Algorithms and computational technology for extremum search in optimal control problems

A.Yu. Gornov,

A.S. Anikin, T.S. Zarodnyuk, E.A. Finkelshtein

Matrosov Institute for System Dynamics and Control Theory of SB RAS

gornov@icc.ru

The 11th International Conference on **Intelligent Data Processing: Theory and Applications (IDP-2016)**

10-14 October 2016 in Barcelona, Spain

Optimal Control Problem,
box control restrictions

$$\dot{x} = f(x, u, t)$$

$$x(t_0) = x^0, \quad t \in T = [t_0, t_1]$$

$$u(t) \in U = \{u \in R^r : \underline{u}_i \le u_i \le \overline{u}_i\}$$

$$I(u) = \varphi(x(t_1)) \to \min$$

Possible approaches to search of a global extremum in optimum control problems

- Reduction to a problem of mathematical programming
- Method of casual multistart
- Tunnel methods
- Methods of integrated representations (Chichinadze-method)
- Methods of casual search
- Convexification methods
- Methods of consecutive search of local extrema
- Evolutionary algorithms
- Methods of the decision of the equation of Krotov-Bellman
- Methods based on approximations of reachable set
- Shepard approximation methods
- Methods of casual coverings

Considerated approaches to search of a global extremum in optimum control problems

- Reduction to a problem of mathematical programming
- Method of casual multistart
- Tunnel methods
- Methods of integrated representations (Chichinadze-method)
- Methods of casual search
- Convexification methods
- Methods of consecutive search of local extrema
- Evolutionary algorithms
- Methods of the decision of the equation of Krotov-Bellman
- Methods based on approximations of reachable set
- Shepard approximation methods
- Methods of casual coverings

Software

OPTCON (OPTimal CONtrol)

- **MAPR** (1980-1986, ЭВМ БЭСМ-6)
- **KONUS** (1986-1990, ЕС ЭВМ)
- **OPTCON-I** (1989-2001, MS-DOS)
- **OPTCON-II** (2002-2003, Интернет сервер)
- **OPTCON-III** (2004-2010, Windows 95/98/2000/XP/VISTA/LINUX)
- **OPTCON-IV** (2011, Windows/LINUX)
- **OPTCON-F** (2015-2016, Windows/LINUX)

Tjatjushkin A.I.
Zholudev A.I.
Erinchek N.M.
Pinegina T.N.
Zarodnyuk T.S.
Podkamenniy D.V.
Madzhara T.I.

Daneeva A.V.
Anikin A.S.
Golomolzhina T.A.
Veyalko I.A.
Finkelshtein E.A.
Dorzhieva A.B.
Khandarov F.V.

Theoretical foundations of proposed algorithms

- The necessary conditions of optimality (Pontryagin maximum principle, the linearized maximum principle);
- Theory and methods of phase estimation (approximation methods of integral funnels and reachable set);
- The theory of global search in finite-dimensional extreme problems (multistart methods, tunneling methods, methods of grids);
- Lipschitz optimization theory (methods of coverings);
- Convexification theory (R.V.Gamkrelidze, A.A.Tolstonogov, B.Sh.Morduhovich).

Theoretical foundations of proposed algorithms

- The theory of dynamical programming (Hamilton-Jacobi equations);
- The theory of evolutionary programming (genetic algorithms);
- The theory of deterministic global optimization (stochastic approximation methods, Monte-Carlo methods);
- Approximation theory (operator of Shepard).

Methods of local extremum search

- Algorithms based on the maximum principle;
- Reduced gradient algorithm;
- Conjugate gradient algorithm;
- Nesterov’s ravine algorithm;
- Spectral Projected Gradient of Yevtushenko;
- Quasi-Newton algorithm BFGS;
- Quasi-Newton algorithm DFP;
- Newton algorithm;
- Algorithm of coordinate-wise search;
- Powell's search algorithm.

Global search algorithms (one-dimentional variant)

- Algorithm of Strongin;
- Algorithm of Yevtushenko;
- Brent's algorithm;
- Algorithm of "parabol";
- "Spline" - algorithm;
- Zhiglyavskii algorithms (, 3 versions).

Phase estimation algorithms for solving of optimal control problems

- Approximation algorithms of support function (A.I.Tyatyushkin);
- Stochastic approximation algorithms;
- Volume maximization algorithm;
- Reduction algorithms to the sequence of the optimal control problems;
- Algorithm are based on the maximum principle ("Ratszinskii's method");
- Algorithms are based on the boundary equation of the integral funnel.

Implemented algorithms

- Algorithm of random multistart;
- Algorithm of grids;
- Consecutive discretization algorithm;
- Convexification method;
- Curvilinear search algorithm;
- Algorithm of random coverings;
- Shepard's algorithm;
- Genetic algorithm;
- Algorithm is based on the maximum principle;
- Tunneling algorithm;

Areas of application

- Flight dynamics;
- Space navigation;
- Mechanics;
- Robotics;
- Electrical power engineering;
- Chemistry;
- Materials science;
- Quantum physics;
- Economics;
- Ecology
- Geography;
- Medicine;
- Criminal science;
- Materials science;
- Seismology, etc.

Applied problems
Flight dynamics

- Investigation of critical aircraft dynamics in general flight maneuver and sortie modes;
- Task of optimal maneuvers synthesis in the frame “aircraft against radar”;
- Problem of landing a heavy aircraft (“Buran”) to the maximum range.

- *- Chelyabinsk higher military school of navigators;*
- *- Ramenskoe instrument-making design office;*
- *- State research Institute of aviation systems.*

Applied problems
Flight dynamics

- Computation of aircraft maneuver for protection against missiles attacking from the rear hemisphere;
- Optimization of spatial maneuvers of the helicopter and modes for engine failure.

- *- Sukhoi experimental design office;*
- *- Gromov flight research institute;*
- *- Kamov helicopter factory.*

Applied problems
Space navigation

- Task of orbital spacecraft orientation;
- Problem of the spacecraft landing to the Earth, Moon, Mars.

- *- Central specialized design bureau "Progress";*
- *- Rocket-space corporation "Energy".*

Applied problems
Electrical power engineering

- Optimization of operation modes of electric networks with DC.

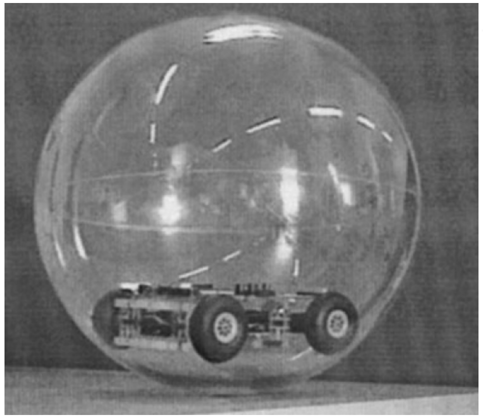
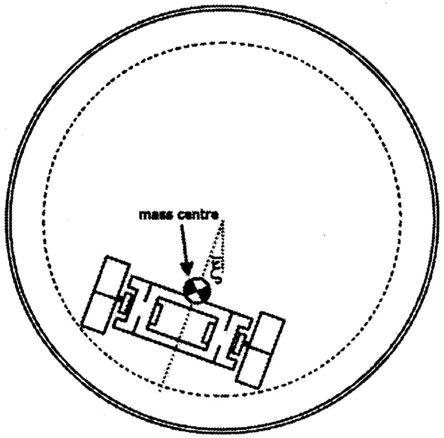
- *- Melentiev energy systems institute.*

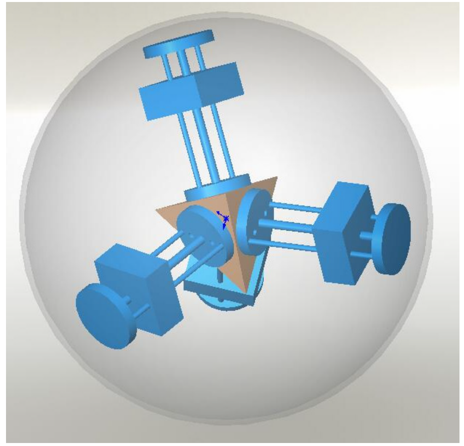
Applied problems
Robotics

- Optimization of anthropogenic robot movement;
- Analysis of spherical robot dynamics.

- *- Irkutsk state transport university;*
- *- Kyushu university, Fukuoka, Japan.*

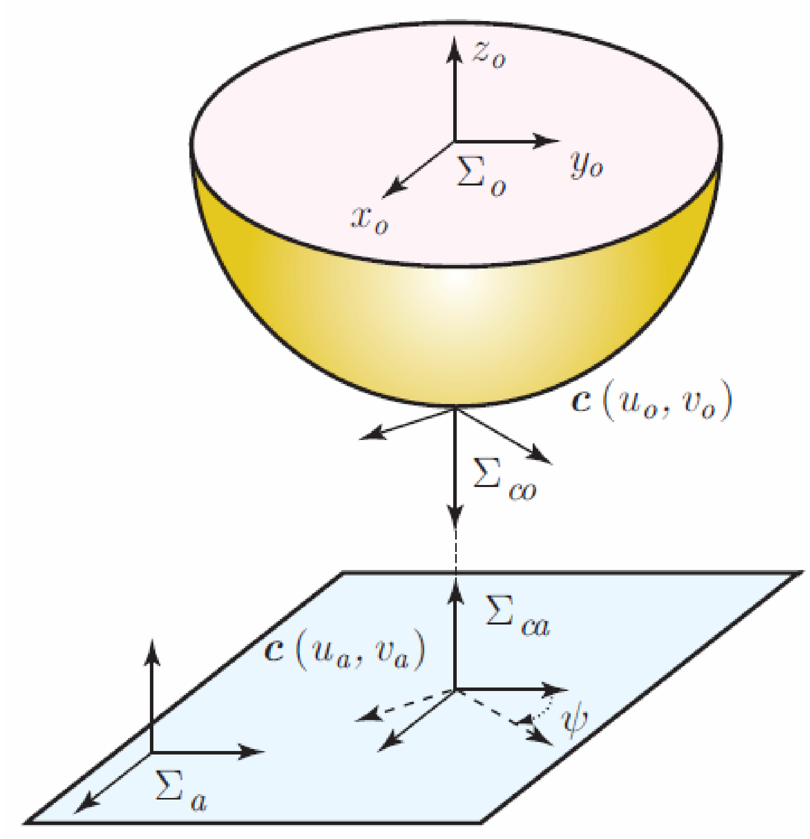
Examples of spherical robots

[1] A. Halme, T. Schonberg, and Y. Wang. Motion control of a spherical mobile robot. In Advanced Motion Control, 1996. AMC'96-MIE. Proceedings., 1996 4th International Workshop on, volume 1, pages 259{264. IEEE, 1996.
[2] J. Alves and J. Dias. Design and control of a spherical mobile robot. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering,217(6):457{467, 2003.
[3] V. Joshi, R. Banavar, Motion analysis of a spherical mobile robot, Robotica, 2009, 343-353, doi: 10.1017/ S0263574708004748.
[4] Shengju Sang, Jichao Zhao, Hao Wu, Shoujun Chen, Qi An: Modeling and simulation of a spherical mobile robot. Comput. Sci. Inf. Syst. 7(1): 51-62 (2010).

The geometric meaning



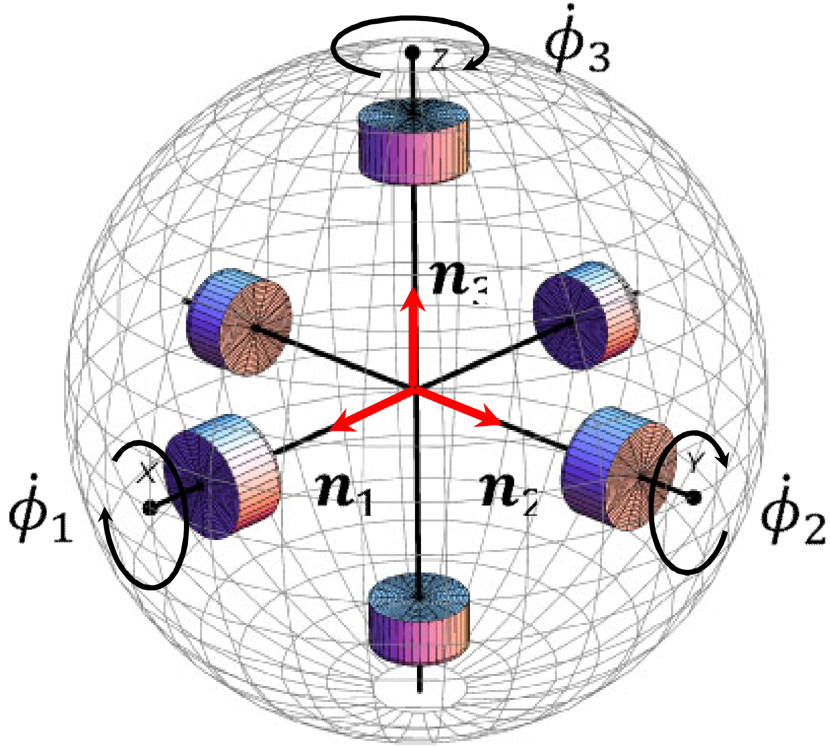
The inertial coordinate system $\sum_a$ and the coordinate system $\sum_o$ related to the scope of its geometrical center.

The auxiliary coordinate system $\sum_{ca}$ and $\sum_{co}$

The problem for spherical mobile robot optimal control with three-dimensional control

Dynamic model

- $\dot{x} = G(x)J^{-1}(x)J_r \sum_{k=1}^{n} n_k(x)u_k$,
- state and control are defined as
- $x \triangleq [u_a, v_a, u_o, v_o, \psi]^T$,
- $u \triangleq [\dot{\varphi}_1, \dot{\varphi}_2, \dot{\varphi}_3]^T$,
- a $\varphi_i, i = \overline{1,3}$ denote the angles of engines rotation.



Finkelstein E.A. (ISDCT), Svinin M.M. (Mechanical Engineering Department, Faculty of Engineering, Kyushu University, Fukuoka, Japan)

Dynamic model

$$G = \begin{bmatrix} 0 & -R & 0 \\ R & 0 & 0 \\ \sin\psi/\cos v_o & \cos\psi/\cos v_o & 0 \\ \cos\psi & -\sin\psi & 0 \\ \sin\psi\tan v_o & \cos\psi\tan v_o & 1 \end{bmatrix},$$

vectors $\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3$ are the columns of the matrix

$$R = \begin{bmatrix} \cos u_o\cos\psi + \sin u_o\sin v_o\sin\psi & \cos v_o\sin\psi & -\sin u_o\cos\psi + \cos u_o\sin v_o\sin\psi \\ -\cos u_o\sin\psi + \sin u_o\sin v_o\cos\psi & \cos v_o\cos\psi & \sin u_o\sin\psi + \cos u_o\sin v_o\cos\psi \\ \sin u_o\cos v_o & -\sin v_o & \cos u_o\cos v_o \end{bmatrix}$$

The inertia matrix of the system is defined as

$$J = \begin{bmatrix} (2/3m_o + M)R^2 & 0 & 0 \\ 0 & (2/3m_o + M)R^2 & 0 \\ 0 & 0 & 2/3m_oR^2 \end{bmatrix} + (2J_p + J_r)E,$$

where M is a total mass of the robot, m_o and m_r denote the mass of the spherical shell and a separate rotor.

Statement of the optimal control problem

Optimality criterion for control norm

$$J_1 = \int_0^T \boldsymbol{u}^T \boldsymbol{u} \, \mathrm{d}t,$$

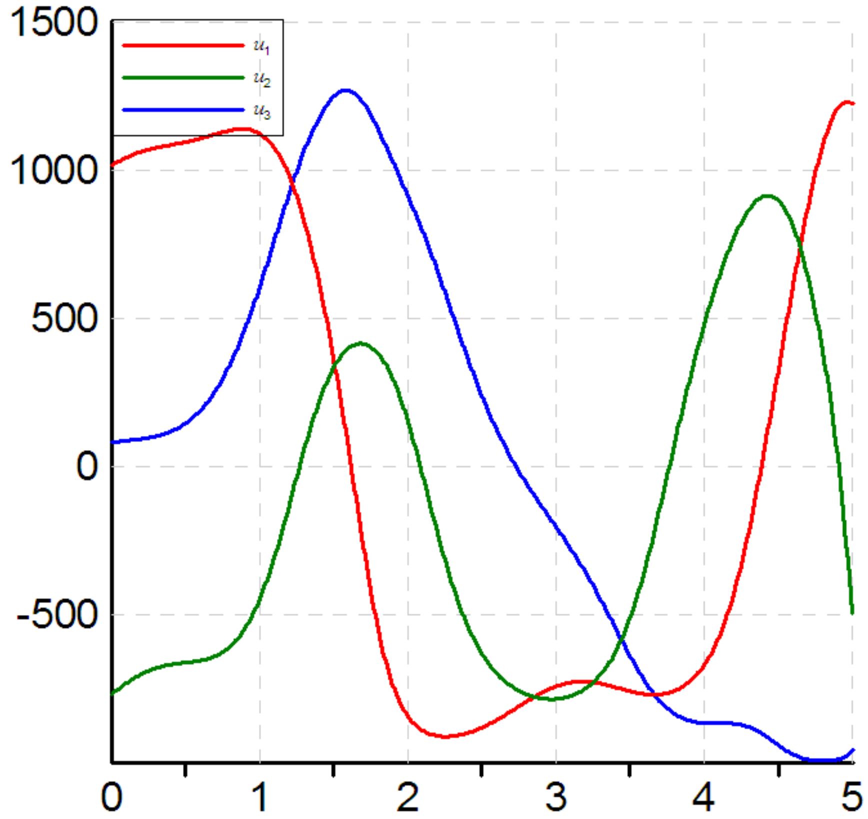
The initial state of the system is defined as

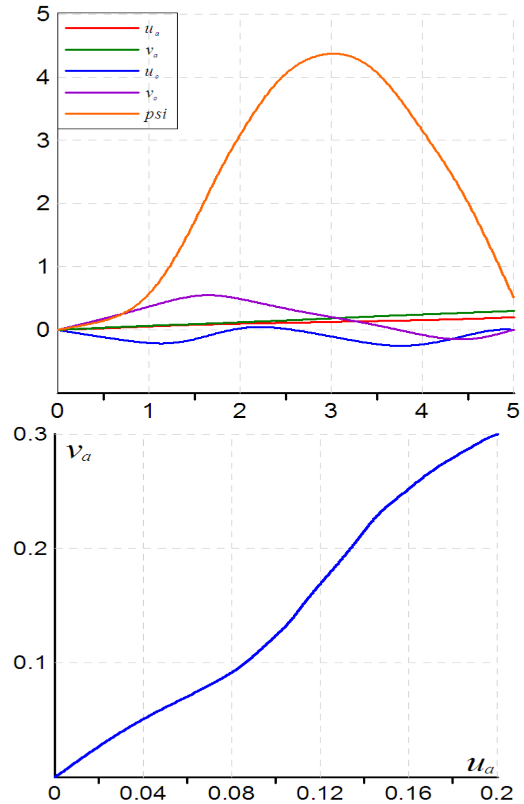
$$\boldsymbol{x}(\mathbf{0}) = [\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}]^T,$$

The final state $\boldsymbol{x}(\boldsymbol{T})$ is fixed.

The computing experiments

Transfer problem from point $x(0) = [0,0,0,0,0]^T$ to point $x(T) = [0{,}2,\ 0{,}3,\ 0,\ 0, \frac{\pi}{6}]$.

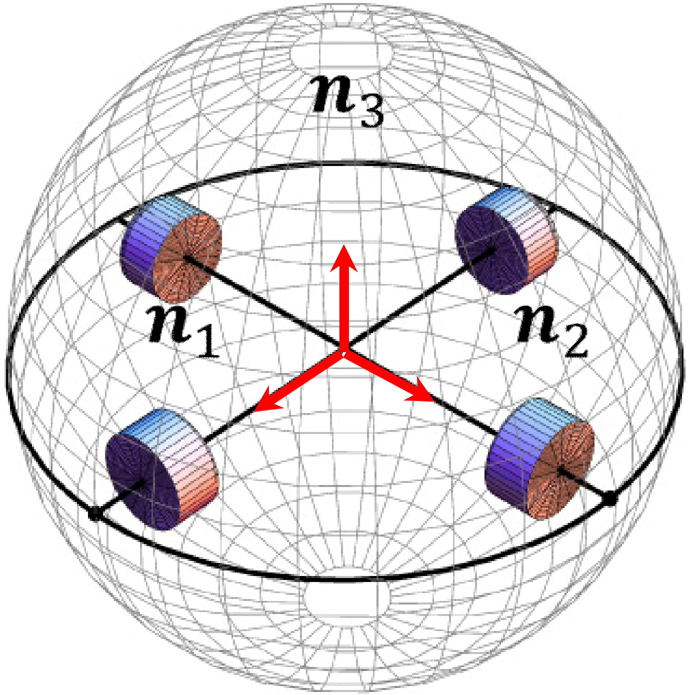




The trajectories of the phase variables and the trajectory of the contact point on the plane

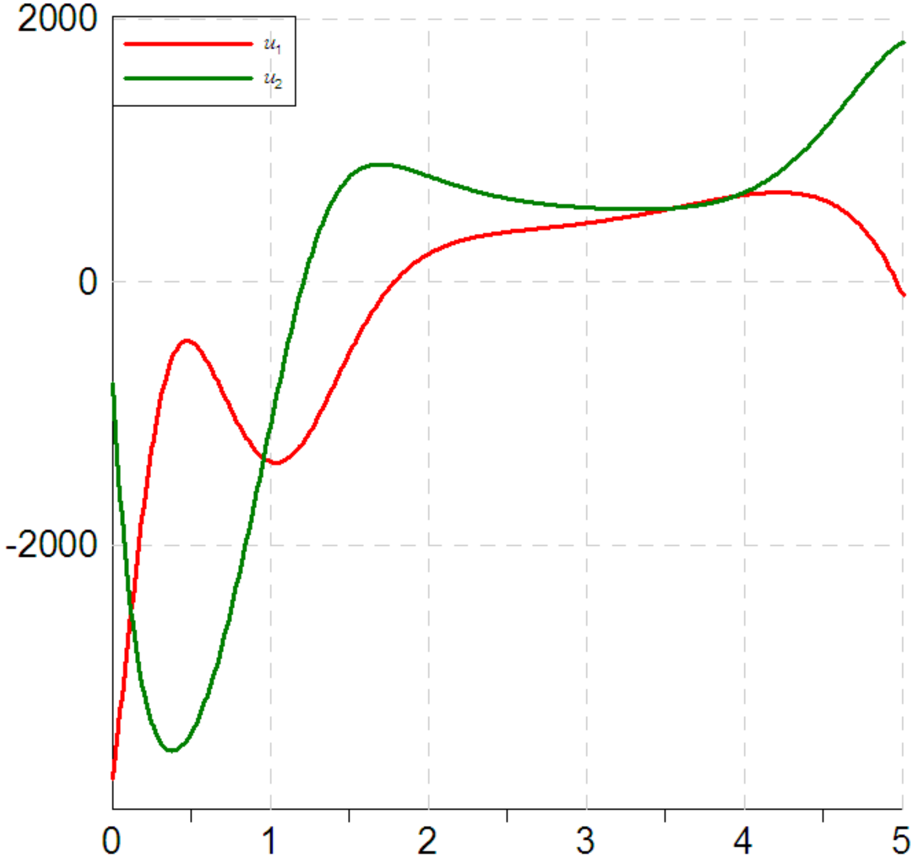
The problem for spherical mobile robot optimal control with two controls

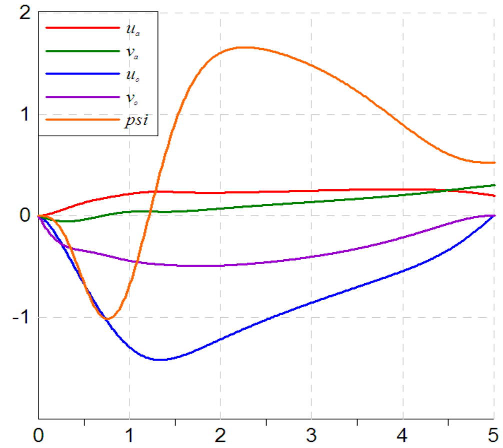
- $\dot{x} = G(x)J^{-1}(x)J_r \sum_{k=1}^{2} n_k(x)u_k$,
- $u \triangleq [\dot{\varphi}_1, \dot{\varphi}_2]^T$
- If the contact point lies on the equator
- $u_o = \pm(2k+1)\frac{\pi}{2}, k = 0, 1, 2, \dots$,
- then the plane of the rotors arrangement becomes perpendicular of the touching plane. The system becomes locally uncontrollable, there is a physical singularity.

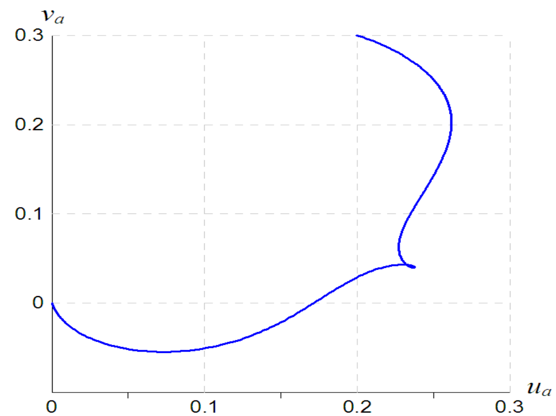


The computing experiments

- Transfer problem from point $x(0) = [0,0,0,0,0]^T$ to point $x(T) = [0{,}2,\ 0{,}3,\ 0,\ 0,\frac{\pi}{6}]$.







The trajectories of the phase variables and the trajectory of the contact point on the plane

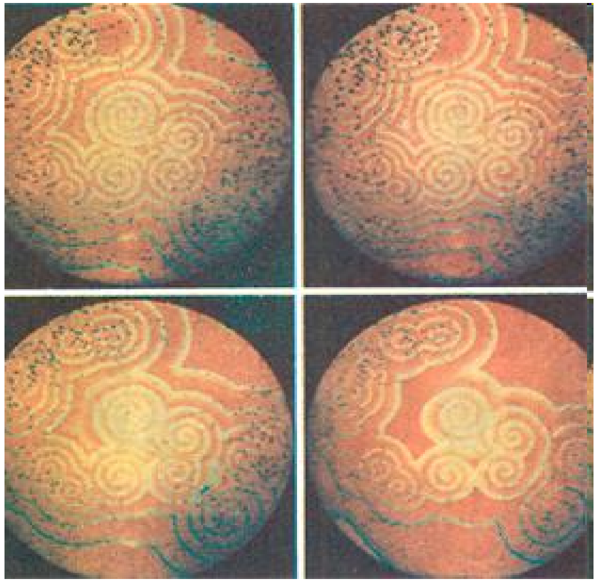
Applied problems
Chemistry

- Model identification and search of oscillating heterogeneous catalytic reactions;
- Software developing for analyzing electrocardiogram data of experimental animals.

- *- Boreskov catalysis institute, Novosibirsk, Russia;*
- *- Vorozhtsov Novosibirsk institute of organic chemistry.*

The phenomenon of self-oscillations

•The processes of spatial and temporal self-organization very often observed in various catalytic systems.



•In homogeneous catalysis these phenomena were discovered by B. P. Belousov in 1951.

•In heterogeneous catalysis for the first time oscillations of the reaction rate observed in the oxidation of CO on platinum in the early 70-ies of the last century.

