

**NEW APPROACH TO THE DETERMINATION OF
INTERMOLECULAR POTENTIAL FUNCTION FROM
THE SECOND VIRIAL COEFFICIENT BY TIKHONOV'S
REGULARIZATION METHOD**

Gulnara M. Kuramshina

Department of Physical
Chemistry
Faculty of Chemistry,
Moscow State University
Moscow 119991, Russia

Outlines:

1. Introduction

2. Intramolecular potential: molecular force field

a) inverse nonlinear ill-posed problem

b) Examples

b) Open problems

3. Intermolecular potential: non-parametric approach

a) Examples

b) Open problems

Macromolecules are everywhere: they form the **material basis of all biological structures and processes**. Life and all biological processes, the occurrences within living cells, within organisms, and their interactions with the environment, are all due essentially to the chemistry of macromolecules. Along with natural macromolecules such **as proteins, DNA, or cellulose** we find in our modern environment a great number of synthetic macromolecules like polyethylene, polystyrene, or Teflon.

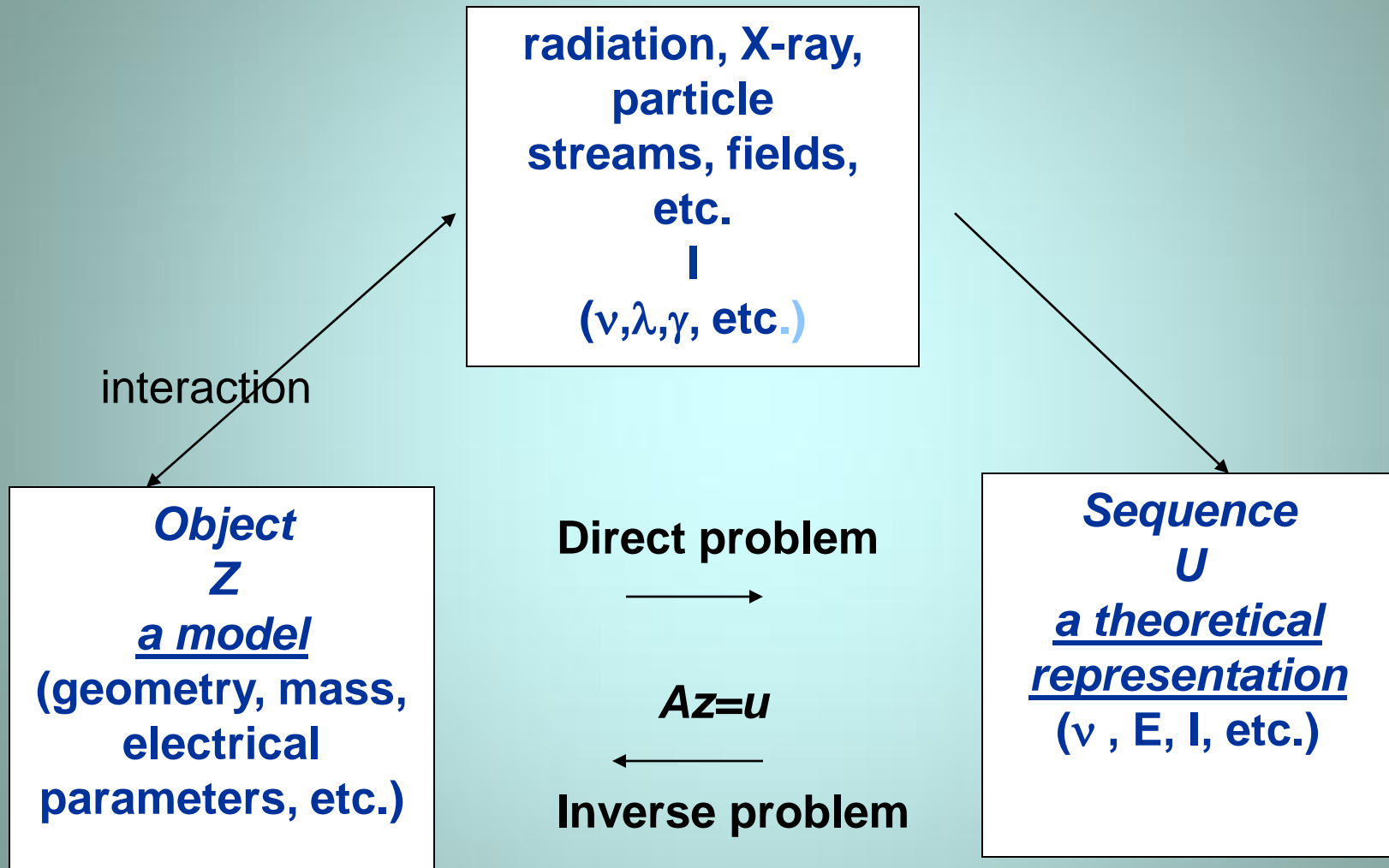
When macromolecules are formed out of many, usually identical small molecular units, called monomers, then we term them ***polymers***.

