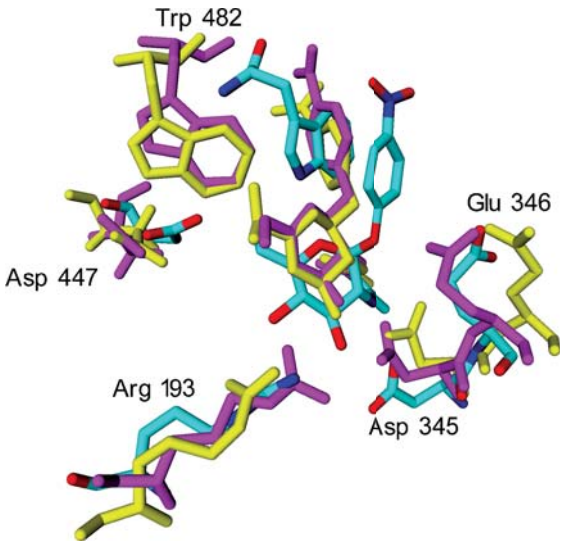


Fig. 1. Compare of substrates: yellow—p-Nitrophenyl 2-acetamido-2,4-dideoxy-α-L-threo-hex-4-enodialdo-1,5-pyranoside, magenta—p-Nitrophenyl 2-acetamido-2,4-dideoxy-α-L-threo-hex-4-enopyranoside and colors by elements—standard substrate p-Nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside: overlay of enzymes with bound substrates. Amino acid positions differing from WT are shown.

Programs

Gromacs 3.2.1–4.0.7
Yasara 7.5.14
Gaussian 03

Acknowledgments

We gratefully acknowledge support from the Ministry of Education, Youth and Sports of the Czech Republic (MSM6007665808, LC06010), Academy of Sciences of the Czech Republic (AVOZ60870520), Grant Agency of the Czech Republic (203/08/0114 and KJB101120805), Ministry of Education of the Czech

Republic (MSM 0021620835) and joint Czech–US National Science Foundation International Research Cooperation (ME09016 and NSF INT03-09049). Access to METACentrum supercomputing facilities was provided under the research intent MSM 6383917201.

Impacted Publications Acknowledging the Use of Metacentrum

1. Strawn, R., Melicherčík, M., Green, M., Stockner, T., Carey, J., Ettrich, R.: Symmetric allosteric mechanism of hexameric Escherichia coli arginine repressor exploits competition between L-arginine ligands and resident arginine residues, PLoS Comput. Biol. (*accepted for publishing*).
2. Sovová, Ž, Kopecký, jr., V., Pazderka, T., Hofbauerová, K., Rozbeský, D., Vaněk, O., Bezouška, K., Ettrich, E.: A structural analysis of natural killer cell receptor proteins 1 (NKR-P1) reveals the long loop region to be a conserved key feature involved in ligand specifity, J. Mol. Model. (*submitted*)
3. Slámová, K., Gažák, R., Bojarová, P., Kulik, N., Ettrich, R., Pelantová, H., Sedmera, P., Křen, V.: 4-Deoxy-substrates for β-N-acetylhexosaminidases: how to make use of "wobbling specificity", Glycobiology (*in press*)
4. Kopecký, V., Kohoutová, J., Lapkouski, M., Hofbauerová, K., Sovová, Ž., Kutá-Smatanová, I., Revuelta, J. L., Arellano, J. B., Ettrich, R.: Drop coating deposition Raman—a powerful new tool demonstrated on extrinsic PsbP protein of PSII from Spinacea oleracea (*submitted*)

Theoretical Investigation of Chemical and Physical Properties of Microporous Materials

Petr Nachtigall[1], Ota Bludský[2], Miroslav Rubeš[2], Lukáš Grajciar[1], and Iva Voleská[3]

[1] Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University in Prague, Albertov 6, 128 43 Prague 2, Czech Republic
[2] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám. 2, CZ-16610 Prague, Czech Republic
[3] Department of Physical Chemistry, Faculty of Chemical Technology, University of Pardubice, Nám. Čs. Legií 565, CZ-53210 Pardubice, Czech Republic

Our research is focused on the theoretical investigation of physical and chemical properties, including catalytic activities, of various microporous materials such as zeolites or metal-organic frameworks (MOF's). It is our goal to increase our understanding of these materials at the atomistic level, focusing on the relation between material structure and activity, in particular. Our computational work is done in close collaboration with experimental groups, e.g., Doc. Bulánek (University of Pardubice), Prof. Čejka (Heyrovsky Institute of Physical Chemistry), Prof. Areán (Spain).

For a reliable representation of microporous materials it is required that the adopted model accounts for all relevant interactions in the system. Microporous crystalline materials, like zoelites or MOF's, are typically modeled within the framework of the periodic density functional theory (DFT) calculations. We usually employ the GGA type exchange-correlation functional (PBE) and PAW approach implemented in VASP program package. This model provides reliable description of the material itself; however, for the description of the interaction of material with adsorbed molecule the DFT calculations suffer from the fact that weak inter-molecular interactions are not adequately described. Therefore, we have developed the DFT/CC method [1] that accounts for weak interactions and we use this for the description of interactions between molecules and microporous materials.

Due to the large unit cell size of common zeolites and MOF's (typically 2000–5000 Å [3]) periodic DFT calculations are demanding on computational resources. In addition to our own computational resources (located in IOCB and newly also in the Department of Physical and Macromolecular Chemistry) we greatly acknowledge the computational resources provided by MetaCentrum without which we would not be able to finish as many project as we did. The following computer programs have been used for this project: VASP, Molpro, Turbomol, and Gaussian.

The main results of our group obtained with the help of MetaCentrum computational resources are summarized in the following text; this work has resulted in 9 contributions to the leading journals in the field of physical chemistry published or submitted in 2009. Note that some of the papers appearing in 2009 were actually computed during 2008. Our work has been supported by following grants: GA ČR 203/09/0143 (2009–2012), 7. RP EU NMP-LA-2009-228862 (2009–2013), and MSMT LC512 (2005–2010).

Interaction of probe molecules with cationic sites (both extra-framework metal cations and Brønsted acid sites were investigated) in molecular sieves was investigated. Stabilities of adsorption complex and vibrational dynamics of adsorbed molecules were investigated. The focus has been on the characterization of adsorption complexes with a probe molecule (CO), with green house gases (CO_2) and with H_2 molecule. The following topics have been investigated:

1. Vibrational dynamics of Brønsted OH groups in H-FER and their interaction with CO and N_2 molecules. [2] A new model for precise calculations of stretching frequency of Brønsted OH groups in zeolites has been proposed and tested. This model is based on DFT optimization of system geometry (using periodic DFT) and subsequent use of ω/r correlation employing the precise frequencies obtained at the CCSD(T) level. Calculated frequencies for free Brønsted groups in zeolite as well as for those involved in the intermolecular complexes with adsorbed CO or N_2 molecules are in very good agreement with experimental data. A follow-up study of the acid strength of Brønsted acid sites in zeolites for zeolites with increasing Al (and H^+) content was also performed. [3]
2. The structure and stability of nitrosyl complexes formed in Cu-FER zeolite were investigated using a periodic DFT model. [4] The reliability of both DFT methods and cluster models when describing Cu^+ interaction with NO molecules was examined. The relative stabilities of mononitrosyl complexes on various Cu^+ sites in Cu-FER are governed by the deformation energy of the particular site. Three types of dinitrosyl complexes with different coordination on the Cu^+ cation were identified: (i) four-fold tetrahedral, (ii) four-fold square-planar and (iii) three-fold trigonal-planar complexes. The most stable dinitrosyl complex, formed when the two NO molecules interact with Cu^+ via the N atom, has a tetrahedral coordination on Cu^+. The cyclic adsorption complex, having a square-planar arrangement of ligands on Cu^+ and interaction via O atoms, is only about 10 kJ/mol less stable than the N-down dinitrosyl complex. This cyclic dinitrosyl complex is suggested to be the key intermediate in the deNOx process taking place in Cu-zeolites. The mechanism of the deNOx process has been investigated in a follow-up study [5] that shows that the rate-determining step is a re-attachment of N_2O to Cu-O specie.
3. Based on calculations employing the periodic DFT model, large isosteric heats of CO_2 adsorbed on alkali metal exchanged FER were attributed to the formation of the linearly bridged CO_2 adsorption complexes formed between a pair of Na^+ cations located in cationic positions. [6] The CO_2 adsorption

complexes formed on such dual cation sites are ca 10 kJ/mol more stable than the CO_2 adsorption complexes where CO_2 interacts only with a single Na cation.

4. Carbon monoxide adsorption on the LTA-type zeolite Ca-A was studied by a combination of variable-temperature infrared spectroscopy and computational methods involving periodic density functional calculations and correlation between stretching frequency and bond length of adsorbed CO (ω_{CO}/r_{CO} correlation). [7] Based on agreement among calculated and experimental results, the main adsorption species identified were bridged $Ca^{2+} \cdots CO \cdots Ca^{2+}$ complexes formed on dual cation sites constituted by a pair of nearby Ca^{2+} cations.
5. Adsorption complexes of CO in Mg-FER zeolites were investigated by a combination of periodic DFT and experimental IR methods. [8] Experimental data were interpreted based on a very good agreement between theoretical and experimental results. The most intense band in the spectra centered at 2205 cm^{-1} was assigned to monocarbonyl complexes formed on Mg^{2+} cations located in the six-member ring separating two adjacent FER cages (P6 site); it is concluded that Mg^{2+} cations exchange preferably into this site. An unusually high site-specificity of ν_{CO} observed for CO complexes formed in Mg-FER can be exploited in the investigation of preferable exchange sites of Mg^{2+} in zeolites and in the investigation of the localization of Al pairs in zeolites.
6. Interaction of H_2 with Ca^{2+} cations in zeolite Ca-LTA was investigated by a combination of periodic DFT method and experimental VTIR technique. [9] Calculations showed that adsorption complexes can be formed on two types of adsorption sites: Ca^{2+} in the center of six-member rings and Ca^{2+} in eight-member rings. Only the former adsorption complexes were experimentally observed in IR spectra; they are characterized by H–H stretching frequency of 4083 cm^{-1} and $\Delta H^0 = -12$ kJ/mol in agreement with calculated results.
7. Adsorption complexes on isolated Si(OH)Al Brønsted acid sites involve an adsorption enthalpy (ΔH^0) in the range of -33 to -36 kJ/mol, about half of which is due to the weak intermolecular interactions between CO_2 and zeolite framework. [10] Calculated adsorption enthalpies are in good agreement with the experimentally found value of $\Delta H^0 = -30$ kJ/mol.

Computational work outlined above has helped in understanding the localization and coordination of extra-framework cations in microporous materials. We believe that this work, performed in close collaboration with experimental groups, has increased our understanding of the relation between the cation coordination and its adsorption and catalytic activity. Such knowledge is essential for the design of materials with optimal properties for particular applications.

References

1. Bludsky, O.; Rubes, M.; Soldan, P.; Nachtigall, P. *J. Chem. Phys.* **2008**, *128*, 114102.

2. Nachtigall, P.; Bludsky, O.; Grajciar, L.; Nachtigallova, D.; Delgado, M. R.; Arean, C. O. *Phys. Chem. Chem. Phys.* **2009**, *11*, 791.
3. Grajciar, L.; Arean, C. O.; Pulido, A.; Nachtigall, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1497.
4. Pulido, A.; Nachtigall, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1447.
5. Pulido, A.; Nachtigall, P. *ChemCatChem* **2009**, *1*, 449.
6. Pulido, A.; Nachtigall, P.; Zukal, A.; Dominguez, I.; Cejka, J. *J. Phys. Chem. C* **2009**, *113*, 2928.
7. Pulido, A.; Nachtigall, P.; Delgado, M. R.; Arean, C. O. *Chemphyschem* **2009**, *10*, 1058.
8. Bulanek, R.; Voleska, I.; Ivanova, E.; Hadjiivanov, K.; Nachtigall, P. *J. Phys. Chem. C* **2009**, *113*, 11066.
9. Arean, C. O.; Palomino, G. T.; Carayol, M. R. L.; Pulido, A.; Rubes, M.; Bludsky, O.; Nachtigall, P. *Chem. Phys. Lett.* **2009**, *477*, 139.
10. Pulido, A.; Delgado, M. R.; Bludsky, O.; Rubes, M.; Nachtigall, P.; Arean, C. O. *Energy & Environmental Science* **2009**, *2*, 1187.

Simulation of Stress Wave Propagation

V. Pelikán, P. Hora, A. Machová, and O. Červená
pelikan@cdm.it.cas.cz

Institute of Thermomechanics, Dolejškova 5, 18000, Prague

1 Introduction

In this article we describe extensive 3D simulations of elastic wave propagation in the bcc iron, which are based on the molecular dynamics methods. Basically it is about proper behavior testing of the MD simulations by the comparison of simulated stress waves velocities with the continuum theory. This research is a continuation of the topics mentioned in [1] and [2]. Is described a behavior of the stress waves caused by surface impulsion and inner as well. In these simulations was used a N-body potential of Finnis-Sinclair type for the metals [4].

The tasks of this type have a physical sense only if they are not devalued by the stress waves reflections from the free surfaces of the investigated sample. For this reason the models must be large enough and the simulations on these models can be realized only with the massive application of the parallel programming techniques.

2 Description of the Experiment

All MD simulations were made on a two types of samples (see fig. 1):

a) the cube with the edge of 400 atoms (i.e., total 127 521 199 atoms),
b) the infinite plate, 200 atoms thick in the y axis direction.
The *infinity* in the other two directions was reached by the application of the periodic boundary conditions on the square sample with the edge of 600 atoms (i.e., 143 640 000 atoms).

In both cases it was absolutely homogeneous, ideally surface relaxed bcc crystal of the iron with the grid constant $a_0 = 2.8665$ Å at the temperature 0 K.

On these samples were progressively made four quite different classes of the tests:

1. all-area excitation in the y axis direction on the outer sides, orthogonal to the y axis,
2. local excitation in the y axis direction in the middle of the outer sides, orthogonal to the y axis,
3. all-area excitation in the y axis direction in the central plane, orthogonal to the y axis,
4. all-directions local excitation of the sample eight central atoms.

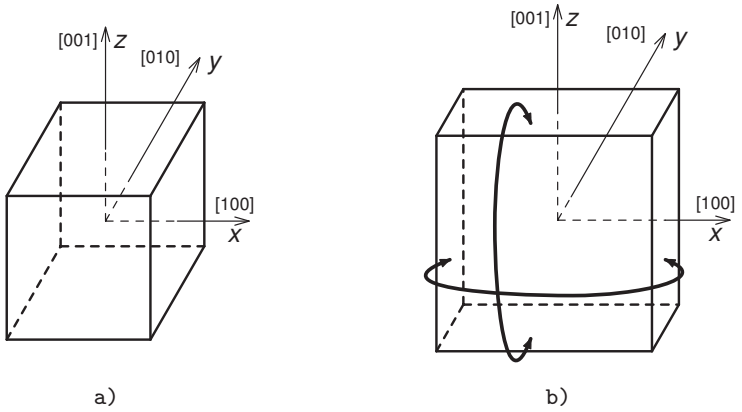


Fig. 1. The geometry of used samples.

Each test was thought in four modifications. Two of them were about the time dependence of the excitation: step excitation with the linear slope (fig. 2a) and step excitation by the Heaviside function (fig. 2b).

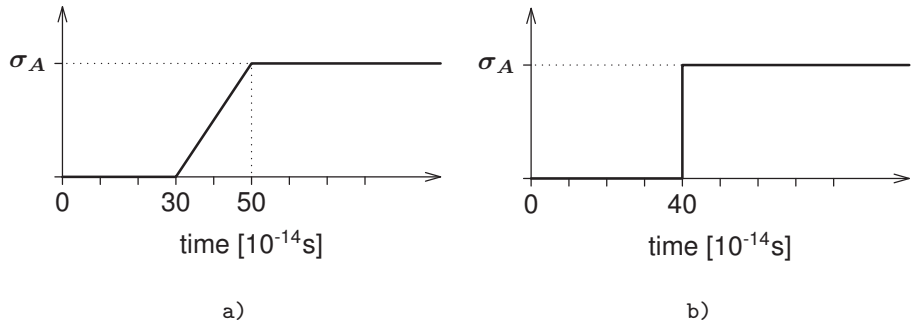


Fig. 2. The time dependences of the excitations.

The other two modifications were about the excitation orientation: tension (explosion) and compression (implosion).

In every simulation step was monitored the total energetic balance (the kinetic energy, the potential energy and the work of external forces) and the total number of the atom interactions as well. In every test the total number of the interactions was constant (889 775 586 with the cube and 1 003 320 000 with the infinite plate). It means that all the experiments dealt with the elastic deformations.

All the tests were made with the usage of the same excitation stress $\sigma_A =$ 1.35 GPa.

In all cases was used the same simulation step $1 \cdot 10^{-14}$ s. The tests were made between the steps 0 and 1000 for the cube and between the steps 0 and 600 for the infinite plate.