Kinetic model of methane oxidation on the nickel surface

Consider the system describing the methane oxidation on the nickel surface

$$x_1' = k_1 \boldsymbol{P_{CH4}}(1-S) - k_2 x_1 - k_3(1-S)x_1,$$
$$x_2' = (k_3 x_1 - k_4 x_2)(1-S),$$
$$x_3' = (k_4 x_2 - k_5 x_3)(1-S),$$
$$x_4' = (k_5 x_3 - k_6 x_4)(1-S),$$
$$x_5' = (k_3 x_1 + k_4 x_2 + k_5 x_3 + k_6 x_4)(1-S) - 2k_{10} x_5^2 - k_{16} x_5 x_7 - k_{17} x_5 x_8 - k_{18} x_5 x_{10},$$
$$x_6' = k_6 x_4 (1-S) - k_9 x_6 x_7 - k_{12} x_6 x_8,$$
$$x_7' = 2k_7 \boldsymbol{P_{O2}}(1-S)^2 - 2k_8 x_7^2 - k_9 x_{6_{x7}} - k_{11} x_7 - k_{14} x_9 x_7 - k_{16} x_5 x_7,$$
$$x_8' = k_{11} x_7 k_{12} x_6 x_8 - k_{15} x_9 x_8 - k_{17} x_5 x_8 - 4k_{19} \boldsymbol{P_{CH4}} x_8^4,$$
$$x_9' = k_9 x_6 x_7 + k_{12} x_6 x_8 - k_{13} x_9 - k_{14} x_9 x_7 - k_{15} x_9 x_8,$$
$$x_{10}' = k_{16} x_5 x_7 + k_{17} x_5 x_8 - k_{18} x_5 x_{10}.$$

Kinetic model of methane oxidation on the nickel surface

Definition area: $0 \le x_i \le 1$, $S = \sum_{i=1}^{10} x_i \le 1$,

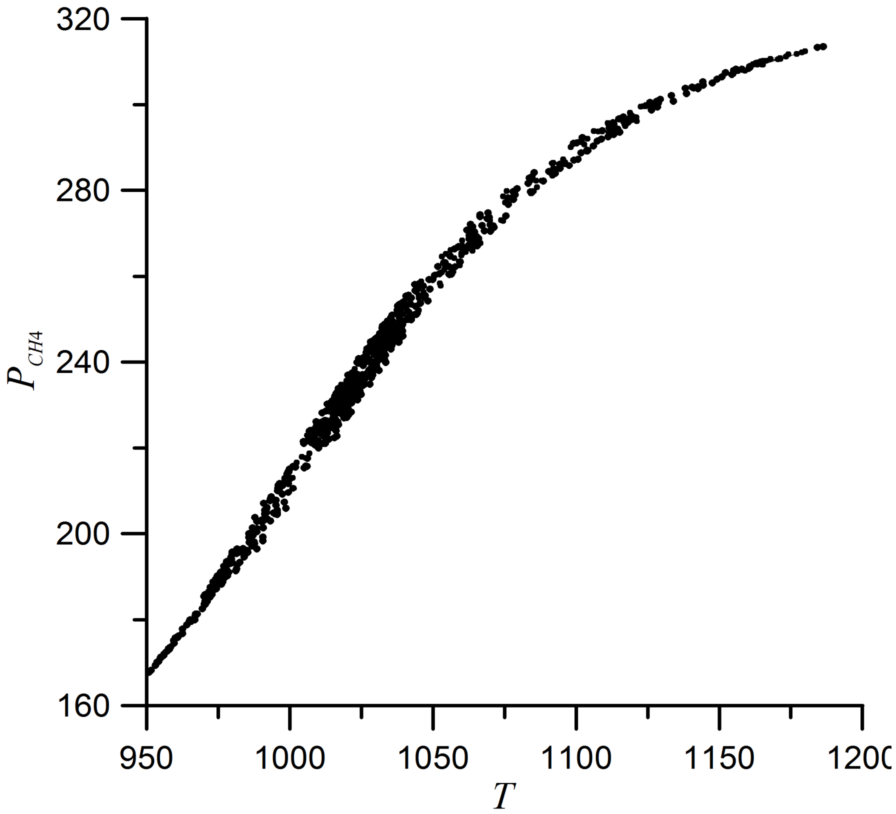
$$k_j = k_j^0 \exp\left\{-\frac{E_j}{R\boldsymbol{T}}\right\},$$

where R, k_j^0, E_j defined and fixed, temperature $\boldsymbol{T}$ and partial pressure $\boldsymbol{P_{CH_4}}, \boldsymbol{P_{O_2}}$ set by admissible intervals: $T \in [800, 1200]$,

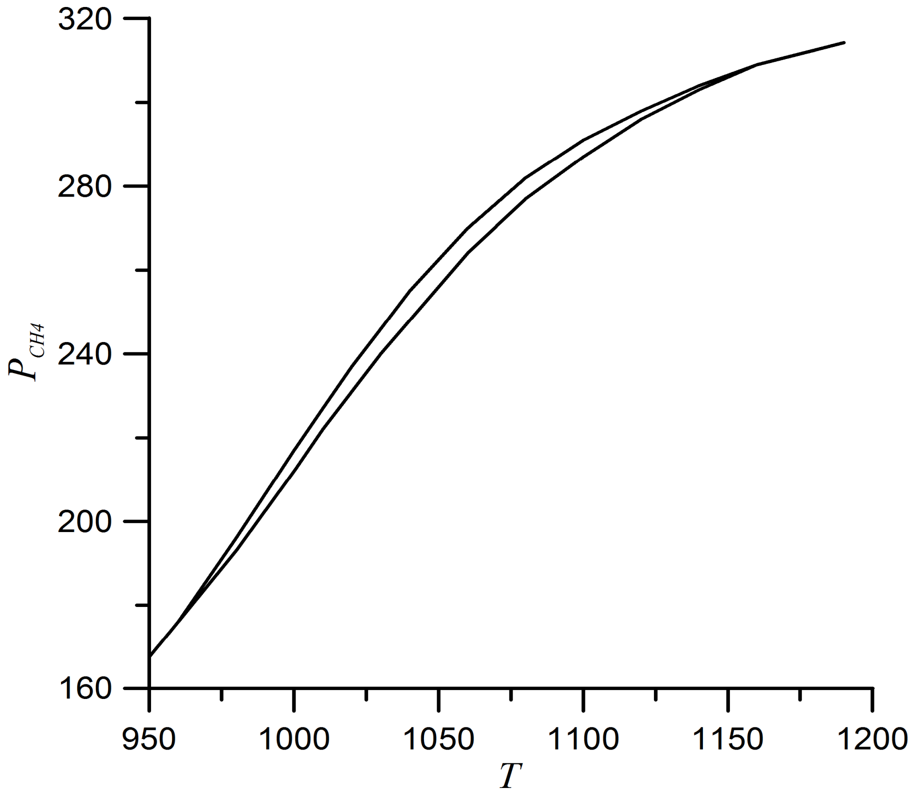
$P_{CH_4} + P_{O_2} = 329$ Торр, $\frac{P_{CH_4}}{P_{O_2}} \in [1,30]$.

The domain of oscillations existence

Stochastic approximation

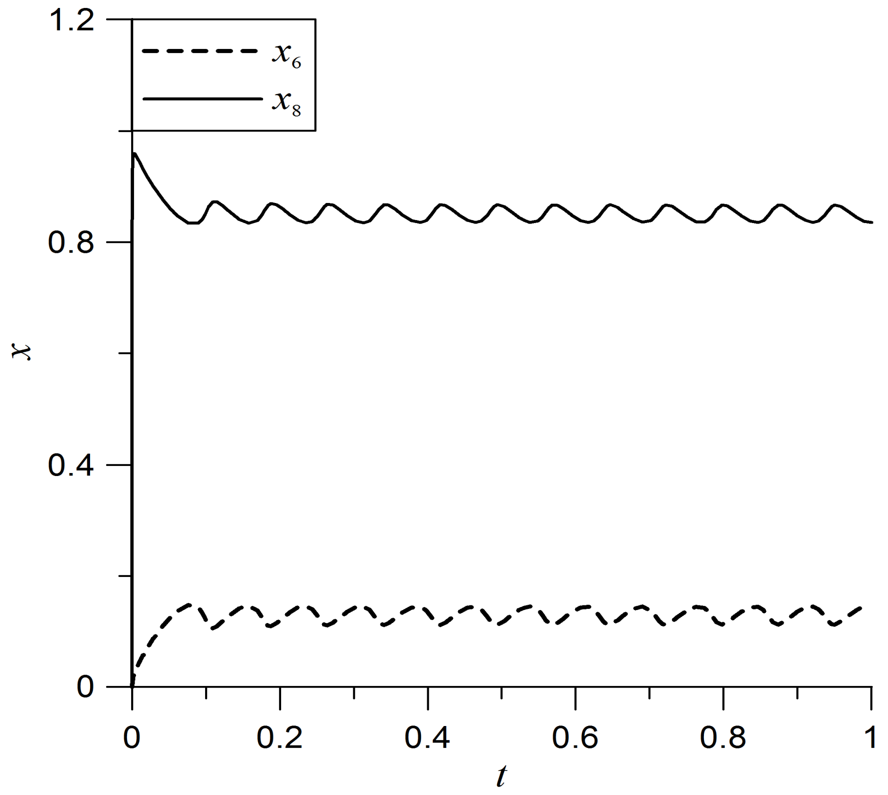


Border approximation

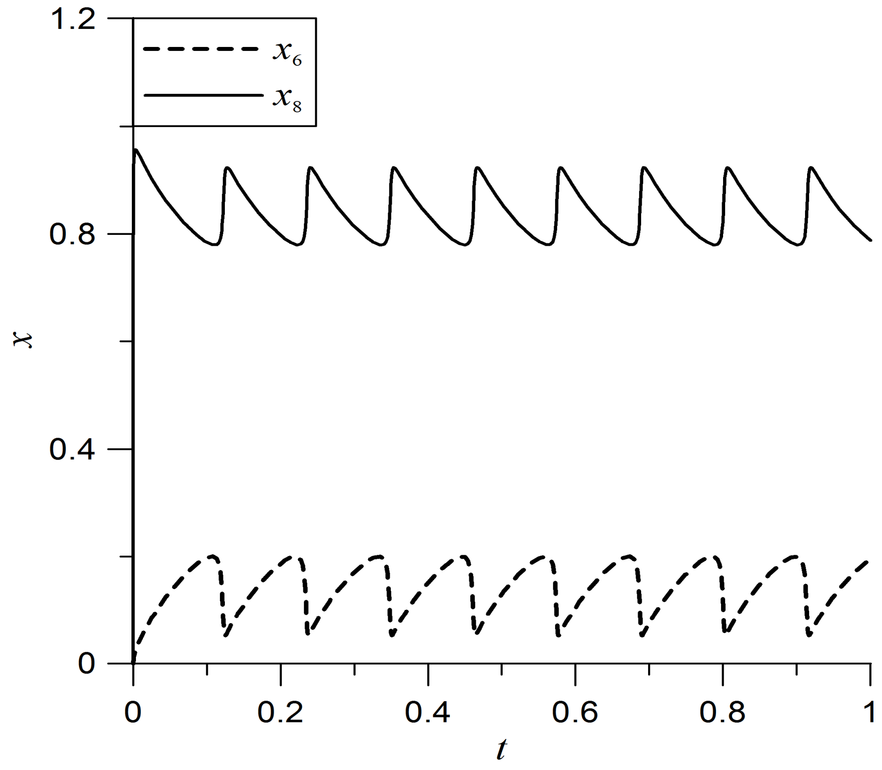


Trajectories of the system variables

$T = 1056, P_{CH_4} = 261, P_{O_2} = 68$

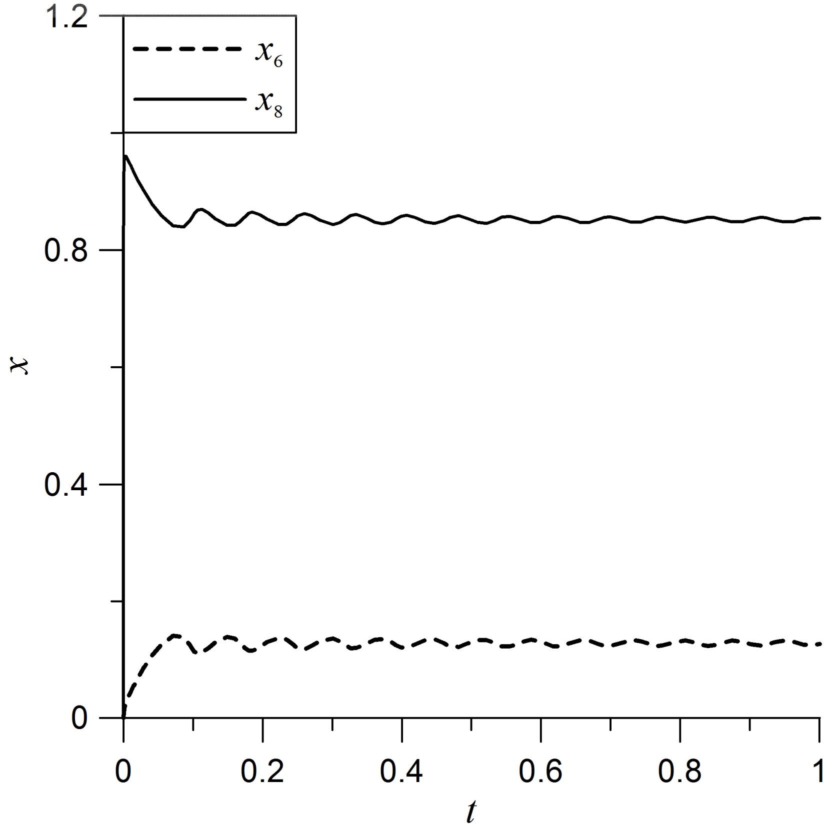


$T = 1056, P_{CH_4} = 266, P_{O_2} = 63$



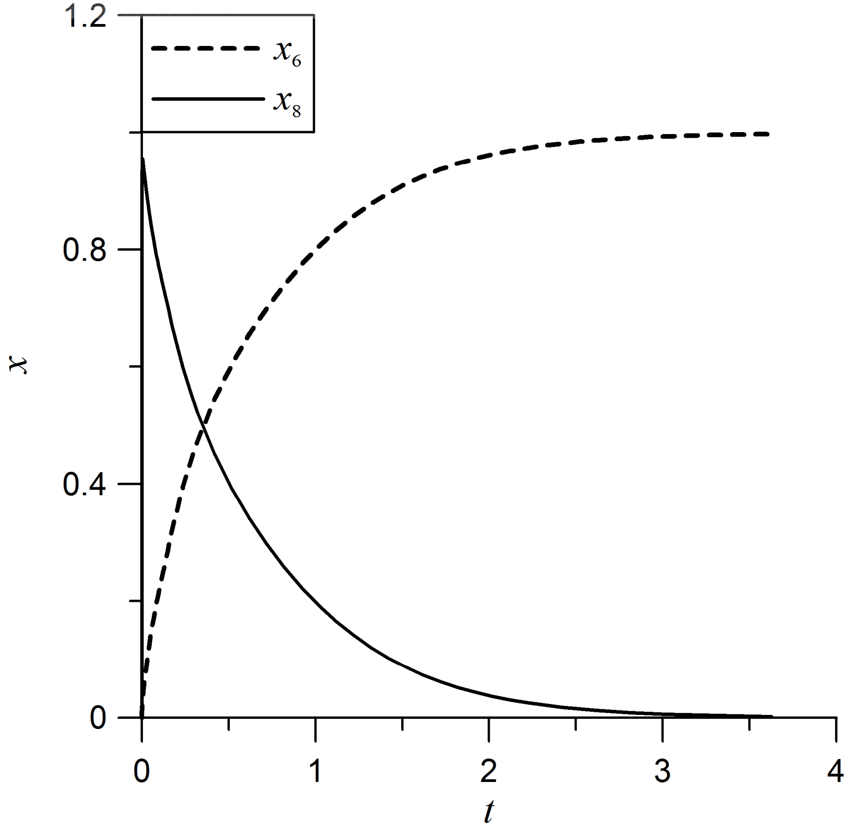
When the parameter P_{CH_4} goes through the lower boundary of the region the oscillations become damped.

$T = 1056, P_{CH_4} = 260, P_{O_2} = 69$



When the parameter P_{CH_4} goes through the region upper boundary variables $\boldsymbol{x_5}$ and $\boldsymbol{x_{10}}$ aim at -∞, variables $\boldsymbol{x_6}$ and $\boldsymbol{x_8}$ cease to oscillate.

$T = 1056, P_{CH_4} = 268, P_{O_2} = 61$



The study of self-oscillations in *CO* oxidation on platinum in terms of the temperature inhomogeneity. Initial ODE approximation

$$\dot{z}_1 = k_1 P_{co}(1 - z_1^3) - k_2 z_1 - k_3 z_1 z_2 + D$$

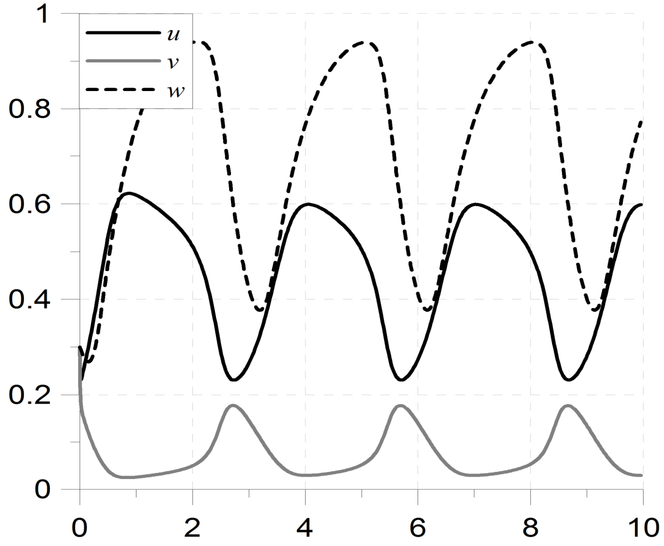
$$\dot{z}_2 = k_4 P_{o_2}\big(s_1 z_3 + s_2(1 - z_3)\big) \cdot (1 - z_1 - z_2)^2 - k_3 z_1 z_2$$

$$\dot{z}_3 = k_5 \left(\frac{1}{1+\exp\left(\frac{u_0 - z_1}{\delta u}\right)} - z_3 \right)$$

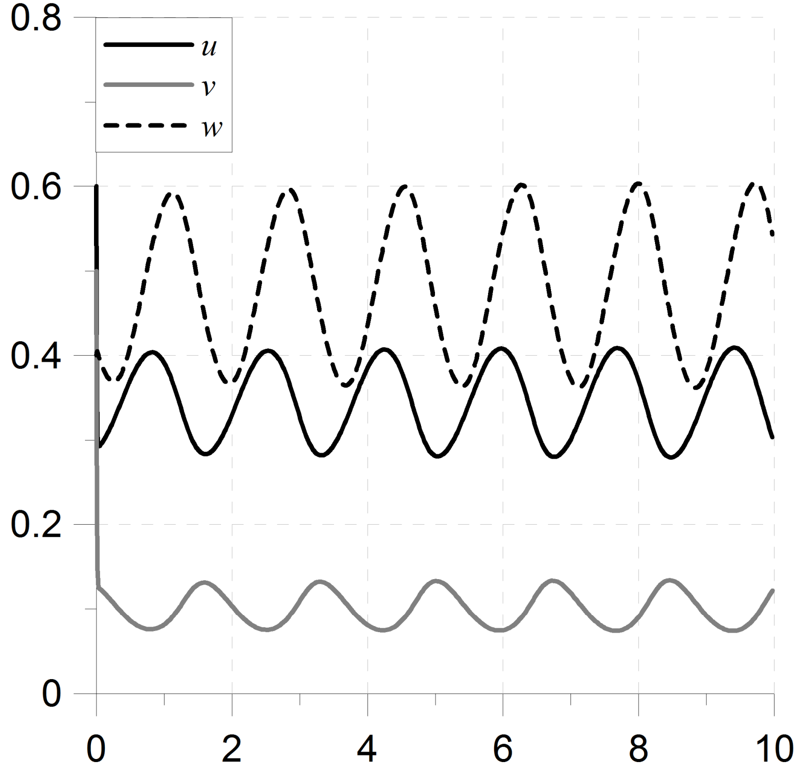
z_1, z_2, z_3 are levels of CO, oxygen and platinum.
Parameters k_2, k_3, k_5 depends on the T value of temperature at which the reaction takes place, pressure P_{CO} and P_{O_2} are the parameters of the model, the remaining parameters are constant.

Trajectories of the system variables

- $x(0) = (0, 0, 0), T = 543.5$

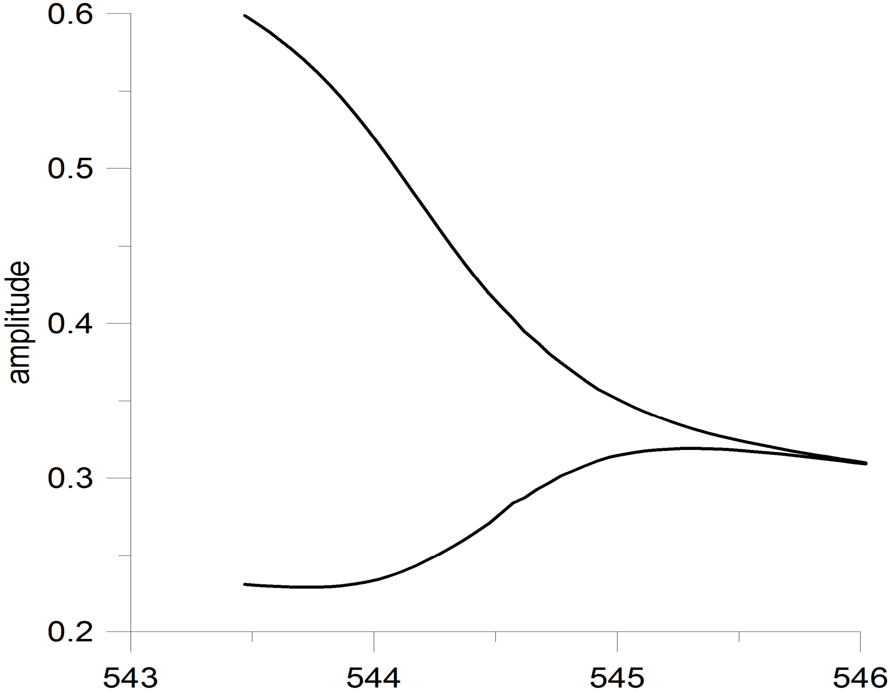


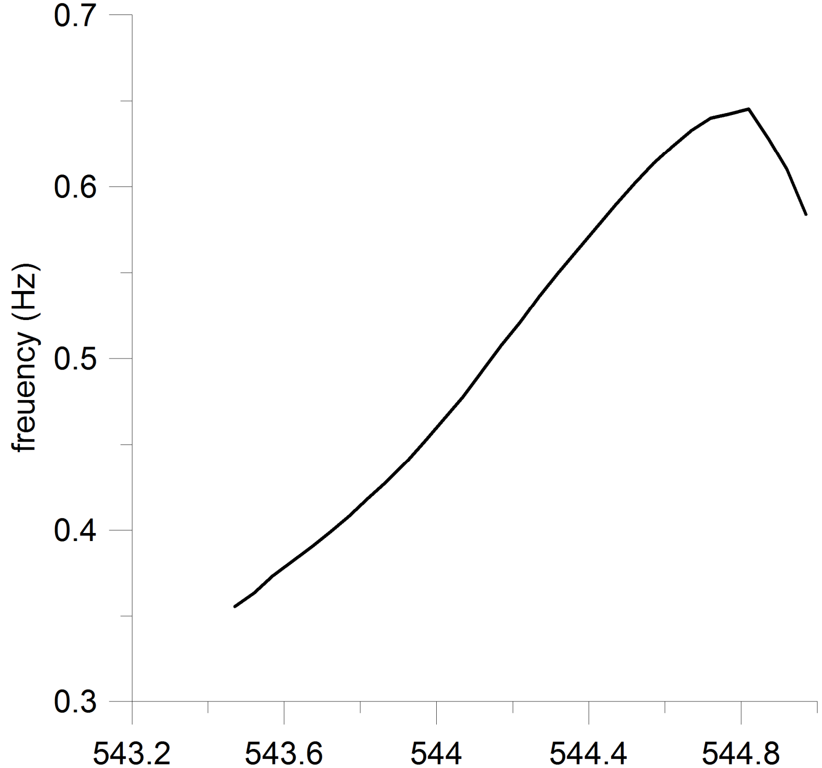
- $x(0) = (0.6, 0.5, 0.4), T = 544.5$



Dependence of the oscillations on temperature

Plots of amplitude (left) and frequency (right) of oscillations in the concentration of CO from the reaction temperature.





Dynamic control system with partial derivatives

$z_1(x, t)$, $z_2(x, t)$, $z_3(x, t)$ are spatial variables

$$\frac{\partial z_1}{\partial t} = k_1 P_{co}(1 - z_1^3) - k_2 z_1 - k_3 z_1 z_2 + D\,\frac{\partial^2 z_1}{\partial x^2}$$

$$\frac{\partial z_2}{\partial t} = k_4 P_{o_2}\big(s_1 z_3 + s_2(1 - z_3)\big) \cdot (1 - z_1 - z_2)^2 - k_3 z_1 z_2$$

$$\frac{\partial z_3}{\partial t} = k_5\left(\frac{1}{1+\exp\left(\frac{u_0 - z_1}{\delta u}\right)} - z_3\right)$$

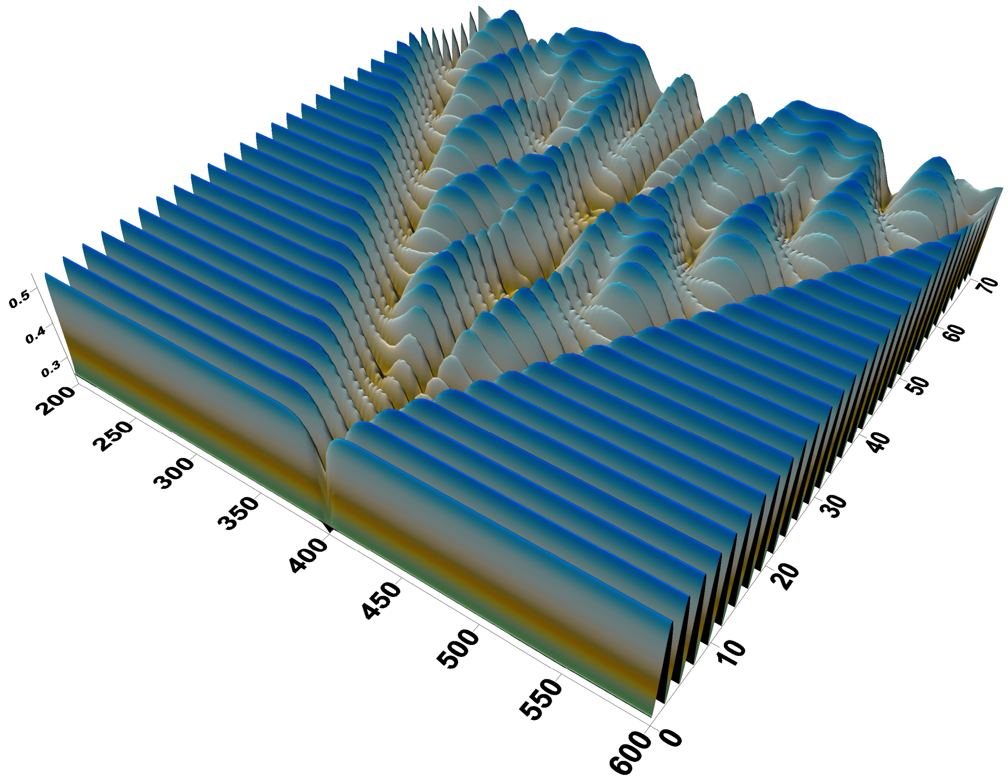
Parameters $\square_2$, $\square_3$, $\square_5$ depends on the $\square$ value of temperature at which the reaction takes place, pressure $\square_{\square\square}$ and $\square_{\square_2}$ are the parameters of the model, the remaining parameters are constant.

$$T(x) = T_0 + \Delta T \exp\left(-\frac{(x - x_0)^2}{2\sigma^2}\right), \qquad x_0 = 400, \sigma = 12, T_0 = 543.47, \Delta T = 2.5$$

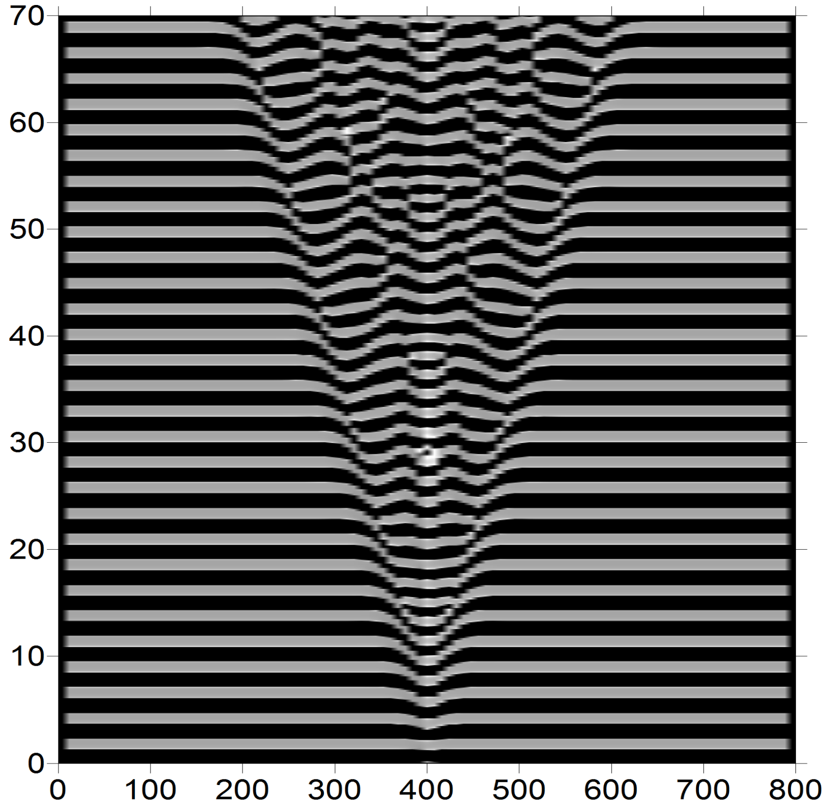
Computational experiments. Surface plot

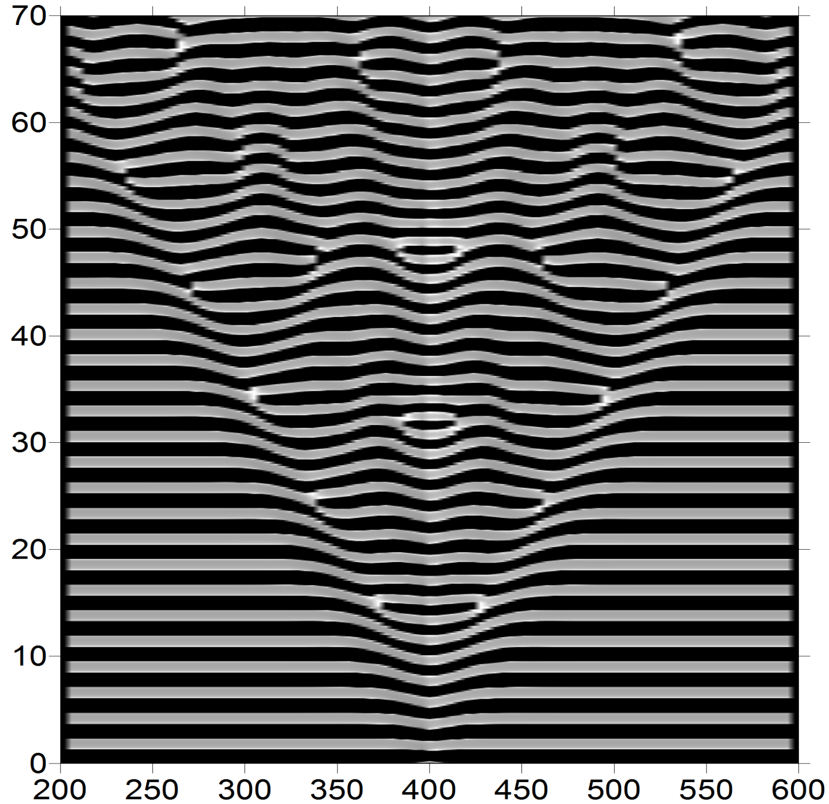
The surface profile of the *CO* concentration in (*x*, *t*) space.