Macromolecules can also be formed when identical or different molecules group together to give new units under the influence of intermolecular forces, especially Van der Waals forces; one then obtains ***supermolecules, molecular clusters***, and ***inclusion compounds***. Molecules can also form larger units through covalent binding, yielding ***supramolecules***

When their functions, e.g. in biological systems, can be carried out only through the form of these specifically organized macromolecules, they are also called molecular functional units.

In molecular physics, **small molecules** have initially occupied the most prominent place for good reasons. The physicist wishes to determine, understand and calculate the physical properties of the objects investigated in the most precise and complete manner possible.

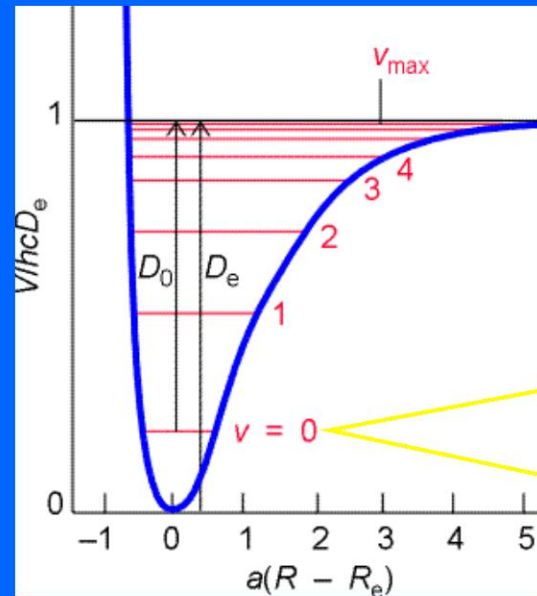
General scheme of structural research by physical methods



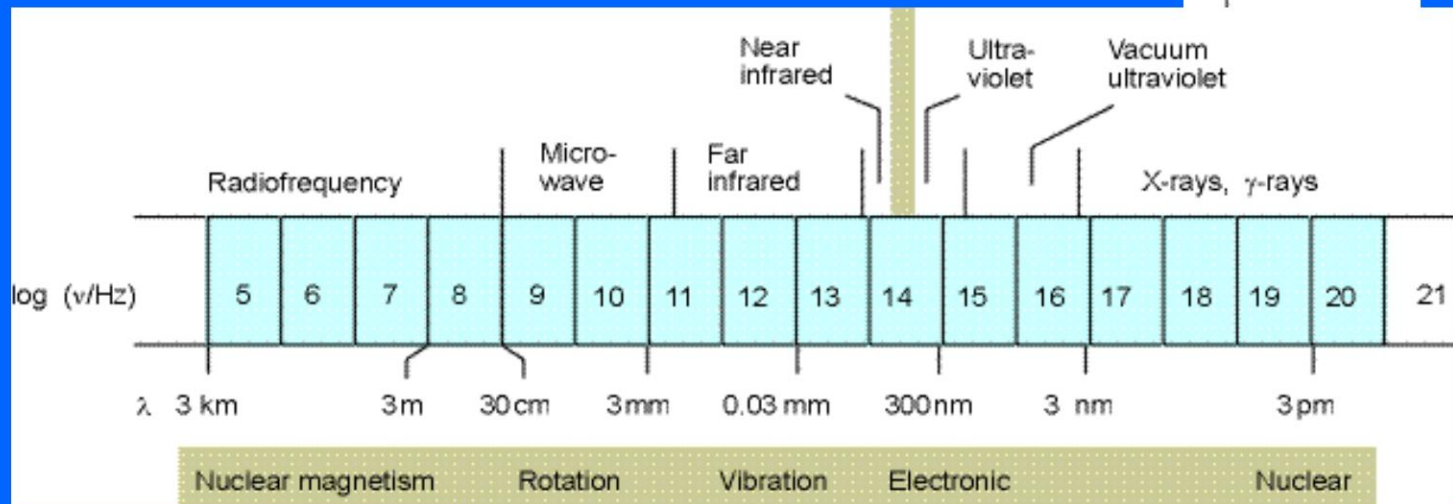
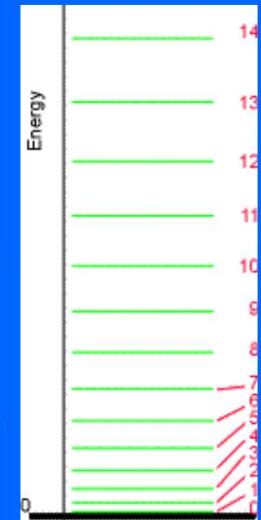
Electronic transitions