These calculations were made on the cluster MINOS of the University of West Bohemia in Pilsen. The implementation of one simulation step takes in average 54 seconds for the cube case and 78 seconds for the infinite plate case.

3 Wave Propagation in BCC Iron

In this part is outlined the behavior of stress waves in infinite anisotropic media. The field equations of elastodynamics [3]:

$$\begin{aligned}
&\text{equations of motion:} && \nabla \cdot \mathbf{T} = \rho \ddot{\mathbf{u}} \leftrightarrow T_{ij,j} = \rho \ddot{u}_i \\
&\text{constitutive relations:} && \mathbf{T} = \mathbf{C} \cdot \mathbf{S} \leftrightarrow T_{ij} = C_{ijkl} S_{kl} \\
&\text{strain-displacement rel.:} && \mathbf{S} = \tfrac{1}{2}\left\{\nabla \mathbf{u} + (\nabla \mathbf{u})^{\mathbf{T}}\right\} \leftrightarrow S_{kl} = \tfrac{1}{2}\left\{u_{k,l} + u_{l,k}\right\}
\end{aligned} \tag{1}$$

where $\mathbf{T}$ is the stress 2-tensor field, $\mathbf{S}$ is the strain 2-tensor field, $\mathbf{C}$ is the elasticity 4-tensor, and $\mathbf{u}$ is the displacement vector field, and ρ is the material density.

The traction across any surface with normal $\mathbf{n}$ is given by: $\mathbf{T} \cdot \mathbf{n} \leftrightarrow T_{ij} n_j$. In general, tractions and displacements must be continuous across boundaries.

From angular momentum balance considerations of non-polar media, we know that the stress tensor $\mathbf{T}$ is symmetric, ie: $T_{ij} = T_{ji}$. From its definition, the strain tensor $\mathbf{S}$ is symmetric as well: $S_{kl} = S_{lk}$. It follows therefore, that the elasticity 4-tensor has the following minor symmetries: $C_{ijkl} = C_{jikl}$; $C_{ijkl} = C_{ijlk}$. From thermodynamic considerations, the elasticity 4-tensor also has the following major symmetry: $C_{ijkl} = C_{klij}$. In view of these symmetries, there are at most 21 independent elastic stiffness constants for the most anisotropic material. With increasing levels of material symmetry, the number of independent elastic stiffness constants decreases, with only 3 for cubic crystals, and only 2 for isotropic materials.

For computational purposes, it is often simpler to consider the six independent components of the stress and strain tensors stacked up as six-dimensional column vectors:

$$\begin{aligned}
\mathbf{T} &= [T_{11}\ T_{22}\ T_{33}\ \ T_{23} = T_{32}\ \ \ T_{31} = T_{13}\ \ \ T_{12} = T_{21}]^T, \\
\mathbf{S} &= [S_{11}\ S_{22}\ S_{33}\ 2S_{23} = 2S_{32}\ 2S_{31} = 2S_{13}\ 2S_{12} = 2S_{21}]^T,
\end{aligned}$$

where we recognize that we are using the engineering definition rather than the tensorial definition of shear strains (hence the factor two in some of the above). We will use capital subscripts to remind us that we are using contracted notation.

3.1 Normal Modes in Unbounded Media

Let us first consider a homogenous unbounded linear elastic anisotropic medium and figure out whether we can propagate plane harmonic waves in these. In an

unbounded medium we do not have to worry about boundary conditions and so we just have to satisfy the field equations above. Let us seek plane harmonic waves given by the following displacement field:

$$\mathbf{u}(\mathbf{r},t) = \mathbf{U}\exp\left\{ik(\mathbf{l}\cdot\mathbf{r} - vt)\right\}, \tag{2}$$

where $\mathbf{U}$ is the displacement amplitude vector, $\mathbf{l}$ is a unit vector along the propagation direction of the wave, $\mathbf{r} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$ is the position vector, $k = 2\pi/\lambda$ is the wavenumber, λ is the wavelength, and v is the phase velocity of the wave. The angular frequency of the harmonic wave is related to the wavenumber and velocity through: $\omega = kv = 2\pi v/\lambda$. The $\mathbf{e}_i$'s are unit vectors along the 1,2 and 3 directions, and we will use the symbol i for the imaginary number.

Putting the above in the field equations (1), we get:

$$C_{ijkl}l_l l_j k^2 U_k = \rho\omega^2 U_i,$$

which we can re-write as:

$$\left\{\Gamma_{ik} - \delta_{ik}\rho v^2\right\} U_k = 0, \tag{3}$$

where

$$\Gamma_{ik} = C_{ijkl}l_j l_l \tag{4}$$

is called the Christoffel matrix, and (3) is called the Christoffel equation. We note that (3) is an eigenvalue problem. It is independent of the frequency (and there are no boundary conditions to be satisfied here), therefore we expect non-dispersive plane waves in unbounded anisotropic media.

We can easily see that the Christoffel matrix is symmetric, and under some nonrestrictive conditions on the elastic stiffnesses, we can show that it is positivedefinite. From the spectral theorem for positive-definite symmetric matrices, therefore, we are guaranteed three positive, real eigenvalues for $\mathbf{\Gamma}$. This is good news because it means that the wave phase velocities v, which are just the square root of these eigenvalues divided by the density, are guaranteed to be real and so will represent propagating modes. The eigenvalues are obtained by solving the secular equation:

$$\det\left\{\Gamma_{ik} - \delta_{ik}\rho v^2\right\} = 0. \tag{5}$$

Again from the spectral theorem, corresponding to each of these eigenvalues $v^{(i)}$ there is at least one real eigenvector and, furthermore, we can always find three orthogonal eigenvectors which we shall denote by: $\mathbf{U}^{(i)}$. Therefore, what we find is that in any homogeneous anisotropic material, we can always propagate three types of plane harmonic waves along any chosen propagation direction $\mathbf{l}$. In general, these three waves will have different phase velocities $v^{(i)}$ and the corresponding particle displacement vectors $\mathbf{U}^{(i)}$ will be mutually orthogonal. Each of these modes is called a normal mode of propagation. We call the direction of the displacement vector, the polarization direction of the wave. Note, however, that the particle displacement vector need not be parallel or perpendicular to the

propagation direction in general. If it so happens that the polarization direction of a wave is parallel to the propagation direction, we call it a pure longitudinal wave. Waves with polarization direction normal to the propagation direction are pure shear waves. If the polarization directions are neither parallel nor perpendicular to the propagation direction, the waves are neither pure longitudinal nor pure shear. In such cases, the mode whose polarization makes the smallest angle to the propagation direction is called a quasilongitudinal wave, and the other two are called quasi-shear waves.

For a general anisotropic material, it is not easy to simplify the secular equation analytically for arbitrary propagation directions, even though it may be possible to obtain analytically tractable expressions for special cases of propagation along certain material symmetry directions. In general, however, one seeks numerical solutions for the anisotropic problem.

Note that the Christoffel matrix is most readily calculated using contracted notation by defining a 3×6 propagation direction–cosine matrix:

$$\mathbf{L} = \begin{bmatrix} l_1 & 0 & 0 \\ 0 & l_2 & 0 \\ 0 & 0 & l_3 \\ 0 & l_3 & l_2 \\ l_3 & 0 & l_1 \\ l_2 & l_1 & 0 \end{bmatrix}.$$

Then, by direct expansion you can show that $\mathbf{\Gamma} = \mathbf{L}^T\mathbf{C}\mathbf{L}$, i.e.,

$$\mathbf{\Gamma} = \begin{bmatrix} l_1 & 0 & 0 & 0 & l_3 & l_2 \\ 0 & l_2 & 0 & l_3 & 0 & l_1 \\ 0 & 0 & l_3 & l_2 & l_1 & 0 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56} \\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66} \end{bmatrix} \begin{bmatrix} l_1 & 0 & 0 \\ 0 & l_2 & 0 \\ 0 & 0 & l_3 \\ 0 & l_3 & l_2 \\ l_3 & 0 & l_1 \\ l_2 & l_1 & 0 \end{bmatrix}.$$

Corresponding to each propagation direction in an anisotropic material, the eigenvalue problem (5) will in general have three (normal mode) solutions with three phase velocities with mutually orthogonal polarizations. We can plot these phase velocities as a function of propagation direction to get three three-dimensional surfaces. If we plot:

$$\mathbf{V}^{(i)}(\mathbf{l}) = v^{(i)}\mathbf{l},$$

we get what are called three phase-velocity surfaces, more commonly called normal surfaces. For an isotropic materials, these are just three spheres, with the one corresponding to the longitudinal mode being on the outside, and the two spheres corresponding to the transverse modes being degenerate.

It is actually more useful to plot slowness surfaces in so-called k-space with axes: kl_i/ω (which is the reciprocal of velocity, hence *slowness*):

$$\mathbf{S}^{(i)}(\mathbf{l}) = \frac{\mathbf{l}}{v^{(i)}}.$$

There are again three *sheets* representing the three slowness surfaces. Any direction in this space represents the propagation direction, and the distance to the slowness surface gives the reciprocal of the phase velocity of the associated mode in this direction. For an isotropic material these are spheres, the inner one now representing the longitudinal mode, and the two shear slowness surfaces being degenerate. The situation is more complicated for anisotropic materials.

Bulk wave velocities along a cubic crystalline axes [100], [110] and [111] in bcc iron are summarized at the tab. 1. These velocities calculated from the continuum theory were absolutely satisfactorily verified by the MD simulations.

In fig. 3 is shown a cross section of the slowness surfaces by the plane $x-z$. Blue color shows the fastest quasi-longitudinal mode, green represents pure shear mode and red quasi-shear mode.

Velocity	$\langle 100 \rangle$	$\langle 110 \rangle$	$\langle 111 \rangle$
c_L $[\mathrm{m\,s^{-1}}]$	5550	6266	6487
c_T $[\mathrm{m\,s^{-1}}]$	3832	2494	3007
c_t $[\mathrm{m\,s^{-1}}]$	3832	3832	3007

Table 1. Velocities of the wave propagation in a significant directions.

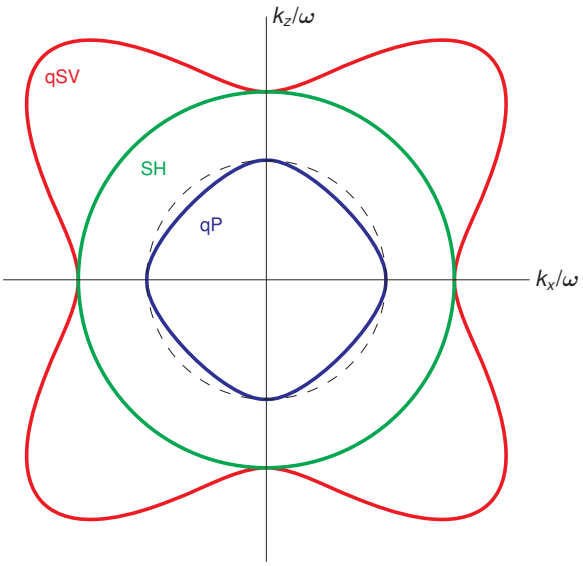


Fig. 3. A cross section of the slowness surfaces by the plane $x-z$.

4 Results and Discussion

For the illustration of reached results are in this chapter mentioned maps of the velocity magnitude for the chosen classes of the tests. By the imaging of the velocity magnitude, which are proportional to the kinetic energy, become the wave head of the stress waves best visible.

Regarding to the cubic symmetry of the thought materials it is possible to draw only the partial areas (see fig. 4) of the whole sample.

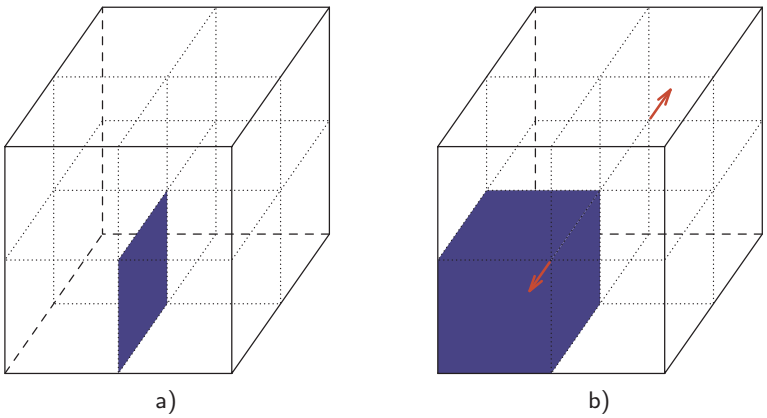


Fig. 4. The scheme of shown details used in the following figures.

4.1 Fast All-area Tension on the Outer Side of the Cubic Sample

In the fig. 5 is displayed a distribution of the velocity magnitude for a fast (see fig.2b) all-area tension on the outer side of the cubic sample at the times: 2, 4, 6 and 8 ps from the beginning of the simulation, i.e., in the simulation steps 200, 400, 600 and 800. In the figure is displayed left lower quadrant of the central crystal layer (see fig. 4a) along main direction of the stress wave propagation. In the figure for t=8 ps is drawn a zoom, its detail, together with the waves polarization, is displayed on the bottom part of the figure. Data shown along the axes of this figure gives (for easier orientation) an atomic crystal layer order.

In the fig. 5 it is possible to monitor the main wave front of longitudinal wave, the wave front of shear wave created from the bottom side and the wave front of conical wave. On the detail of the figure it is possible to clearly recognize longitudinal wave front (with the same polarization as a direction of propagation) and creating shear and conical (von Schmidt) waves as well (with the polarization upright to the direction of propagation).

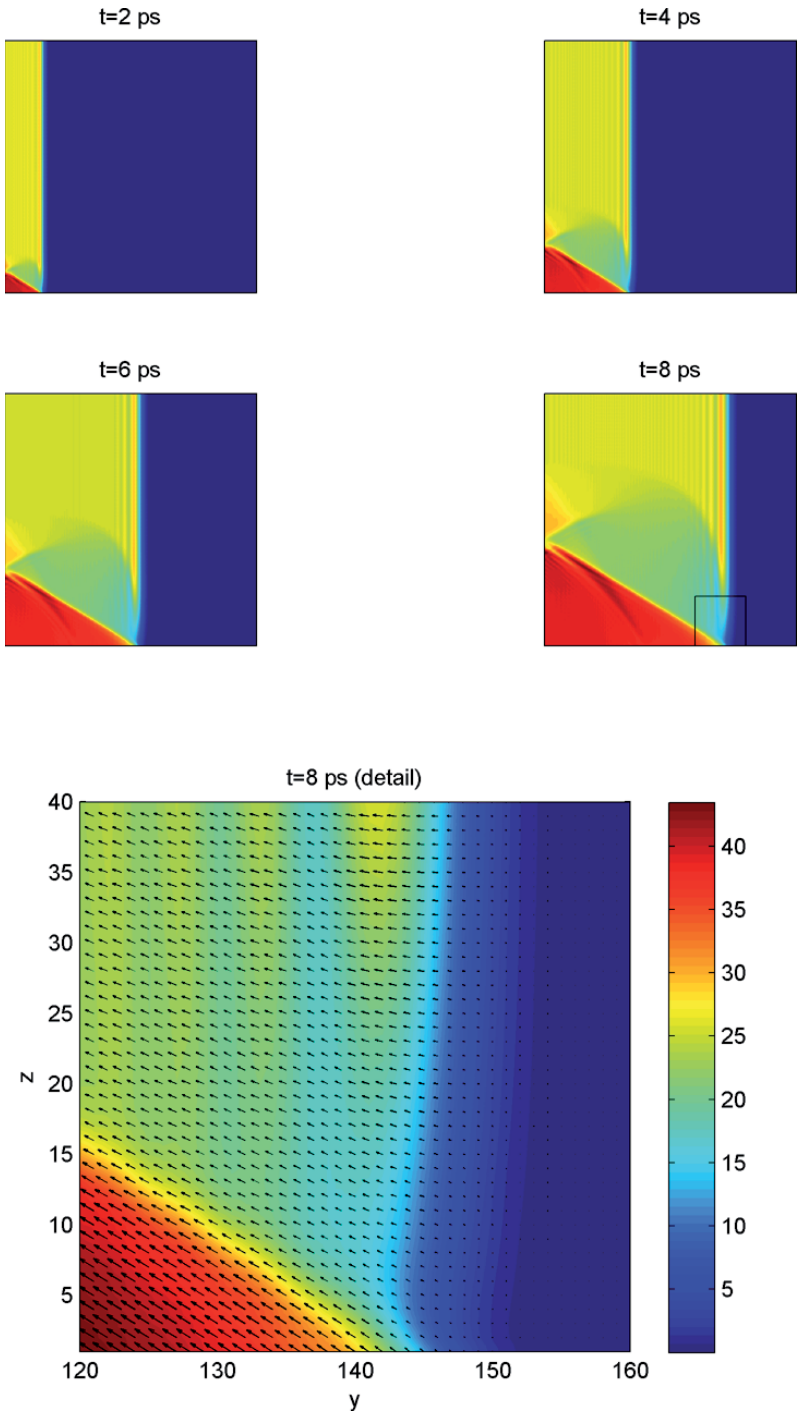


Fig. 5. Fast all-area tension on the outer side of the cubic sample.