$P_{CO} = 4.48 \cdot 10^{-5}$
$P_{O_2} = 1.1 \cdot 10^{-4}$



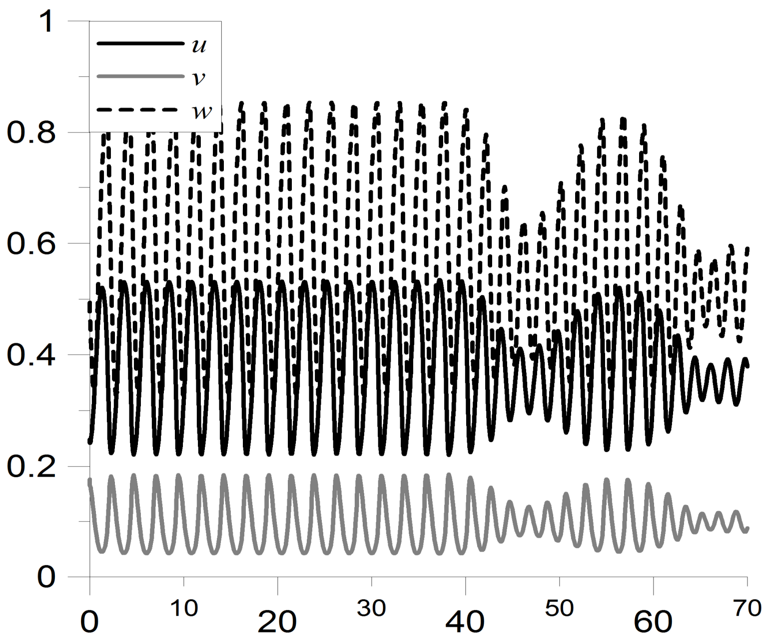
Computational experiments. Shaded surface map



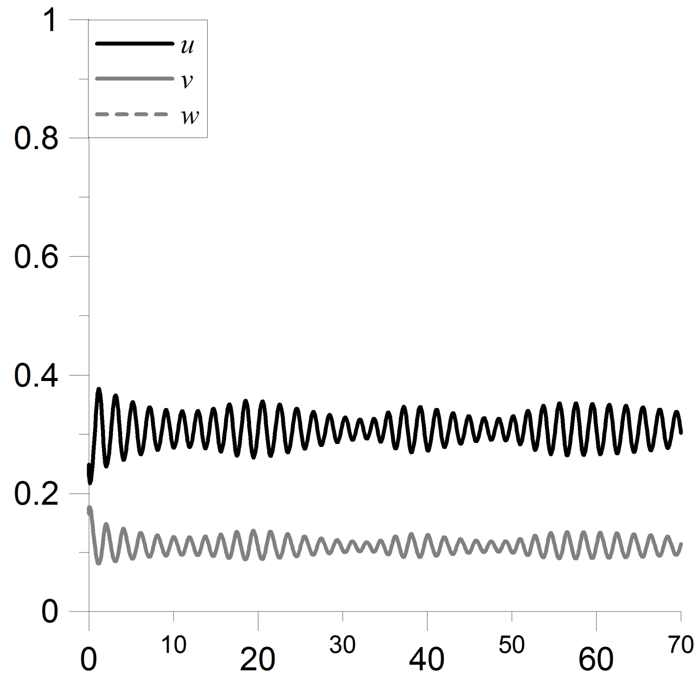


Computational experiments. Separate variables cut

$x = 260, T = 543.47,$
$z_{1\_max} = 0.53, z_{1\_min} = 0.22$



$x = 400, T = 544.97,$
$z_{1\_max} = 0.35, z_{1\_min} = 0.26$



Applied problems
Materials science

- Optimization problem of composite structures;
- Calculations for designing space engines of new type.

- *- Design Technological Institute of Digital Techniques, Novosibirsk*

Applied problems
Quantum physics

- Calculation of the basic operations of the quantum computer;
- Identification of distributed dynamic models of electrons spin epitaxy;
- Modeling of strained heterostructures in quantum dots “silicon-germanium”.

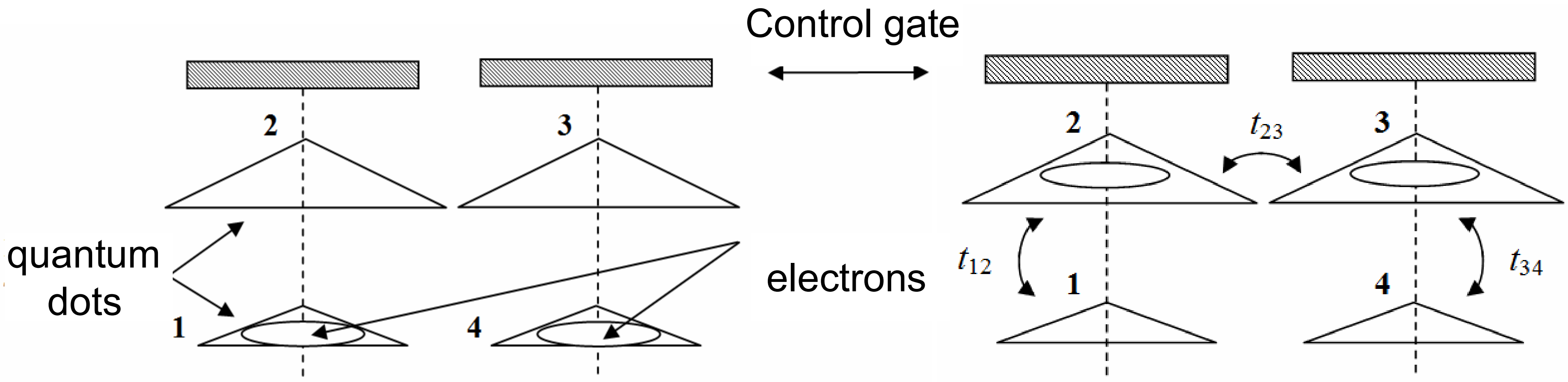
- *- Rzhanov institute of semiconductor physics, Novosibirsk*.

Quantum computer

- A.V. Rzhanov institute of semiconductor physics SB RAS (A.V. Dvurechensky, A.F. Zinov'eva, A.V. Nenashev);
- Institute for System Dynamics and Control Theory SB RAS (A.Yu. Gornov, T.S. Zarodnyuk);
- Employment of quantum dots (QD) as basic elements of a quantum computer has a lot of advantage: we can control of their geometric characteristics and location;
- It is proposed to use vertically combined layers of QD in Ge-Si system for carrying out of quantum computation.

Two cells of the quantum computer

- It is considered two cells of a quantum computer which are based on four tunnel-coupled semiconductor QD;



- The size and the density of QD is determined to existence of sufficient tunnel couple in top layer for implementation of quantum logical operations;

- The exchange operation of information (SWAP) implements due to movement particles to a top layer.

Applied problems

Quantum physics

$$\dot{x} = \begin{pmatrix} 0 & M \\ -M & 0 \end{pmatrix} \cdot x, \quad x(t_0) = x^0$$

$$M = \frac{m}{h} \begin{pmatrix} \frac{J}{4} + \beta\Delta H & \frac{g - \Delta g/2}{2}\beta H_C & \frac{g + \Delta g/2}{2}\beta H_C & 0 \\ \frac{g - \Delta g/2}{2}\beta H_C & -\frac{J}{4} + \Delta g\left(\frac{hw}{2g} + \frac{\beta\Delta H}{2}\right) & \frac{J}{2} & \frac{g + \Delta g/2}{2}\beta H_C \\ \frac{g + \Delta g/2}{2}\beta H_C & \frac{J}{2} & -\frac{J}{4} - \Delta g\left(\frac{hw}{2g} + \frac{\beta\Delta H}{2}\right) & \frac{g - \Delta g/2}{2}\beta H_C \\ 0 & \frac{g + \Delta g/2}{2}\beta H_C & \frac{g - \Delta g/2}{2}\beta H_C & \frac{J}{4} - \beta\Delta H \end{pmatrix}$$

$\Delta H(t)$ is a control function of such structure

$u(t) = a_4 + a_2 \cos(a_1 t + a_3)$, $a_1 \in [0,200]$ is frequency of oscillation,

$a_2 \in [0,3]$ amplitude, $a_3 \in [0,2\pi]$ initial phase,

$a_4 \in [-3,3]$ is non-zero constant shifting one of the electrons closer to the resonance.

Applied problems
Quantum physics

$$I = \sum_{i=1}^{32} x_i^2(t) + 4 - 2\sqrt{A^2 + B^2} \to \min$$

A and B for the operation of "φ-turn" states as following:

$$A = \cos\frac{\varphi}{2}\left(x_1 + x_{10} + x_{19} + x_{28}\right) + n_x \sin\frac{\varphi}{2}\left(x_6 + x_{13} + x_{24} + x_{31}\right) +$$

$$+ n_y \sin\frac{\varphi}{2}\left(x_2 - x_9 + x_{20} - x_{27}\right) + n_z \sin\frac{\varphi}{2}\left(x_5 - x_{14} + x_{23} - x_{32}\right)$$

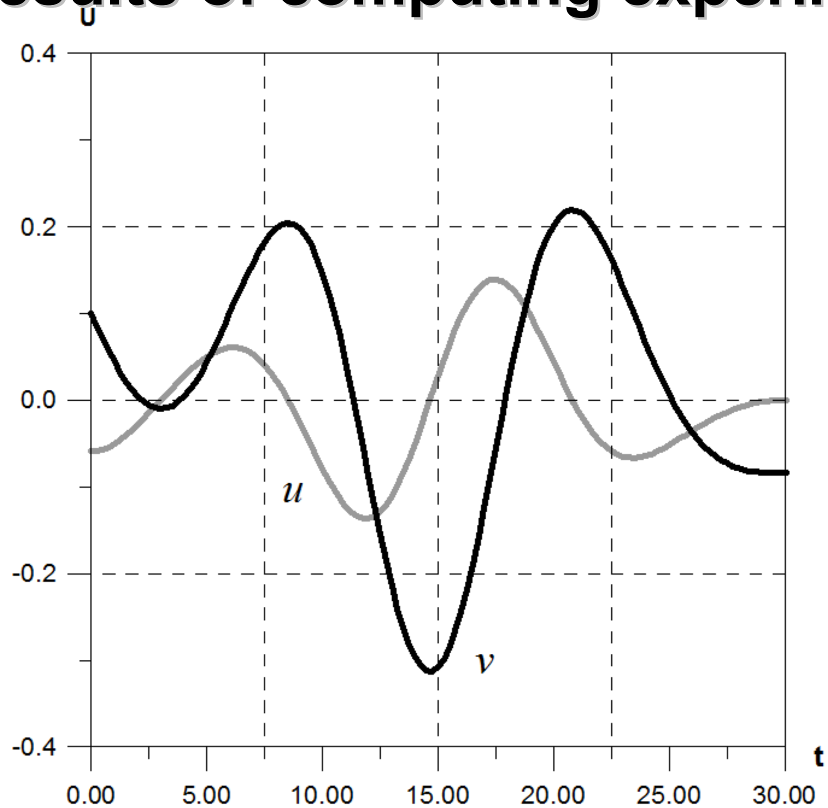
$$B = \cos\frac{\varphi}{2}\left(x_5 + x_{14} + x_{23} + x_{32}\right) + n_x \sin\frac{\varphi}{2}\left(-x_2 - x_9 - x_{20} - x_{27}\right) +$$

$$+ n_y \sin\frac{\varphi}{2}\left(x_6 - x_{13} + x_{24} - x_{31}\right) + n_z \sin\frac{\varphi}{2}\left(-x_1 + x_{10} - x_{19} + x_{28}\right)$$

Components n_x, n_y, n_z of $\vec{n}$ vector determines rotation around the axes X, Y and Z respectively.

The optimal control problem for system of quantum dots

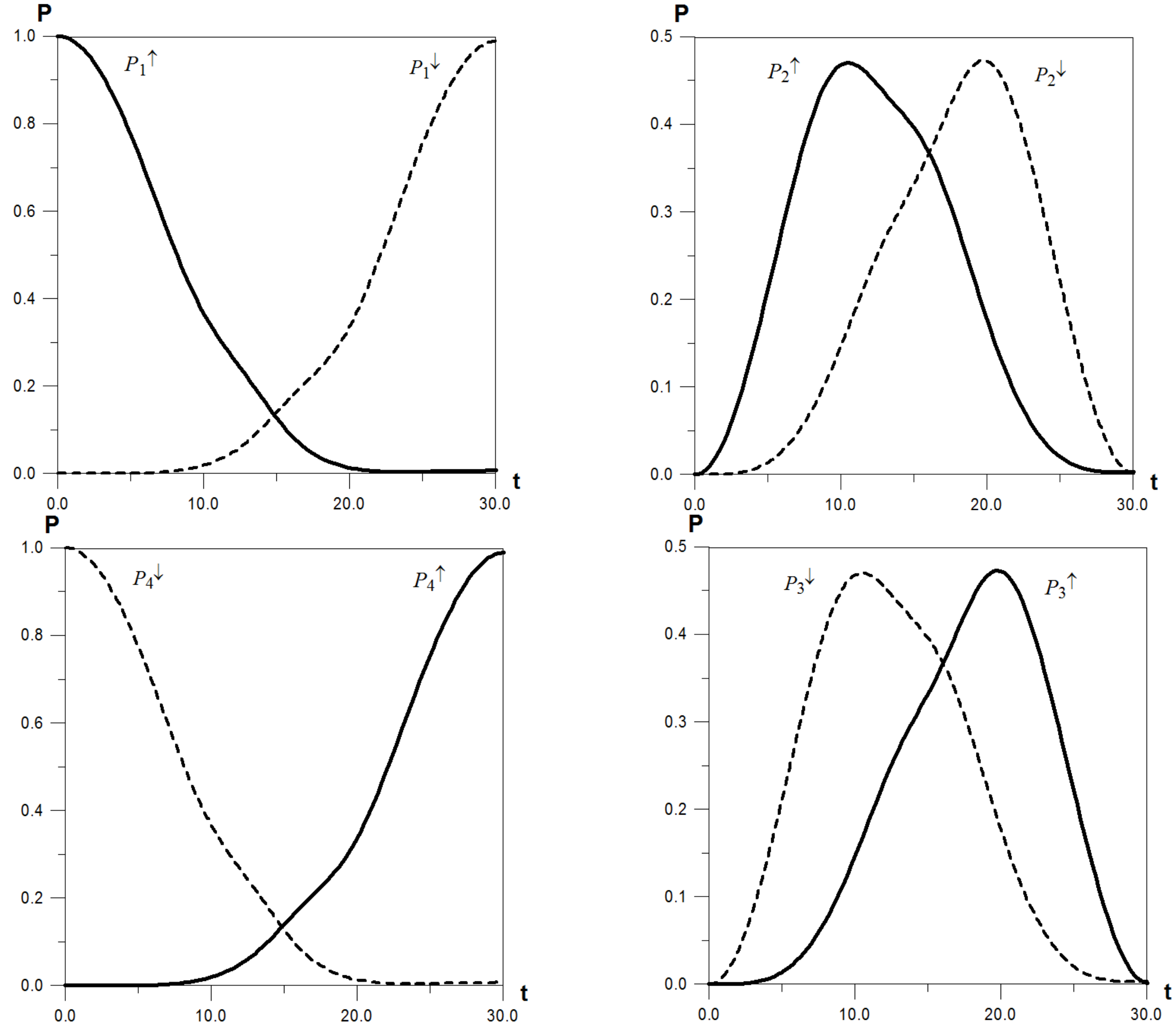
The results of computing experiments



Optimal control $u = u(t)$

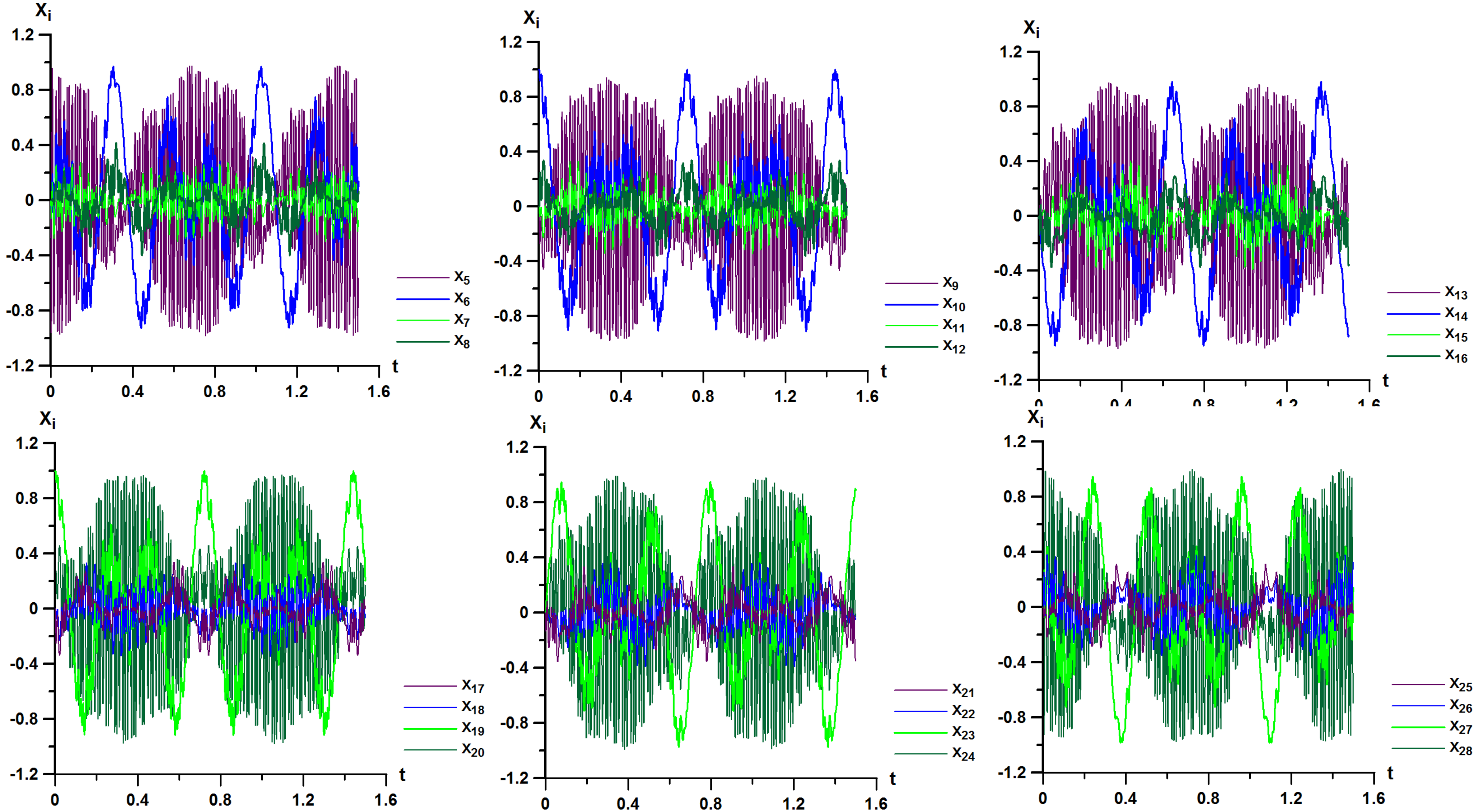
and и controlled voltage impuls $v = v(t) = \dfrac{du}{dt}$

Probability of finding electrons in quantum dots



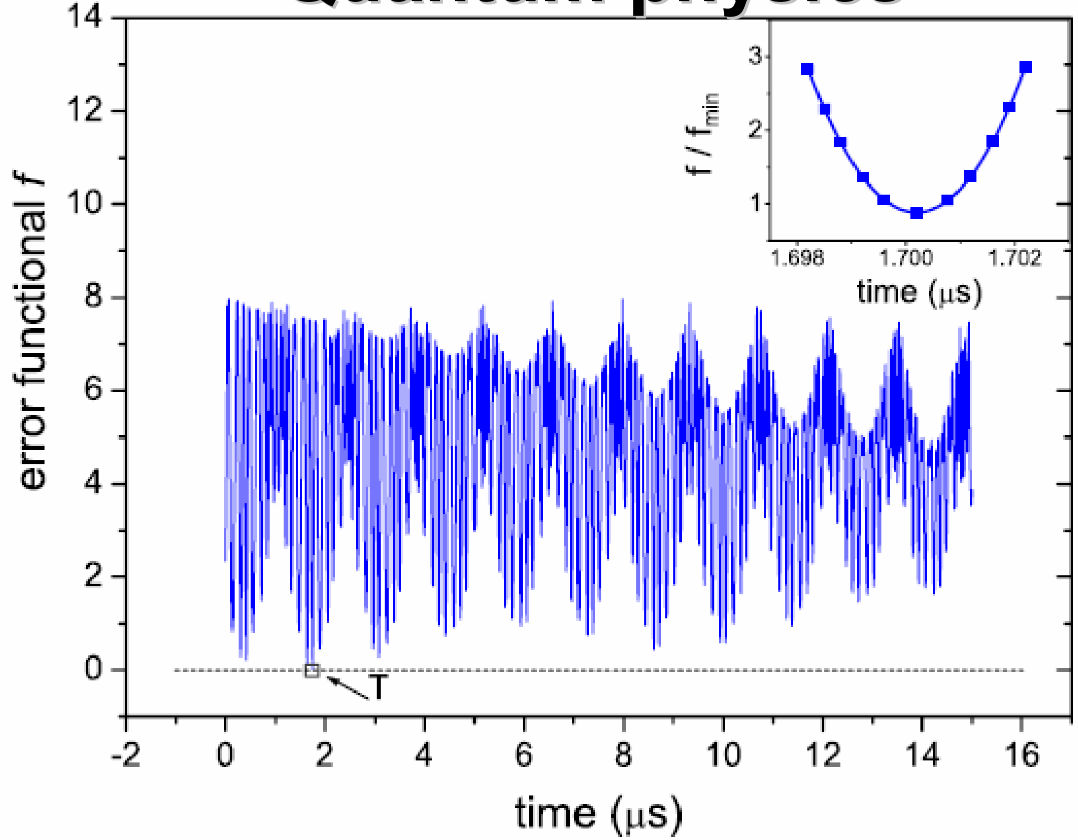
The probability variations of finding electrons with spin "down“ $P_k^{\downarrow}$("up“ $P_k^{\uparrow}$) in each of the four quantum dots ($k = \overline{1,4}$)

Computing experiments



Optimal trajectories of the system, allowing to realize quantum logical storage operation

Applied problems
Quantum physics



Dependence of the error functional for the operation of "rotate around the specified axis for $\varphi = \pi/2$"

[**A.V. Nenashev, A.F. Zinovieva, A.V. Dvurechenskii, A.Yu. Gornov, T.S. Zarodnyuk.** Quantum logic gates from time-dependent global magnetic field in a system with constant exchange // Journal of Applied Physics **117**, 113905 (2015)].

Applied problems
Ecology

- Modeling of lands desertification processes for the steppe regions of Kalmykia and Mongolia;
- Optimization problem forest utilization on the territory of Irkutsk region;
- Modelling of biotransformation processes of organic substances in forest ecosystems of the Baikal region.

- - *Sochava institute of geography, Irkutsk;*
- *- Siberian institute of plant physiology and biochemistry, Irkutsk.*

Applied problems
Medicine

- Evaluation of medico-social factors of fertility and oncological diseases of the Irkutsk region population;
- Investigation of socially important environmental health problems of the Baikal and Arctic regions population;

- *- East-Siberian institute of medico-ecological research, Angarsk;*
- *- Irkutsk diagnostic and treatment center;*
- *- Irkutsk state medical university.*

Applied problems
Biology

- Identification of significant factors in biomolecular networks.

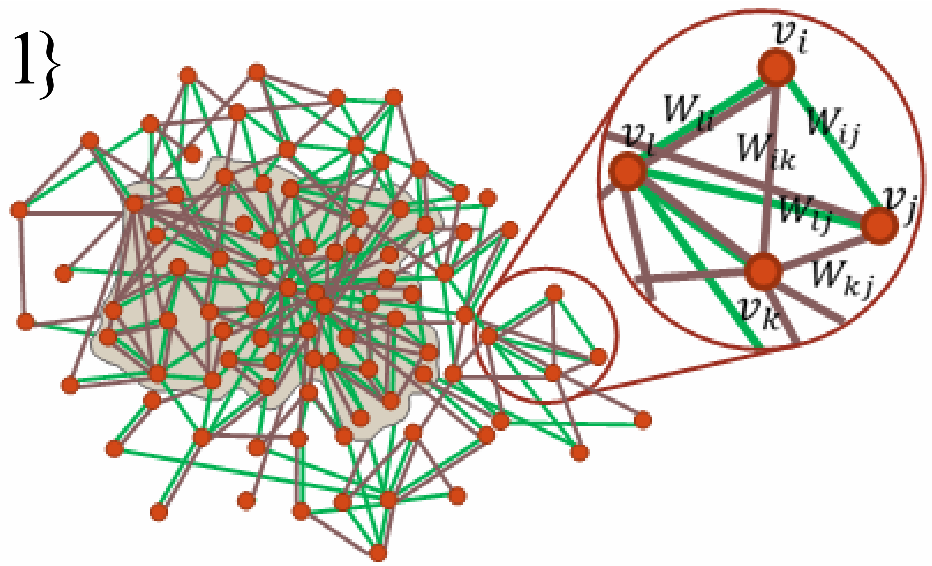
- Mongolian national institute.

Identification of significant factors in biomolecular networks

$$xAx \rightarrow \max$$

$$\Delta = \{x \in R^k : x_i \geq 0, e^T x = 1\}$$

The matrix A consists of 1000 rows and 1000 columns.
The numerical solution is non-null elements of vector x.



[B. Amgalan, H. Lee. DEOD: uncovering dominant effects of cancer-driver genes based on a partial covariance selection method // Bioinformatics (2015) 31 (15): 2454-2460].

Applied problems
Criminal science

- Optimization problem of anti drug addiction and crime investment programs;
- Modeling of youth crime processes in the Irkutsk region.

- - *Vienna university of technology, Austria;*
- *- East-Siberian institute of the Ministry of internal Affairs of Russia.*

Applied problems
Economics

- Calculation of investment programs of Kabansky region in Buryatia;
- Modeling market equilibrium for the grain market of Mongolia.

- *- Mongolian national institute;*
- *- Sochava institute of geography, Irkutsk.*

Applied problems
Transportation

- Calculations of urban traffic equilibrium.

- *- Keldysh institute of applied mathematics, Moscow;*
- *- Moscow Institute of Physics and Technology .*

Applied problems
Seismology

- Modeling and estimation of seismic resistance of buildings.