Vibrational transitions



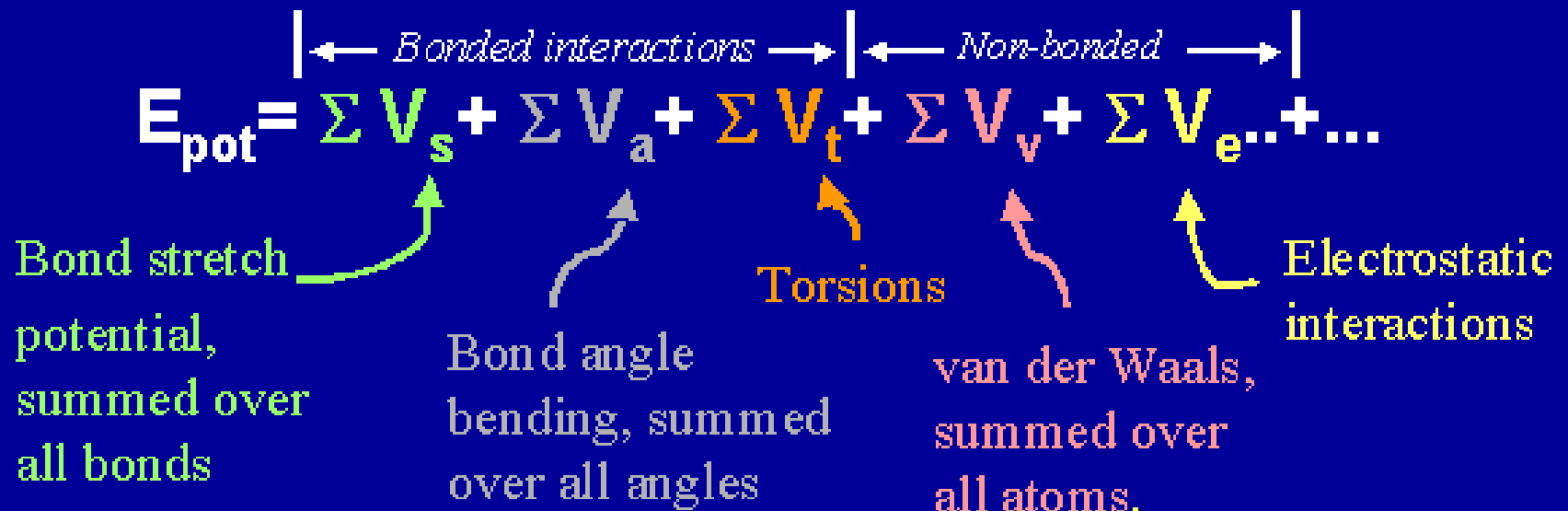
Rotational transitions



QM+MM

Molecular design and molecular modeling for the large (nano or polymer, cluster etc.) molecular structures are based on the **two major methods** used to describe the interactions within a system - **quantum mechanics** and **molecular mechanics**. These results could be followed by the techniques that use the same energy models, including energy minimization, molecular dynamics, Monte Carlo simulations and conformational analysis. Both in molecular mechanics and molecular dynamics the molecular potential energy (and its force field) is represented as the sum of contributions due to bond stretching, bond bending, energy changes accompanying internal rotation about single bonds (all describing the “internal” molecular vibrations, and this part is closely connected with empirical force field calculations), van der Waals attractions and repulsions between nonbonded atoms, and electrostatic interactions due to polar bonds.

•Molecular Mechanics Force Field Or Potential Functions



- ◆ A Linear Combination of Independent Terms
- ◆ Simple Mechanical Functions-- like stretching and bending-- for bonded atoms
- ◆ Attraction and Repulsion Terms-- like van der Waals terms-- for non-bonded atoms.