4.2 Fast Local Tension in the Middle of Sample Surface

The complex view of the simulation state at the 10 ps gives a fig. 6. It is a part of the cubic sample, as it is schematically shown on the fig. 4b). Once again it is a imaging of the particular atoms velocity magnitudes. In to the figure are drawn a wave fronts of the particular group velocities.

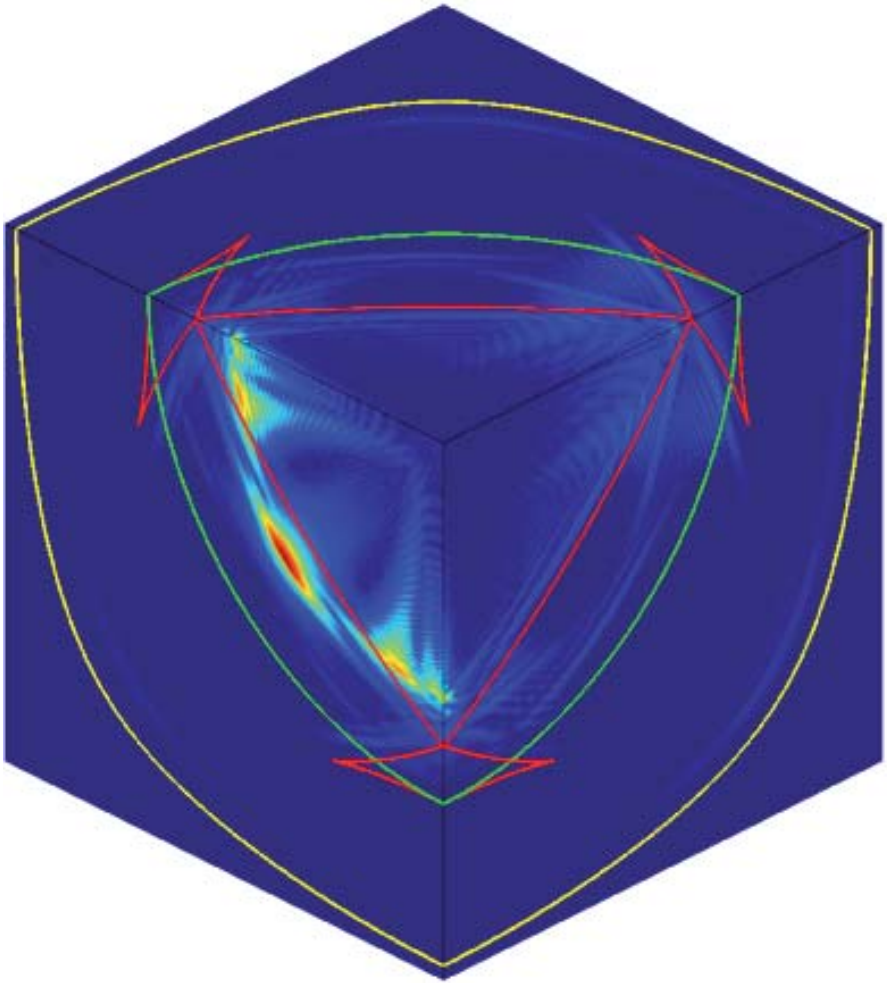


Fig. 6. Fast local tension in the center of sample surface at the time 10 ps.

5 Used Program

The molecular dynamics calculations were performed using the own code in Fortran90 by means of MPI library same as in [5] and [6].

Acknowledgement

The work was supported by the Institute Research Plan AV0Z20760514 and by the grants GA CR No 101/07/0789 and GA AS CR KJB200760802. The access to the MetaCentrum clusters provided under the research intent MSM6383917201 is highly appreciated.

References

1. Pelikán, V.; Hora, P.; Machová, A.:
   Parallel programming in molecular dynamics simulation. [in czech]
   In Computational mechanics 2003, ZČU Pilsen, Czech Republic, pp 351–358
2. Pelikán, V.; Hora, P.; Machová, A.:
   Large model simulations in molecular dynamics problems. [in czech]
   In Computational mechanics 2004, ZČU Pilsen, Czech Republic, pp 397–404
3. Brdička, M.; Samek, L.; Sopko, B.:
   Continuum mechanics. [in czech]
   Academia, Praha, 2000
4. Machová, A.; Ackland, G.J.:
   Dynamic overshoot in α-iron by atomistic simulations.
   Modelling Simul. Mater. Sci. Eng., Vol. 6., 1998, pp 521–542
5. Pelikán, V.; Hora, P.; Machová, A.; Červená, O.:
   The simulation of a wave propagation in a bcc iron crystal with a crack.
   Applied and Computational Mechanics. Vol. 1, 2007, pp 225–232, ISSN 1802-680X.
6. Pelikán, V.; Hora, P.; Machová, A.; Červená, O.:
   The influence of the excitation pulse shape on the stress wave propagation in a bcc iron crystal.
   Applied and Computational Mechanics. Vol.2, 2008, pp 323–334, ISSN 1802-680X

Summary of Recognition and Searching in the Czech Holocaust Testimonies With Relation to the MALACH Project

Aleš Pražák, Josef Psutka, Pavel Ircing, Josef V. Psutka, and Jan Švec
{aprazak,psutka,ircing,psutka_j,svec}@kky.zcu.cz

Department of Cybernetics
University of West Bohemia
306 14, Plzen, Czech Republic*

1 Introduction

The whole story began in 1994 when, after releasing "Schindler's List", Steven Spielberg was approached by many survivors of the World War II Holocaust who wanted him to listen to their stories of the Holocaust. Inspired by these request, Spielberg decided to start the Survivors of the Shoah Visual History Foundation (VHF) so that as many survivors as possible could tell their stories and have them saved. In his original vision, he wanted the VHF (which later became USC Shoah Foundation Institute of Visual History and Education) to perform several tasks, including collecting and preserving the Holocaust survivors' testimonies and cataloging those testimonies to make them accessible.

The "collecting" part of the mission has been completed in 1997, resulting into what is believed to be the largest collection of digitized oral history interviews on a single topic: almost 52,000 testimonies (interviews) in 32 languages, a total of 116,000 hours of video. About half of the collection is in English, and about 4,000 of English interviews (approx. 10,000 hours, i.e. 8% of the entire archive) have been extensively annotated by subject-matter experts (subdivided into topically coherent segments, equipped with a three-sentence summary and indexed with keywords selected from a pre-defined thesaurus). This annotation effort alone required approximately 150,000 hours (75 person-years) and proved that a manual cataloging of the entire archive is unfeasible.

This finding prompted the proposal of the MALACH (Multilingual Access to Large Spoken Archives) project [1] whose aim was to use automatic speech recognition (ASR) and information retrieval techniques for access to the archive of spontaneous, accented and highly emotional speech of holocaust survivors and thus circumvent the need for manual annotation and cataloging. The MALACH project was carried out between 2002-2007 (in cooperation with the VHF, IBM,

* This work has been funded by the NSF (USA) under the Information Technology Research (ITR) program, NSF IIS Award No. 0122466, and by the Ministry of Education of the Czech Republic under projects MSM235200004 and LN00A063 and by the project AVČR No. 1QS101470516.

JHU Baltimore, University of Maryland, CU in Prague and UWB in Pilsen) with the financial support of the NSF (US National Science Foundation). The goal of our Department of Cybernetics at the University of West Bohemia was originally only to prepare the ASR training data for several Central and Eastern European languages (namely Czech, Slovak, Russian, Polish and Hungarian); over the course of the project, we gradually became involved in essentially all the research areas, at least for the Czech language.

Although a great deal of work was done not all objectives were fully fulfilled [2, 3] in the MALACH project. We have continued in the effort to fulfill the MALACH project visions even after the end of the project itself. The state-of-the-art techniques of acoustic and language modeling were applied to build up a Large Vocabulary Continuous Speech Recognition (LVCSR) system that allows very fast access to the Czech part of archive allowing to detect segments of interviews containing pronounced words, clusters of words presented in pre-defined time intervals, and also words that were not known to the LVCSR system.

2 Automatic Speech Recognition

2.1 Data Preparation

Testimonies of the Czech holocaust corpus as well as other languages are stored at the Shoah Foundation Institute (SFI) at the University of Southern California as video interviews. The speech of each interview participant (the interviewer and interviewee) was usually recorded in quiet rooms via lapel microphones that recorded speech on separate channels. The speech quality is however often quite poor from the ASR point of view. There is frequent whispered or emotional speech along with many disfluencies and non-speech events as crying, laughter, etc. The quality and fluency of speech was often affected by the age of speakers. The age of the oldest survivor was 94; the average age of all speakers was 75 years. The speaking rate was also quite variable, ranging from 64 to 173 words per minute, with an average rate of 113. Using of many colloquial (non-grammatical) words by speakers is undesirable too.

The average length of a Czech testimony is 1.9 hours. Each testimony was divided and stored at SFI as half-hour parts in MPEG-1 video files. For the further processing the audio streams were extracted. The audio track was stored at 128 kb/sec stream in 16-bit resolution and 22.05 kHz sampling rate.

For preparing the acoustics, only 15 minute segment per each speaker was transcribed using the speech annotation software Transcriber. For testing, entire testmonies were transcribed. The transcribers also marked several non-speech events. The ratio between males and females in terms of the number of speakers and the amount of transcribed speech is shown in Table 1.

2.2 Acoustic Modeling

The acoustic models in our system are based on the state-of-the-art Hidden Markov Models (HMM) architecture. Standard 3-state left-to-right models with

	Training data		Development data		Testing data	
	Male	Female	Male	Female	Male	Female
Speakers	168	232	5	5	5	5
Hours transcribed	42	58	10	10	10	10

Table 1. Transcribed speech data

a mixture of multiple Gaussians in each state are used. Triphone dependencies (including the cross-word ones) are taken into account. As follows from the Table 1, the models are trained with 100 hours of transcribed speech, using the HTK toolkit [4].

The data was parameterized as 15 dimensional PLP cepstral features including their delta and delta-delta derivatives [5] . These features were computed at the rate of 100 frames per second. Cepstral mean subtraction was applied per speaker. The number of clustered states and number of Gaussian mixtures per state was optimized using development data and had more than 6k states and 16 mixtures per state (almost 100k Gaussians). A silence model was trained by borrowing Gaussians from all non-speech HMMs in proportion to their state and mixture occupancies. The resulting silence model contained 128 mixtures per state and was found to be useful in rejecting non-speech events during recognition.

Speaker-adaptive models (SAT) were trained via fMLLR, for each training speaker. After fMLLR transforms for training speakers were computed against the original speaker-independent model, the original model was then re-estimated using the affinely transformed features. This process was repeated few times to converge. The DT model was developed from SAT model via four training iterations based on MMI-FD objective function [6]. Because the speaker identity is available, it can be used to improve the recognition. All training data were split to three clusters (male-speakers female-speakers and interviewer) for DT adaptation. This DT adaptation was done via two iterations of DT-MAP on SAT-DT acoustic model [7].

2.3 Language Modeling

When building a language model for a specific ASR task, the first choice that has to be made is the selection of appropriate training data. Our initial experiments [8] revealed that even though the transcripts of the acoustic model training data are constitute a rather small corpus from the language modeling point of view (approx. 1.1M tokens), there are by far more suitable for the task than much larger, but "out-of-domain" text corpora (comprising, for example, newspaper articles). However, if a more sophisticated technique is used for an additional data selection, it's possible to further improve the recognition performance. Therefore we investigated the possibility of using automatic methods to select sentences from the Czech National Corpus CNC (approx. 400M tokens) that are similar in language usage, lexicon and style to the sentences in the

training data transcriptions (see [9] for details). This in-domain selection from CNC contains 82MB of text (16M tokens).

During the preparation of the training corpus, the annotators were instructed to use the orthographic transcription of colloquial words (i.e., not to "standardize" them artificially) in order to keep the transcriptions close to what was actually said. This procedure is very beneficial for the acoustic modeling as the resulting transcription is very close to the actual phonetic realization of the word in question. On the other hand, it in some cases results into an explosion of different surface representations of the same standard word form, which harms the quality of the language models as the already sparse text training data become even sparser.

In order to exploit both the advantage of close orthographic transcription of colloquial words in acoustic modeling and the benefit of standard word forms in language modeling, we decided to "standardize" the pronunciation lexicon. We went through the lexicon built from the original (orthographic) transcriptions and added a corresponding standard form to each colloquial word form, creating a new 3-column lexicon. Such approach allows us to automatically "standardize" the transcripts and consequently use the resulting parallel corpora (original and standardized) for counting of the relative frequencies of the individual colloquial variants. The column with standard word forms then constitutes the lexicon of the language model (which is of course estimated using the standardized transcripts). The "colloquial" column is used for acoustic modeling only. The standardization has a positive effect not only on the lexicon size (which is obvious) but also on the test set Out-Of-Vocabulary (OOV) rate (i.e., the percentage of tokens from the test set that are not present in the lexicon) and, most importantly, on the speech recognition performance.

Finally, an interpolated language model has been created with the ratio 2:1 (transcriptions to the CNC). The resulting trigram language model with modified Kneser-Ney smoothing contains 252k words (308k phonetical variants). Language models were estimated using the SRI Language Modeling Toolkit (SRILM) [10].

2.4 Word and Phoneme Lattices

Due to the stereo speech signal (the interviewer on one channel and interviewee on the other) and to the conversation character of the data the special algorithm was introduced in order to reduce the huge amount of speech data. Only parts, where at least one speaker was talking, were further processed. In the case, where interviewer and interviewee were cross-talking, both channels were recognized separately. This task was quite challenging, because there occurred echoes even though the speakers had lapel microphones. During recording, the speech of interviewer and interviewee mixed together so that each speaker was recorded in both channels, only with different level of energy.

The LVCSR system was designed to work in two passes. In the first pass, clustered DT adapted acoustic models was automatically adapted to each of

550 speakers, using a bigram language model. This automatic iterative fMLLR+MAP adaptation [11] used only speech segments with posterior probabilities over 0.99. Word lattices were then generated based on information about word transitions performed during the second pass recognition by LVCSR system. The lattices were built up retroactively, from the last recognized word by adding other most probable word hypotheses (alternatives) of the recognized words according to the desired depth (number of concurrent hypotheses) of word lattices. For searching through word lattices, the posterior probability computed by forward-backward algorithm was assigned to each hypothesis in the word lattice. Normalized acoustic likelihoods and a trigram language model were used during the lattices computation. Due to the effect of the segmentation of the word graph [12], posterior probabilities for different hypotheses of the same word were summed.

Phoneme lattices were generated in the same manner, based on information about phoneme transitions performed during the recognition by phoneme recognizer without use of any language model. This recognizer was built for each speaker on its acoustic model adapted by the first pass of LVCSR system.

The parameters for the LVCSR system and phoneme recognizer were optimized on the development data. The recognition accuracy reached 71.44 % at the word level and 70.38 % at the phoneme level. These results were enumerated on the testing data comprising 63,205 words with 2.39 % OOV words.

3 MetaCentrum Involvement

The whole Czech part of the archive (550 testimonies of Czech Holocaust survivors) was processed using MetaCentrum computation facilities at the end of year 2009. The acoustic tracks of the testimonies were split to the 1,109 parts (the interviewer and interviewee for each testimony) and recognized in two passes as described above. 2,218 tasks, 2 processors per each, were computed at MetaCentrum computation facilities in about 7 days (during the Yuletide). We have used recognizer developed at the Department of Cybernetics at the University of West Bohemia optimized for multi-processor systems.

4 Searching

The resulting application for searching in the archive of the Czech holocaust testimonies searches only keywords and/or key-phrases but is extremely fast and fully interactive. The system using SQL database that is able to store the huge number of records. The searching algorithm depends on the searched word. First, a lemma of the searched word is generated. Then if the lemma is found in the vocabulary the word lattice search is performed. It is also possible to search for all possible forms (from the lemma) of the searched vocabulary word. This behavior can be optionally disabled if the user wants to search only the exact word form. If the searched word is an out-of-vocabulary word the phonetic lattice search is performed. The phonetic transcription of the word is generated

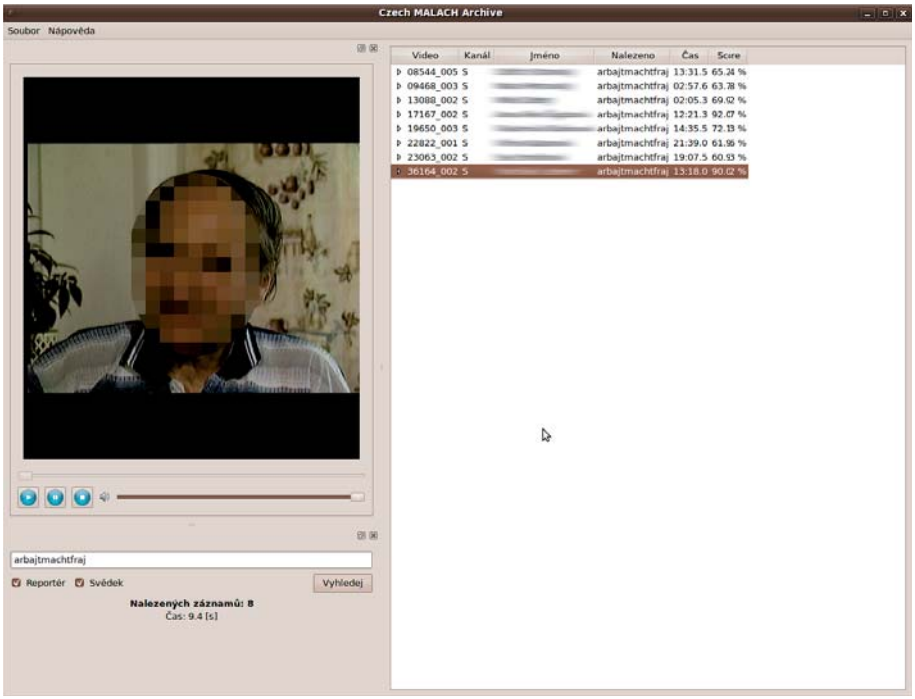


Fig. 1. Application for searching in the archive of the Czech holocaust testimonies.

and the overlapping trigrams of phones are computed. Then the database query selects the records corresponding to these trigrams. The results are grouped by a corresponding audio track and ordered by the time. If the searched word occurs in the audio track at the given time there must be a cluster of the trigrams. The algorithm we use does not strictly require the presence of all trigrams from the searched word. The score of the word occurrence is computed from the score of indexed trigrams and the total number of found trigrams.