- *- Institute of earth crust, Irkutsk.*

Modeling and estimation of seismic resistance

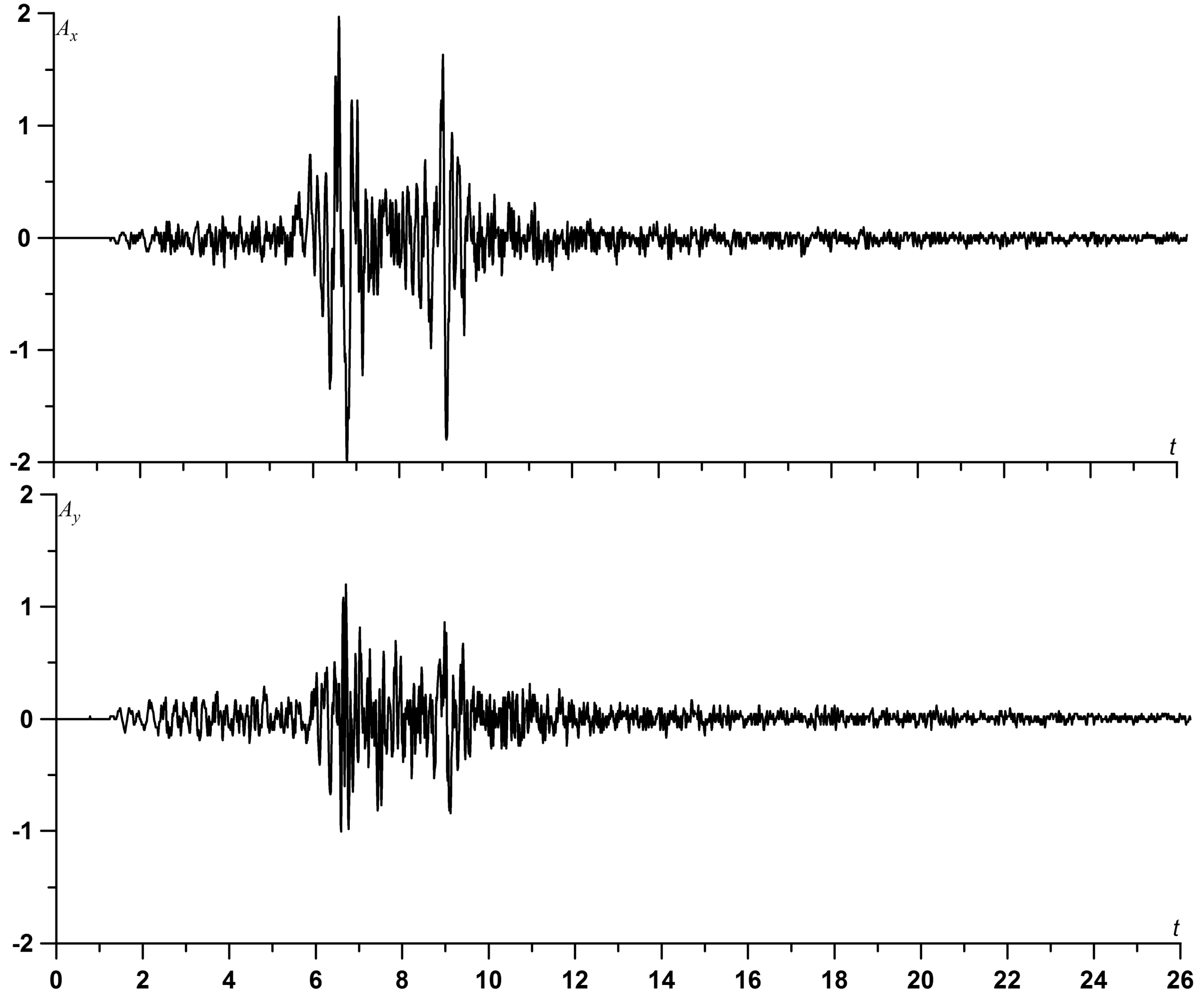
$$\dot{y}_1 = y_2 \qquad \dot{y}_2 = -\frac{(y_5 + C_x y_2)}{M} - \ddot{X}(t) \qquad \dot{y}_3 = y_4 \qquad \dot{y}_4 = -\frac{(y_6 + C_y y_4)}{M} - \ddot{Y}(t)$$

$$\begin{matrix}\dot{y}_5 \\ \dot{y}_6\end{matrix} = \begin{cases} D \cdot (y_2, y_4)^T, \text{if } \dfrac{y_5^2}{R_{TX}^2} + \dfrac{y_6^2}{R_{TY}^2} \le 1, \\ f(y_5, y_6) \cdot (y_2, y_4)^T, \text{if } \dfrac{y_5^2}{R_{TX}^2} + \dfrac{y_6^2}{R_{TY}^2} > 1 \text{ and } \dfrac{y_5^2}{(R'_{TX})^2} + \dfrac{y_6^2}{(R'_{TY})^2} < 1, \\ 0, \text{if } \dfrac{y_5^2}{(R'_{TX})^2} + \dfrac{y_6^2}{(R'_{TY})^2} \ge 1. \end{cases}$$

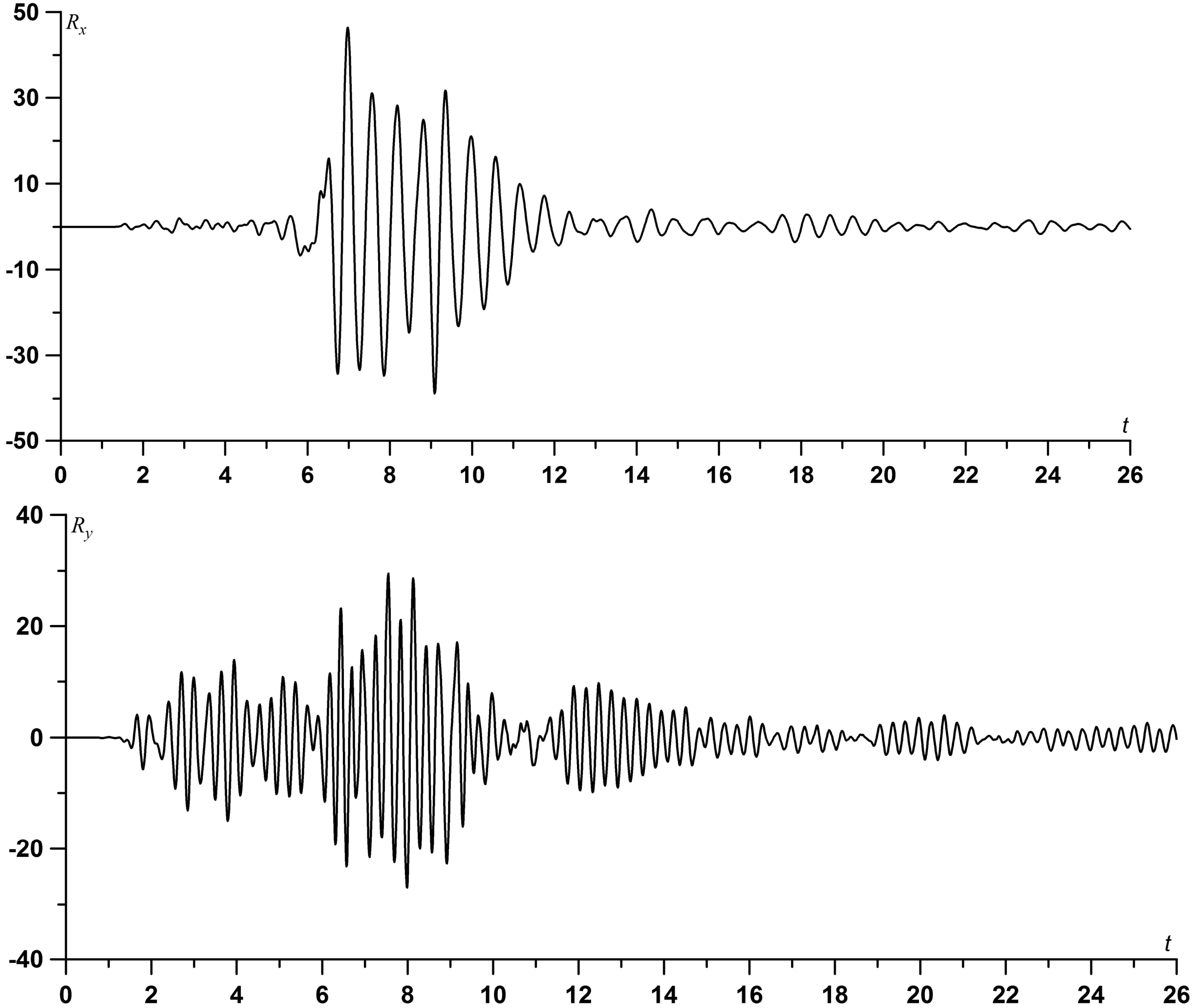
The generalized stiffness parameters obtained from the elastic calculation in SCAD software are D_{11} = 2272, D_{12} = D_{21} = -397, D_{22} = 8982.

The attenuation coefficients C_x = 26, C_y = 35.
The yield strength of the building frame were determined by methods of limit equilibrium theory. R_{TX} = 190 and R_{TY} = 63.
The limit of elastic-plastic work relies R'=105%$\cdot R$
M = 208/9.81

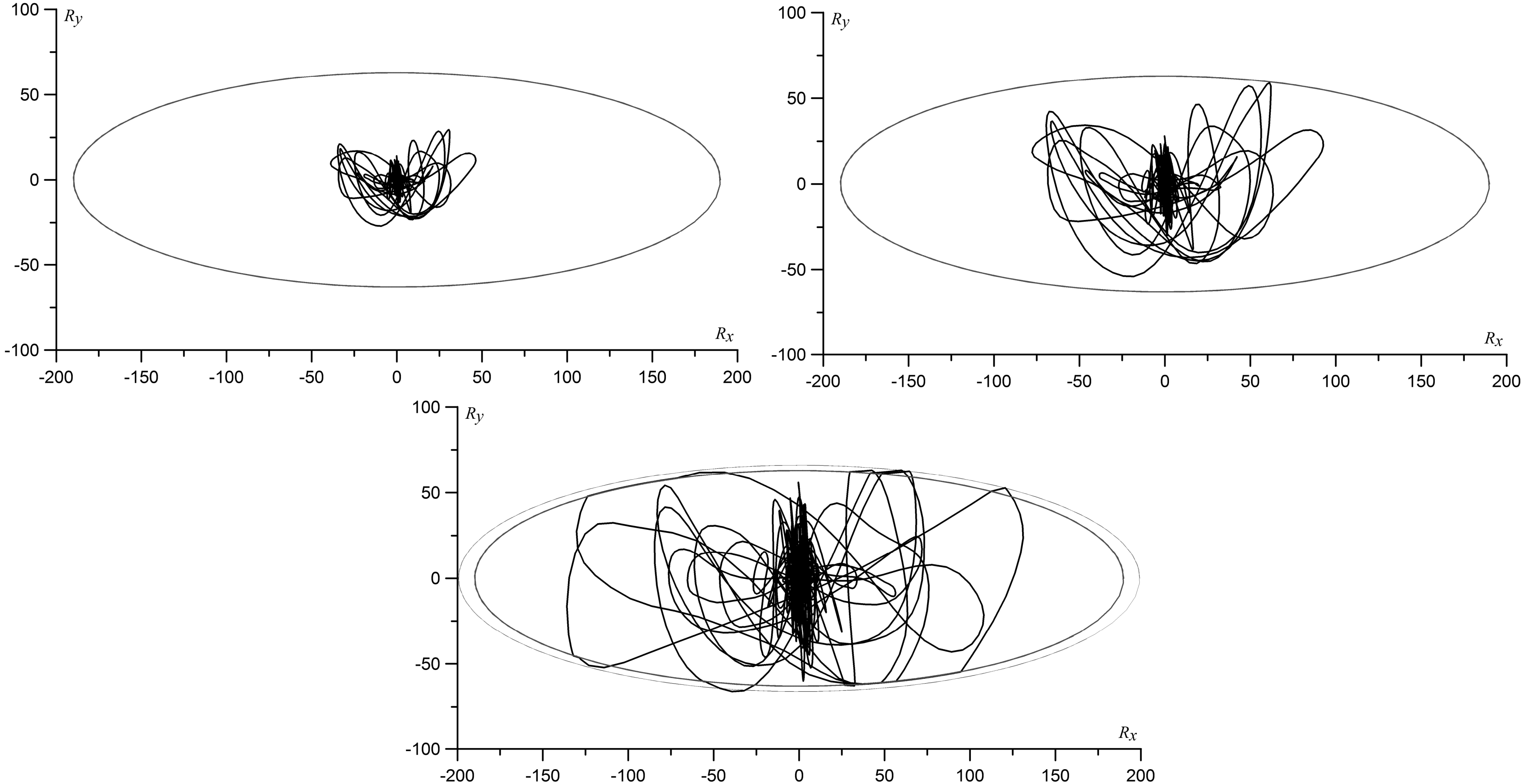
Accelerogram of earthquake near Baikal 27.08.2008 (Talaya station), scaling by 8 points



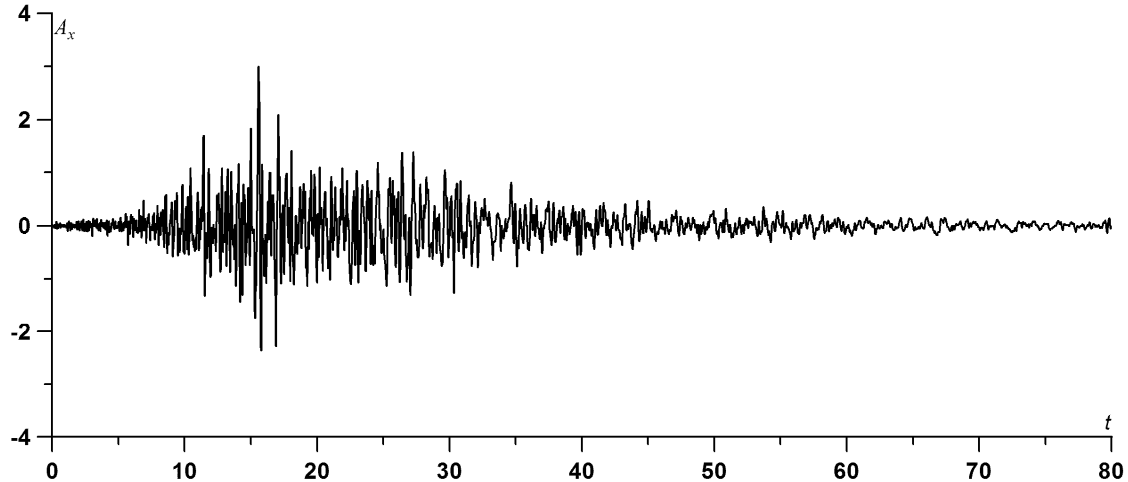
Graphs of the reactions projection R_x and R_y

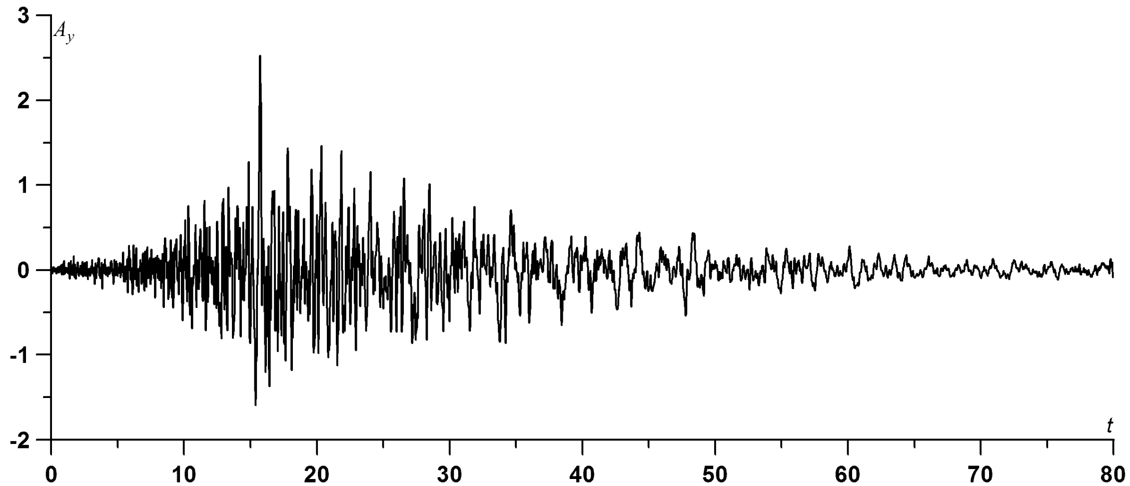


Reaction dynamics on the plane (R_x, R_y) and yield surface. 8, 9 and 10 points

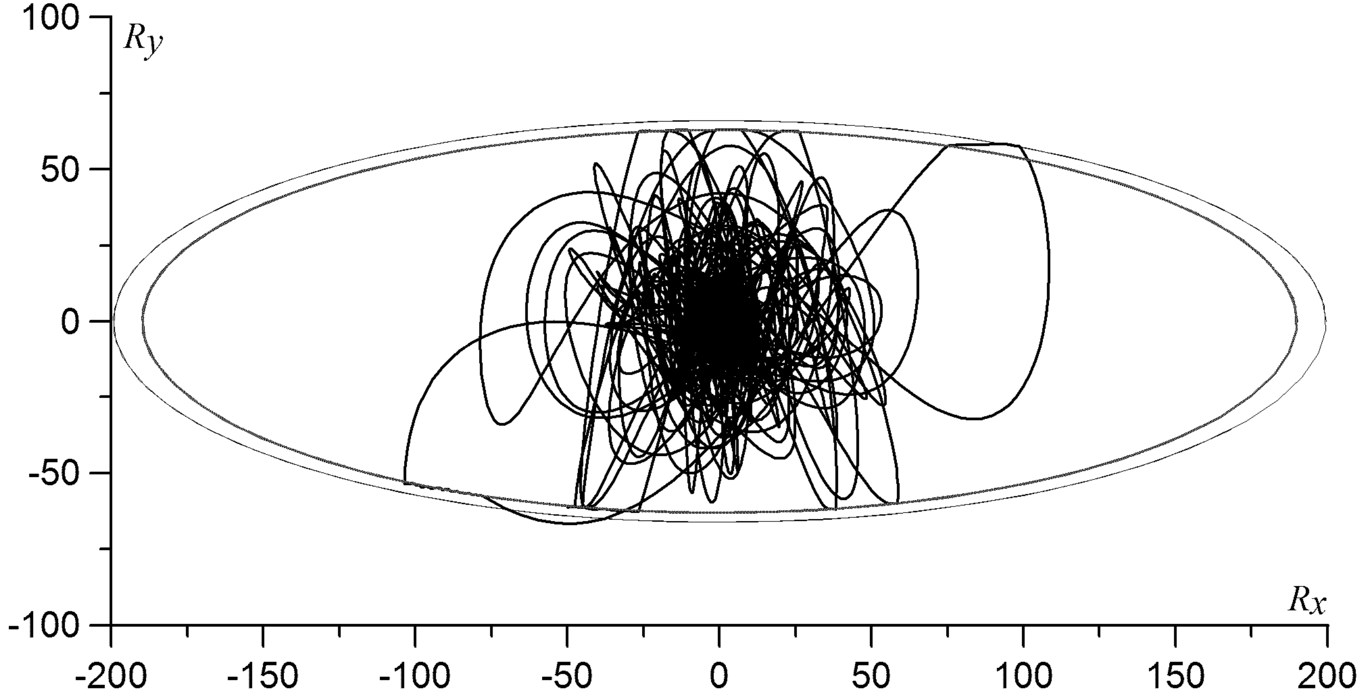


Accelerogram of earthquake near Limon City, Kosta-Rica, 22.04.1991, scaling by 8 points

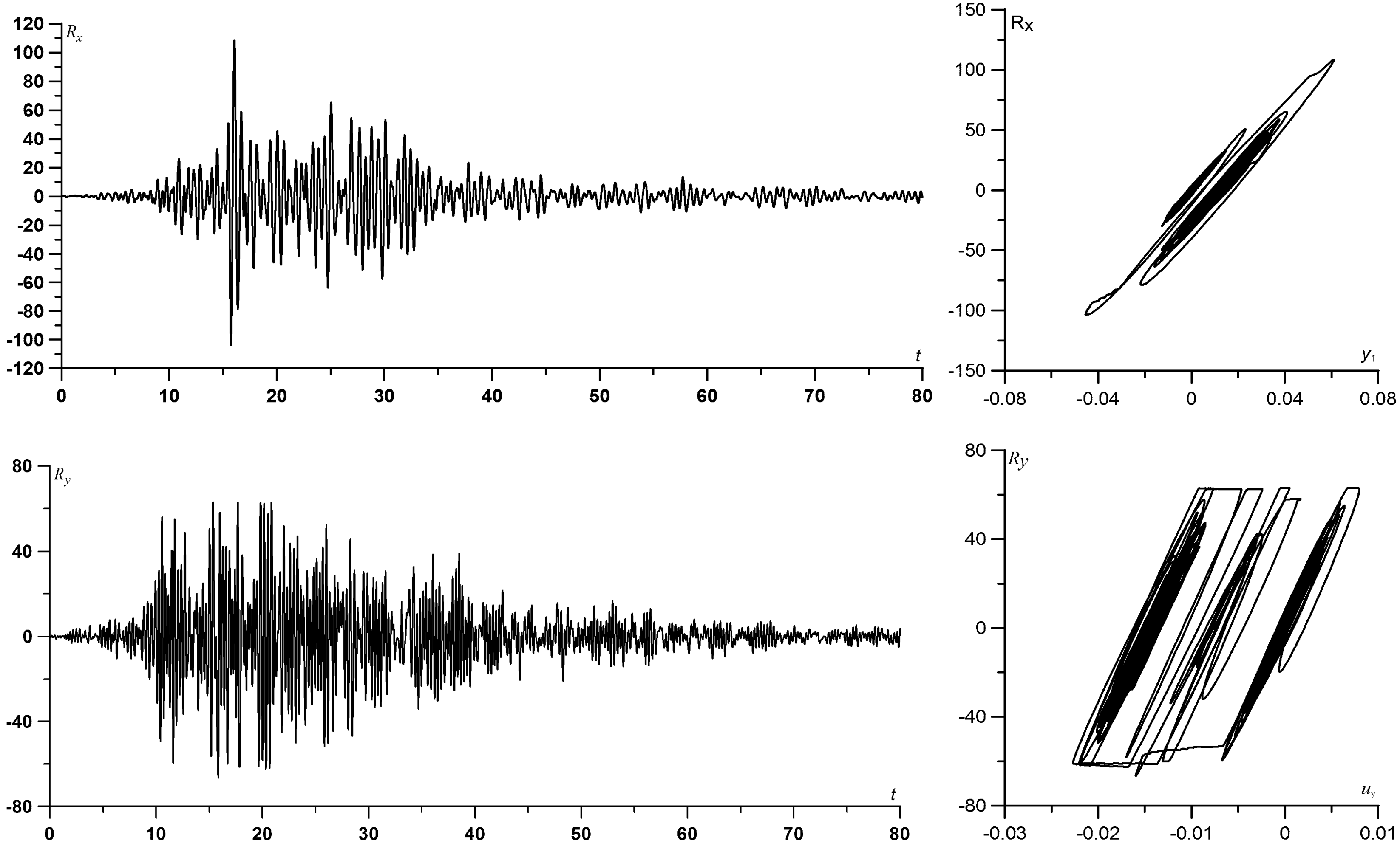




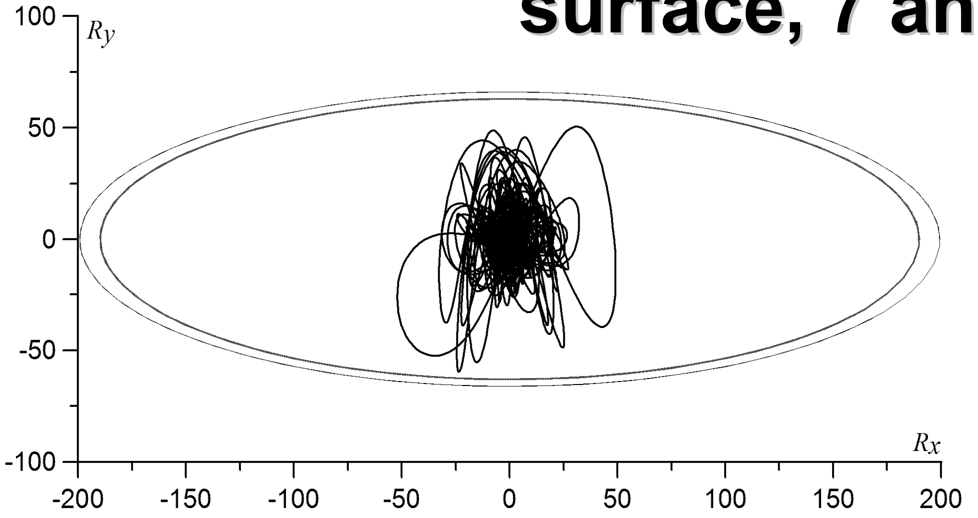
Reaction dynamics on the plane (R_x, R_y) and yield surface, scaling by 8 points

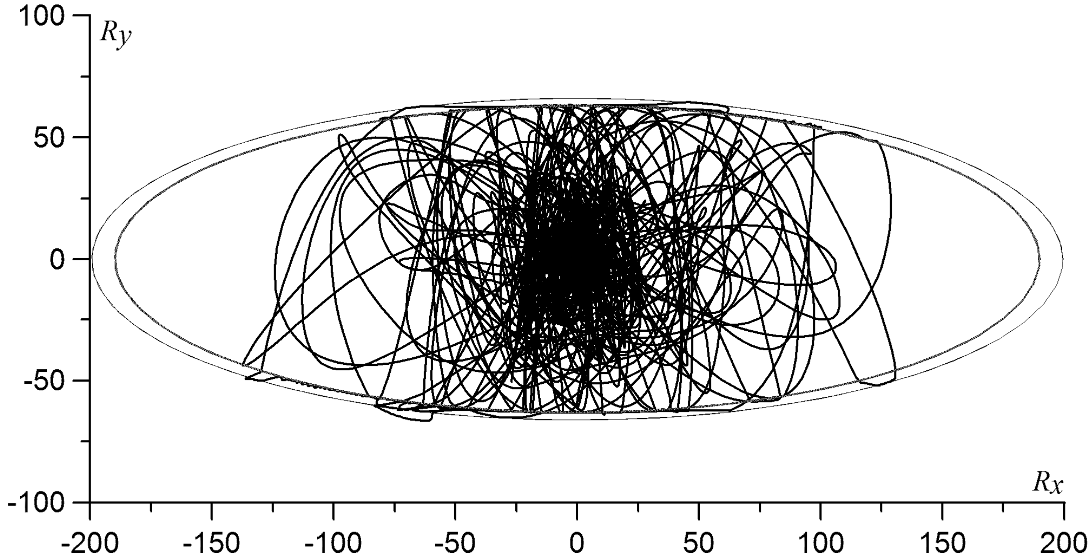


Graphs of the reactions projection R_x and R_y, their dependence on movement



Reaction dynamics on the plane (R_x, R_y) and yield surface, 7 and 9 points





Algorithms for the solution of huge quasiseparable optimization problems

- Morse potential optimization;

- Keating potential optimization;

- Huge-Scale separable convex optimization problem;

- PageRank problem.

Morse potential optimization

A **cluster** is a structure consisting of a finite number of atoms or molecules. Occupies an intermediate position between the individual particles and the bulk material.

The interaction between the elements of the clusters described various potential functions defined multidimensional potential energy surface.

Finding the minima of the potential allows to obtain stable atomic-molecular configurations.

Such simulation in some cases replaces the field experiments.

The Cambridge Energy Landscape Database (The Cambridge Cluster Database): http://www-wales.ch.cam.ac.uk/CCD.html

Кластер – структура, состоящая из конечного числа атомов или молекул. Занимают промежуточное положение между отдельными частицами и объемным веществом.

Взаимодействие между элементами таких кластеров описывается различными функциями потенциалов, задающих (многомерные) поверхности потенциальной энергии (ППЭ).

Нахождение минимумов (стационарных точек) таких потенциалов (поверхностей) позволяет получать устойчивые атомно-молекулярные конфигурации.

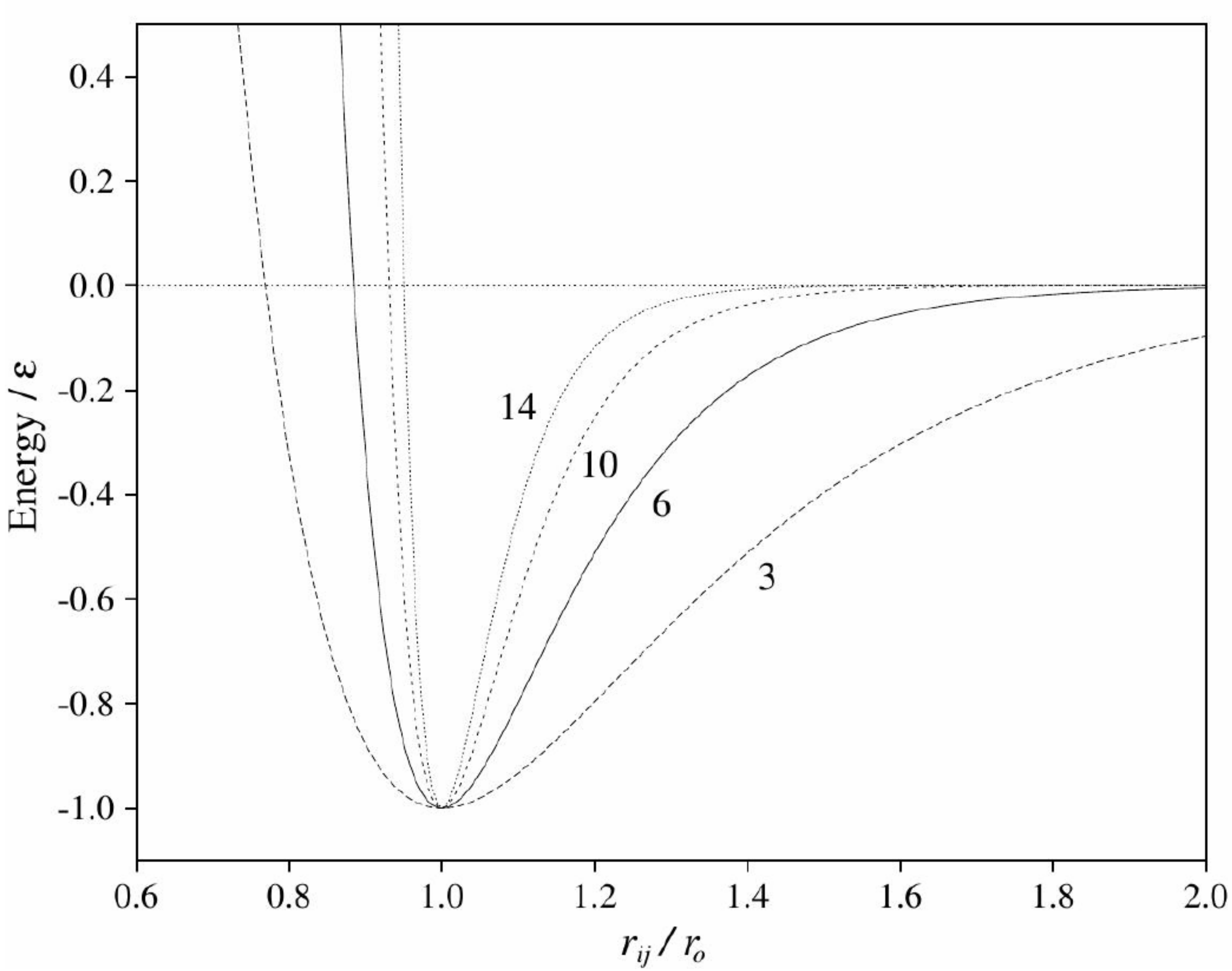
Подобное моделирование в ряде ситуаций заменяет натурные (физические) эксперименты.

Morse potential function

$$V_M = \sum_{i=1}^{n} \sum_{j>i}^{n} \left(\left(e^{\rho_0(1-r_{ij})} - 1\right)^2 - 1 \right)$$

Morse potential function

with different ρ values



Morse potential optimization

- A global optimization problem.