QM + RQMFF + RIMP

For the calculations of vibrational spectra of the large size molecular systems such as polymers, nanostructures, biological systems, clusters, etc. the next scheme can be proposed:

- 1) preliminary quantum mechanical analysis of moderate size molecules (fragments of large molecular systems) chosen as key or model molecules (**QM**) ;
- 2) joint treatment of *ab initio* and experimental data on vibrational spectra, structural electron diffraction (ED) and microwave (MW) data for model molecules within the regularization theory, determination the equilibrium geometry parameters and harmonic force constants (**RQMFF**);
- 3) determination of intermolecular potential parameters by means of stable numerical methods; **RIMP**
- 4) organization of a database on structural data, force field parameters and intermolecular potentials transferable in a series of related compounds;
- 5) synthesis (construction) of a large molecular system from separate fragments included in the database and calculation of its vibrational spectra and thermodynamical functions.

Vibrational problems

- Normal coordinate analysis (G. Herzberg, E.B. Wilson, M.A.Elyashevich) principal equation

$$\blacksquare \mathbf{GFL} = \mathbf{L}\Lambda \quad (1)$$

- presents :
- (a) **direct problem** – determination of normal vibration frequencies (matrix Λ) and forms (amplitude vectors \mathbf{L}) with known matrix of kinematic coefficients \mathbf{G} and matrix of force constants \mathbf{F} ,
- U is a potential energy, $\{q_{ij}\}$ is a set of generalized coordinates.
$$\mathbf{F} = \{F_{ij}\} = \left\{ \frac{\partial^2 U}{\partial^2 q_{ij}} \right\} \Big|_{q_0} \cdot$$
- (b) **inverse problem** – force field determination, which is obviously ill-posed.

Well-posed and Ill-posed mathematical problems (J.Hadamard)

- The problem

- $Az = u$ (1)

- where z and u are some elements of normed spaces Z and U (u is the the set of experiwental data, z is the set of properties of the object to be determined); A is an operator establishing the correspondence between z and u , is called well-posed if it satisfies the following three conditions:
 1. it is solvable for any u from U
 2. the solution of (1) is unique
 3. the solution is stable with respect to small perturbations of u , i.e. results in only small deviations of z .
- In this case one can write $z = A^{-1}u$ (instead of $z = Ru$, i.e. and inverse operator exists ($R = A^{-1}$)). Otherwise the problem is ill-posed. The majority of physical problems belong in mathematical sense to ill-posed problems.

Variational approach for constructing regularizing algorithms

- Consider nonlinear ill-posed problems in finite-dimensional case and use Tikhonov's approach for constructing regularizing algorithms (proposed at first by A.N.Tikhonov and developed later in publications by A.V.Goncharsky, A.S.Leonov, and A.G.Yagola). Consider an operator equation

- $$Az = u, z \in D \subseteq Z, u \in U \quad (1)$$

- where D is a nonempty set of constraints, Z and U are finite-dimensional normed spaces, A is a class of operators from D into U .

A general formulation of Tikhonov's scheme of constructing a regularizing algorithm for solving the main problem:

for the operator Eq. (1) on D find an element z^* for which

$$\|Az^* - u\| = \min \{ \|Az - u\| : z \in D \} \equiv \mu \quad (2)$$

We call this problem as the quasisolution problem for Eq. (1). In the case $D=Z$ problem (2) gives pseudosolution of (1). If the measure of incompatibility μ is equal to zero, then the solutions of (2) are the solutions of Eq.(1) on D . The quasisolution problem (2) may be ill-posed. Namely, problem (2) is not solvable for some equations of the form (1). The solution of (2) may be nonunique and unstable in the metrics of Z with respect to perturbations of the data (A, u) .

INVERSE VIBRATIONAL PROBLEM

- **Problem 1**

- It is required to obtain

$$\bar{F}_n = \arg \min \|F - F^0\|,$$

$$F \in \left\{ F : F \in D, \|AF - \Lambda\| = \mu \right\}$$

- where

$$\mu = \inf \|AF - \Lambda\|, F \in D$$

- F is unknown force constant matrix, A is nonlinear operator that corresponds the symmetric F with the set Λ of all experimental information available: vibrational frequencies of the molecule (and its isotopomeres, if available), the Coriolis constants, the mean square amplitudes, etc. This set of data may be presented as a vector in finite-dimensional space. Consider matrix F as a vector of finite-dimensional space \mathbf{Z} , consisting either of the elements of the matrix F or the quantities by means of which this matrix can be parametrized.

$$M^\alpha[F] = \|A_h F - \Lambda_\delta\|^2 + \alpha \|F - F^0\|^2$$

$\|F - F^0\|^2$ is a stabilizer.