The time required to search through the whole archive strongly depends on the searched word itself, mainly on the number of occurrences. For in-vocabulary words the time needed to search the whole archive is typically between 5 and 10 seconds. The out-of-vocabulary words are typically searched between 30 and 60 seconds.

5 Publications Prepared Using MetaCentrum Resources

1. Aleš Pražák, Pavel Ircing, and Luděk Müller. Language Model Adaptation Using Different Class-Based Models. In *Proceedings of the 12th International Conference on Speech and Computer (SPECOM 2007).*

References

1. Byrne, W. and Doerman, D. and Franz, M. and Gustman, S. and Hajič, J. and Oard, D. and Picheny, M. and Psutka, J. and Ramabhadran, B. and Soergel, D. and Ward, T. and Zhu, W. : Automatic recognition of spontaneous speech for access to multilingual oral history archives . IEEE transactions on speech and audio processing, vol. 4, p. 420-435, 2004.
2. Psutka, J. and Ircing, P. and Psutka, J.V. and Hajič, J. and Byrne, W. and Mírovský, J. : Automatic transcription of Czech, Russian and Slovak spontaneous speech in the MALACH project . Interspeech Lisboa 2005, p. 1349-1352, ISCA, Bonn, 2005.
3. Psutka, J. and Ircing, P. and Hajič, J. and Radová, V. and Psutka, J. and Byrne, W. and Gustman, S. : Issues in annotation of the Czech spontaneous speech corpus in the MALACH project . Fourth international conference on language resources and evaluation, p. 607-610, European Language Resources Association, Lisbon, 2004.
4. Young, S. J., et al., "The HTK Book", Entropic Inc., 1999.
5. Hermansky, H.: Perceptual linear predictive (PLP) analysis of speech. J. Acoustic. Soc. Am.87, 1990.
6. Povey D.: Discriminative Training for Large Vocabulary Speech Recognition. Ph.D. thesis, Cambridge University, Department of Engineering, 2003.
7. Vaněk, J. and Psutka, J.V. and Zelinka, J. and Pražák, A. and Psutka, J. : Discriminative training of gender-dependent acoustic models. Lecture Notes in Artificial Intelligence, Lecture Notes in Artificial Intelligence, 5729, p. 331-338, Springer, Berlin , 2009.
8. J. Psutka, P. Ircing. J. V. Psutka, V. Radová, W. Byrne, J. Hajič, S. Gustman, B. Ramabhadran: Automatic Transcription of Czech Language Oral History in the MALACH Project: Resources and Initial Experiments. Proceedings of TSD 2002, Brno, 2002.
9. Psutka, J. and Ircing, P. and Psutka, J.V. and Radová, V. and Byrne, W. and Hajič, J. and Mírovský, J. and Gustman, S. : Large vocabulary ASR for spontaneous Czech in the MALACH project . EUROSPEECH 2003 PROCEEDINGS, p. 1821-1824, ISCA, Geneva, 2003.
10. Stolcke, A.: SRILM - An Extensible Language Modeling Toolkit. In: International Conference on Spoken Language Processing (ICSLP 2002), Denver, USA, 2002.
11. Zajíc, Z., Machlica, L., Müller, L.: Refinement Approach for Adaptation Based on Combination of MAP and fMLLR. Lecture Notes in Computer Science, vol. Volume 5729/2009, p. 274-281, 2009.
12. Wessel, F. and Schlüter, R. and Macherey, K. and Ney, H. : Confidence measures for large vocabulary continuous speech recognition . IEEE Transactions on Speech and Audio Processing, vol. 9, no. 3, IEEE, 2001.

Study of Edge Plasma Turbulence in Tokamaks

Jakub Seidl

Institute of Theoretical Physics, Faculty of Mathematics and Physics,
Charles University in Prague,
V Holešovičkách 2, 180 00 Prague 8–Holešovice, Czech Republic
jakub.seidl@email.cz

1 Introduction

Understanding the role of plasma turbulence in strong energy and particle transport across magnetic field lines in tokamak edge region is a problem of primary importance for future fusion devices. Cross-field fluxes that originate in the region of confined plasma carry outwards significant amounts of energy, thus featuring high risk of damaging tokamak first wall, divertor target and other plasma facing components. In the same time, impurities released by these components spread through the boundary region into main plasma, cooling it down by their radiation and decreasing the fusion rate due to dilution of the fuel.

It is generally agreed that the strong radial transport of plasma particles and energy, that is much larger than expected by taking into account collisional diffusion only, is closely connected with existence of turbulent structures called blobs, that were experimentally observed in edge region of various fusion devices. These blobs and the turbulence itself originate as a result of interchange instability at the vicinity of last closed flux surface on tokamak outer midplane and they consequently propagate into the region of open magnetic field-lines—scrape-off layer (SOL).

Behaviour of individual particles in such turbulent field may significantly differ depending on particle species and charge. Particles of the plasma itself, i.e. electrons and cold ions, may be successfully treated using drift approximation that reduces their cyclotron gyration leaving only drifts of the guiding center. On the other hand, orbits of heavier particles with Larmor radius not negligible, e.g., plasma impurities or warm ions, have to be treated more carefully taking into account also higher order terms or, as in our case, using the full equations of motion.

2 Numerical Modeling of Plasma Turbulence

A major part of our current effort is dedicated to simulations of the plasma turbulence itself. We use and further develop 2D electrostatic edge turbulent model ESEL [1]. ESEL solves self-consistent time evolution of reduced-fluid equations for plasma density, vorticity, temperature and potential. The simulated turbulence is driven by the interchange instability caused by non-homogeneous magnetic field and density gradient in the edge. Since parallel dynamics is considered

non-turbulent, the fluid equations are restricted to the 2D plane perpendicular to magnetic field-lines and dynamics along the field-lines is simplified by estimating characteristic rates of parallel damping for every quantity.

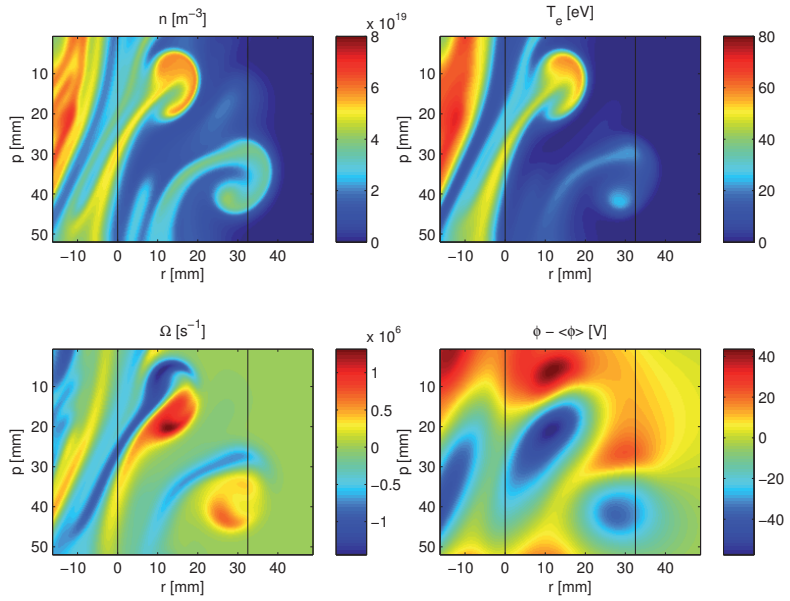


Fig. 1. Time snapshot of turbulent fields simulated by ESEL model showing characteristic "mushroom-like" shape of turbulent blobs in density n and temperature T_e. Snapshot of vorticity Ω reveals two counter-rotating lobes and potential perturbation $\phi - <\phi>$ demonstrates bipolar nature of the blob. Black vertical lines delimit region of SOL with confined plasma on its left side and wall shadow region (with strong parallel plasma losses) on the right.

The results of the model show presence of bipolar turbulent structures in SOL, consisting of two counter-rotating lobes (see Fig. 1). The structures carry significant amount of density and energy and move radially outwards from the region of confined plasma. In the past, ESEL model was successfully compared with experimental measurements on tokamak TCV in regimes with high parallel collisionality. However, when comparing the model with low-collisionality regimes on other tokamaks several disagreements were found. Therefore, the aim of this project is study and modification of parallel dynamics in ESEL model in order to get closer agreement with experimental measurements also for low-collisionality regimes. This will bring a better understanding of physics underlaying turbulent transport in tokamak edge region. First results comparing different closures of divergence of parallel current were already presented in [6].

3 Behaviour of Plasma Impurities in a Turbulent Field

ESEL model discussed in previous section is based on reduced fluid equations. This way, the dynamics is reduced to the lowest-order drifts and such approach is valid only for particles with Larmor radius r_L significantly smaller than characteristic size of turbulence structures. Therefore, we performed several simulations focused on investigation of behaviour of heavy and weakly charged plasma impurities in a field of background turbulent potential. Since density of impurities is significantly lower than density of plasma, we assumed that impurities do not influence the turbulence itself and test-particle approach was used.

Assuming simple periodic and stationary "egg-crate" model of the background potential, it has been shown earlier [2] that for sufficiently large ratio of particle mass and charge $R = m/q$ particle dynamics may be described as super-diffusion (Lévy walk). Such particles alternate periods of being trapped inside turbulent structures with jumps between them and high probability of long jumps then results in a strong particle transport.

In order to investigate behaviour of heavy particles in a more realistic conditions, we have replaced the "egg-crate" background potential by time-evolving 2D field of floating potential experimentally measured in edge region of tokamak CASTOR in a plane perpendicular to magnetic field-lines. Our simulations [3] then confirmed presence of super-diffusion, however, only in poloidal direction. The transport in radial direction did not transit to a super-diffusive state as a result of strong radial electric fields present in the measured potential. Strength of radial particle transport was found increasing with R, while the poloidal one was decreasing at the same time.

Lévy walk dynamics was found in both mentioned forms of turbulent plasma potential, nevertheless, both of these models included only the lowest spatial harmonics of the real potential field. Therefore, our current effort leads to exploitation of results of ESEL code as the background potential for movement of plasma impurities. We expect that its high temporal and spatial resolution will allow detailed investigation of mechanism of particle trapping and dragging by turbulent structures.

4 Used Programs

- *Application software installed in METACenter:* MATLAB
- *User installed programs:* ESEL [1], developed in Risø National Laboratory for Sustainable Energy, Technical University of Denmark, Denmark.

References

[1] Garcia, O.E. et al.: Interchange turbulence in the TCV scrape-off layer, Plasma Phys. Control. Fusion, 48, L1-L10, 2006.
[2] Krlín, L. et al.: Stochastic (E × B) diffusion of ions in a spatially periodical potential field, Plasma Phys. Control. Fusion, 41, 339-353, 1999.

Publications Prepared with the Use of MetaCenter

[3] Seidl J., Krlín L., Pánek R., Pavlo P., Stöckel J., Svoboda V.: Simulations of Anomalous Ion Diffusion in Experimentally Measured Turbulent Potential, Eur. J. Phys. D, 54, 399-407, 2009.
[4] Seidl J., Krlín L.: Interchange Driven Turbulence and Bellan Instability in Tokamak Scrape-off Layer, WDS'09 Proceedings of Contributed Papers: Part III Physics, Prague, Matfyzpress, 64-70, 2009.

Contributions at International Conferences

[5] Seidl, J., Krlín, L., Pánek, R., Pavlo, P., Stöckel, J., Svoboda V.: Simulations of Anomalous Ion Diffusion in Experimentally Measured Turbulent Potential, 23rd Symposium on Plasma Physics and Technology, 2008, Prague.
[6] Nielsen, A.H., Seidl, J., Horacek, J., Garcia, O.E., Naulin, V., Rasmussen, J.J.: Investigation into parallel dynamics in the ESEL model, 2nd EFDA Transport Topical Group Meeting, 2009, Culham.

Thermodynamic, Magnetic and Mechanical Properties of Advanced Materials

M. Šob[1,2], J. Pavlů[1,2], and J. Vřešťál[1,2]

[1] Department of Chemistry, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic
{sob, houserova, vrestal}@chemi.muni.cz
[2] Institute of Physics of Materials, Academy of Sciences of the Czech Republic, v.v.i., Žižkova 22, 616 62 Brno, Czech Republic
{mojmir, houserova}@ipm.cz

1 Introduction

In 2009, our scientific activities utilizing computational facilities of MetaCenter were devoted to theoretical studies of mechanical, magnetic and thermodynamic properties of advanced materials. The particular subjects are listed below, together with the most important results achieved.

2 Principal Results Obtained in the Year 2009

2.1 Thermodynamics and Phase Diagrams in Advanced Metallic Materials

Main task in the year 2009 was to finish a combined ab initio/CALPHAD study of the energetics of formation of C14, C15 and C36 Laves-phases in the systems Cr-Nb, Cr-Ta [1] and Cr-Zr [2]. Further, we have also started to study the systems Cr-Ti and Cr-Hf. The results obtained were used for assessment of thermodynamic properties and phase diagram calculations using a new physical model of Laves phases. An extended account of our methodology was presented in an invited paper in a monograph [7], where we also employed the results on Cr-based systems as illustrative examples. Our findings were presented in invited lectures at the 5th Conference of Asian Consortium on Computational Materials Science (ACCMS-5), Hanoi (invited talk) [11] and at the 2009 Autumn Annual Meeting of The Japan Institute of Metals, Kyoto (Foreign Scholar Special Invited Lecture) [12] as well as in a poster contribution at the 4th Int. Conf. on Advanced Materials and Nanotechnology, Dunedin, New Zealand [18] and oral contributions at the XXXVIIIth CALPHAD Conference (Prague) [21, 24], 13th Österreichische Chemietage Vienna: Czech, Slovak & Austrian Chemical Societies (Vienna) [29, 30], European Congress on Advanced Materials and Processes (EUROMAT) 2009 (Glasgow) [31], XIth Workshop of the Associated Phase Diagram and Thermodynamic Committee (APDTC) [33], 2009 Fall Conference, The Korean Institute of Metals and Materials, Daegu, South Korea [34] and 2009 Materials Research Society Fall Meeting, Boston, MA [35].

2.2 Magnetism, Electronic Structure and Atomic Configuration of Nanowires

We performed ab initio spin-density functional calculations of the electronic and magnetic properties of Fe, Ni and Mn nanostructures with a geometry varying between a straight linear wire and a three-dimensional nanorod. In the case of Mn, we included both collinear and noncollinear, commensurate and incommensurate magnetic configurations. With decreasing tension along the axis of the nanostructure we found a series of transitions first from dimerized to periodic and zigzag wires, then to a planar triangular stripe, and further to a nanorod consisting of a periodic stacking of triangular antiprims (Fe, Ni) or of distorted octahedra (Mn). In all Fe nad Ni nanostructures atoms are in a high-moment state, with magnetic moments of about 3.1 μ_B for Fe and about 1 μ_B for Ni. A transition to a low-spin or nonmagnetic state is initiated at a fixed critical value of the interatomic distance, independent of dimension and coordination number. The analysis of the electronic structure shows that already for the one-dimensional nanostructures the ratio between exchange splitting and magnetic moment is close to the universal value $I = \Delta/M = 1\,\mathrm{eV}/\mu_\mathrm{B}$ established for bulk itinerant magnets. All Mn nanostructures are also in a high-moment state, with absolute values of the local magnetic moments per atom varying between 2.96–3.79 μ_B. Collinear and noncollinear magnetic structures are energetically nearly degenerate, if the geometric and magnetic degrees of freedom are relaxed simultaneously. Compression of the nanostructures leads to a decrease in the magnetic moments. These results were published in extended papers in Physical Review B [4, 6] and presented at the Psi-k Workshop on Magnetism in Complex Systems, Vienna [20] and at the International Conference on Magnetism, Karlsruhe [26].