- An astronomical number of local extrema. For example, for a cluster of 147 atoms experts provide estimates of the order of 10^{60}.

- The current state: “large clusters”, consisting of more than 200 atoms (600 variables).

•Задача глобальной оптимизации.

•Астрономические число локальных экстремумов. Например, для кластера из 147 атомов эксперты дают оценки порядка 10^{60}.

• Современное состояние задачи – «большие кластеры», состоящие более чем из 200 атомов (600 переменных).

Morse potential optimization

The Cambridge Cluster Database
D.J. Wales, J.P.K. Doye, A. Dullweber, M.P. Hodges, F.Y. Naumkin, F. Calvo, J. Hernandes-Rojas and T.F. Middleton.
www-wales.ch.cam.ac.uk/CCD.html

Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China.
staff.ustc.edu.cn/~clj

Jorge Marques
Department of Chemistry Research in Computational Chemistry and Molecular Modeling
University of Coimbra, Portugal.

Applied optimization methods

Local optimization techniques

The main (universal) methods

- Conjugate Gradient;
- L-BFGS;

Additional methods

- Cauchy;
- Powell;

Special methods

- Polyak;

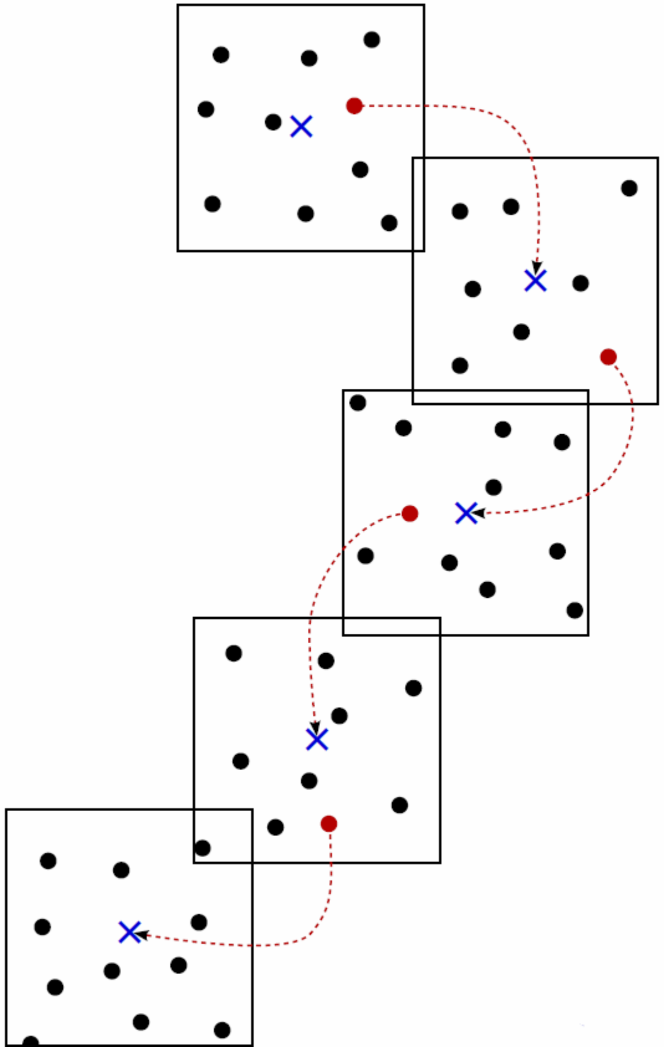
Global optimization techniques

- Multi-start;
- MSBH-Monotonic Sequential Basin-Hopping;

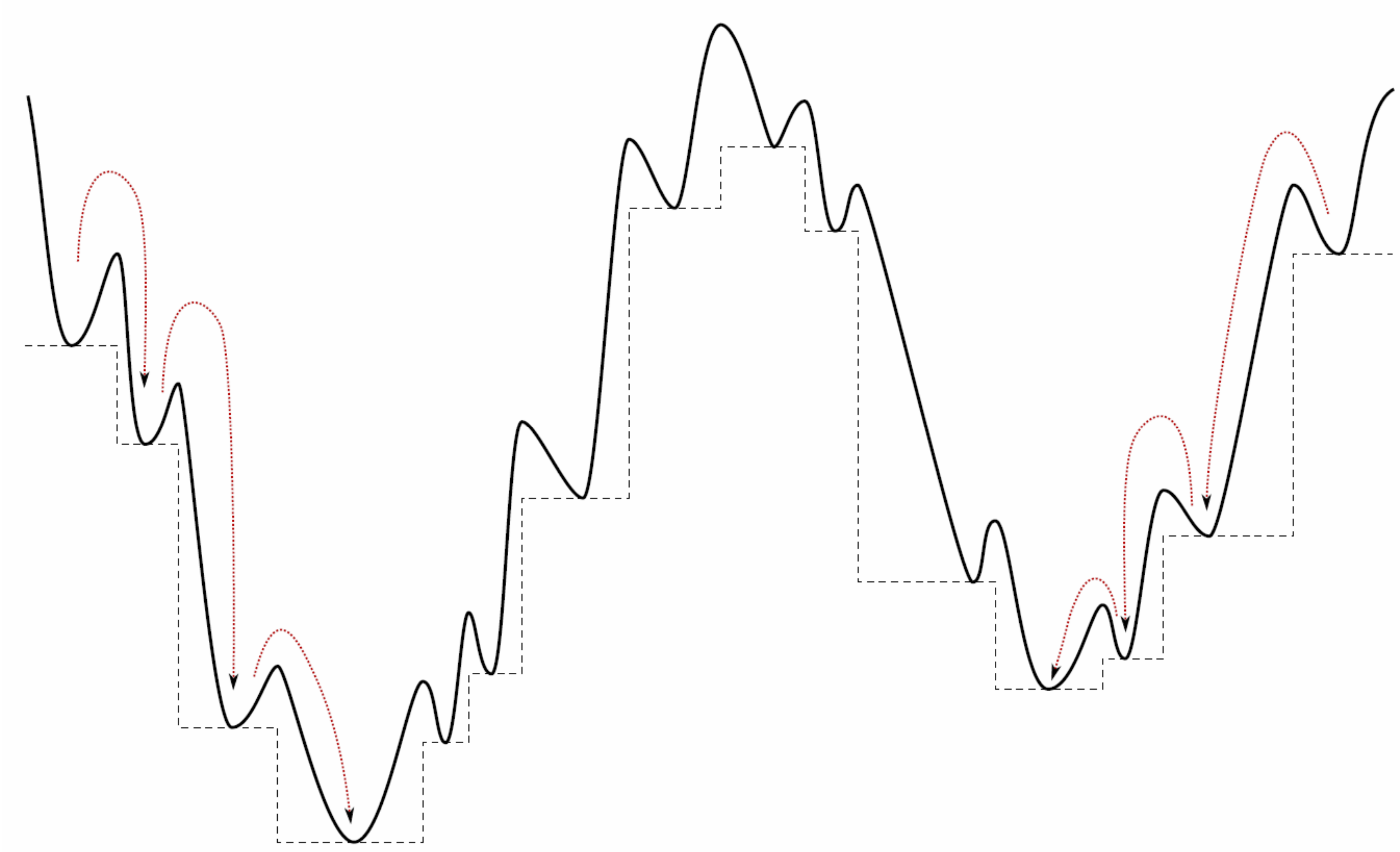
-“Big-Bang”;

-“Forest”

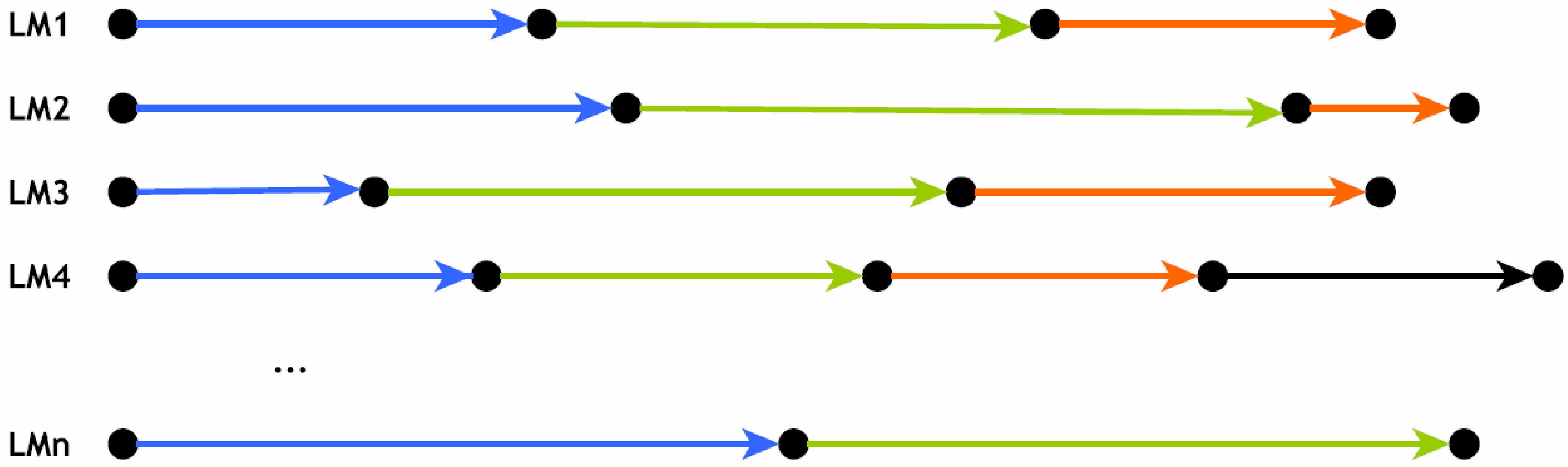
MSBH method



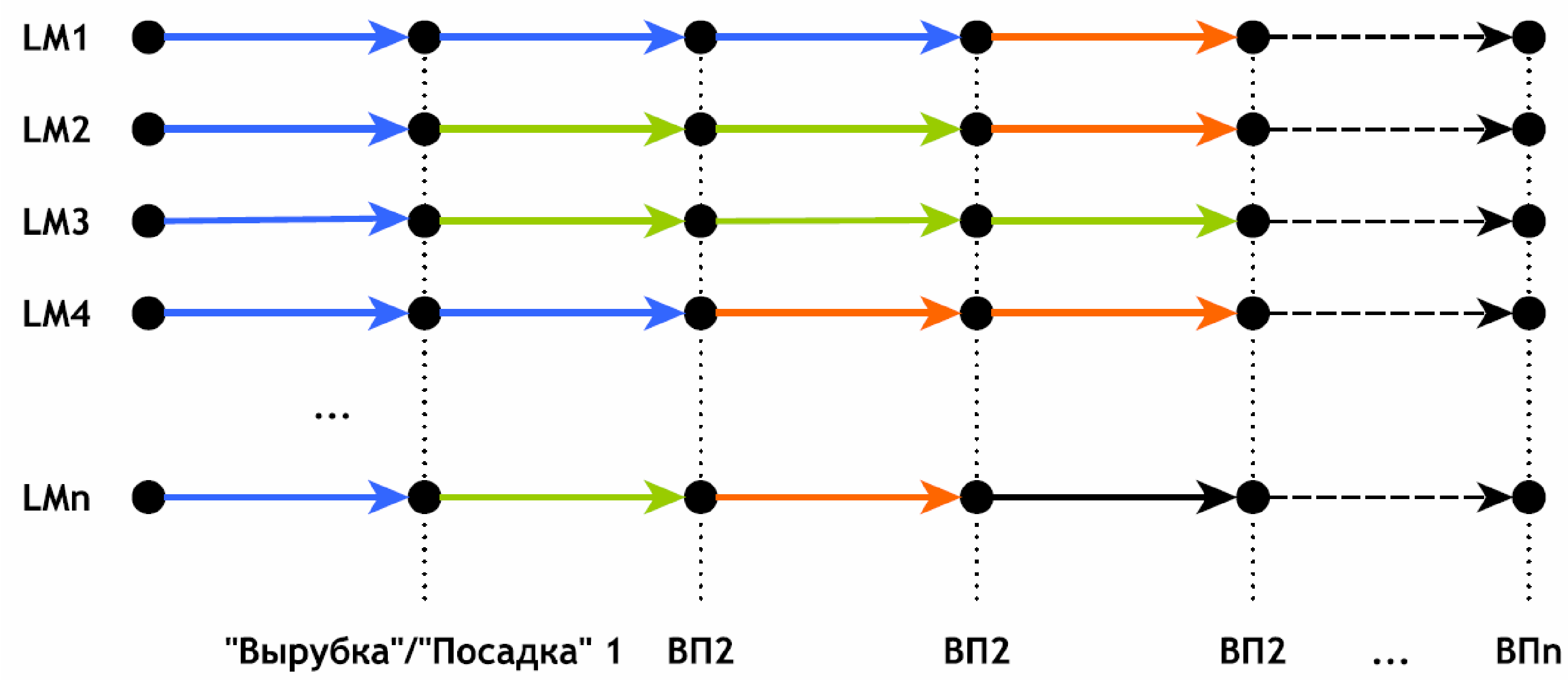
MSBH method



Forest method



Forest method



Критерии “вырубки” (рестарта) локального спуска:

- Достижение стационарной точки – проверка нормы градиента и т.д.
- Время работы – рестарт “слишком старых” ветвей
- Близость к другому экземпляру (локальному спуску)
- “Успешность” работы – рестарт экземпляров, имеющих слишком высокое значение оптимизируемой функции
- Изначально разрабатывается для параллельной реализации
- Локальные спуски разбиваются на участки, с фиксированным временем работы (“кванты”)
- Простая синхронизация
- Может быть реализован на аппаратных платформах типа GPGPU (Nvidia CUDA, OpenCL, ...)

Computing Experiments MSBH/Forest

n	UK (CCD)	ISDCT
20	-97.417393	-97.41739307417
80	-690.577890	-690.5778902004155952
147	-1531.498857	-1531.49885718999575761

n is a number of atoms.

Computing Experiments MSBH/Forest

n	CN	ISDCT
150	-1570.956895	-1570.956894507743300
155	-1639.571558	-1639.571558368015758
160	-1705.794373	-1705.794372516992553
165	-1774.727689	-1774.727688598778741
170	-1842.786500	-1842.786499541551848
175	-1911.754684	-1911.754684452901074
180	-1979.907966	-1979.907965818779076
185	-2048.617785	-2048.617785496087890
190	-2119.104888	-2119.104888297832076
195	-2189.107474	-2189.107474368099702
200	-2260.148943	-2260.148943425931975

Computing Experiments MSBH/Forest

n	CN	ISDCT
205	-2329.258501	-2329.25850195624831
210	-2400.884161	-2400.88416410538582
215	-2473.351504	-2473.22663177961703 7
220	-2544.094288	-2543.33035786210166 4
225	-2616.672973	-2616.67297273232043 2
230	-2691.174648	-2691.17464820874693 0
235	-2767.215086	-2767.21508589343966 4
240	-2839.054430	-2839.09992474870296 1

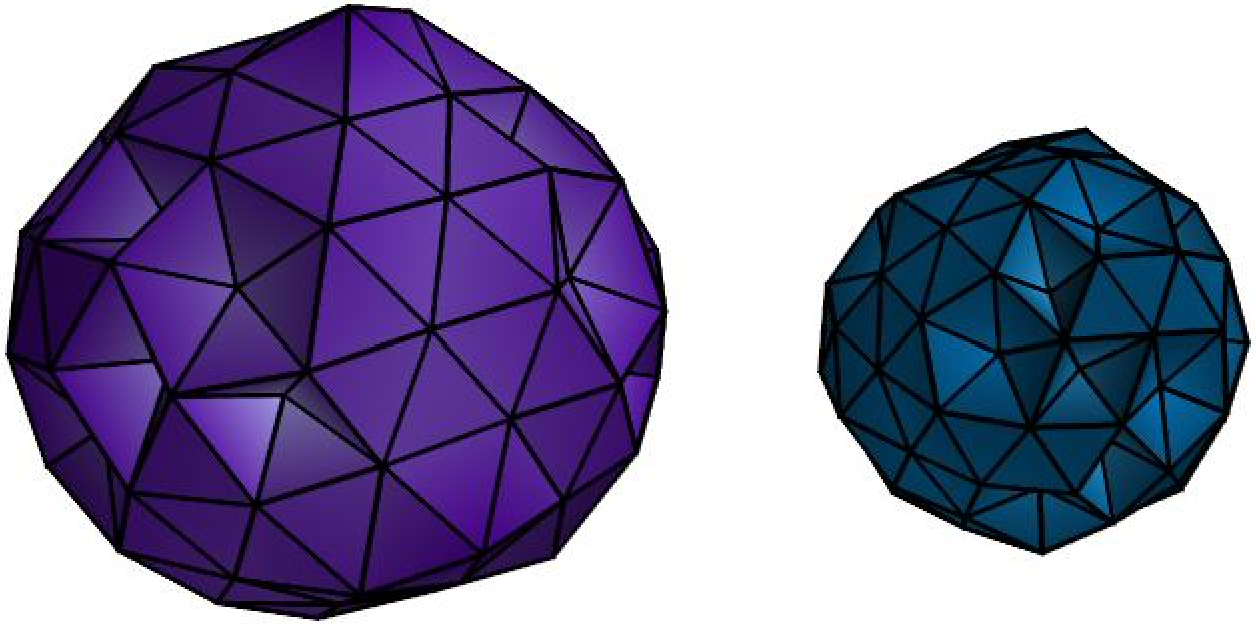
Computing Experiments MSBH/Forest, n = 240

CN	-2839.054430
PT	-2839.099925
ISDCT	-2839.09992474870296l

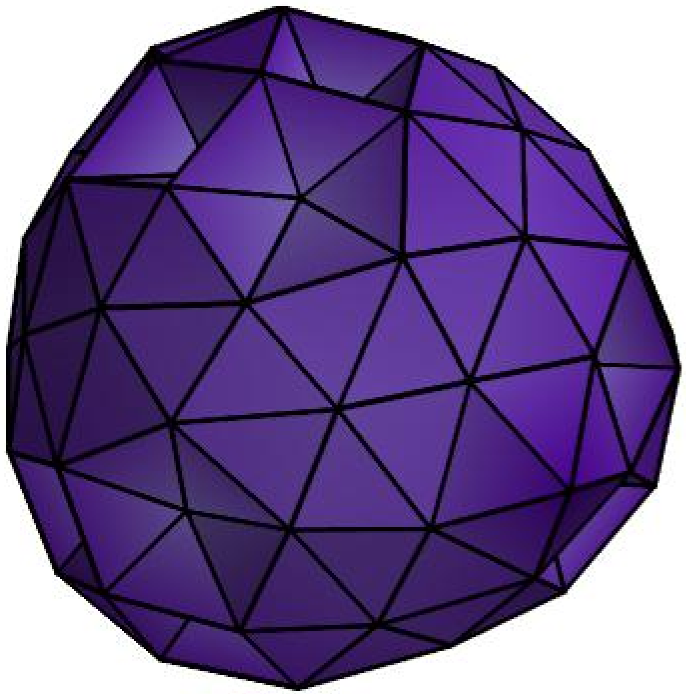
Computing Experiments
Forest

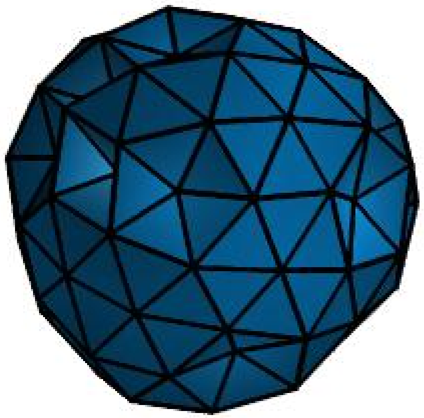
n	ISDCT
241	-2852.938110154795595
242	-2866.778881123787869
243	-2882.570361711906116
244	-2897.072046040393616
245	-2910.707949591107536
246	-2924.517707573666030
247	-2940.293677679405846
248	-2955.679013678213323
249	-2971.203337281702716
250	-2985.771711424368277
251	-2999.469988809987626
252	-3013.550134756378611