$\|\Lambda - \Lambda_\delta\| \leq \delta$, δ is an experimental accuracy. $h \geq 0$ is an operator error. α - parameter of regularization (generalized discrepancy principle)

In the Tikhonov regularizing procedure, one can increase the stability and accuracy of the calculated solution F_α by using

- an extended set of experimental data (**H,D,T**)
- an improved choice of the stabilizer matrix F^0 (**QM**)
- an improved choice of the constraints set D (**QM + model**)

*D is a given set of a priori constraints (supposed to be closed).
A set of a priori constraints may arise from several types of
limitations on force constant values:*

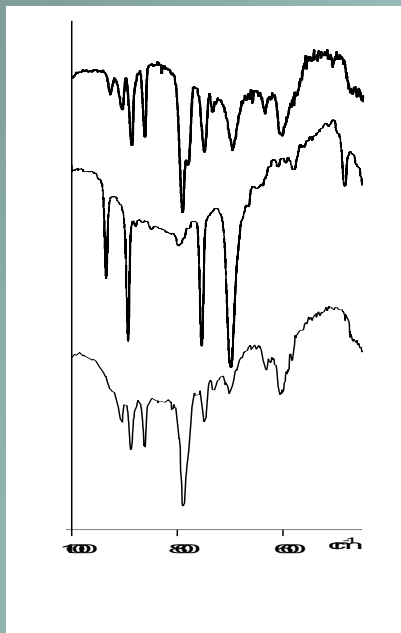
- 1. *some force constants* may be stipulated on a priori grounds *to be a zero*;
- 2. some force constants may be stipulated to satisfy inequalities $a_{ij} \leq f_{ij} \leq b_{ij}$, where a_{ij} , b_{ij} -are certain known values;
- 3. some force constants may be stipulated to conform to be *equal in a series of related molecules* (or conformers);
- 4. the final solution may be stipulated to conform to Pulay's scaled force constant matrix, which may also be considered as a kind of constraint. In this case, set D is specified as:
- $D = \{F : F = BF^0B\}$, $B = \text{diag}\{\beta_1, \dots, \beta_n\}$ (β_i are the scaling parameters).

REGULARIZING PROCEDURES FOR SOLVING THE GENERAL INVERSE PROBLEM OF STRUCTURAL CHEMISTRY

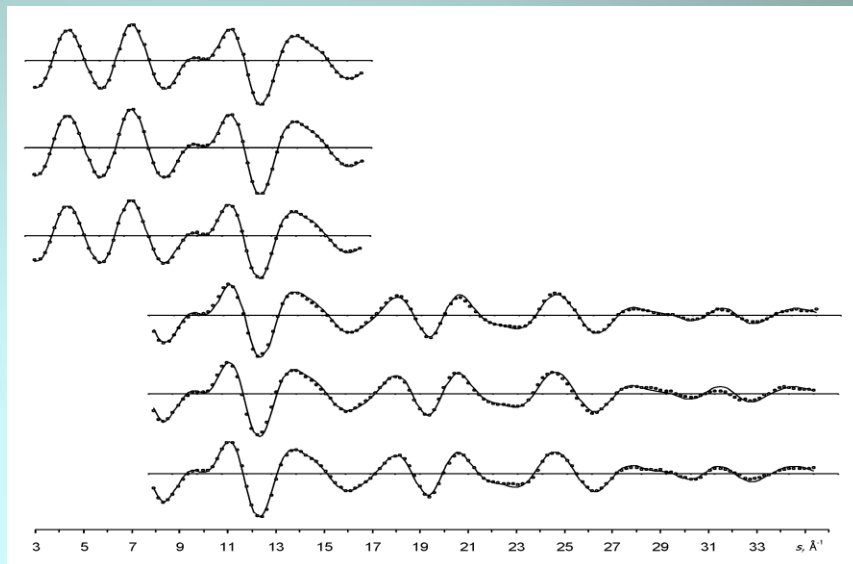
The model is used to predict experimentally measured values, such as vibrational frequencies ω , electron diffraction intensity $M(s)$, rotational constants $\{A, B, C\}$ obtained from microwave molecular spectra, etc. All of these values are functions of geometric (R) and force field (F) parameters. With experimental data and parameters represented as elements of normalized finite-dimensional spaces, we can formulate the problem of simultaneous refinement of the force field and equilibrium geometry of the molecule as a system of non-linear equations

$$\begin{cases} \omega(F, R) = \omega_{\text{exp}}, \\ M(s, F, R) = M_{\text{exp}}(s), \\ \{A, B, C\}(R, F) = \{A, B, C\}_{\text{exp}} \end{cases}$$

on a set of predefined constraints $F \in D_F, R \in D_R$. This system can be extended to include additional experimental evidence when available (for example, data for isotopic species of a molecule sharing the same force field and equilibrium geometry).