2.3 Magnetism, Electronic Structure and Atomic Configuration of Grain Boundaries

Grain boundaries (GBs) are important elements of microstructure in polycrystalline solids, which have been widely used as engineering materials. Till now, however, fundamental interactions that determine the structure, stability and other important properties of GBs have not been fully understood, especially in magnetic materials. Theoretical calculations are, in this respect, very helpful as they can provide microscopic information that is often hardly accessible experimentally. In our studies, we employed an ab initio pseudopotential technique and concentrated on selected coherent GBs, such as Σ=5 (210) and Σ=19 (331), in nickel. We investigated first their stability and magnetism of atoms in the vicinity of GBs. The GB stability appears not to be significantly influenced by magnetism. On the other hand, we can observe an enhancement of atomic magnetic moments in the vicinity of grain boundaries. Furthermore, we examined an interaction of studied GBs with point defects. Namely, these are vacancies and sulphur and antimony impurities which tend to segregate at GBs in Ni. Both the grain boundary energy and segregation enthalpy are affected by

the presence of a vacancy. In some cases, the so-called vacancy delocalization at GBs, which is demonstrated as a substantial reduction of the vacancy free volume, is observed. Finally, the magnetic properties of GBs are also influenced by the interaction with point defects. These findings indicate that GB imperfections need to be seriously taken into account when calculating GB properties. Further, we launched a systematic study of segregation and embrittling potency of sp elements of the 3rd, 4th and 5th period (Al, Si, P, S, Ga, Ge, As, Se, In, Sn, Sb and Te) at the Σ=5(210) grain boundary (GB) in fcc nickel and the segregation of these impurities at the (210) free surface (FS). Full relaxation of the geometric configuration of the GB and FS without and with impurities has been performed and the effect of impurities on the distribution of magnetic moments has been analysed. We determined the embrittling potency energy from the difference between the GB and FS binding energies on the basis of the Rice-Wang model; here a positive/negative value of the embrittling potency energy means that the solute atom has the embrittling/strengthening effect on the GB. These results were published in the Proceedings of the 9th Workshop of Physical Chemists and Electrochemists, Brno [14] and in the Proceedings of the 5th Doctoral Conference on Multiscale Design of Advanced Materials, Brno [15], presented in an invited talk at the Psi-k Workshop on Magnetism in Complex Systems, Vienna [9] and in the contributions at the International Conference on Magnetism, Karlsruhe [25] and at the 2009 Materials Research Society Fall Meeting, Symposium LL: Multiphysics Modeling in Materials Design, Boston, MA [36] and also at several prominent academic and industrial institutions [38, 40–42].

2.4 Ab Initio Study of Structure and Extended Defects in $MoSi_2$

Mechanical properties of molybdenum disilicide have been investigated extensively in recent years. Owing to its high melting temperature, low density and high creep strength, it belongs to promising materials in high-temperature structural applications. We have simulated a tensile test in $MoSi_2$ with $C11_b$ structure along the [001] direction from first principles using the full-potential linearized augmented plane wave method. A full relaxation of all external and internal structural parameters was performed, and influence of each relaxation process on energetics, strength and behavior of interatomic bonds was investigated in detail. It turned out that for a correct description of the phenomena studied, as, e.g., tension-compression asymmetry or the behavior of the interatomic bonds, full relaxation of all structural parameters must be performed. Further, we studied the gamma-surfaces for the (013) and (110) planes in $MoSi_2$ calculated by employing the density functional based method as implemented in the VASP code. While there is only one minimum on the (110) gamma-surface, three distinct minima have been found on the (013) gamma-surface. These minima, which determine three types of possible stacking faults on the (013) plane, are not symmetry dictated and thus the fault vectors are to a great extent controlled by the details of the interatomic bonding in $MoSi_2$. The results obtained were published in a top academic journal [3] as well as in the Proceedings of the Symposium

on Advanced Intermetallic-Based Alloys for Extreme Environment and Energy Applications (2008 MRS Fall Meeting, Boston, MA) [13] and presented at the 15th International Conference on Strength of Materials, Dresden [28].

2.5 Relation of Mechanical and Magnetic Properties in Metals and Intermetallics

Magnetic solids constitute a basis of many technologically important materials, however, very little is known how their magnetic behavior changes when a high-strain deformation is applied (as it is, for example, in heavily deformed regions of extended defects, such as grain boundaries, dislocation cores, crack tips etc.). We have studied the effect of magnetism on ideal tensile strength and structural stability of iron, cobalt and nickel and of selected magnetic intermetallic compounds at high-strain tetragonal and trigonal deformation. The total energies were calculated by spin-polarized full-potential LAPW method and were displayed in contour plots as functions of tetragonal or trigonal distortion c/a and volume; borderlines between various magnetic phases were shown and stability of higher-energy phases was discussed. The anisotropy of calculated tensile strength was explained in terms of higher-symmetry structures encountered along the deformation paths studied. The calculated phase boundaries were used to predict the lattice parameters and magnetic states of iron, cobalt and nickel overlayers on various (001) and (111) substrates. Importance of magnetic ordering on the theoretical tensile strength of these overlayers was also revealed. Whereas magnetism does not play an important role in stabilization of the ground-state structure of nickel and Ni-intermetallics, the magnetic effects in iron and Fe-intermetallics are vital. These results were presented in a keynote talk at the Workshop on Probing of Limits of Strength, University of California at Berkeley [12] and in a contribution at the 15th International Conference on Strength of Materials, Dresden [27].

2.6 Summary of Presentation of Results: Publications, Invited Talks, Conference Presentations

In total, in 2009 our results have been published in 6 papers in top academic journals [1–6], in one monograph [7] and in four Proceedings of international and national conferences and have been presented in one keynote [10] and four invited [8–9, 11–12] talks at important international conferences, in 20 other contributions at conferences, scientific meetings and schools [17–36] and in 6 talks at scientific institutions abroad [37–42].

2.7 Organization of an International Meeting

Prof. J. Vřešťál served as a Chairman and Dr. J. Pavlů and Prof. M. Šob as Members of the Organizational Committee of the XXXVIIIth Int. Conf. on Phase Diagrams and Computational Thermodynamics (CALPHAD), which was organized in Prague in the period of May 17–22, 2009 [44, 45]. The conference was

attended by 150 leading scientists from the whole world. It surveyed the present state of this research field and promoted establishing new personal contacts and scientific collaborations.

3 Software and Financial Support

The software for ab initio calculations of electronic structure (LMTO-ASA, WIEN2k–FLAPW, VASP–pseudopotential code) used for solution of the above-mentioned tasks was provided by the cooperating Institute of Physics of Materials, Academy of Sciences of the Czech Republic, Brno, and was implemented at the MetaCenter of the Masaryk University under the Research Project No. MSM6383917201.

Further, our research was supported by the Grant Agency of the Czech Republic (Projects No. 202/09/1786, 106/07/1078 and 106/05/H008), by Ministry of Education of the Czech Republic (Projects COST OC147, COST OC164 and COST OC09011), Grant Agency of the Academy of Sciences of the Czech Republic (Project No. IAA100100920) and by the Research Projects AV0Z20410507 (Academy of Sciences of the Czech Republic) and MSM0021622410 (Ministry of Education of the Czech Republic). The access to the computing facilities of the MetaCenter of the Masaryk University, Brno, provided under the Research Project No. MSM6383917201, is gratefully acknowledged.

4 Publication and Presentations of Results

4.1 Publications in Refereed Journals and in Scientific Monographs

[1] J. Pavlů, J. Vřešťál, M. Šob: Re-Modelling of Laves Phases in the Cr-Nb and Cr-Ta Systems Using First-Principles Results. CALPHAD **33** (2009), 179-186.
[2] J. Pavlů, J. Vřešťál, M. Šob: Stability of Laves Phases in the Cr-Zr System. CALPHAD **33** (2009), 382-387.
[3] M. Šob, M. Friák: The effect of structural degrees of freedom on bonding and strength characteristics of molybdenum disilicide. Intermetallics **17** (2009), 523-528.
[4] M. Zelený, M. Šob, J. Hafner, Magnetism in low-dimensional nanostructures—from nanowires to nanorods. Phys. Rev. B **79** (2009), 134421 (10 pp).
[5] M. Šob, D. Legut, M. Friák: Reply to the Comment on "Why is Polonium Simple Cubic and So Highly Anisotropic?" Phys. Rev. Lett. **102** (2009), 079702 (1 pp).
[6] M. Zelený, M. Šob, J. Hafner, Noncollinear magnetism in manganese nanostructures. Phys. Rev. B **80** (2009), 144414 (19 pp).
[7] M. Šob, A. Kroupa, J. Pavlů, J. Vřešťál: Application of Ab Initio Electronic Structure Calculations in Construction of Phase Diagrams of Metallic Systems with Complex Phases. In: Solid Phase Transformations II, ed. by J. Čermák and I. Stloukal, published in Solid-State Phenomena **150** (2009), 1-28.

4.2 Invited Talks at International Conferences

[8] M. Šob, M. Friák, M. Zelený, D. Legut: Magnetism of 3d metals and their intermetallics at high-strain deformation. Int. Conf. On Study of Matter at Extreme Conditions, Miami, Florida, USA, March 28–April 2, 2009 (invited talk).

[9] M. Šob, M. Čák, M. Všianská, J. Kuriplach, J. Hafner: Ab initio study of magnetism of clean and decorated grain boundaries in iron and nickel. Psi-k Workshop on Magnetism in Complex Systems, Vienna, Austria, April 16-19, 2009 (invited talk).

[10] M. Šob: The effect of magnetism on ideal strength and structural stability in metals and intermetallics. Workshop on Probing of Limits of Strength, University of California at Berkeley, California, USA, Aug. 10-12, 2009 (keynote talk).

[11] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: Application of Ab Initio Results in Modelling Phase Diagrams Containing Complex Phases. The 5th Conference of Asian Consortium on Computational Materials Science (ACCMS-5), Hanoi, 7-11 September 2009 (invited talk).

[12] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: Ab-Initio Based Construction of Phase Diagrams in Systems with Complex Phases. 2009 Autumn Annual Meeting of The Japan Institute of Metals, Kyoto, 15-17 September 2009 (Foreign Scholar Special Invited Lecture).

4.3 Contributions in Proceedings of International and National Conferences—Published

[13] M. Čák, M. Šob, V. Paidar, V. Vitek: Stacking Faults and Dislocation Dissociation in $MoSi_2$, in Advanced Intermetallic-Based Alloys for Extreme Environment and Energy Applications, edited by M. Palm, B.P. Bewlay, M. Takeyama, J.M.K. Wiezorek, Y-H. He (Mater. Res. Soc. Symp. Proc. Volume 1128, Warrendale, PA, 2009), paper No. 1128-U07-10.

[14] M. Všianská, M. Šob: Ab initio study of segregation of impurities at grain boundaries and surfaces in nickel. In: 9th Workshop of Physical Chemists and Electrochemists, ed. L. Trnková and R. Kizek, Faculty of Science, Masaryk University, Brno & Faculty of Agronomy, Mendel University, Brno 2009, pp. 149-150.

[15] M. Všianská, M. Šob: Segregation of impurities at grain boundaries in nickel. In: Multiscale Design of Advanced Materials 2009 (Proceedings of Doctoral Conference), eds. I. Dlouhý, J. Švejcar, M. Šob, Institute of Physics of Materials, Academy of Sciences of the Czech Republic, Brno 2009, pp. 89-94.

[16] P. Elstnerová, M. Friák, J. Neugebauer: Ab initio studie kalcitu substituovaného hořčíkem a fosforem. In: Multiscale Design of Advanced Materials 2009 (Proceedings of Doctoral Conference), eds. I. Dlouhý, J. Švejcar, M. Šob, Institute of Physics of Materials, Academy of Sciences of the Czech Republic, Brno 2009, pp. 5-12.

4.4 Unpublished Presentations at International and National Conferences and Schools

[17] M. Šob, M. Zelený, M. Friák: Phase transitions in Fe, Co and Ni overlayers on (001) and (111) substrates. 4th Int. Conf. on Advanced Materials and Nanotechnology, Dunedin, New Zealand, Feb. 8-12, 2009 (oral).
[18] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: Application of ab initio results in modelling phase diagrams containing complex phases. 4th Int. Conf. on Advanced Materials and Nanotechnology, Dunedin, New Zealand, Feb. 8-12, 2009 (poster).
[19] D. Legut, M. Friák, M. Šob: A prediction of pressure-induced lattice instability in polonium—a challenge for experimental research. Int. Conf. On Study of Matter at Extreme Conditions, Miami, Florida, USA, March 28-April 2, 2009 (oral).
[20] M. Zelený, M. Šob, J. Hafner: Noncollinear magnetism in Mn nanostructures. Psi-k Workshop on Magnetism in Complex Systems, Vienna, Austria, April 16-19, 2009 (oral).
[21] J. Pavlů, J. Vřešťál, M. Šob: Energetics and magnetism of Cr-based sigma phases. XXXVIIIth Int. Conf. On Phase Diagrams and Computational Thermodynamics, Prague, Czech Republic, May 17-22, 2009 (oral).
[22] M. Friák, M. Šob: Ab initio analysis of pressure-induced bcc-hcp transformation in iron. XXXVIIIth Int. Conf. On Phase Diagrams and Computational Thermodynamics, Prague, Czech Republic, May 17-22, 2009 (oral).
[23] M. Všianská, D. Legut, M. Šob: Ab initio Study of Stability of In-Sn Alloys. XXXVIIIth Int. Conf. On Phase Diagrams and Computational Thermodynamics, Prague, Czech Republic, May 17-22, 2009 (oral).
[24] J. Pavlů, J. Vřešťál, M. Šob: Stability of Laves phases in the Cr-Ti and Cr-Hf systems. XXXVIIIth Int. Conf. On Phase Diagrams and Computational Thermodynamics, Prague, Czech Republic, May 17-22, 2009 (oral).
[25] M. Šob, M. Čák, M. Všianská, J. Kuriplach, J. Hafner: Magnetism of grain boundaries in iron and nickel. International Conference on Magnetism, Karlsruhe, Germany, July 26-31, 2009 (poster).
[26] M. Zelený, M. Šob, J. Hafner: Magnetism of manganese in low-dimensional nanostructures—nanowires and nanostripes. International Conference on Magnetism, Karlsruhe, Germany, July 26-31, 2009 (poster).
[27] M. Šob, M. Zelený, M. Friák: Strength and stability of Fe, Co and Ni overlayers at the (001) and (111) substrates. 15th International Conference on Strength of Materials, Dresden, Germany, Aug. 16-21, 2009 (oral).
[28] V. Paidar, M. Čák, M. Šob, V. Vitek: Theoretical analysis of dissociation of $1/2\langle 331]$ dislocations in $MoSi_2$. 15th International Conference on Strength of Materials, Dresden, Germany, Aug. 16-21, 2009 (oral).
[29] J. Pavlů, J. Vřešťál, M. Šob: Magnetism in Cr-based Sigma Phases. 13th Austrian Chemistry Days 2009, Vienna, Austria, 24-27 August 2009 (oral).
[30] J. Pavlů, J. Vřešťál, M. Šob: Phase Equilibria with Laves Phases in the Cr-Ti and Cr-Hf Systems. 13th Austrian Chemistry Days 2009, Vienna, Austria, 24-27 August 2009 (oral).

[31] J. Pavlů, J. Vřešťál, M. Šob: Stability of Cr-containing Laves Phases. European Congress and Exhibition on Advanced Materials and Processes (EUROMAT 2009), Glasgow, Great Britain, Sept. 7-10, 2009 (oral).
[32] D. Legut, M. Šob: Deformation-induced phase transformations in Ni_3Al and Fe_3Al. The 8th European Symposium on Martensitic Transformations (ESOMAT 2009), Prague, Czech Republic, Sept. 7-11, 2009 (poster).
[33] J. Pavlů, J. Vřešťál, M. Šob: Recent progress in using of ab initio calculations for determination of stability of Cr-containing Laves phases, XIth Workshop of the Associated Phase Diagram and Thermodynamic Committee (APDTC). Ljubljana, Slovenija, Sept. 18-20, 2009 (oral).
[34] J. Vřešťál, J. Pavlů, M. Šob: Multiscale Modeling of the Stability of Intermetallic Phases, 2009 Fall Conference, The Korean Institute of Metals and Materials, Daegu, South Korea, Oct. 22-23, 2009 (oral).
[35] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: A Combined Ab Initio/CALPHAD Modeling of Phase Diagrams in Systems with Complex Phases. 2009 Materials Research Society Fall Meeting, Symposium LL: Multiphysics Modeling in Materials Design, Boston, MA, Nov. 30 - Dec. 4, 2009.
[36] J. Kuriplach, O. Melikhova, M. Šob: Magnetic Grain Boundaries in Nickel. 2009 Materials Research Society Fall Meeting, Symposium LL: Multiphysics Modeling in Materials Design, Boston, MA, Nov. 30 - Dec. 4, 2009.