150 atoms cluster: CN, ISDCT

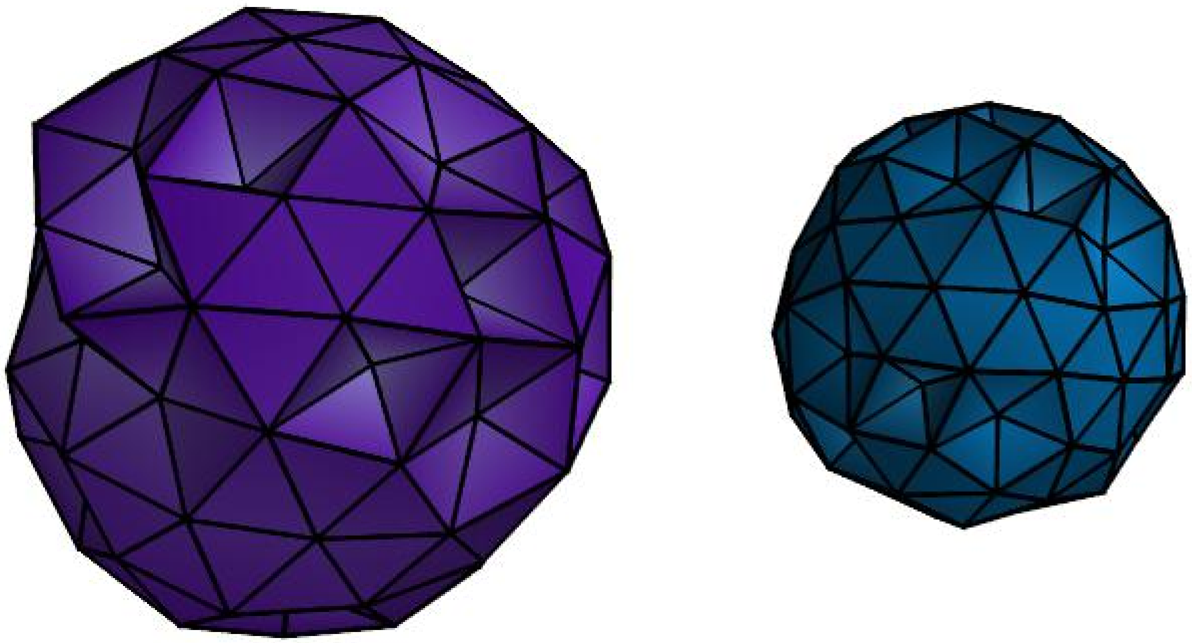


155 atoms cluster: CN, ISDCT

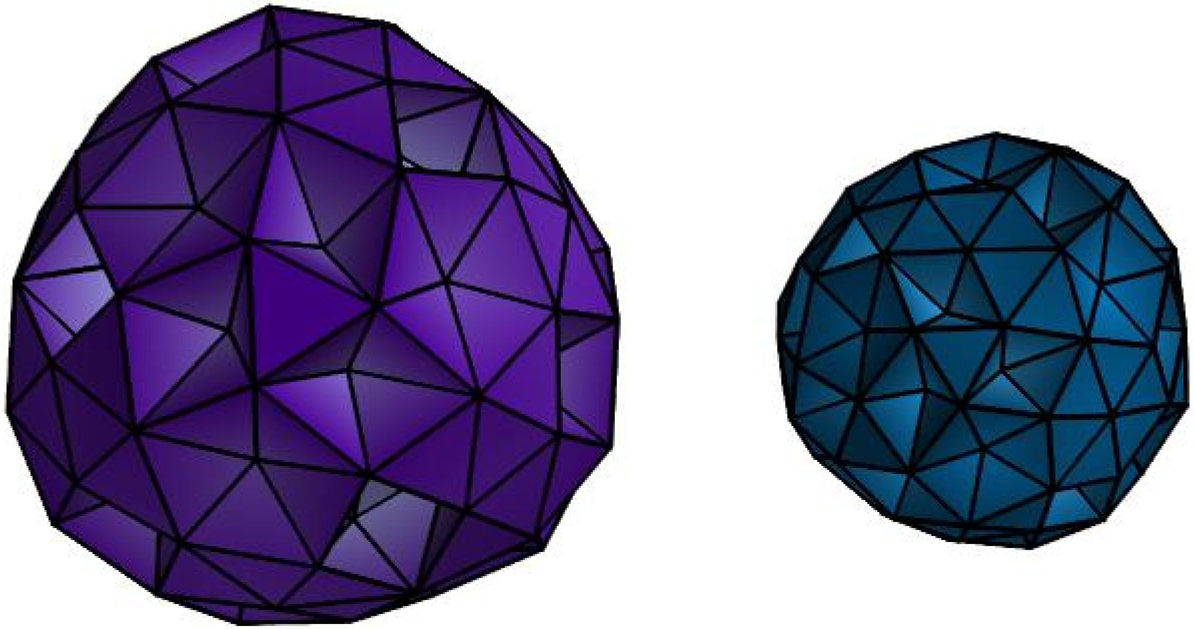




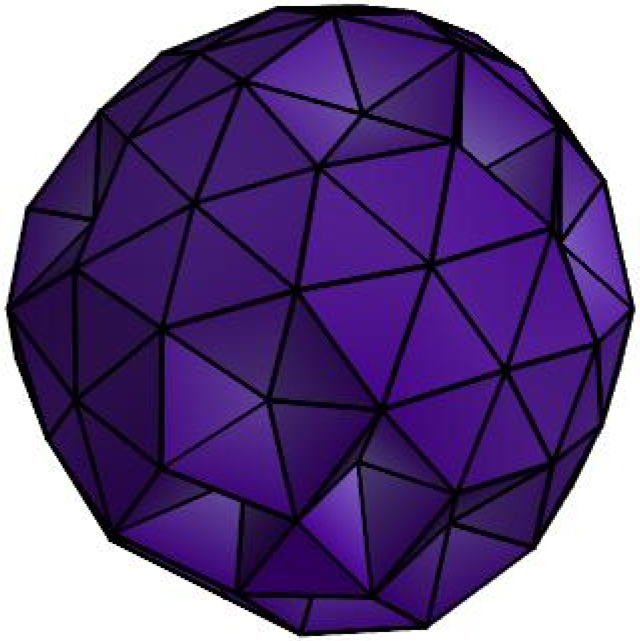
160 atoms cluster: CN, ISDCT

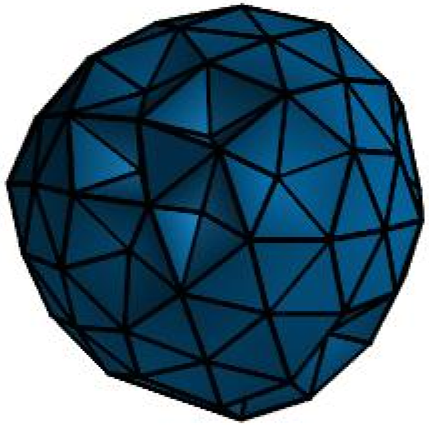


165 atoms cluster: CN, ISDCT

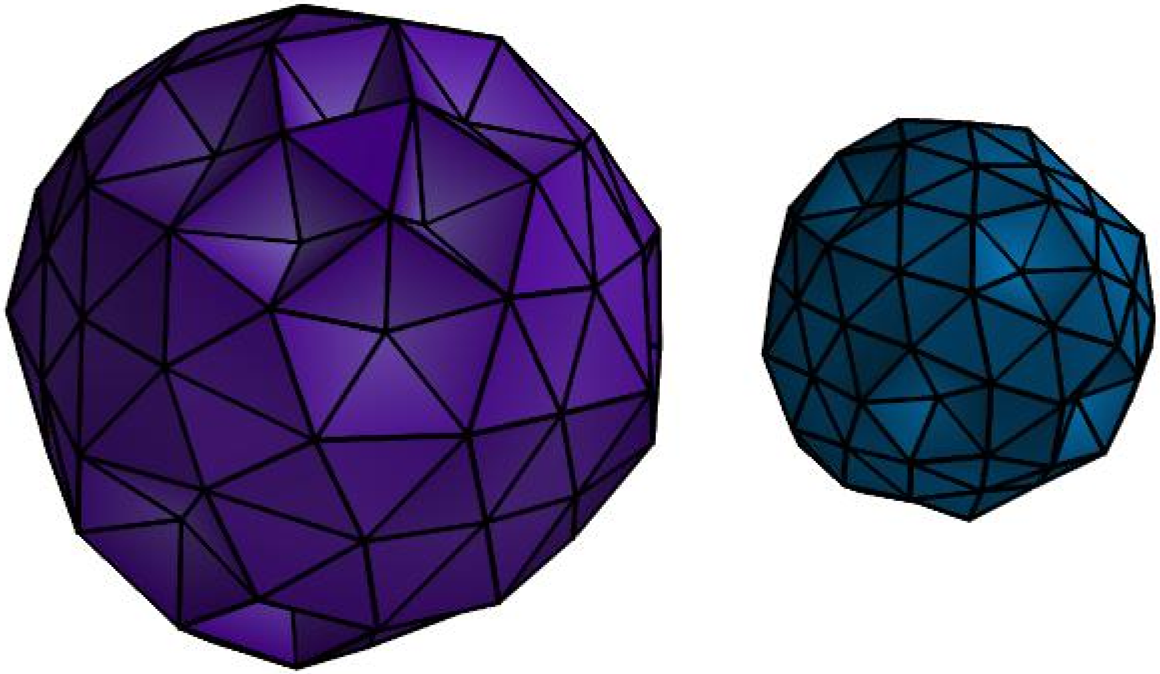


170 atoms cluster: CN, ISDCT

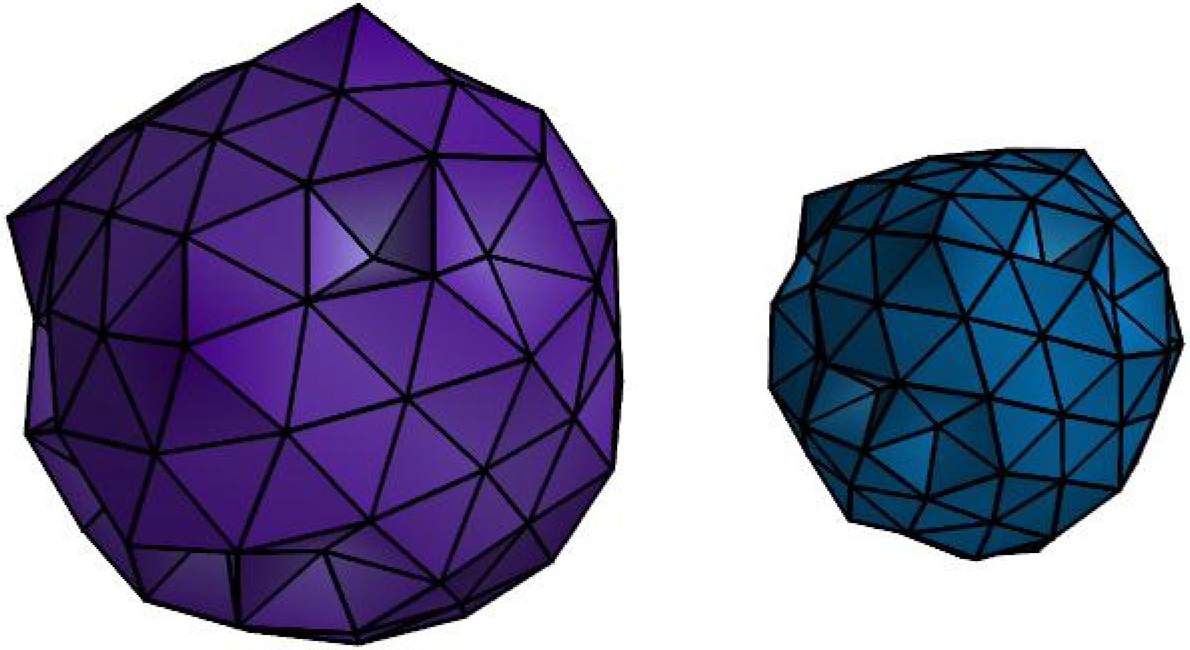




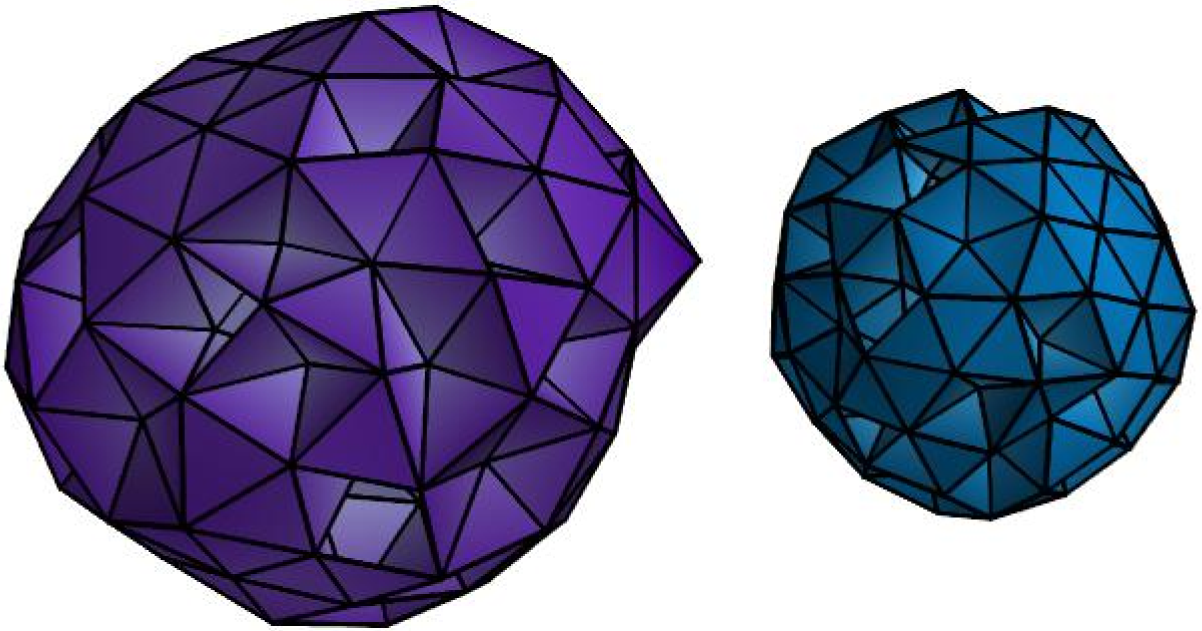
175 atoms cluster: CN, ISDCT



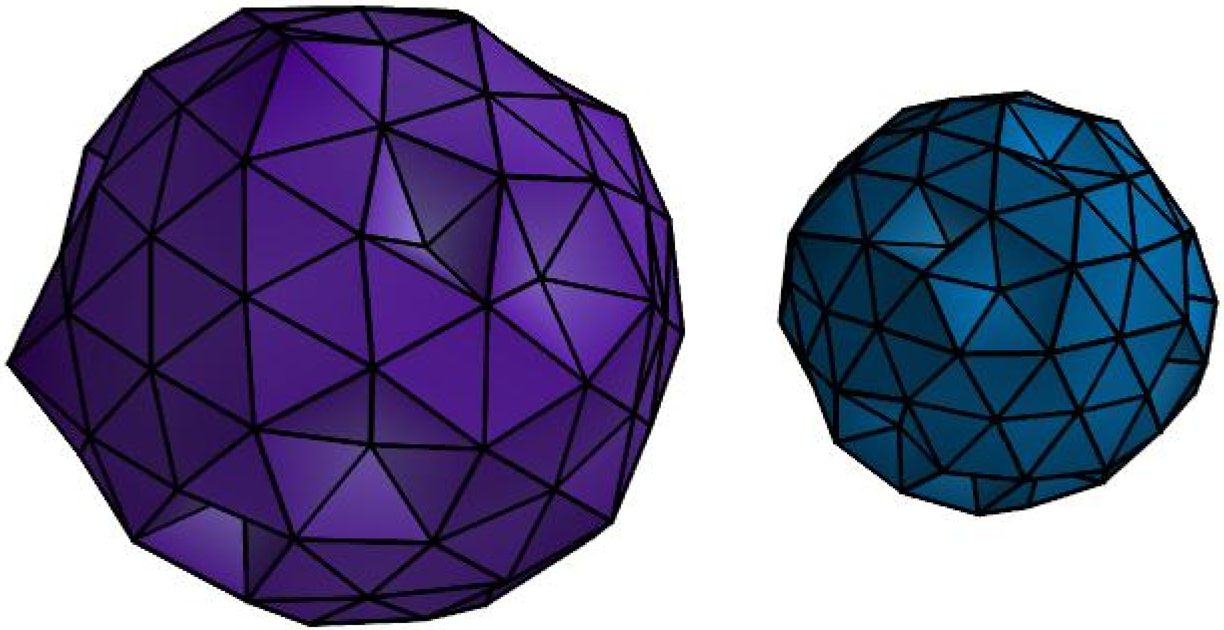
180 atoms cluster: CN, ISDCT



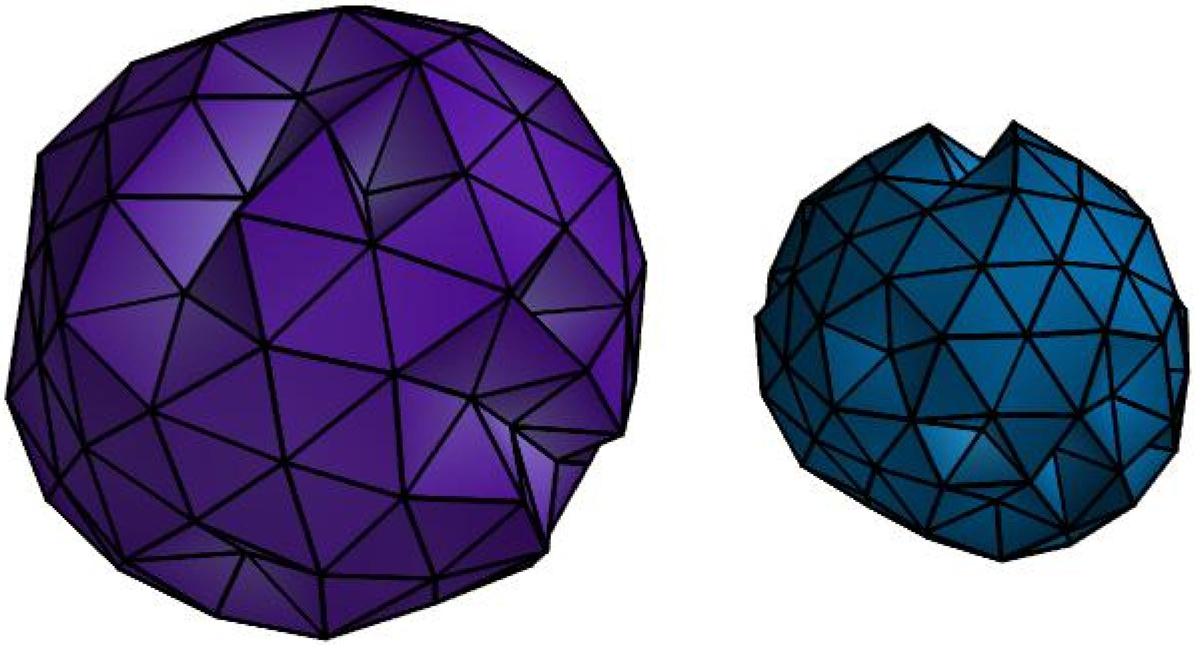
185 atoms cluster: CN, ISDCT



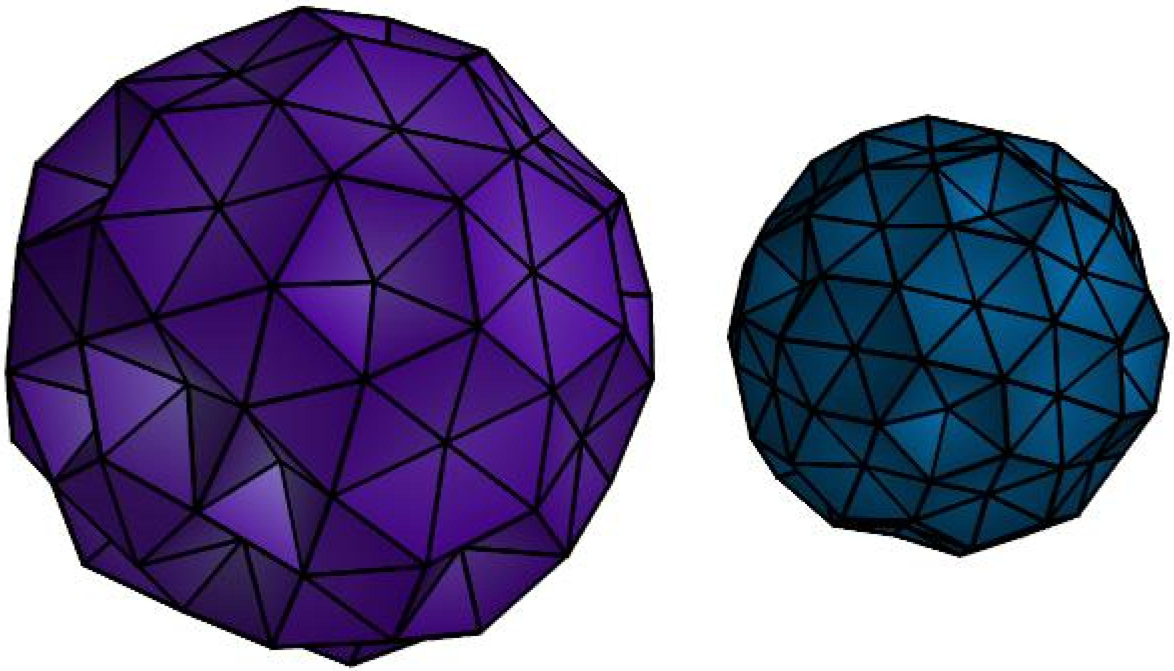
195 atoms cluster: CN, ISDCT



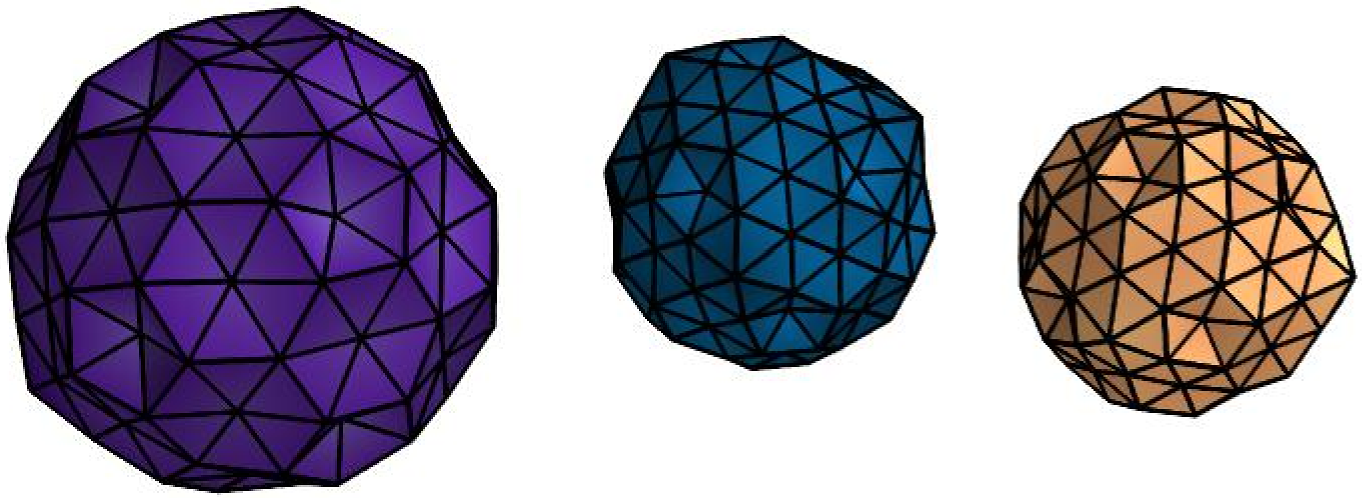
200 atoms cluster: CN, ISDCT



230 atoms cluster: CN, ISDCT

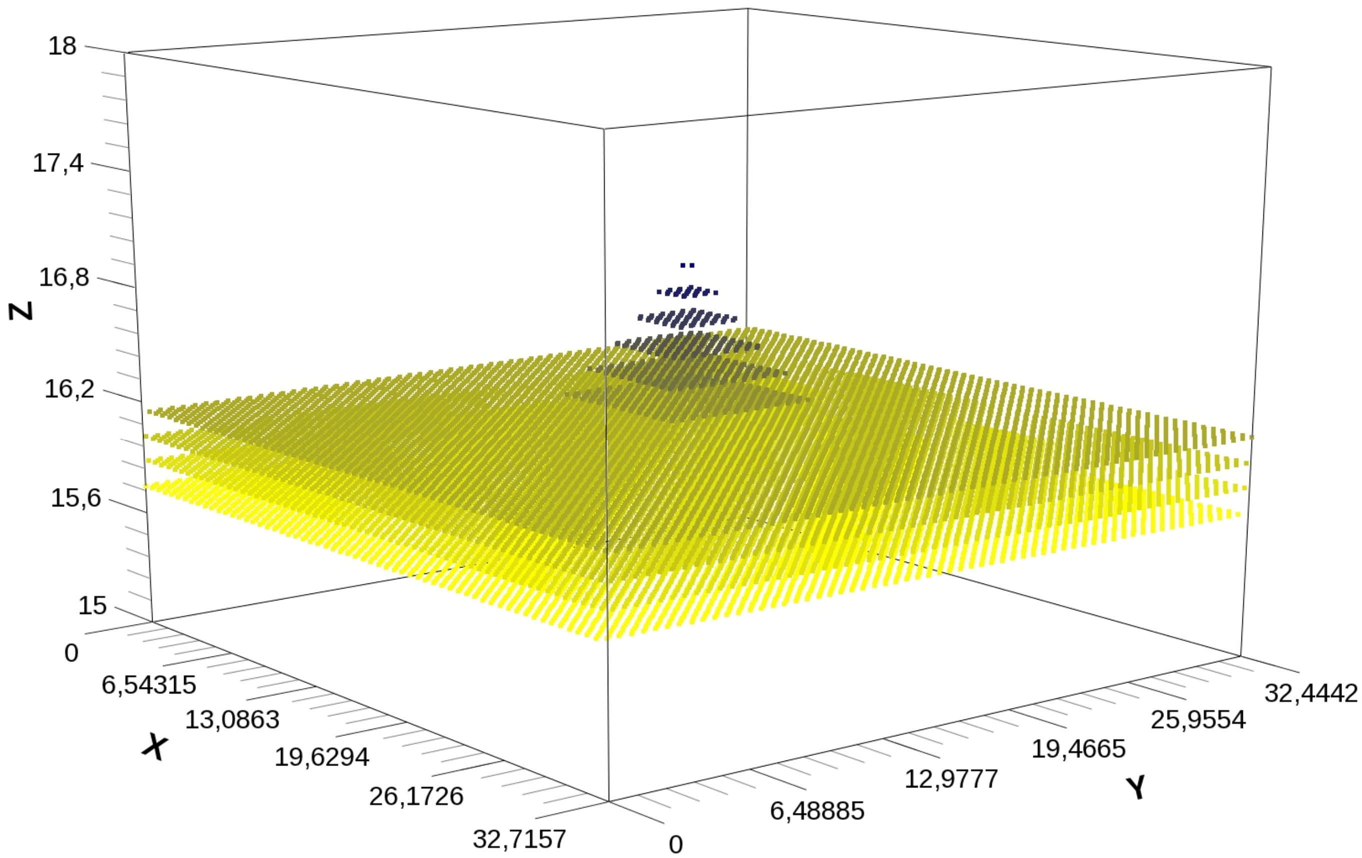


240 atoms cluster: CN, ISDCT, PT



Keating potential optimization.

Form of a quantum dot Si-Ge



Keating potential function

$$E = \sum_{i=1}^{n} \left[\frac{3}{16} \sum_{j=1}^{4} \frac{\alpha_{ij}}{d_{ij}^2} \{ ||r_i - r_j||_2^2 - d_{ij}^2 \}^2 + \right.$$

$$\left. + \frac{3}{8} \sum_{j=1}^{4} \sum_{k=j+1}^{4} \frac{\beta_{ijk}}{d_{ij} \cdot d_{ik}} \left\{ \langle r_i - r_j, r_i - r_k \rangle + \frac{d_{ij} \cdot d_{ik}}{3} \right\}^2 \right]$$

n is the number of atoms in the crystal lattice;

- D_{ij}, d_{ik}, α_{ij}, β_{ijk} are constants set;
- $R_i = (x_i, x_{2i}, x_{3i})$ radius vector of the i-th node (optimized variables).

Features of the problem:

- High dimensionality of 10^5 variables and more.
- The high demands on the result accuracy.

Tested optimization methods:

- Cauchy method;
- Conjugate Gradient Method;
- Newton's Method.

Difficulties with Newton's method implementation

The dimension of these problems depends of physical limits of the Hessian matrix size which flows from the available memory.

The high computational complexity due to the required long time for solving problem of such dimension

Hessian matrix

$$
\left|\begin{array}{ccccc}
\dfrac{\partial f}{\partial x_1 \partial x_1} & \dfrac{\partial f}{\partial x_1 \partial x_2} & \dfrac{\partial f}{\partial x_1 \partial x_3} & \cdots & \dfrac{\partial f}{\partial x_1 \partial x_n} \\
\dfrac{\partial f}{\partial x_2 \partial x_1} & \dfrac{\partial f}{\partial x_2 \partial x_2} & \dfrac{\partial f}{\partial x_2 \partial x_3} & \cdots & \dfrac{\partial f}{\partial x_2 \partial x_n} \\
\dfrac{\partial f}{\partial x_3 \partial x_1} & \dfrac{\partial f}{\partial x_3 \partial x_2} & \dfrac{\partial f}{\partial x_3 \partial x_3} & \cdots & \dfrac{\partial f}{\partial x_3 \partial x_n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\dfrac{\partial f}{\partial x_n \partial x_1} & \dfrac{\partial f}{\partial x_n \partial x_2} & \dfrac{\partial f}{\partial x_n \partial x_3} & \cdots & \dfrac{\partial f}{\partial x_n \partial x_n}
\end{array}\right|
$$

Storage of a dense matrix requires about n^2 memory cells.

Sparse Hessian matrix

$$\begin{vmatrix} \frac{\partial f}{\partial x_1 \partial x_1} & \frac{\partial f}{\partial x_1 \partial x_2} & 0 & \cdots & \frac{\partial f}{\partial x_1 \partial x_n} \\ \frac{\partial f}{\partial x_2 \partial x_1} & \frac{\partial f}{\partial x_2 \partial x_2} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_n \partial x_1} & 0 & 0 & \cdots & \frac{\partial f}{\partial x_n \partial x_n} \end{vmatrix}$$

Storage of a sparse matrix requires less then n^2 memory cells.

Methods of sparse matrix storage

- Diagonal scheme for storing circuit tape matrices;
- Profile storage scheme of symmetric matrices;
- Connected scheme of sparse storage;
- Sparse line format;

and a number of other methods, and various their modifications.

Методы хранения разреженных матриц

• Диагональная схема хранения ленточных матриц.
• Профильная схема хранения симметрических матриц.
• Связные схемы разреженного хранения.
• Разреженный строчный формат,

а так же ряд других методов и различные их модификации.

Applied method of sparse matrix storage

Индексы :	$I_{1,1}, I_{1,2} \ldots I_{1,L}$	$I_{2,1}, I_{2,2} \ldots I_{2,L}$	$\ldots$	$I_{n,1} \ldots I_{n,L}$
Значения :	$V_{1,1}, V_{1,2} \ldots V_{1,L}$	$V_{2,1}, V_{2,2} \ldots V_{2,L}$	$\ldots$	$V_{n,1} \ldots V_{n,L}$

L is a maximum number of nonzero elements in the Hessian line. For considered problem L = 51.

$$I_{i,1} = i$$

$$V_{j,k} = \frac{\partial f}{\partial x_j \partial x_m}, \; m = I_{j,k}$$

The ratio of dense and sparse matrices size

n	M	M_{sparse}	M_{sparse}/M
10	$800\ b$	$6\ Kb$	7.65
10^2	$78.1\ Kb$	$59.8\ Kb$	$7.65 \cdot 10^{-1}$
10^3	$7.6\ Mb$	$598.7\ Kb$	$7.65 \cdot 10^{-2}$
10^4	$762.9\ Mb$	$5.8\ Mb$	$7.65 \cdot 10^{-3}$
10^5	$74.5\ Gb$	$58.4\ Mb$	$7.65 \cdot 10^{-4}$
10^6	$7.3\ Tb$	$583.6\ Mb$	$7.65 \cdot 10^{-5}$
10^7	$727.6\ Tb$	$5.7\ Gb$	$7.65 \cdot 10^{-6}$

$M = 8\, n^2$ the size of a dense array (in bytes).

$M_{sparse} = 51\,(8 + 4)\, n$ the size of a sparse matrix (in bytes).

Features of used storage format

Selected format has a number of positive features:

- a fixed amount of memory used order $2 \cdot L \cdot n$ cells;
- a small number of occupied cells, about 4-5 % when the dimension of the problem is 10^5 ;
- a quick access to the elements of the main diagonal;
- ease of implementation procedures for sparse matrix multiplication
- on a tight vector.

Выбраный формат имеет ряд положительных особенностей :

- Фиксированный размер используемой памяти порядка $2 \cdot L \cdot n$ ячеек
- Малое число незанятых ячеек –
  4 - 5% при размерности задачи 10^5
- Быстрый доступ к элементам главной диагонали
- Простота реализации процедуры умножения разреженной матрицы на плотный вектор

Computing experiments

Newton's method modification

n	The value of the function before optimization	The value of the function after optimization	t, s	Number of iterations
98304	$1.744335 \cdot 10^{-3}$	$2.678759 \cdot 10^{-23}$	190	11
139968	$1.744335 \cdot 10^{-3}$	$1.240275 \cdot 10^{-22}$	260	11

Computing experiments

Conjugate gradient method

n	The value of the function before optimization	The value of the function after optimization	t, s	Number of iterations
24000	$1.742574 \cdot 10^{-3}$	$5.230097 \cdot 10^{-18}$	5.6	434
81000	$1.742574 \cdot 10^{-3}$	$6.477517 \cdot 10^{-18}$	43.8	803
201600	$1.742574 \cdot 10^{-3}$	$7.544217 \cdot 10^{-18}$	109	1145
421824	$1.742574 \cdot 10^{-3}$	$9.643463 \cdot 10^{-18}$	312	1527
648000	$1.742574 \cdot 10^{-3}$	$2.538089 \cdot 10^{-17}$	564	1747
1536000	$1.742574 \cdot 10^{-3}$	$7.920582 \cdot 10^{-17}$	2003	2349
10535424	$1.742574 \cdot 10^{-3}$	$6.006104 \cdot 10^{-15}$	13204	2154
21233664	$1.742574 \cdot 10^{-3}$	$9.104004 \cdot 10^{-15}$	16009	1960

Tests were carried out on a computer system containing 10 cores Intel Xeon X5670

Huge-Scale separable convex optimization problem

The classification of local optimization problems on the number of variables proposed by the Yu.E. Nesterov:

- “Small” – up to 100 variables
- “Medium” – from 10^3 to 10^4 variables
- “Large” – from 10^5 to 10^7 variables
- “Huge” – more than 10^8 variables

Huge-Scale separable convex optimization problem

Difficulties

- the number of variables – memory limitations
- the computational complexity – time limit
- the required amount of computation – time limit

$\downarrow$

parallel computing

Required memory

- float - 4 bytes per cell
- double - 8 bytes per cell

Required memory

the vector size of n elements

n	float		double	
10^2	0.39	КБ	0.78	КБ
10^3	3.91	КБ	7.81	КБ
10^4	39.06	КБ	78.13	КБ
10^5	390.63	КБ	781.25	КБ
10^6	3.81	МБ	7.63	МБ
10^7	38.15	МБ	76.29	МБ
10^8	381.47	МБ	762.94	МБ
10^9	3.73	ГБ	7.45	ГБ
10^{10}	37.25	ГБ	74.51	ГБ
10^{11}	372.53	ГБ	745.06	ГБ
10^{12}	3.63	ТБ	7.28	ТБ

Memory (RAM) - the main hardware limitations for many modern Huge-Scale optimization problem.

Test optimization problem 1

$$f(x) = \sum_{i=1}^{n} (x_i^2 + x_i^6)$$

This test function is convex, separable, the minimum value is known ($f_{\min} = 0$).

The calculation of values of the function $f(x)$ and its gradient $\nabla f(x)$ performed in parallel on different CPU cores, for large-scale problems it is impossible for one compute node due to physical limitations on the amount of RAM.

Computing experiments

Computational experiments were carried out with

•Computing cluster MBC-100K of Interdepartmental Supercomputer Center.
RAM – 1 Gb, 1 CPU.

•Computing cluster MBC-100K of Keldysh Institute of Applied Mathematics RAS.

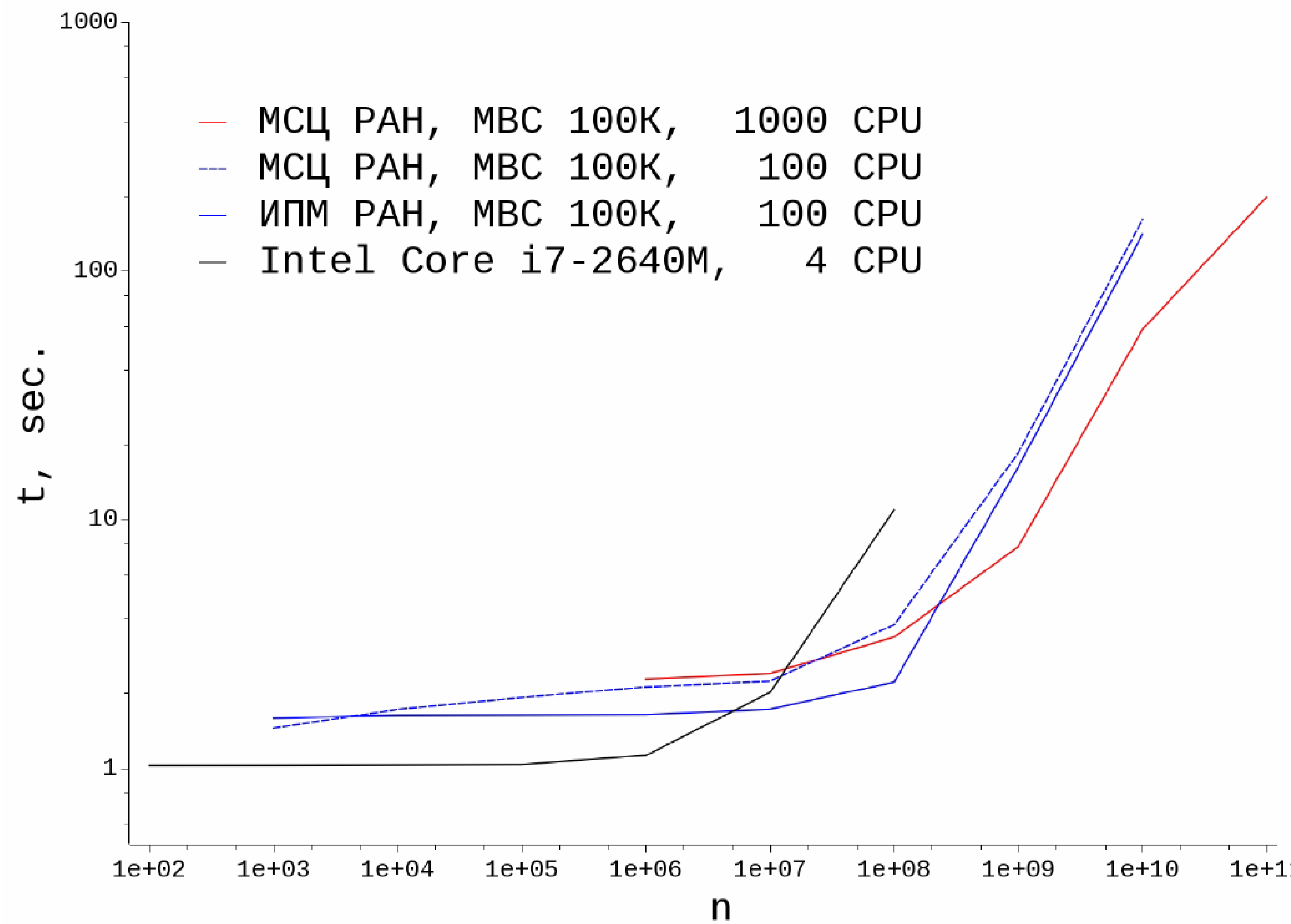
•Computing cluster “Academician V.M. Matrosov”, unit “Tesla”.
RAM – 250 Gb, 32 CPU.