Infrared
spectrum



Electron diffraction
intensity curves, $sM(s)$

Experimental data

Due to experimental errors, lack of experimental data and model limitations, this system of equations (that can be also treated as a finite-dimensional non-linear operator equation) usually fails to define unique solution, often proves to be incompatible and does not provide stability with respect to the errors of input data. To avoid these unfavorable features characteristic to the ill-posed problems, it is necessary to implement a regularizing algorithm for its solution.

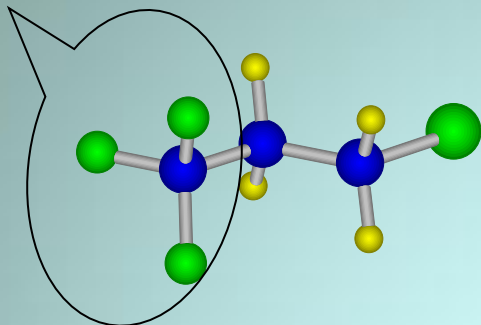
We suggest to use a regularizing algorithm based on optimization of the Tikhonov's functional

$$M^\alpha(F,R) = \|\omega(F,R) - \omega_{\text{exp}}\|^2 + \|M(s,F,R) - M_{\text{exp}}(s)\|^2 + \|\{A,B,C\}(F,R) - \{A,B,C\}_{\text{exp}}\|^2 + \alpha \left\{ \|F - F^0\|^2 + \|R - R^0\|^2 \right\}$$

where in the last ("stabilizer") term F^0 and R^0 represent parameters of *ab initio* force field and equilibrium geometry, respectively. With the appropriate choice of regularization parameter α (that depends on the experimental errors characterized by some numerical parameter δ), it proves possible to obtain approximations converging to a normal pseudosolution of the system (1) when experimental errors tend to zero. These approximations are obtained as extrema $\{F_{\alpha}, R_{\alpha}\}$ of functional (2).

Synthesis of MFF: transferable force fields

$F(\text{CF}_3^-)$

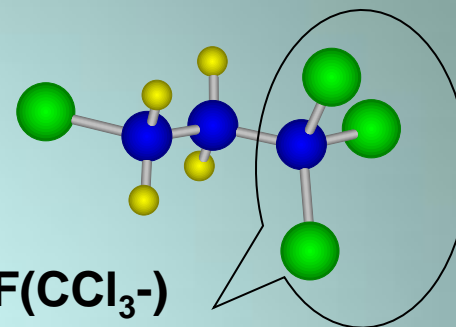


Trans

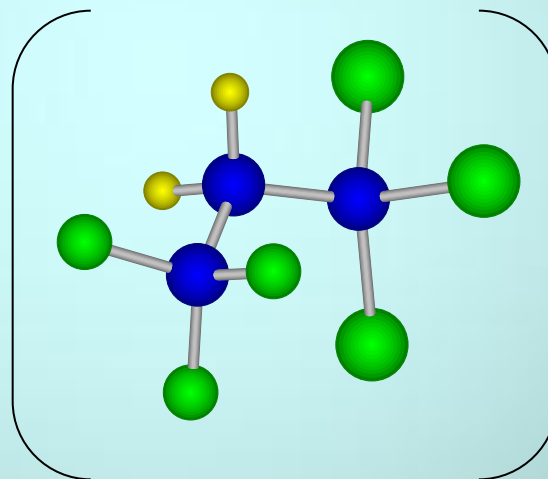


and

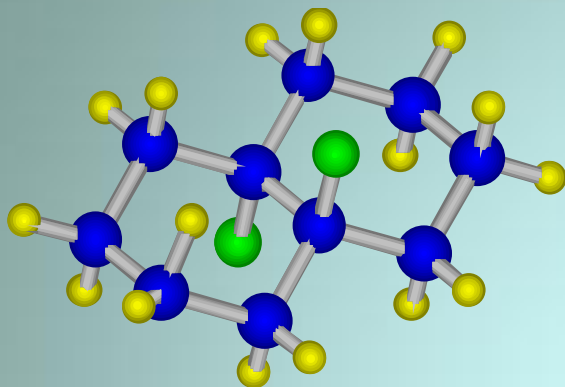
$F(\text{CCl}_3^-)$



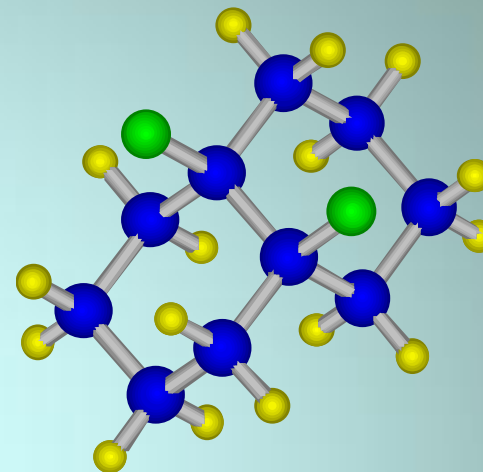
F



Perfluorobicyclo[4,4,0]-decane (perfluorodecaline)



trans
 C_{2h}



cis
 C_1

Constraints (18 atoms molecules):