4.5 Presentations at Scientific Institutions

[37] M. Šob: The role of ab initio electronic structure calculations in multiscale modelling of materials. Auckland University of Technology, Centre for Advanced Manufacturing Technology, Auckland, New Zealand, Feb. 18, 2009, and The University of Waikato, School of Science and Engineering, Hamilton, New Zealand, Feb. 20, 2009.
[38] M. Šob, M. Čák, M. Všianská, J. Kuriplach, J. Hafner: Ab initio study of magnetism of clean and decorated grain boundaries in iron and nickel. National Center for Electron Microscopy, Lawrence Berkeley National Laboratory, Berkeley, USA, Aug. 6, 2009.
[39] M. Šob, M. Friák, D. Legut, J. Kuriplach, I. Turek, V. Vitek: The application of ab initio electronic structure calculations in multiscale modeling of materials. Center for Theoretical Physics and Institute of Physics (VAST), Hanoi, Vietnam, Sept. 4, 2009.
[40] M. Šob, M. Čák, M. Všianská, J. Kuriplach, J. Hafner: Ab initio study of magnetism of clean and decorated grain boundaries in iron and nickel. Dept. of Mechanical Engineering and Science, Graduate School of Engineering, Kyoto University, Japan, Sept. 18, 2009.
[41] M. Šob, M. Čák, M. Všianská, J. Kuriplach, J. Hafner: Ab initio study of magnetism of clean and decorated grain boundaries in iron and nickel. Hitachi Ltd., Advanced Research Laboratory, Kokubunji-shi, Tokyo, Japan, Sept. 24, 2009.
[42] M. Šob, M. Čák, M. Všianská, J. Kuriplach, J. Hafner: Ab initio study of magnetism of clean and decorated grain boundaries in iron and nickel. Institute

of Industrial Science (IIS), The University of Tokyo, Tokyo, Japan, Sept. 25, 2009.

4.6 Proceedings of International and National Conferences

[43] I. Dlouhý, J. Švejcar, M. Šob (eds.): Multiscale Design of Advanced Materials (Proceedings from Doctoral Conference, Brno, Dec. 3, 2009), Institute of Physics of Materials of the Academy of Sciences of the Czech Republic, Brno, 2009, 157 pp., ISBN 978-80-254-6070-2.

4.7 Books of Abstracts from International and National Conferences

[44] J. Vřešťál, A. Kroupa, M: Šob, P. Brož, J. Sopoušek, J. Leitner, F. Chmelík, A. Zemanová, J. Pavlů (eds.): XXXVIIIth Int. Conf. On Phase Diagrams and Computational Thermodynamics (CALPHAD)—Book of Abstracts, Masaryk University and Institute of Physics of Materials, Academy of Sciences of the Czech Republic, Brno, 2009, 205 pp.

4.8 Organization of an International Conference

[45] XXXVIIIth Int. Conf. on Phase Diagrams and Computational Thermodynamics (CALPHAD), Prague, Czech Republic, May 17-22, 2009 (J. Vřešťál, J. Pavlů and M. Šob served as Members of the Organizational Committee).

Inter-chain Charge Carrier Mobility in Conjugated Polymers Doped with Polar Additives

Petr Toman

Institute of Macromolecular Chemistry, Academy of Sciences of the Czech Republic,
Heyrovský Sq. 2, 162 06 Praha 6, Czech Republic
toman@imc.cas.cz

1 Introduction

Conjugated polymers are promising materials for low-cost and easy-processing optoelectronic devices. Their semiconducting properties can be utilized in the construction of gas sensors, solar cells, field-effect transistors, electroluminescent diodes, etc. The mobility of charge carriers in conjugated polymers is one of the most important parameters determining the performance characteristics (e.g. maximum current or switching time) of organic optoelectronic devices.

Multistable molecular systems have recently attracted a lot of research interest because of their potential application in future molecular-scale logic and memory elements. An example of bistable molecular systems are photochromic molecules undergoing reversible changes of several electronic properties like absorption spectra or dipole moment when irradiated by light of a suitable wavelength.

In our project, we have studied photoswitching of the inter-chain hole transport in MEH–PPV doped with the photochromic additive 6-nitro-1',3',3'-trimethylspiro[2*H*-1-benzopyran-2,2'-indoline]. Upon irradiation with light of an appropriate wavelength, this molecule undergoes a ring-opening reaction from the closed form of spiropyran (SP) to the open merocyanine form (MR). The reversible SP $\leftrightarrow$ MR reaction is accompanied by a charge redistribution resulting in a significant increase of the dipole moment of the molecule.

2 Determination of the On-Chain Charge Carrier States

In order to calculate the hole states, the polymer chains are modeled as sequences of N sites corresponding to the repeating units. The length N corresponds to the average distance between chemical defects breaking the conjugation rather than the real polymer chain length. For simplicity, all polymer chains in our model possess the same length N. The hole states on such a chain can be described within the tight-binding approximation by the Hamiltonian

$$H = \sum_{n=1}^{N} \left[\varepsilon_n a_n^+ a_n - b_{n,n+1} \left(a_{n+1}^+ a_n + a_n^+ a_{n+1} \right) \right] \tag{1}$$

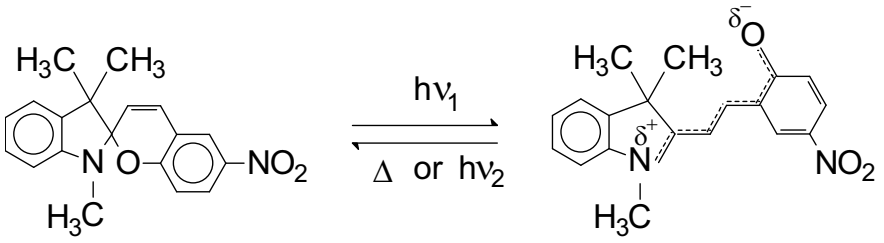


Fig. 1. Photochromic reaction of the SP ↔ MR molecule.

where a_n and a_n^+ are the annihilation and creation operators for a hole located at the n-th site, ε_n is the energy of this hole, and $b_{n,n+1}$ is the transfer integral between the sites n and $n+1$. Distributions of both molecular parameters ε_n and $b_{n,n+1}$ are influenced by the random structure of polymer chains and their surrounding.

On the basis of the quantum chemical calculations one can expect at least two-fold increase of the energetic disorder during the SP → MR photochemical reaction. The energetic disorder of ε_n was calculated for several values of the dipole moment of the additive m. Since the mutual interaction of the additive molecules is neglected, the standard deviation of the ε_n distribution is proportional to the dipole moment m of the additive and to the square root of the additive concentration c.

Using the Hamiltonian (1), solution of the time-independent Schrödinger equation yields the hole states $|\varphi_i\rangle$ together with their energies E_i. Thus, the energy spectrum of the density of states of the valence band and the extent of the delocalization of the charge carrier states $|\varphi_i\rangle$ along the polymer chain can be determined. An increasing energetic disorder $\sigma(\varepsilon_n)$ leads to the broadening of the originally sharp valence band edge and the formation of the tail states in the gap (see Fig. 2). These states, with their relatively low density and weak connectivity, behave as dipolar hole traps. Moreover, the energetic disorder increases localization of the charge carrier states and consequently decreases their connectivity.

3 Model of the Inter-Chain Hopping

The simplified material structure considered in the modeling is shown in Fig. 3. The inter-chain transfer is slow in comparison with the charge carrier thermalization, which occurs typically in times of several picoseconds. Thus, it is possible to expect full charge carrier thermalization over all states of the given chain A between two subsequent inter-chain hops (see Fig. 4). Therefore, our model assumes the following steps of the inter-chain charge carrier transport: charge carrier moving to any possible state on the chain A, charge carrier thermaliza-

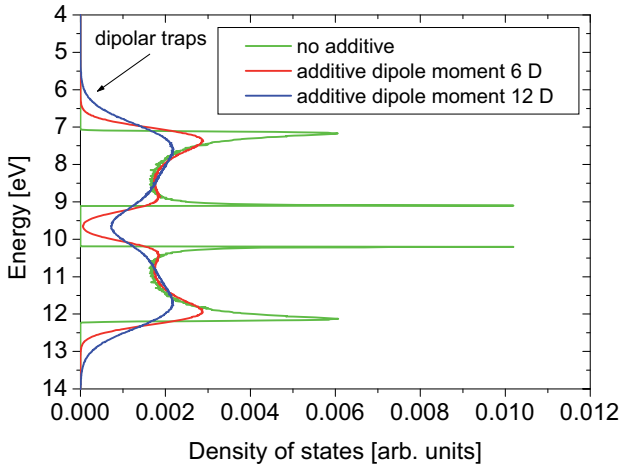


Fig. 2. Density of states in the pure and doped polymer (additive concentration $c = 4 \times 10^{-4}\text{Å}^{-3}$).

tion over all its possible states on the chain A, and finally charge carrier hop to any possible state on one of the nearest neighboring chains B.

Generally, fermions thermalize to the Fermi–Dirac distribution. However, under usual experimental conditions it is very improbable that there is more than one free charge carrier on a given conjugated chain segment. Hence, it is possible to consider thermalization to the Boltzmann distribution

$$p_i(E_i) = \frac{\exp\left(\frac{-E_i}{kT}\right)}{Z(T)}, \tag{2}$$

where $p_i(E_i)$ is the probability of occupation of the state i at temperature T, E_i is the energy of state i, and $Z(T)$ is partition function over all states of the chain A.

The rate for hole hopping between an initial state i with the energy E_i on the chain A and a final state j with the energy E_j on an adjacent chain B can be calculated according to semi-classical Marcus relation as

$$\upsilon_{i\to j} = \frac{J_{ij}^2}{\hbar}\sqrt{\frac{\pi}{\lambda_{ij}kT}}\exp\left(\frac{-(E_i - E_j - \lambda_{ij})^2}{4\lambda_{ij}kT}\right) \tag{3}$$

where λ_{ij} is the charge carrier reorganization energy and J_{ij} is the effective charge transfer integral. The molecular parameters can be in principle calculated using standard quantum chemical methods.

From the charge transfer rates $\upsilon_{i\to j}$ between two states and the thermalized occupation probability p_i of the initial state the charge transfer rates $\upsilon_{A\to B}$

between two adjacent chains A and B can be easily calculated

$$\upsilon_{A\to B} = \sum_{i\in A, j\in B} p_i(E_i)\upsilon_{i\to j}, \tag{4}$$

whereas summation goes through all states of the respective chain. Determination of the charge transfer rates $\upsilon_{A\to B}$ makes possible to solve the master equation and consequently calculate the hole mobility $\mu(\omega)$ according to Kubo formula.

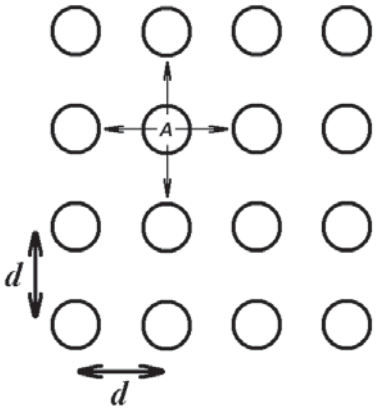


Fig. 3. Simplified polymer chain alignment considered in the inter-chain hole mobility modeling. Circles represent parallel equidistantly placed polymer chains. Much smaller additive molecules randomly dispersed among them are not shown. The possible charge carrier hops from a given chain A are denoted by arrows.

4 Results

The proposed model of the inter-chain hopping predicts a strong dependence of the hole mobility in MEH–PPV on the local energetic disorder (see Figs. 5 and 6). Doubling of the local energetic disorder causes the decrease of the inter-chain hole mobility by up to several orders of magnitude. These results confirm the possibility to construct an efficient optoelectrical switch based on a polymer doped with a photochromic polar additive. It should be noted, that the magnitudes of the calculated mobilities are relative only; however the scaling does not affect the ratio of the mobilities of the pristine and doped material.

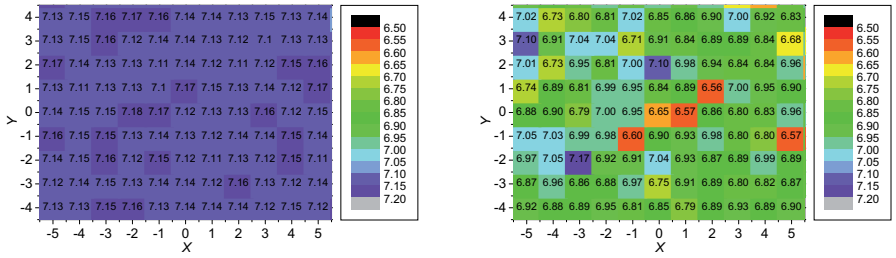


Fig. 4. Map of the average energy (in eV) of the thermalized hole located on particular polymer chains calculated for the pristine (a) and doped (b) polymer, respectively. Polymer chains are depicted as color squares. Chains acting as hole traps are indicated by red and orange colors. Coordinates X and Y show the chain positions (see Fig. 3).

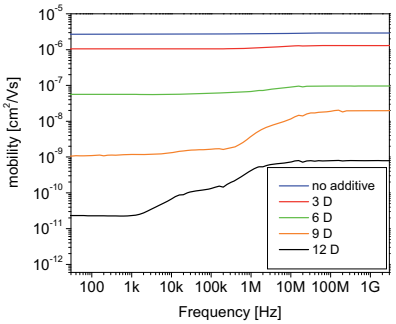


Fig. 5. Frequency dependence of the hole mobility $\mu(f)$ calculated for different values of the additive dipole moment.

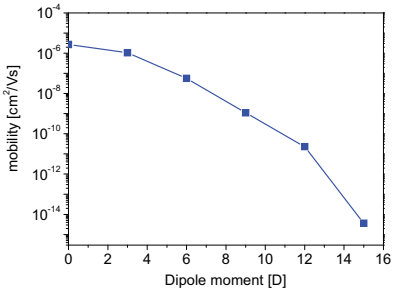


Fig. 6. Dependence of the low-frequency mobility on the additive dipole moment.

Grant Support

- Czech Science Foundation (203/06/0285).
- Grant Agency of the Academy of Sciences of the Czech Republic (IAA401770601).
- Ministry of Education, Youth, and Sports of the Czech Republic (Czech–Polish collaboration project MEB050815).

Used Programs and Applications

Gaussian 03, Portland Group Fortran Compiler

List of Publications Dedicated to MetaCentrum

1. P. Toman, S. Nešpůrek, M. Weiter, M. Vala, J. Sworakowski, W. Bartkowiak, M. Menšík: "Influence of Dipolar Species on Charge Transport in poly[2-methoxy-5-(2'-ethylhexyloxy)-p-phenylene vinylene]", *Polym. Adv. Technol.* **17**, 673–678 (2006).
2. S. Nešpůrek, P. Toman, A. Fujii, K. Yoshino: "Poly[(diphenylsilanediyl)ethynediyl]: structure and optical and electroluminescent properties", *J. Appl. Polym. Sci.* **105 (1)**, 208–214 (2007).
3. P. Toman, S. Nešpůrek: "Modeling of hole transport in poly[2-methoxy-5-(2'-ethylhexyloxy)-p-phenylene vinylene] doped with polar additives", *Mol. Cryst. Liq. Cryst.* **496**, 25–38 (2008).
4. P. Toman, S. Nešpůrek, M. Weiter, M. Vala, J. Sworakowski, W. Bartkowiak, M. Menšík: "Model of the Influence of Energetic Disorder on Inter-chain Charge Carrier Mobility in poly[2-methoxy-5-(2'-ethylhexyloxy)-p-phenylene vinylene]", *Polym. Adv. Technol.* **20**, 263–267 (2009).
5. P. Toman, S. Nešpůrek, W. Bartkowiak: "Modelling of charge carrier transport in conjugated polymers doped by polar additives", *Materials Science-Poland* **27(3)**, 797–812 (2009).
6. S. Ehala, P. Toman, E. Marklík, V. Kašička: "Application of affinity capillary electrophoresis and density functional theory to the investigation of benzo-18-crown-6-ether complex with ammonium cation", *J. Chromatogr. A* **1216(45)**, 7927–7931 (2009).
7. M. Weiter, J. Navrátil, M. Vala, P. Toman: "Photoinduced reversible switching of charge carrier mobility in conjugated polymers", *Eur. Phys. J. Appl. Phys.* **48**, 10401-p1 – 10401–p6 (2009)
8. S. Ehala, E. Marklík, P. Toman, V. Kašička: "ACE applied to the quantitative characterization of benzo-18-crown-6-ether binding with alkali metal ions in a methanol-water solvent system", *Electrophoresis* **31**, 702–708 (2010).
9. M. Vala, J. Vyňuchal, P. Toman, M. Weiter, S. Luňák Jr.: "Novel, soluble diphenyl-diketo-pyrrolopyrroles: Experimental and theoretical study", *Dyes and Pigments* **84**, 176–182 (2010).