Computing experiments

the running time of algorithm

n	Core i7, 4 CPU	ИПМ, 100 CPU	МСЦ, 100 CPU	МСЦ, 1000 CPU
10^2	1.02559			
10^3	1.02685	1.58932	1.44929	
10^4	1.03031	1.63431	1.72392	
10^5	1.03406	1.63837	1.92358	
10^6	1.12513	1.64386	2.12392	2.27892
10^7	2.01723	1.72646	2.23966	2.40381
10^8	10.93844	2.22012	3.77897	3.36539
10^9		16.29881	18.55662	7.78179
10^{10}		140.30873	160.74543	58.09801
10^{11}				198.05384

Computing experiments



Required memory (1 core)

n	RAM	
10^2	0.78	КБ
10^3	7.81	КБ
10^4	78.13	КБ
10^5	781.25	КБ
10^6	7.63	МБ
10^7	76.29	МБ
10^8	762.94	МБ

Required memory (100 cores)

n	RAM		RAM / 1 CPU core	
10^7	76.29	МБ	76.29	МБ
10^8	762.94	МБ	762.94	МБ
10^9	7.45	ГБ	7.63	МБ
10^{10}	74.51	ГБ	76.29	МБ
10^{11}	745.06	ГБ	762.94	МБ
10^{12}	7,28	ТБ	7.45	ГБ

Computing experiments

- good scalability of the proposed implementation;
- the main limiting factor - the amount of available RAM;
- High-performance for separable problems.

Test optimization problem 2

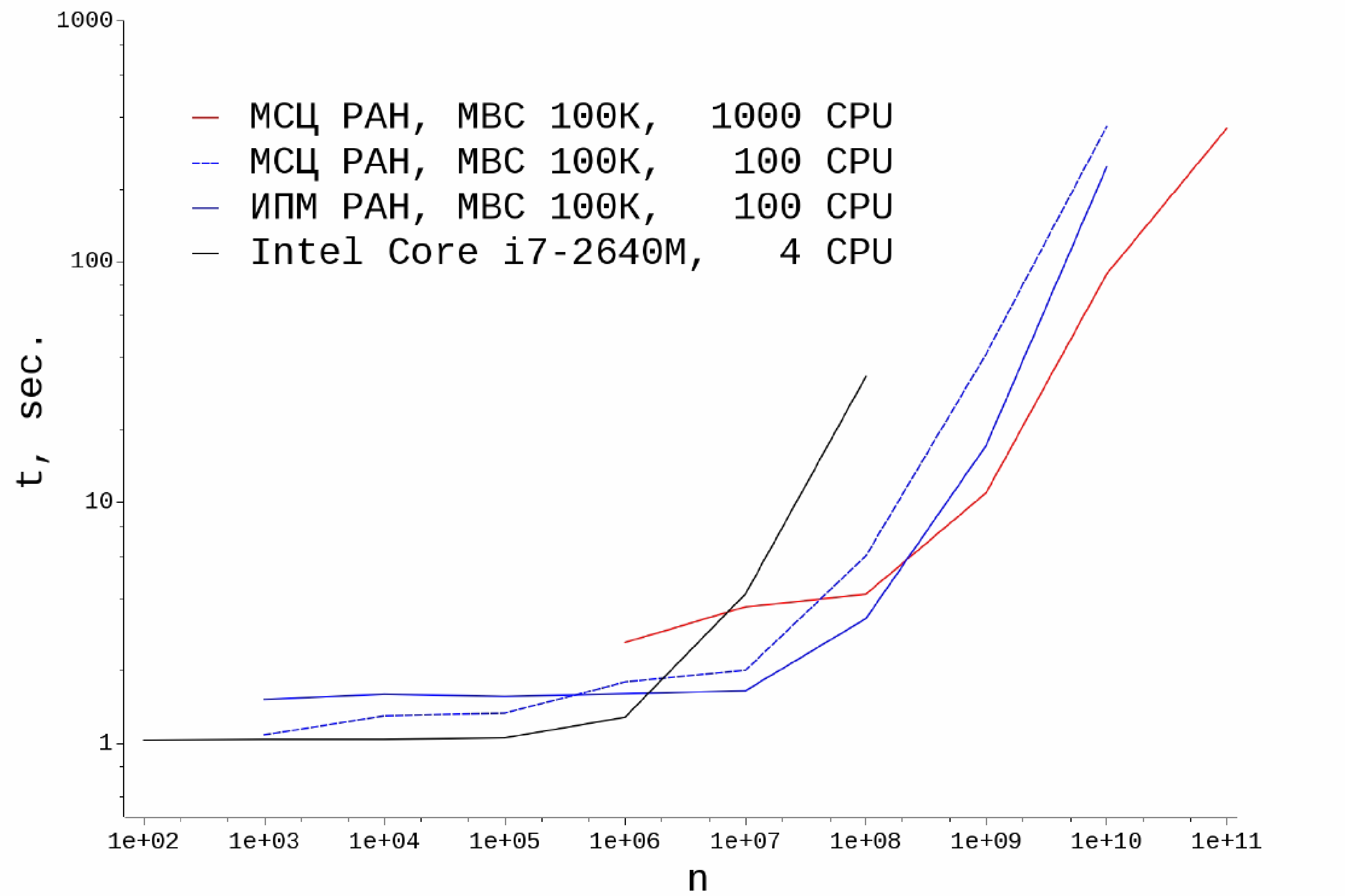
$$f(x) = \sum_{i=1}^{n} x_i^2 + \sum_{i=2}^{n} (x_i - x_{i-1})^2$$

Computing experiments

the running time of algorithm

n	Core i7, 4 CPU	ИПМ, 100 CPU	МСЦ, 100 CPU	МСЦ, 1000 CPU
10^2	1.02845			
10^3	1.03536	1.52187	1.08275	
10^4	1.03566	1.59923	1.30111	
10^5	1.05166	1.56577	1.33428	
10^6	1.28414	1.60642	1.79563	2.62236
10^7	4.17233	1.65246	2.00984	3.68216
10^8	33.29102	3.29295	5.99972	4.15825
10^9		17.26339	41.40902	10.97092
10^{10}		246.75377	363.31317	88.66335
10^{11}				358.11812

Computing experiments



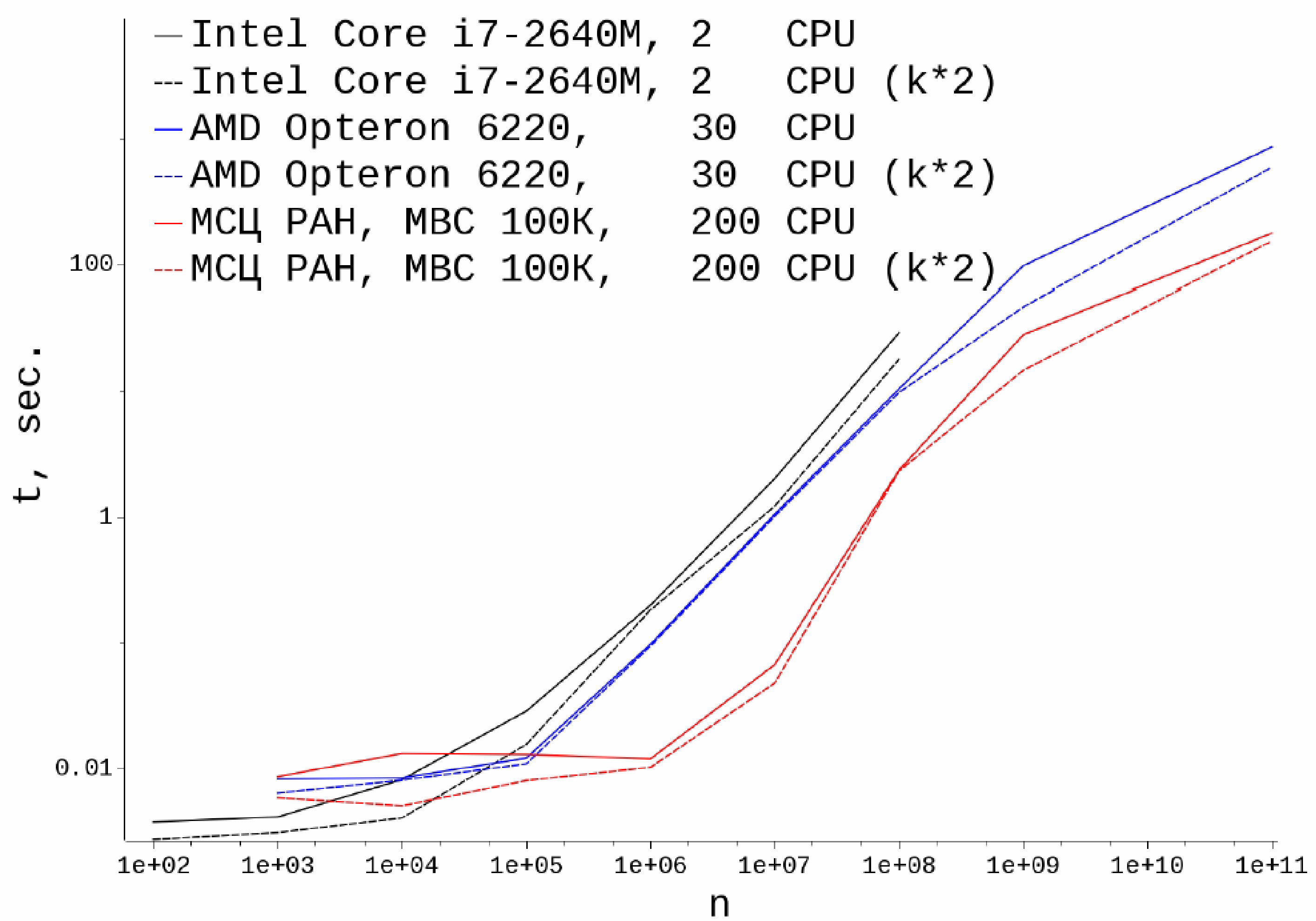
Computing experiments, test 2

Modification of Polyak method, time (s)

n	Core i7, 2 CPU K=1	Core i7, 2 CPU K=2	Matrosov, 30 CPU K=1	Matrosov, 30 CPU K=2	МСЦ, 200 CPU K=1	МСЦ, 200 CPU K=2
10^2	0.003	0.002				
10^3	0.004	0.003	0.008	0.006	0.008	0.005
10^4	0.008	0.004	0.008	0.008	0.013	0.005
10^5	0.028	0.015	0.012	0.010	0.012	0.007
10^6	0.199	0.181	0.097	0.093	0.011	0.010
10^7	2.021	1.218	1.052	1.013	0.066	0.047
10^8	29.261	18.030	10.444	9.794	2.361	2.322
10^9			97.397	46.655	28.134	14.689
10^{10}			857.933	592.157	176.949	153.497

Computing experiments, test 2

Modification of Polyak method, time (s)



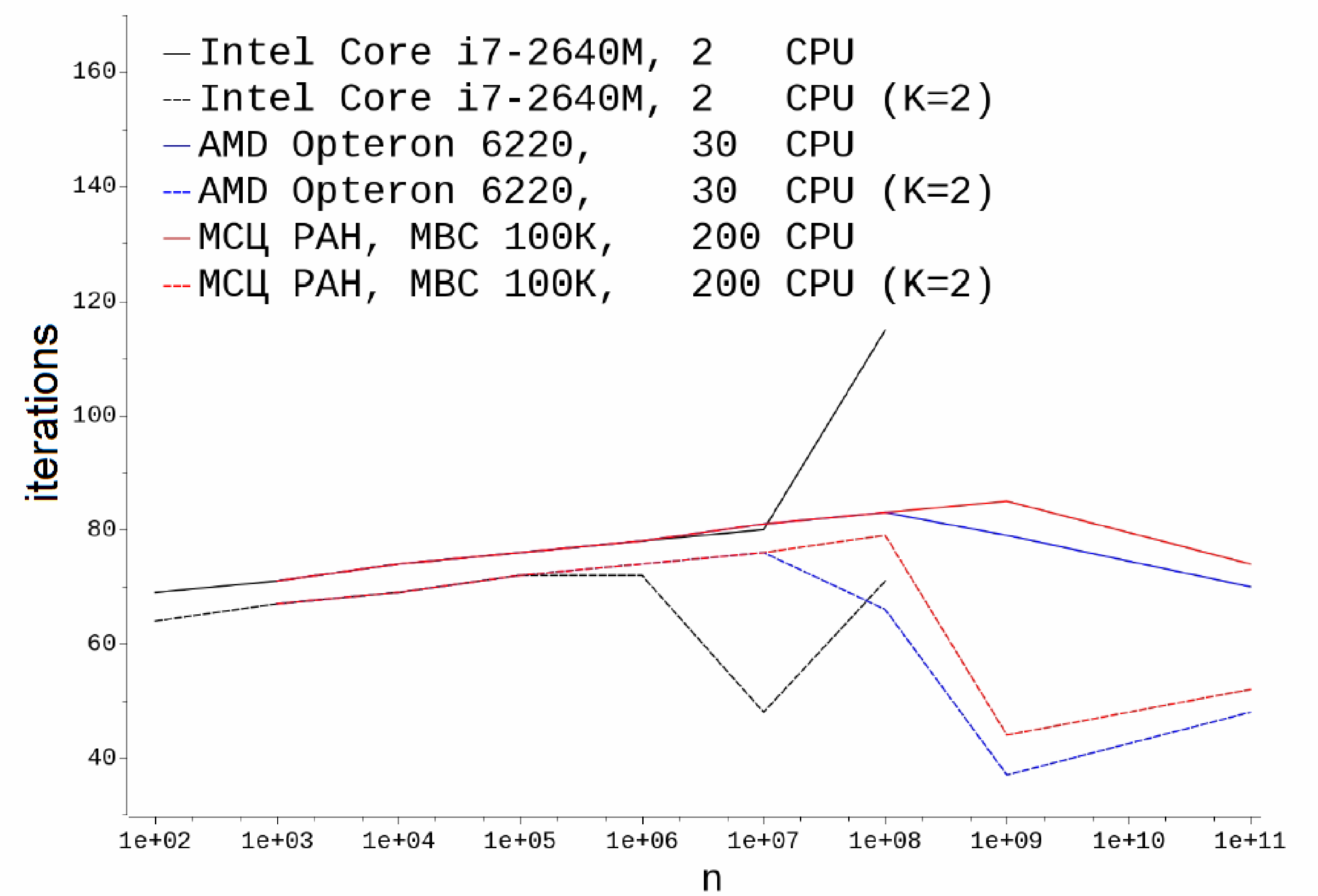
Computing experiments, test 2

Modification of Polyak method, iterations number

n	Core i7, 2 CPU K=1	Core i7, 2 CPU K=2	Matrosov, 30 CPU K=1	Matrosov, 30 CPU K=2	МСЦ, 200 CPU K=1	МСЦ, 200 CPU K=2
10^2	69	64				
10^3	71	67	71	67	71	67
10^4	74	69	74	69	74	69
10^5	76	72	76	72	76	72
10^6	78	72	78	74	78	74
10^7	80	48	81	76	81	76
10^8	115	71	83	66	83	79
10^9			79	37	85	44
10^{10}			70	48	74	52

Computing experiments, test 2

Modification of Polyak method, iterations number



The problem of finding PageRank-vector

$$P^T x = x$$

$$P \in R^{n \times n}, \quad x \in R^n$$

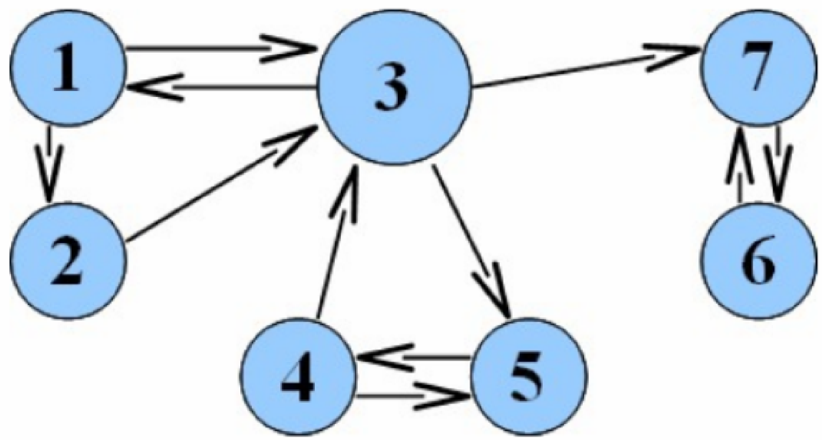
$$\langle x, e \rangle = 1, \quad e = (1, \ldots, 1)^T$$

$$x_i \geq 0, \quad i = \overline{1, n}$$

P is a stochastic matrix that defines the original graph.

It is implemented the Fletcher-Reeves conjugate gradient method for PageRank problem.

PageRank problem



$$P^T = \begin{pmatrix} 0 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1/3 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1/3 & 0 & 0 & 1 & 0 \end{pmatrix}$$

PageRank problem

$$f(x) = \frac{1}{2}||Ax||_2^2 \to \min_{x \in S_n(1)} \quad (1)$$

$$f(x) = ||Ax||_\infty \to \min_{x \in S_n(1)} \quad (2)$$

$$f(x) = \frac{1}{2}||Ax||_2^2 + \frac{\gamma}{2}(\langle x, e \rangle - 1)^2 \to \min \quad (3)$$

where $A = P^T - I$, I is unit matrix, $S_n(1)$ is unit simplex in;

$e = (1,...,1)$;

γ is penalty parameter for missing constrair $\langle x, e \rangle = 1$.

Traditional gradient methods

We make complete ("normal") calculation of the optimized function and its gradient at each iteration.

Computational complexity of order $O(s\ n)$.

Tested implementation (CPU + GPU):

- Conjugate gradient method (**CG**, different versions);
- Conjugate gradient method of Yuri Nesterov;
- Barzilai-Borwein method (**BB**);
- B.T. Polyak method;
- Cauchy method.

Computational experiments

Were performed on system with the following characteristics:

- Intel Core i5-2500K, 16 GB RAM, GeForce GTX 580 (512 CUDA Cores)
- gcc-5.2.1
- CUDA toolkit 7.5

The assembly is made in Release mode.
Compilation flags: $-\mathrm{O2}\ -\mathrm{std} = \mathrm{c}++\ 11\ -\mathrm{mcmodel} = \mathrm{small}$.

Test web-graphs were downloaded from Stanford University website:

- Stanford Large Network Dataset Collection (snap.stanford.edu/data)

- For all the tasks we set $f^* = f_0\, 10^{-4}$, the algorithms were allowed to use unlimited time and iterations. The starting point was set $x_0 = 1/n \cdot e$.

Characteristics of the A matrix

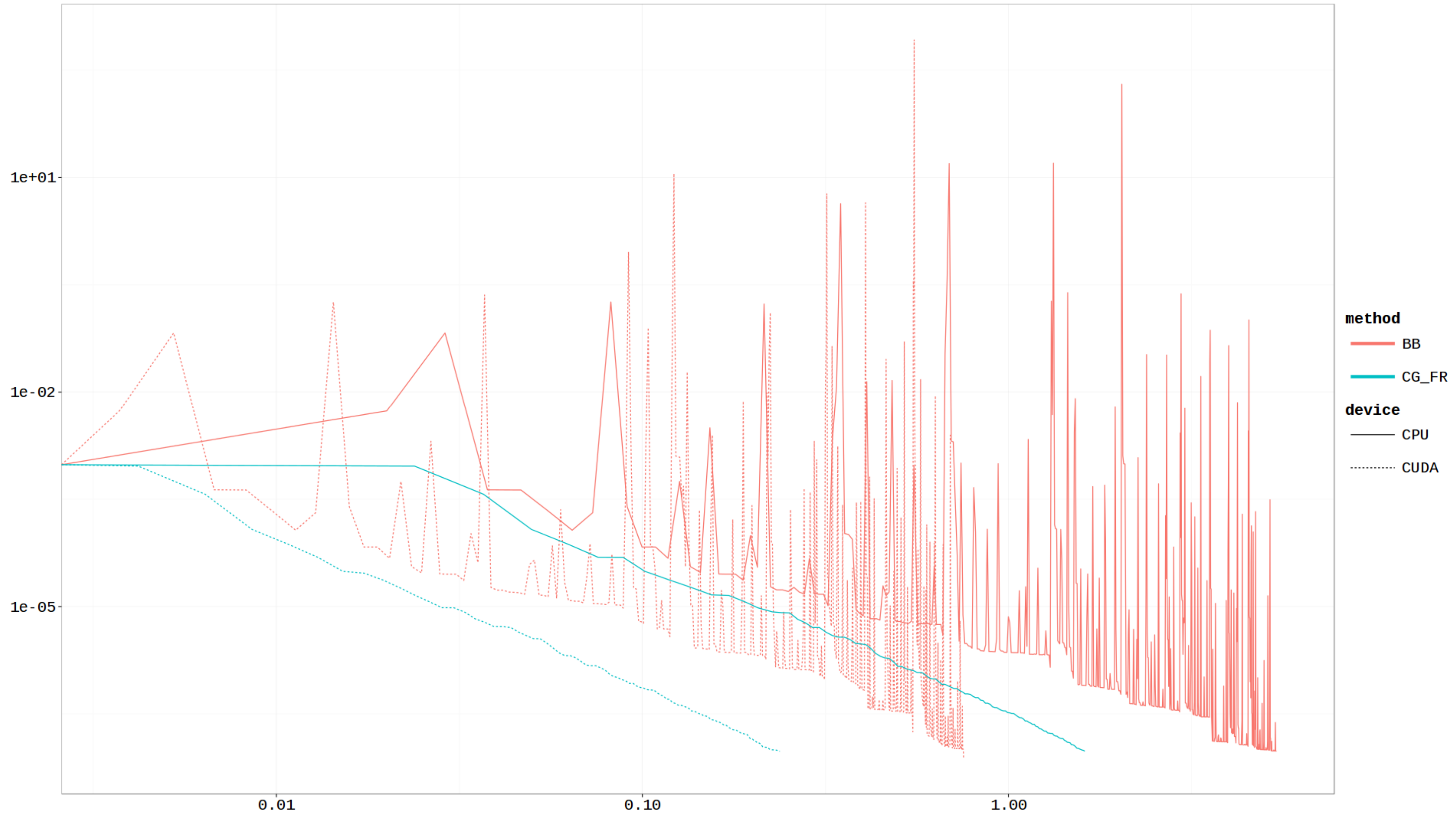
web-graph		Число ненулевых элементов				
		в строке		в столбце		среднее
		мин.	макс.	мин.	макс.	
Stanford,	$n = 281903$	2	38607	1	256	9.2
NotreDame,	$n = 325729$	2	10722	1	3445	5.51
BerkStan,	$n = 685230$	1	84209	1	250	12.09
Google,	$n = 875713$	1	6327	1	457	6.83

Minimization time

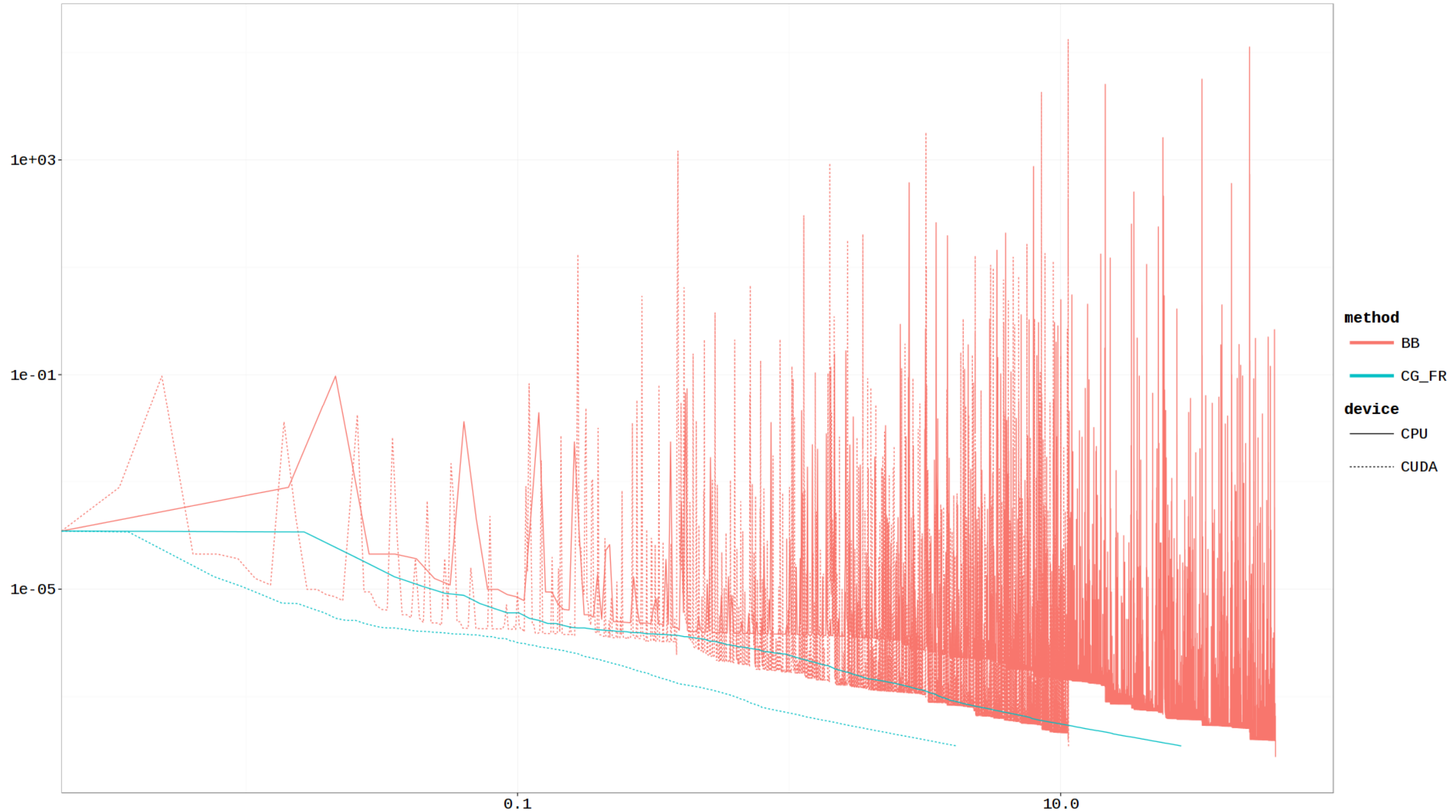
web-graph	CG			BB		
	CPU	GPU	$\frac{CPU}{GPU}$	CPU	GPU	$\frac{CPU}{GPU}$
Stanford	1.61	0.23	*7.00*	5.39	0.75	*7.18*
NotreDame	27.78	4.15	*6.70*	61.81	10.68	*5.78*
BerkStan	5.49	0.90	*6.10*	18.22	3.86	*4.72*
Google	52.47	4.46	*11.76*	176.91	8.76	*20.19*
Суммарно	87.35	9.74	*8.96*	262.33	24.04	*10.91*

Stanford Problem.