List of Conference Presentations Prepared Using MetaCentrum Resources

1. 12th International Conference on Polymers and Organic Chemistry 2006 (POC'06): P. Toman, S. Nešpůrek, M. Weiter, M. Vala, J. Sworakowski, W. Bartkowiak: *Influence of Polar Additives on Charge Transport in MEH-PPV*, July 2-7, 2006, Okazaki, Japan.
2. IX International Conference on Frontiers of Polymers and Advanced Materials (IXth ICFPAM): P. Toman, S. Nešpůrek, M. Menšík, M. Weiter, M. Vala, J. Sworakowski, W. Bartkowiak: *Modeling of charge carrier transport in conjugated polymers doped with polar species*, July 8-12, 2007, Cracow, Poland.
3. 9th International Symposium on Polymers for Advanced Technologies: P. Toman, S. Nešpůrek, M. Weiter, M. Vala, M. Menšík: *Influence of Energetic Disorder on Charge Carrier Mobility in Conjugated Polymers*, October 22–25, 2007, Shanghai, China.
4. 7th International Conference on Electronic Processes in Organic Materials (ICEPOM-7): P. Toman, S. Nešpůrek, M. Menšík: *Modeling of hole transport in poly[2-methoxy-5-(2'-ethylhexyloxy)-p-phenylene vinylene] doped with polar additives*, May 26-30, 2008, Lviv, Ukraine.
5. 11th International Conference Electrical and Related Properties of Organic Solids (ERPOS-11): P. Toman, S. Nešpůrek, M. Weiter, M. Vala, W. Bartkowiak: *Modeling of charge carrier transport in conjugated polymers doped by polar additives*, July 13-17, 2008, Piechowice, Poland.
6. 73rd Prague Meeting on Macromolecules "New Frontiers in Macromolecular Science: From Macromolecular Concepts of Living Matter to Polymers for Better Quality of Life": P. Toman, S. Nešpůrek, M. Weiter, M. Vala, W. Bartkowiak: *Modeling of charge carrier transport in conjugated polymers doped by polar additives*, July 5-9, 2009, Prague, Czech republic.
7. 12th European Polymer Congress (EPF'09): P. Toman, S. Nešpůrek, M. Weiter, M. Vala, W. Bartkowiak: *Modeling of charge carrier mobility in PPV derivatives doped by polar additives*, July 12-17, 2009, Graz, Austria
8. 8th International Conference on Advanced Polymers via Macromolecular Engineering (APME 2009): P. Toman, S. Nešpůrek, M. Weiter, M. Vala, W. Bartkowiak: *Charge carrier transport in conjugated polymers*, October 4-7, 2009, Dresden, Germany.

Evaluation of Feature Space Transforms for Czech Sign-Language Recognition

Jan Trmal and Marek Hrúz
{jtrmal, mhruz}@kky.zcu.cz

Department of Cybernetics
University of West Bohemia
306 14, Plzeň, Czech Republic*

Abstract. In the paper we give a brief introduction into sign language recognition and present a particular research task, where the access to MetaCentrum computing facilities was highly beneficial. Although the problem of signed speech recognition is currently being researched into by many research institutions all around the world, it lacks of a generally accepted baseline parametrization method. Our team introduced a parametrization method based on skin-color detection and object tracking. Because of the relatively high amount of information that is produced during the parametrization process, a method that reduces the unnecessary information while keeping the necessary information is required. Such methods are called *feature space dimension reduction* methods. We used the MetaCentrum facilities to evaluate several methods used widely in the field of acoustic speech recognition and their influence on recognition score of a Czech Sign-Language recognizer.

1 Introduction

For deaf people, sign language is the basic mean of communication—just as speech is for hearing people. Inspired by speech recognition, where the rate of progress has been enormous in the past decade, new ways of communication between deaf people and computers or hearing people are being developed. The main task of automatic sign language recognition is to recognize a sign performed by a signer.

In speech recognition it is obvious that microphone is used as the input device. In sign language recognition more kinds of input devices can be used. Mechanical devices, which measure location of various parts of body, such as data gloves and haptic devices have the advantage in accuracy of measurements. But there is a serious setback—the signer is forced to wear a cumbersome device. Another approach to this problem is to use a camera or even multiple cameras as input devices.

* This research was supported by the Grant Agency of Academy of Sciences of the Czech Republic, under project No. 1ET101470416, the Ministry of Education of the Czech Republic, project No. MŠMT LC536 and project No. ME08106 and by the Grant Agency of the Czech Republic, project No. GAČR 102/08/0707.

Lets consider the output of the camera or multiple cameras. Given the resolution of one picture and the frame frequency, the amount of data produced is huge. This is because the visual stream bears not only the bare information relevant for sign language recognition (SLR) but also additional information about the speaker, clothing, environment, etc. This additional information does not improve the recognition score—quite contrary, the information increase complexity of the task, dimensionality of the feature vectors and slow down the overall research process. The first thing we have to do is to reduce the amount of unnecessary information, while keeping the information useful for the SLR problem. This is the task of visual feature extraction (VFE) system. The VFE system employs computer vision algorithms to isolate (extract) the relevant information.

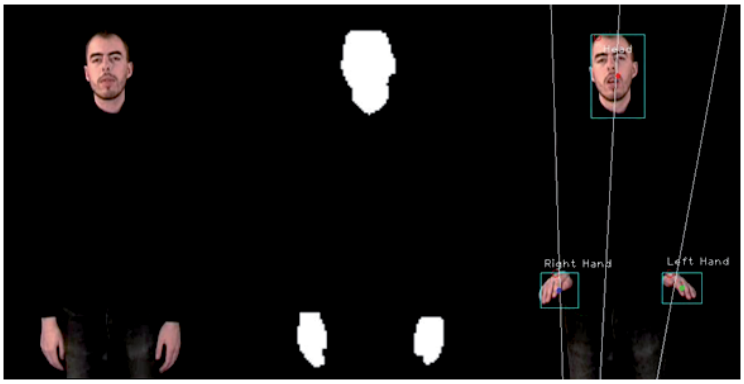


Fig. 1. a) source frame from the camera, b) segmented image c) head and hands tracking

1.1 Visual Feature Extraction System

The sign language consists of manual (i.e., employing hands) and non-manual (i.e., mainly facial expressions) component. We aimed our work against the manual component. The sign linguists distinguish several basic subcomponents of the manual component of the sign—hand shape, palm and hand orientation, location and movement.

To enable successful recognition of a sign, the VFE system must be not only able to find and isolate the hands, but also provide information about location in space and, more importantly, about the progressive change of the hand location. The process of finding the objects of interest in the image is called *object detection* and the process of monitoring of movements is called *object tracking*. In our case, the *objects* are left and right hand and the head of the speaker.

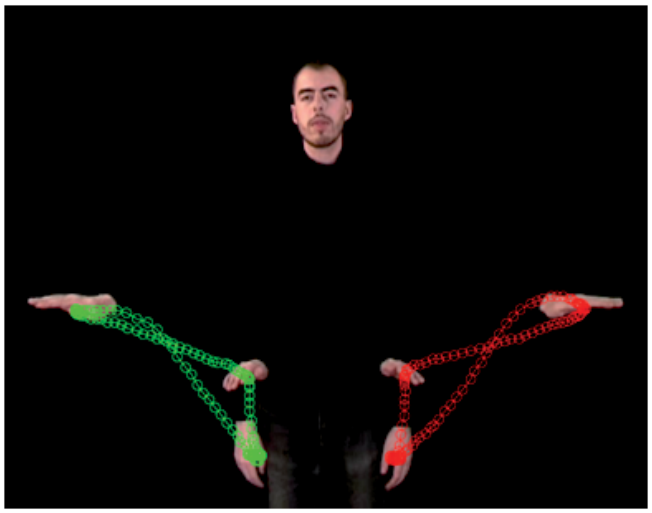


Fig. 2. Hand tracking for one of the signs

Object Detection A common method for detecting parts of a human body is the skin-color detection [4]. Skin-color detection can be combined with motion cues [2] or edge cues [5]. Although the method is widespread, it has several disadvantages. For example it is illumination dependent and there is a large variety in color of human skin. Therefore, an adaptation should be applied to the universal skin-color model [7]. The skin-color detection is the first phase of VFE in our SLR system. Using the adapted, speaker-dependent model, we perform thresholding of the image—for every pixel in the image we determine its likelihood of being a skin-color pixel. If the likelihood is higher than a specific threshold, the pixel is considered to be a part of a skin-color object.

Object Description Every object isolated in the previous phase must be described by some features that correspond to the aforementioned manual subcomponents of the signs. For location and movement description we simply compute the 2D coordinates of the center of mass of the isolated object, for the shape we use the Hu's moments associated to the object.

Object Tracking After the object detection, we ideally obtain the objects representing the left hand, right hand and head. Our VFE system uses separate object tracker for each of these objects. The tracking process itself is based on measure of the distance of the old and new locations of the objects enhanced with occlusion detection system.

Feature Sets In the previous sections we described the VFE system. Using this system, we are able to reduce the amount of data in each input frame to 33 values, 11 for each of the tracked objects:

- x, y - the center of mass of the object
- 7 Hu's moments describing the shape of the object
- the angle of the object relative to the x-axis of the image
- a Boolean value representing whether the object is in occlusion

The occlusion flag is used in post-processing. If an occlusion is detected, the Hu's moments and the angle are linear interpolated between the last values before occlusion and the first values after occlusion. The final step of the post-processing is normalization in the spatial domain. The mean position of the head is considered as the origin and the mean width of the head is considered as one unit. The normalized features (excluding the occlusion flag) are concatenated in the following order: left hand, right hand, and head. For every object, 10 features are obtained, which makes a total count of 30 features for every frame.

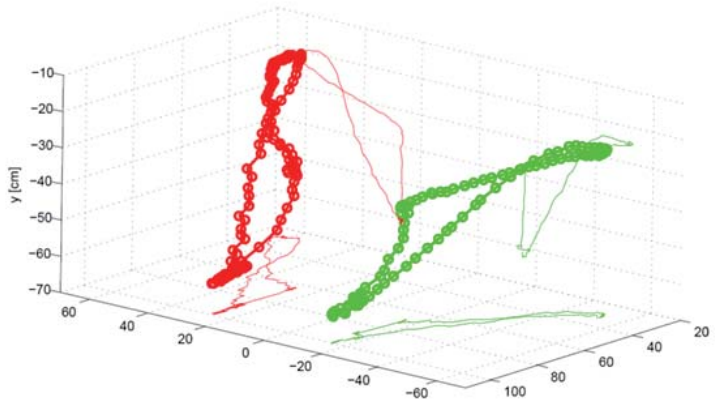


Fig. 3. Trajectory tracking in 3D space. The coordinates origin is located in the mean position of signers head

2 Feature Space Reduction Methods

Features obtained so far by the method described in the previous section are highly correlated and statistically dependent on each other. Also, the number of correlated features can be higher than the size of independent feature set – it can be proved [1] that the Hu system is dependent and incomplete. For our purpose, the Hu's moments are sufficient to describe the contours of the objects, but the dependence points at the possible use of a dimension reduction method.

In addition, it can be shown [3] that the number of the basic sign units can be interpreted as the Cartesian product of 4 sets (corresponding to the basic manual components) with cardinalities of 30, 8, 20, 40—even when no context like "tri-signs" (as an analogy to tri-phones used in speech recognition) is considered. For this reason, the total number of model parameters to be estimated would become extremely large, particularly when one considers the limited size of a training corpus. Even if we had a good model, the total number of parameters is a limiting factor, when recognition in real-time is needed. For this reason, the choice of a suitable projection scheme method is a very important subtask of SL recognition. The projections schemes are usually $m \times n$ matrices, $m \geq n$, where m is the original dimension of feature vectors (30 in our case) and n the new feature vector dimension. The cause of the necessity to choose the right projection matrix only by experiments lies in the problematic definition of *relevant information*. We have investigated and compared 5 popular projection schemes: PCA as the baseline, ICA, LDA, HLDA, and rHLDA (HLDA with more robustly estimated covariance matrices). We will not discuss the methods properties or implementation, for further info, see paper [6].

3 Experiments and MetaCentrum Involvement

We used the MetaCentrum computation facilities to experiment with various possible choices of the constant n and, since our recognizer is based on Hidden Markov Models (HMM), with topology of the HMMs. To perform the experiments, we used the HTK toolkit for HMM modeling and proprietary C++ routines that we used through the Matlab MEX interface. The Matlab was directly available as a user module, but the HTK toolkit was not. After consultation with the MetaCentrum staff, they volunteered to prepare the HTK toolkit to be available in the same way as the Matlab is.

4 Publications Prepared Using MetaCentrum Resources

1. Jan Trmal, Marek Hrúz, Jan Zelinka, Pavel Campr, and Luděk Müller. Feature space transforms for czech sign-language recognition. In *Proceedings of the 9th Annual Conference of the International Speech Communication Association (Interspeech 2008)*, pages 2036–2039. Causal Production Pty ltd., 2008.

References

1. J. Flusser. Moment invariants in image analysis. *Proc. of World Academy of Science, Engineering and Technology*, 11:196–201, 2006.
2. K. Imagawa, S. Lu, and S. Igi. Color-based hand tracking system for sign language recognition. *Proc. Int. Conf. Automatic Face and Gesture Recognition.*, pages 462–467, 1998.

3. S.C.W. Ong and S. Ranganath. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27:873–891, 2005.
4. J. Sherrah and S. Gong. Resolving visual uncertainity and occlusion through probabilistic reasoning. *Proc. British Machine Vision Conf.*, pages 252–261, 2000.
5. J.-C. Terrillon, A. Piplr, Y. Niwa, and K. Yamamoto. Robust face detection and japanese sign language hand posture recognition for human-computer interaction in an "intelligent" room. *Proc. Int. Conf. Vision Interface*, pages 369–376, 2002.
6. Jan Trmal, Marek Hrúz, Jan Zelinka, Pavel Campr, and Luděk Müller. Feature space transforms for czech sign-language recognition. In *Proceedings of the 9th Annual Conference of the International Speech Communication Association (Interspeech 2008)*, pages 2036–2039. Causal Production Pty ltd., 2008.
7. Y. Xiong, B. Fang, and F. Quek. Extraction of hand gestures with adaptive skin color model and its applications to meeting analysis. *Proc. IEEE Int. Symposium on Multimedia 2006*, pages 647–651, 2006.

Relative Production Yields in Homologous Metallofullerene Series: Computations for $X@C_{74}$ and $Z@C_{82}$ Endohedrals

Filip Uhlík[1] and Zdeněk Slanina[2]

[1] Faculty of Science, Charles University, Hlavova 2030/8,
128 00 Prague 2, Czech Republic
[2] Center for Tsukuba Advanced Research Alliance, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan

Abstract. This contribution describes computations for endohedral metallofullerenes $Mg@C_{74}$, $Ca@C_{74}$, $Sr@C_{74}$ and $Ba@C_{74}$ based on encapsulation into the IPR (isolated pentagon rule) C_{74} cage, and for $Al@C_{82}$, $Sc@C_{82}$, $Y@C_{82}$ and $La@C_{82}$ with encapsulation into the C_{2v} IPR C_{82} cage. Their production yields are estimated using the encapsulation Gibbs-energy terms and saturated metal pressures. The results can be related to the ionization potentials of the free metal atoms.

1 Introduction

Although empty C_{74} fullerene [1] is not yet available in solid form, several related endohedral species $X@C_{74}$ have been known like $Ca@C_{74}$ [2, 3], $Sr@C_{74}$ [4], $Ba@C_{74}$ [5] or $La@C_{74}$ [6–8], all based on the isolated pentagon rule (IPR) D_{3h} C_{74} cage. Another common metallofullerene family, $Z@C_{82}$, is based on the IPR C_{2v} C_{82} cage, for example, $Sc@C_{82}$ [9], $Y@C_{82}$ [10] and $La@C_{82}$ [6, 11]. The present paper deals with computational evaluations of the relative equilibrium production yields in the series $X@C_{74}$ and $Z@C_{82}$.

2 Calculations

The geometry optimizations were carried out using density-functional theory (DFT), namely employing Becke's three parameter functional with the non-local Lee-Yang-Parr correlation functional (B3LYP) in the combined basis set of the 3-21G basis for C atoms and the LanL2DZ basis set with the effective core potential for the metal atoms (3-21G∼dz) as implemented in the Gaussian 03 program package [12]. In the optimized B3LYP/3-21G∼dz geometries, the harmonic vibrational analysis was then performed as well as higher-level single-point energy calculations with the standard 6-311G* ($X@C_{74}$) and 6-31+G* ($Z@C_{82}$) basis set for C atoms. As an example, the structure of $Ba@C_{74}$ is shown in Figure 1. The Gibbs energies were evaluated using the rotational-vibrational partition functions constructed from the calculated structural and vibrational

data using the rigid rotator and harmonic oscillator (RRHO) approximation. Although the temperature region where fullerene or metallofullerene electric-arc synthesis takes place is not yet known, there are some arguments to expect it around or above 1500 K. Thus, the computations here are presented for two illustrative temperatures of 1500 and 2000 K.