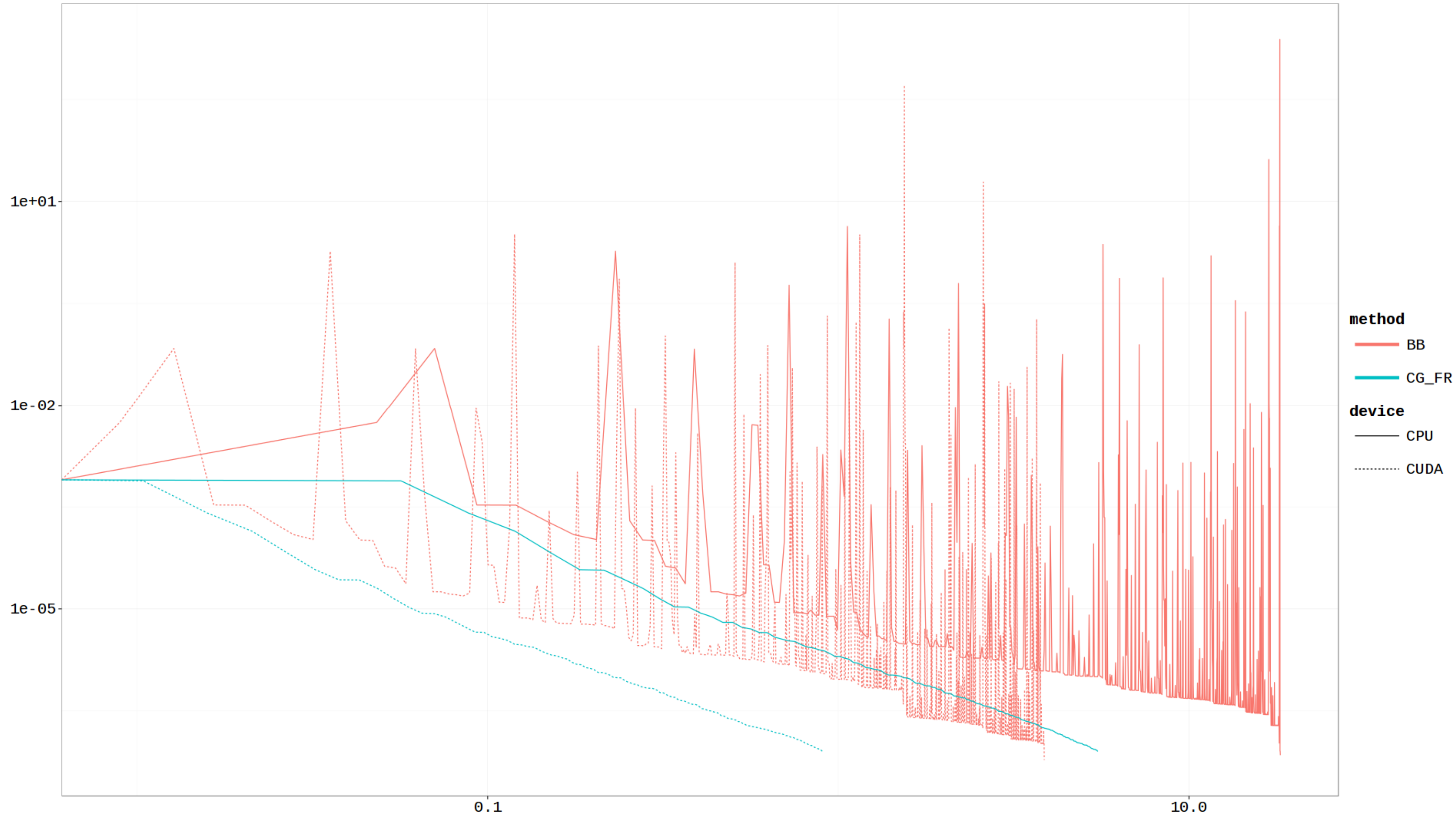
Methods convergence



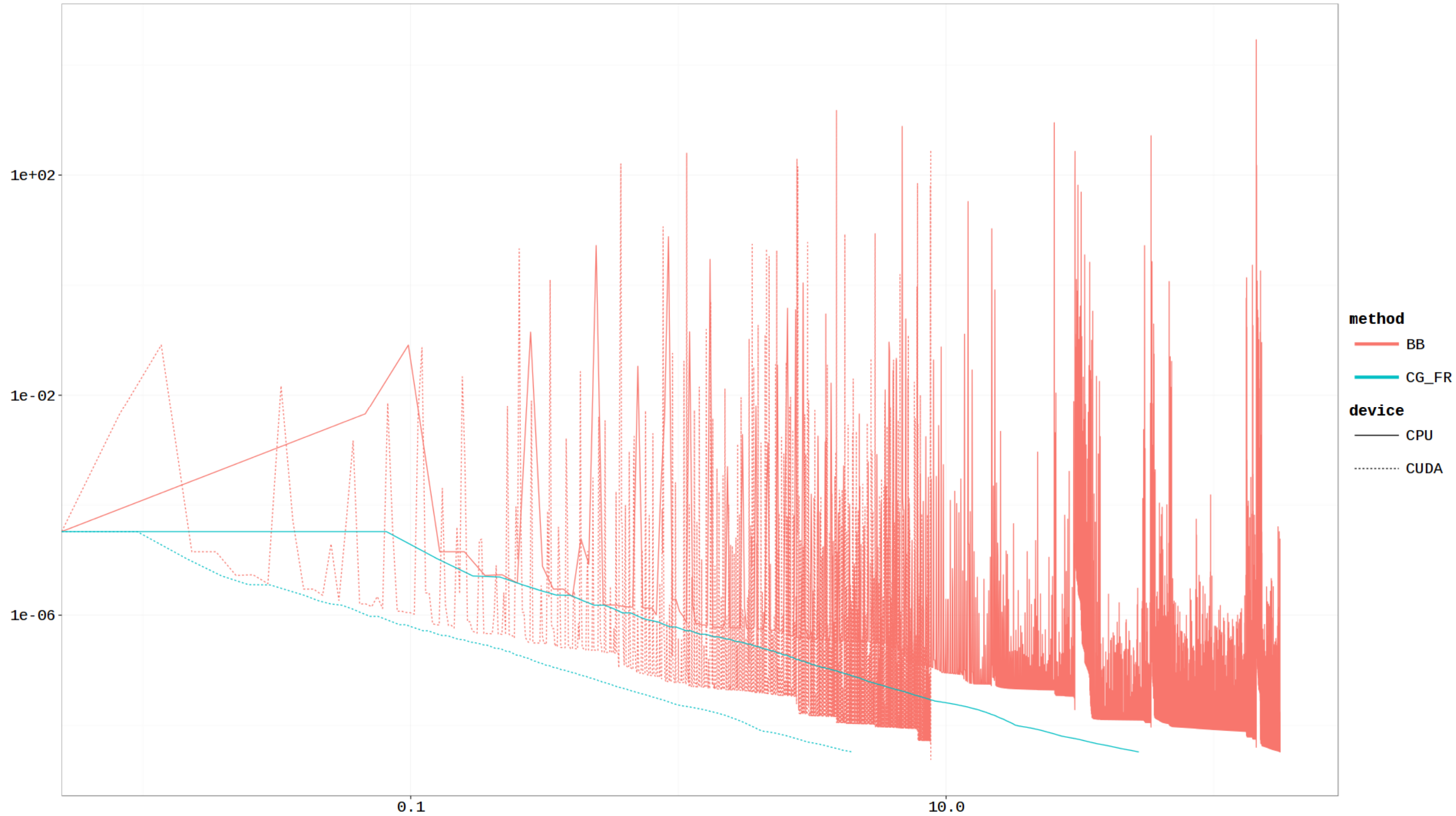
NotreDame Problem.
Methods convergence



BerkStan Problem.
Methods convergence



Google Problem.
Methods convergence



Update methods

The basic idea, which can allow to solve effectively such problems, is to take into account the matrix sparseness when selecting optimization method, and implementing the program.

Considered methods at each iteration minimize not all the components of the vector x, but only several (1-2 variables). This approach is associated with a sparse statement (the matrix A), and with the sparseness of the solution (a vector x) for this class of problems.

This approach allows one to build effective in complexity evaluating methods, but often requires a number of non-trivial steps for efficient software implementations.

«Пересчетные» методы

Основной идеей, позволяющей эффективно решать такие задачи является правильный учет разреженности исходной постановки как на уровне выбираемого метода оптимизации, так и на уровне его последующей программной реализации.

Рассматриваемые методы относятся к покомпонентным, т.е. на каждой итерации производится минимизация не по всем компонентам вектора x, а лишь по небольшой его части (1-2 переменных). Данный подход связан как с разреженностью постановки (матрица A), так и с разреженностью самого решения (вектор x) для рассматриваемого класса задач.

Такой учет фактора разреженности позволяет построить эффективные методы с точки зрения оценки сложности, но зачастую требует выполнения ряда нетривиальных шагов для получения эффективных программных реализаций.

Example of Update Iteration

Accordint to the philosophy of componentwise methods for each iteration, we slightly change the optimized vector $x_{k+1} = x_k + e_k$. Here the vector e_k consists mainly of zeros, so these "full" calculations become too "expensive".
We turn to the updating function and its gradient:

$$b_{k+1} = Ax_{k+1} = A(x_k + e_k) = b_k + Ae_k;\ O(s||e_k||_0)$$

$$g_{k+1} = A^T Ax_{k+1} = A^T Ax_k + A^T Ae_k = g_k + A^T Ae_k;\ O(s^2||e_k||_0)$$

Obviously, the complexity of these operations is substantially less than one when using the traditional approach.

Implemented Methods

Application of described updating ideology allows us to create effective methods for this class of problems, which have significantly better estimates regarding to the "traditional" ones.

We propose 3 of these methods:

- **NL1** – direct gradient method in the 1-norm;
- **FW** – Frank-Wolf method of conditional gradient;
- **GK** – randomized mirror descent in the Grigoriadias-Khachiyan form.

NL1

Direct gradient method

$$x_{k+1} = x_k + h \cdot y_k$$

$$h = \frac{1}{L}(g_{\max} - g_{\min}) = \frac{1}{3}(g_{\max} - g_{\min})$$

$$y = (0, ..., 0, 1^{\max}, 0, ...0, 1^{\min}, 0, ...0), ||y||_0 = 2$$

$$g_{\max} = \operatorname*{argmax}_{i=1,\ldots,n} \ \partial f(x_k)/\partial x^i$$

$$g_{\min} = \operatorname*{argmin}_{i=1,\ldots,n} \ \partial f(x_k)/\partial x^i$$

Here we got 2 function and gradient computation at one iteration.

FW

Frank-Wolf method of conditional gradient

$$x_{k+1} = (1-\gamma_k)x_k + \gamma_k y_k,\ \ \gamma_k = \frac{2}{k+1},\ \ k = 1, 2, ...$$

$$\langle \nabla f(x_k), y \rangle \to \min_{y \in S_n(1)}$$

$$y_k = (0, ..., 0, 1, 0, ..., 0),$$

Where 1 is on the position:

$$i_k = \operatorname*{argmin}_{i=1,...,n}\ \partial f(x_k)/\partial x^i$$

Here we got 1* function and gradient computation at one iteration.

GK

Saddle statement of problem and randomized mirror descent

$$f(x) = ||Ax||_\infty \to \min_{x \in S_n(1)}$$

This problem can be rewritten in a saddle form:

$$\min_{x \in S_n(1)} \max_{||y||_1 \le 1} \langle Ax, y \rangle.$$

As a result, the problem can be rewritten, preserving the properties of sparseness as:

$$\min_{x \in S_n(1)} \max_{\omega \in S_{2n}(1)} \langle \omega, \widetilde{A}x \rangle.$$

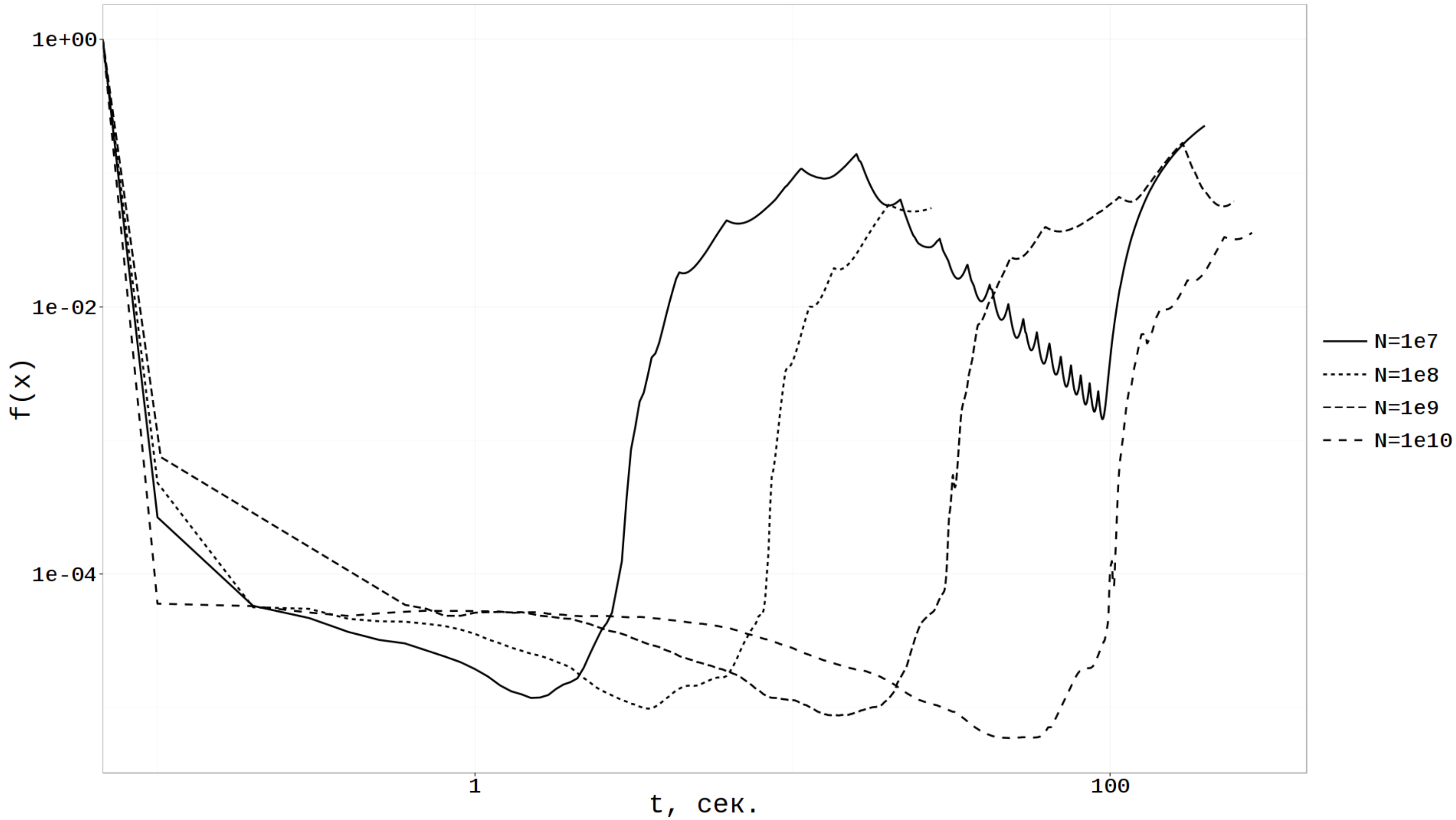
Аникин А.С., Гасников А.В., Горнов А.Ю., Камзолов Д.И., Максимов Ю.В., Нестеров Ю.Е.Эффективные численные методы решения задачи PageRank для дважды разреженных матриц //Труды МФТИ. 2015. Т. 7, № 4, С. 70-91.

Computational experiments

The behavior of these methods was studied on the PageRank problem with matrices of 3 types:

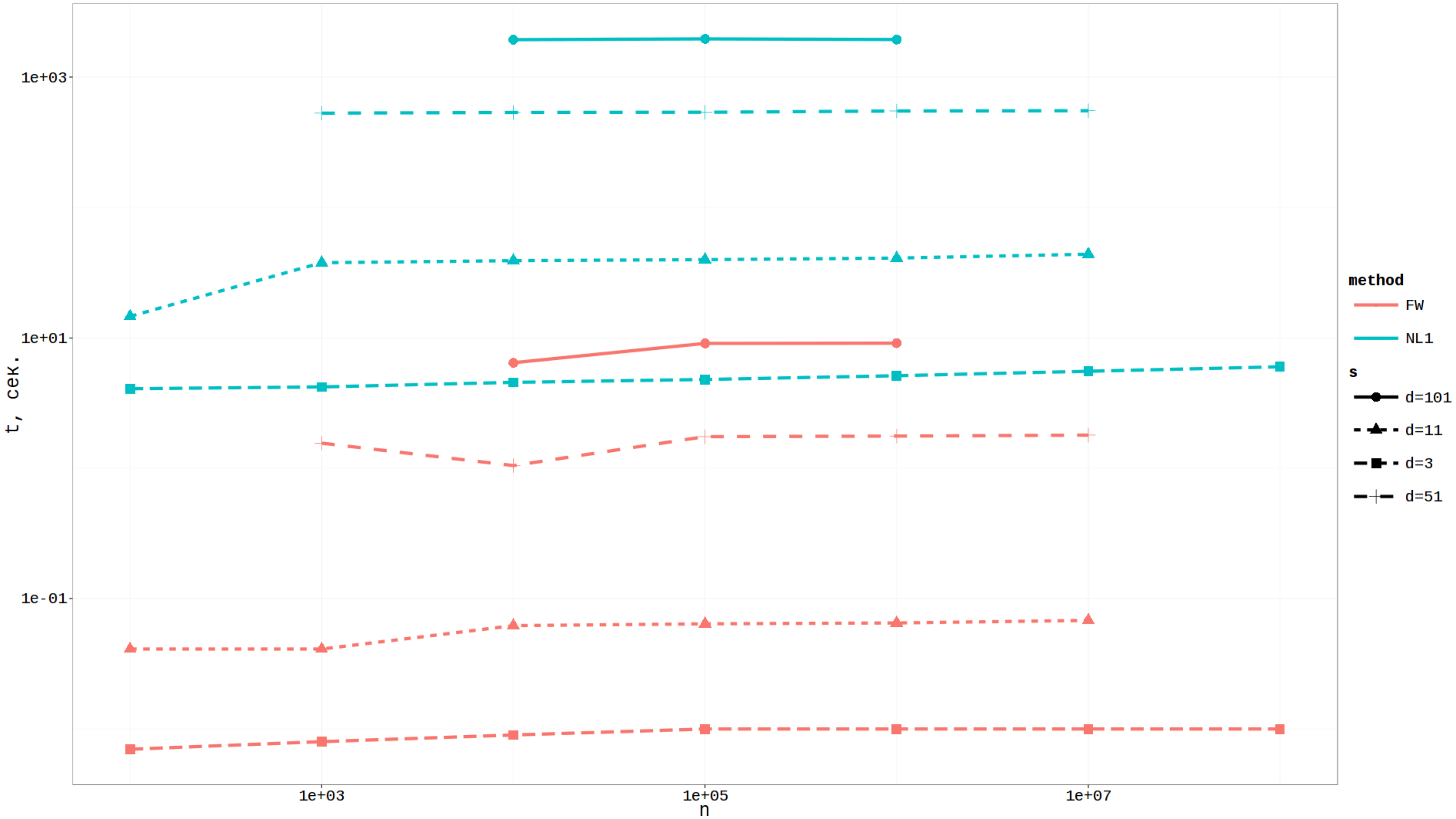
- Diagonal, with given number of the diagonals: $n_d = 1,3,5, \ldots$ Each matrix row / column contains $(n_d - 1)/2 + 1 \leq s \leq n_d$ nonzero elements.
- Randomly generated structure. Each matrix row / column contains exactly s of non-zero elements.
- Stanford University problems. Matrix contain any number of non-zero elements.

GK Method with different *N*



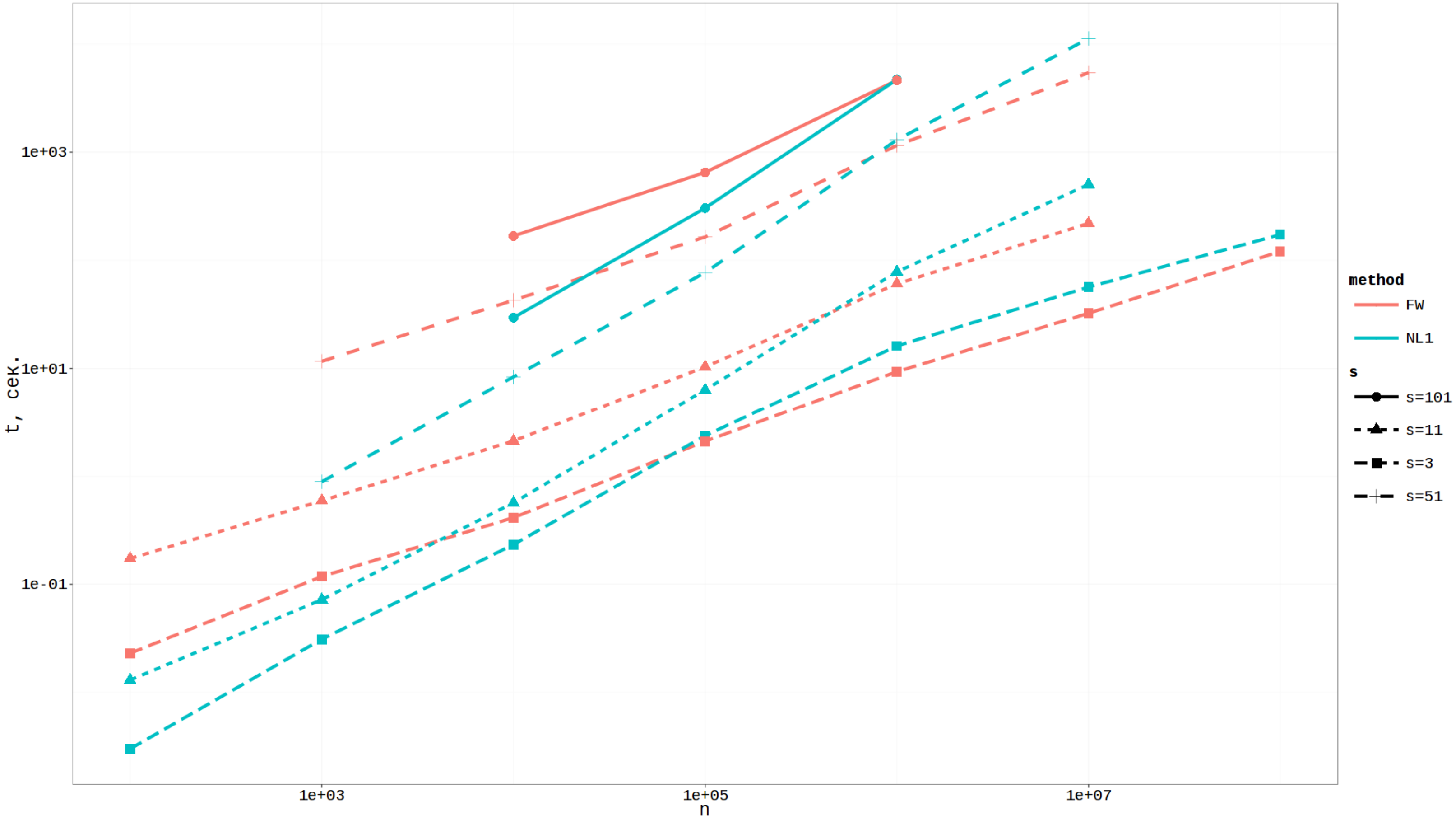
A is random, n=10^2, s=3.

FW vs NL1



A is diagonal.

FW vs NL1



A is random.

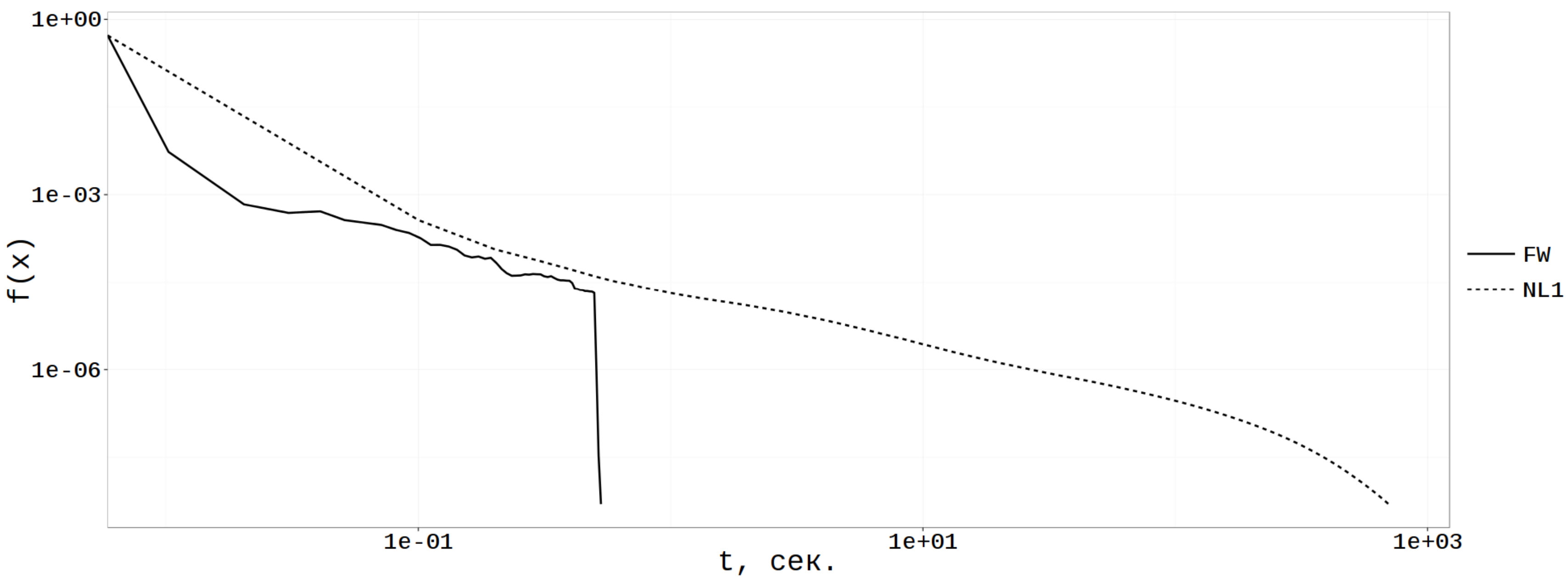
Time (sec.) for solving PageRank problem for web-graphs

		метод NL1		метод FW	
web-граф	n	время	итерации	время	итерации
Stanford	281903	0.145	93152	0.008	14142
NotreDame	325729	700.810	3816436	0.526	38014
BerkStan	685230	38161.847	12315700	0.536	19990
Google	875713	113.643	1083996	0.278	37313

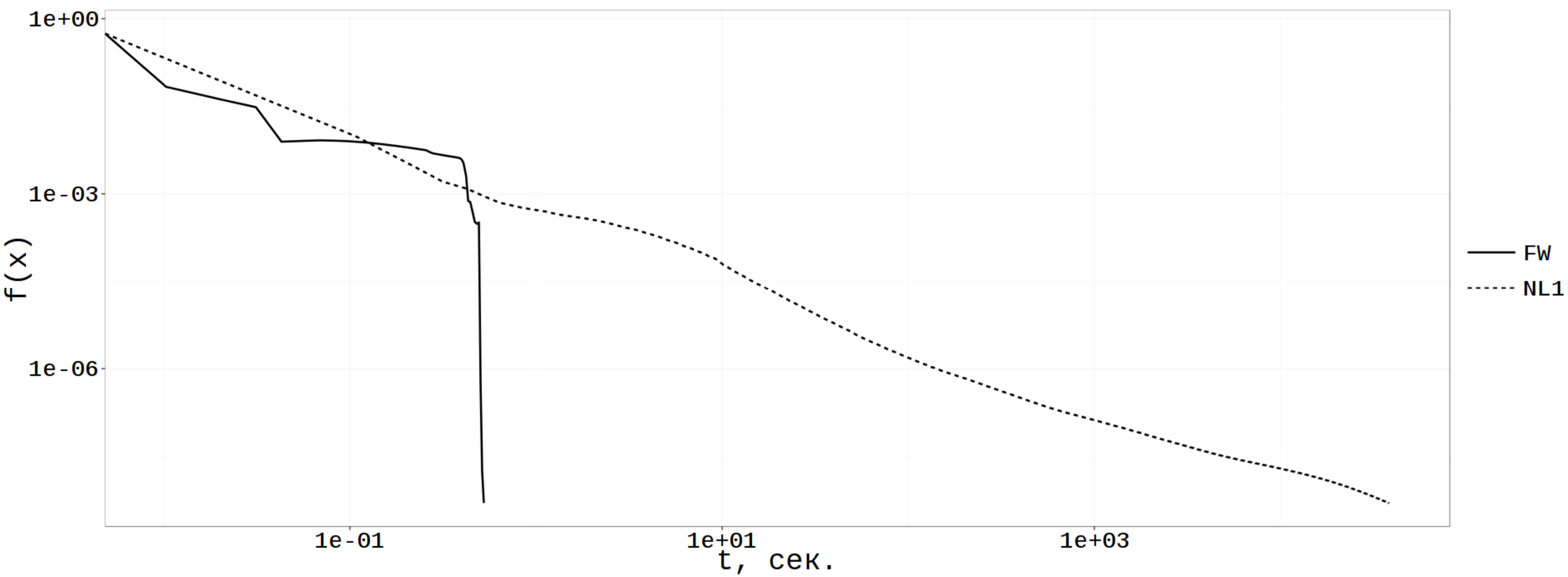
Iteration costs for web-graphs problem

		Stanford		BerkStan	
		NL1	FW	NL1	FW
s_r	мин.	1.0	1.0	1.0	1.0
	макс.	34.0	4.0	84209.0	84209.0
	среднее	3.9	3.9	2278.4	148.6
s_c	мин.	2.0	2.0	1.0	1.0
	макс.	37.0	3.0	244.0	83.0
	среднее	2.9	2.8	15.7	6.2
$s_r \cdot s_c$	мин.	3.0	3.0	2.0	2.0
	макс.	1258.0	12.0	15494456.0	6989347.0
	среднее	11.7	11.3	84304.3	7507.5

Web-NotreDame problem solution



Web-BerkStan problem solution



Conclusions

- The optimization problem of large dimensions **can be** solved (for complex productions - the principle “best-of-known”);

- These problems **should be** studied deeply and actively by a wide range of specialists;

- The correct choice of methods is an important issue, especially for the class of Huge-Scale problems; correct setting of optimization techniques parameters significantly affect their performance and efficiency.

Thanks for your attention!

A.Yu. Gornov,
A.S. Anikin, T.S. Zarodnyuk, E.A. Finkelshtein

Matrosov Institute for System Dynamics and
Control Theory of SB RAS
gornov@icc.ru