3 Results and Discussion

We can consider an overall stoichiometry of a metallofullerene formation:

$$\mathrm{X}(g) + \mathrm{C}_n(g) = \mathrm{X@C}_n(g). \tag{1}$$

The equilibrium composition of the reaction mixture is controlled by the encapsulation equilibrium constants $K_{X@C_n,p}$:

$$K_{X@C_n,p} = \frac{p_{\mathrm{X@C_n}}}{p_{\mathrm{X}} p_{\mathrm{C_n}}}, \tag{2}$$

expressed in the terms of partial pressures of the components. Let us further suppose that the metal pressure p_{X} is actually close to the respective saturated pressure $p_{X,sat}$. While the saturated pressures $p_{X,sat}$ for various metals are known from observations [13], the partial pressure of C_n is less clear as it is obviously influenced by a larger set of processes (though, p_{C_n} should exhibit a temperature maximum and then vanish). Therefore, we avoid the latter pressure in our considerations at this stage. However, if we consider the combined $p_{X,sat}K_{X@C_n,p}$ term:

$$p_{\mathrm{X@C_n}} \sim p_{X,sat}K_{X@C_n,p}, \tag{3}$$

that directly controls the partial pressures of various $\mathrm{X@C}_n$ encapsulates in an endohedral series (based on one common C_n fullerene), we get an applicable scheme. Actually, the considered $p_{X,sat}K_{X@C_n,p}$ term can frequently (though not necessarily) be increasing with temperature. An optimal production temperature could be evaluated in a more complex model that also includes temperature development of the empty-fullerene partial pressure.

Hence, if we want to evaluate production abundances in a series of metallofullerenes like $\mathrm{Mg@C_{74}}$, $\mathrm{Ca@C_{74}}$, $\mathrm{Sr@C_{74}}$ and $\mathrm{Ba@C_{74}}$, just the product $p_{X,sat}K_{X@C_{74},p}$ terms can straightforwardly be used. The results in Table 1 show several interesting features. While for $\mathrm{Mg@C_{74}}$ and $\mathrm{Ca@C_{74}}$ the $p_{X,sat}K_{X@C_{74},p}$ quotient increases with temperature, it is about constant for $\mathrm{Sr@C_{74}}$ for the considered temperatures, and it decreases with temperature for $\mathrm{Ba@C_{74}}$. The behavior results from competition between the decreasing encapsulation equilibrium constants and increasing saturated metal pressures. As the encapsulation enthalpy $\Delta H^{o}_{X@C_n}$ has the most negative value for $\mathrm{Ba@C_{74}}$, its encapsulation equilibrium constant has to exhibit the fastest temperature decrease that already cannot be overcompensated by the temperature increase of the saturated metal pressure so that the $p_{X,sat}K_{X@C_{74},p}$ quotient decreases with temperature in this case. In order to allow for cancellation of various factors introduced by

the computational approximations involved, it is better to deal with the relative quotient $p_{X,sat}K_{X@C_{74},p}/(p_{Ba,sat}K_{Ba@C_{74},p})$. Table 1 shows that the production yield of Sr@C_{74} should be by two or three orders of magnitude smaller than that for Ba@C_{74}. For Ca@C_{74} the production yield for the considered temperatures is computed to be between three and five orders of magnitude lower than for Ba@C_{74} while for Mg@C_{74} even by seven to nine orders of magnitude lower. In principle, an endohedral with lower value of the encapsulation equilibrium constant can still be produced in larger yields if a convenient over-compensation by higher saturated metal pressure can take place.

Although the energy terms are likely still not precise enough, their errors could be comparable in the series and thus, they should cancel out in the relative term $p_{X,sat}K_{X@C_{74},p}/(p_{Ba,sat}K_{Ba@C_{74},p})$. This should be the case of, for example, the basis set superposition error important for evaluation of the encapsulation potential-energy changes. A similar cancellation should also operate for the higher corrections to the RRHO partition functions, including motions of the encapsulate. The motion of the endohedral atom is highly anharmonic, however, its description is yet possible only with simple potential functions. As long as we are interested in the relative production yields, the anharmonic effects should at least to some extent be canceled out in the relative quotient $p_{X,sat}K_{X@C_{74},p}/(p_{Ba,sat}K_{Ba@C_{74},p})$. Incidentally, the computed stability proportions do correlate with qualitative abundances known from observations. For Ba@C_{74} even microcrystals could be prepared [5] so that a diffraction study was possible, while for Sr@C_{74} at least various spectra could be recorded [4] in solution, Ca@C_{74} was studied [3] only by NMR spectroscopy while Mg@C_{74} has never been isolated. Table 2 offers a similar survey for the Z@C_{82} (Z = Mg, Sc, Y, La) series, confirming the highest yield for La@C_{82} and an extremely small population for Al@C_{82}, as expected.

The series of metallofullerene formations with one common cage X@C_n allows for another interesting stability conclusion. Three formal reaction steps can be considered for our illustrative series Mg@C_{74}, Ca@C_{74}, Sr@C_{74}, and Ba@C_{74}: (i) double-ionization of the free metal, (ii) double charging of the empty cage, and (iii) placing the metal dication into the dianionic cage. The (ii) energy is identical for all members of the series, and the (iii) terms should be similar as they are controlled by electrostatics (the metal-cage interactions form ionic, not covalent bonds). Hence, the free-metal ionization potentials should actually represent a critical yield-controlling factor—the computed relative potential-energy changes upon encapsulation $\delta_{rel}\Delta E$ and the relative observed ionization potentials of the free atoms $\delta_{rel}IP$ should according to the above three-step analysis be correlated:

$$\delta_{rel}\Delta E \sim \delta_{rel}IP. \tag{4}$$

This interesting conclusion is documented in Figure 2 that uses both the observed second and first ionization potentials (IP) [14] for X@C_{74} (though the second IP are more relevant here; for methodological reasons a newer functional, MPWB1K/6-31G*∼dz, is used in this case though it does not bring a qualitative difference), and the third and second IP for the Z@C_{82} series. In fact,

this correlation should operate for such a homologous reaction series of metal encapsulations into any type of carbon nanostructures. Moreover, this type of reasoning should step by step explain the fullerene-encapsulation stability islands known throughout the periodic system (though the underlying calculations are quite demanding).

In fact, we are dealing with a special case of clustering under saturation conditions. The saturation regime is a useful simplification—it is well defined, however, it is not necessarily always achieved. Under some experimental arrangements, under-saturated or perhaps super-saturated metal vapors are also possible. This reservation is applicable not only to the electric-arc treatment but even more likely to newly introduced ion-bombardment production technique [15]. Still, eqs. (2) and (3) remain valid, however, the metal pressure has to be described by the values actually relevant. For some volatile metals their critical temperature could be overcome and the saturation region thus abandoned (though practically speaking, this could come into consideration with mercury and cesium). Anyhow, the saturation regime can give a kind of upper-limit estimates of the production yields.

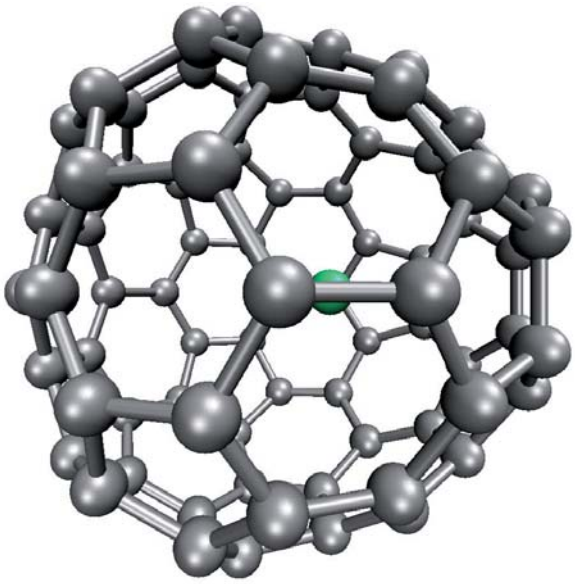


Fig. 1. The structure of $Ba@C_{74}$.

Endohedral	$p_{X,sat}K_{X@C_{74},p}$		$\frac{p_{X,sat}K_{X@C_{74},p}}{p_{Ba,sat}K_{Ba@C_{74},p}}$	
	$T = 1500\,\mathrm{K}$	$T = 2000\,\mathrm{K}$	$T = 1500\,\mathrm{K}$	$T = 2000\,\mathrm{K}$
Mg@C_{74}	1.90×10^{-7}	4.06×10^{-6}	2.14×10^{-9}	3.66×10^{-7}
Ca@C_{74}	2.33×10^{-3}	7.60×10^{-3}	2.62×10^{-5}	6.85×10^{-4}
Sr@C_{74}	0.397	0.293	4.46×10^{-3}	0.0264
Ba@C_{74}	89.0	11.1	1.0	1.0

Table 1. The products of the encapsulation equilibrium constant $K_{X@C_{74},p}$ with the metal saturated-vapor pressure (extracted from available observed data [13]) $p_{X,sat}$ for Mg@C_{74}, Ca@C_{74}, Sr@C_{74}, and Ba@C_{74} computed at two temperatures T.

Endohedral	$p_{Z,sat}K_{Z@C_{82},p}$		$\frac{p_{Z,sat}K_{Z@C_{82},p}}{p_{La,sat}K_{La@C_{82},p}}$	
	$T = 1500\,\mathrm{K}$	$T = 2000\,\mathrm{K}$	$T = 1500\,\mathrm{K}$	$T = 2000\,\mathrm{K}$
Al@C_{82}	1.09×10^{-6}	3.77×10^{-5}	6.03×10^{-8}	4.99×10^{-6}
Sc@C_{82}	5.16	3.66	0.286	0.484
Y@C_{82}	2.57	2.93	0.142	0.388
La@C_{82}	18.1	7.56	1.0	1.0

Table 2. The products of the encapsulation equilibrium constant $K_{Z@C_{82},p}$ with the metal saturated-vapor pressure (extracted from available observed data [13]) $p_{Z,sat}$ for Al@C_{82}, Sc@C_{82}, Y@C_{82}, and La@C_{82} computed at two temperatures T.

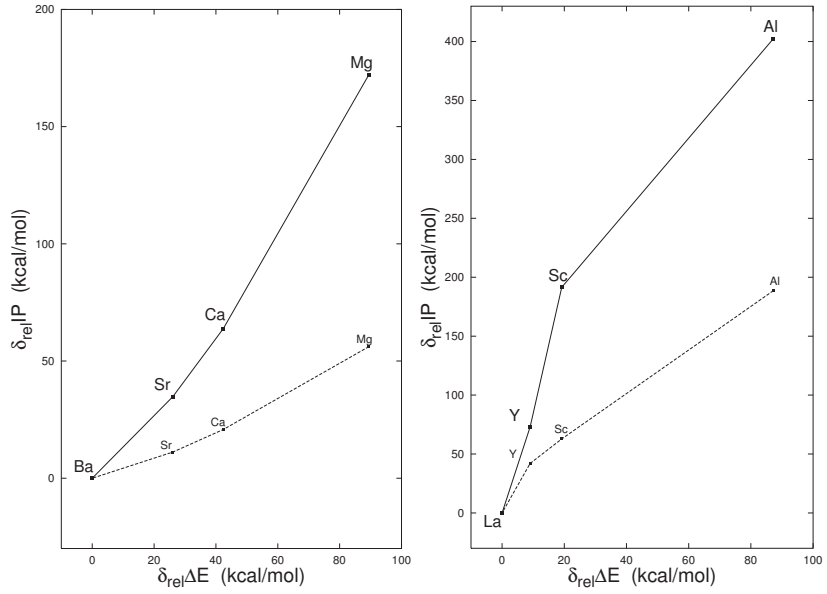


Fig. 2. The computed relative potential-energy changes upon encapsulation $\delta_{rel}\Delta E$ and the observed [14] relative ionization potentials (IP) of the free atoms $\delta_{rel}IP$ for X@C_{74} (left: solid line 2-nd IP, dashed line 1-st IP) and Z@C_{82} (right: solid line 3-rd IP, dashed line 2-nd IP).

4 Acknowledgments

The research has been supported by the Ministry of Education of the Czech Republic (MSM0021620857) and by a Grant-in-aid for the 21st Century COE Program, Nanotechnology Support Project, the Next Generation Super Computing Project (Nanoscience Project), and Scientific Research on Priority Area from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. M. D. Diener, and J. M. Alford, Nature **393**, 668–671 (1998).
2. T. S. M. Wan, H. W. Zhang, T. Nakane, Z. D. Xu, M. Inakuma, H. Shinohara, K. Kobayashi, and S. Nagase, J. Am. Chem. Soc. **120**, 6806–6807 (1998).
3. T. Kodama, R. Fujii, Y. Miyake, S. Suzuki, H. Nishikawa, I. Ikemoto, K. Kikuchi, and Y. Achiba, Chem. Phys. Lett. **399**, 94–97 (2004).
4. O. Haufe, M. Hecht, A. Grupp, M. Mehring, and M. Jansen, Z. Anorg. Allgem. Chem. **631**, 126–130 (2005).
5. A. Reich, M. Panthofer, H. Modrow, U. Wedig, and M. Jansen, J. Am. Chem. Soc. **126**, 14428–14434 (2004).

6. Y. Chai, T. Guo, C. Jin, R. E. Haufler, L. P. F. Chibante, J. Fure, L. Wang, J. M. Alford, and R. E. Smalley, J. Phys. Chem. **95**, 7564–7568 (1991).
7. K. Sueki, K. Akiyama, T. Yamauchi, W. Sato, K. Kikuchi, S. Suzuki, M. Katada, Y. Achiba, H. Nakahara, T. Akasaka, and K. Tomura, Full. Sci. Technol. **5**, 1435–1448 (1997).
8. H. Nikawa, T. Kikuchi, T. Wakahara, T. Nakahodo, T. Tsuchiya, G. M. A. Rahman, T. Akasaka, Y. Maeda, K. Yoza, E. Horn, K. Yamamoto, N. Mizorogi, and S. Nagase, J. Am. Chem. Soc. **127**, 9684–9685 (2005).
9. E. Nishibori, M.Takata, M. Sakata, M. Inakuma, and H. Shinohara, Chem. Phys. Lett. **298**, 79–84 (1998).
10. M. Takata, B. Umeda, E. Nishibori, M. Sakata, Y. Saito, M. Ohno, and H. Shinohara, Nature **377**, 46–48 (1995).
11. T. Akasaka, T. Wakahara, S. Nagase, K. Kobayashi, M. Waelchli, K. Yamamoto, M. Kondo, S. Shirakura, S. Okubo, Y. Maeda, T. Kato, M. Kako, Y. Nakadaira, R. Nagahata, X. Gao, E. van Caemelbecke, and K. M. Kadish, J. Am. Chem. Soc. **122**, 9316–9317 (2000).
12. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian 03, Revision C.01, Gaussian, Inc., Wallingford, CT, 2004.
13. C. B. Alcock, V. P. Itkin, and M. K. Horrigan, Can. Metallurg. Quart. **23**, 309–313 (1984).
14. CRC Handbook of Chemistry and Physics, 85th edition, edited by D. R. Lide, CRC Press, Boca Raton, FL, 2004, pp. 10-183–10-184.
15. E. E. B. Campbell, Fullerene Collision Reactions, Kluwer Academic Publishers, Dordrecht, 2003.

Author Index

Ivana Křenková, David Antoš, and Luděk Matyska (Eds.)

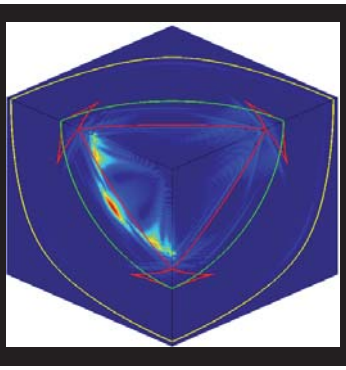
MetaCentrum Yearbook 2009

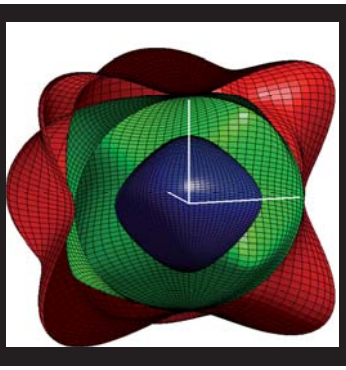
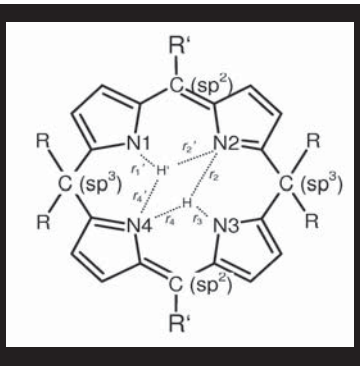
Publisher: CESNET, z. s. p. o., Zikova 4, 160 00 Prague, Czech Republic
Printing: PBtisk, s.r.o., Dělostřelecká 344, Příbram, Czech Republic
Edition: first
Year of publication: 2010

Typesetting, data conversion and design by David Antoš

Cover design: Ivana Křenková

ISBN 978-80-904173-7-3



MetaCentrum Yearbook 2009
ISBN 978-80-904173-7-3

http://meta.cesnet.cz
http://metacentrum.cz

ISBN 978-80-904173-7-3