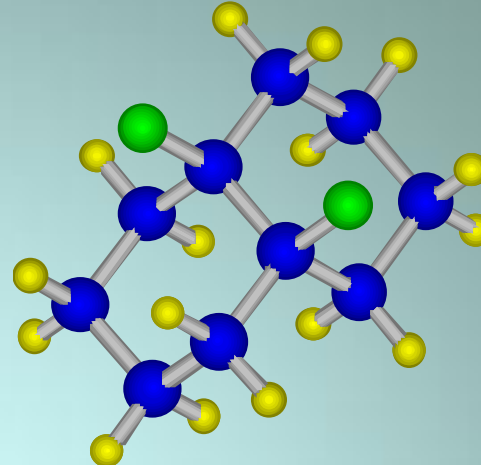
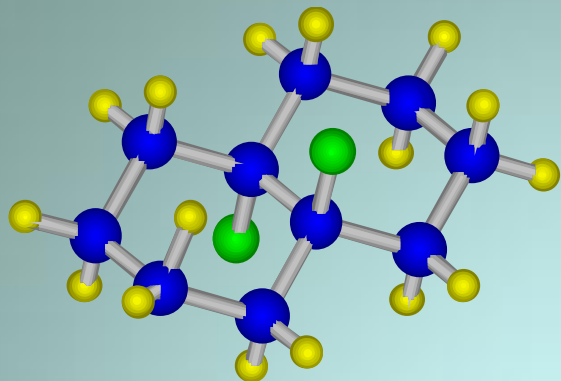
$F_{ij}=0:$

1535(trans)

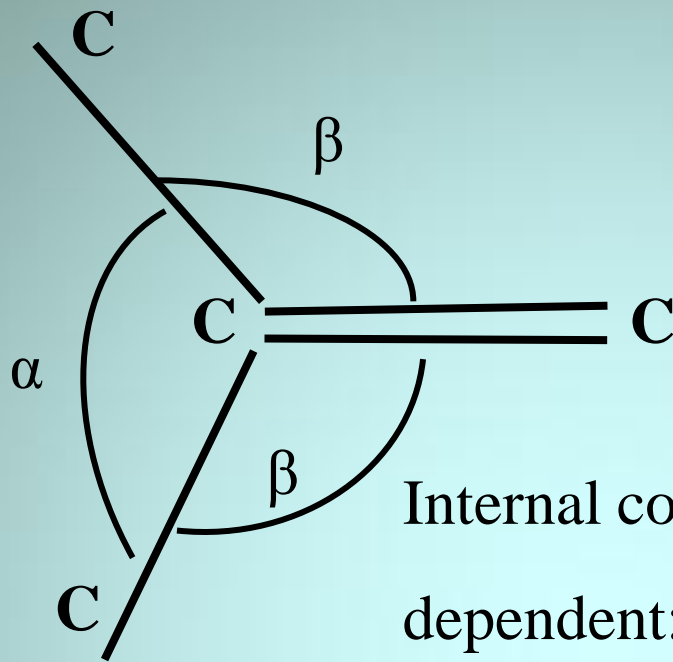
3029(cis)

$F_{ij}^t = F_{kl}^c$: 1040

Joint calculation of force fields: $N(\text{constr}) = 5604$; R^{4095}



<i>trans</i>		<i>cis</i>	
observed	calcul.	observed	calcul.
1092	1096	1076	1079
1085	1088	1068	1069
1052	1054	1062	1061
1024	1027	1040	1038
1020	1019	1011	1009
986	988	986	983
982	980	947	949
944	947	862	863
785	782	819	820
762	764	776	778
688	690	677	677



$$R(\text{C}-\text{C})=1.54 \text{ \AA}$$

$$R(\text{C}=\text{C})=1.39 \text{ \AA}$$

$$\alpha + \beta + \beta = 2\pi$$

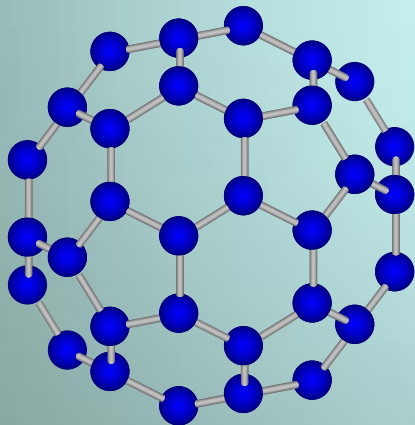
Internal coordinates:

dependent: $n=7$; $R(\text{C}=\text{C})$, $2R(\text{C}-\text{C})$, α , 2β , χ

independent ($3N-6$): $n=6$ (totally, 9 sets)

$$R(\text{C}=\text{C}), 2R(\text{C}-\text{C}), 2\beta, \chi$$

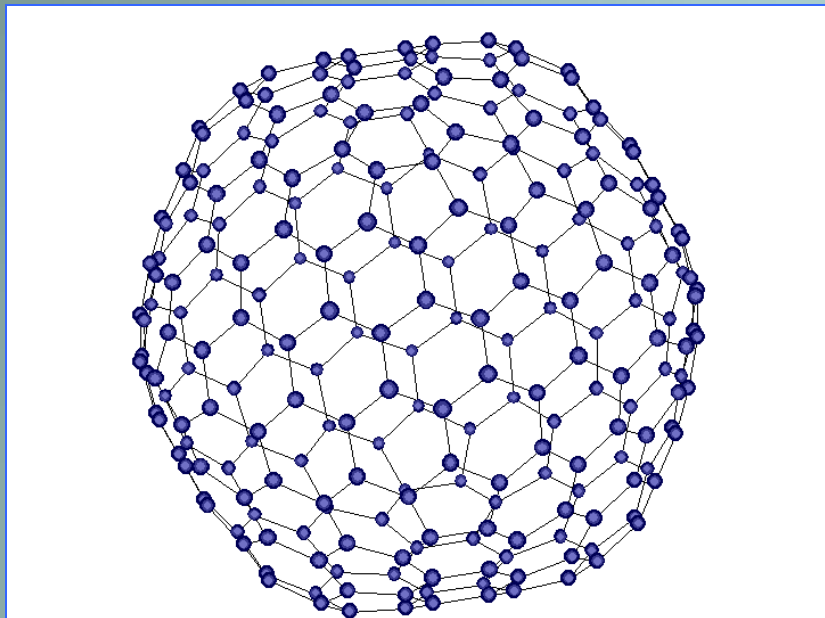
$$R(\text{C}=\text{C}), 2R(\text{C}-\text{C}), \alpha, \beta, \chi$$



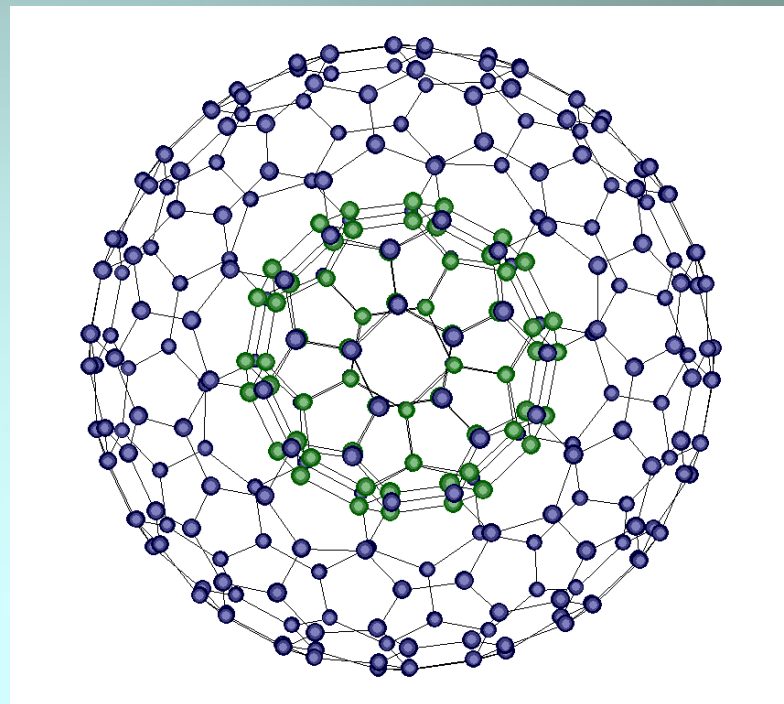
$$\text{C60: } n(\text{vibs})=60 \times 3 - 6 = 172$$

$$n_f = 173 \times 86 = 14878$$

$$\text{Redundant: } 270 \text{ coords, } n_f = 36585$$



**FULLERENE
C240**

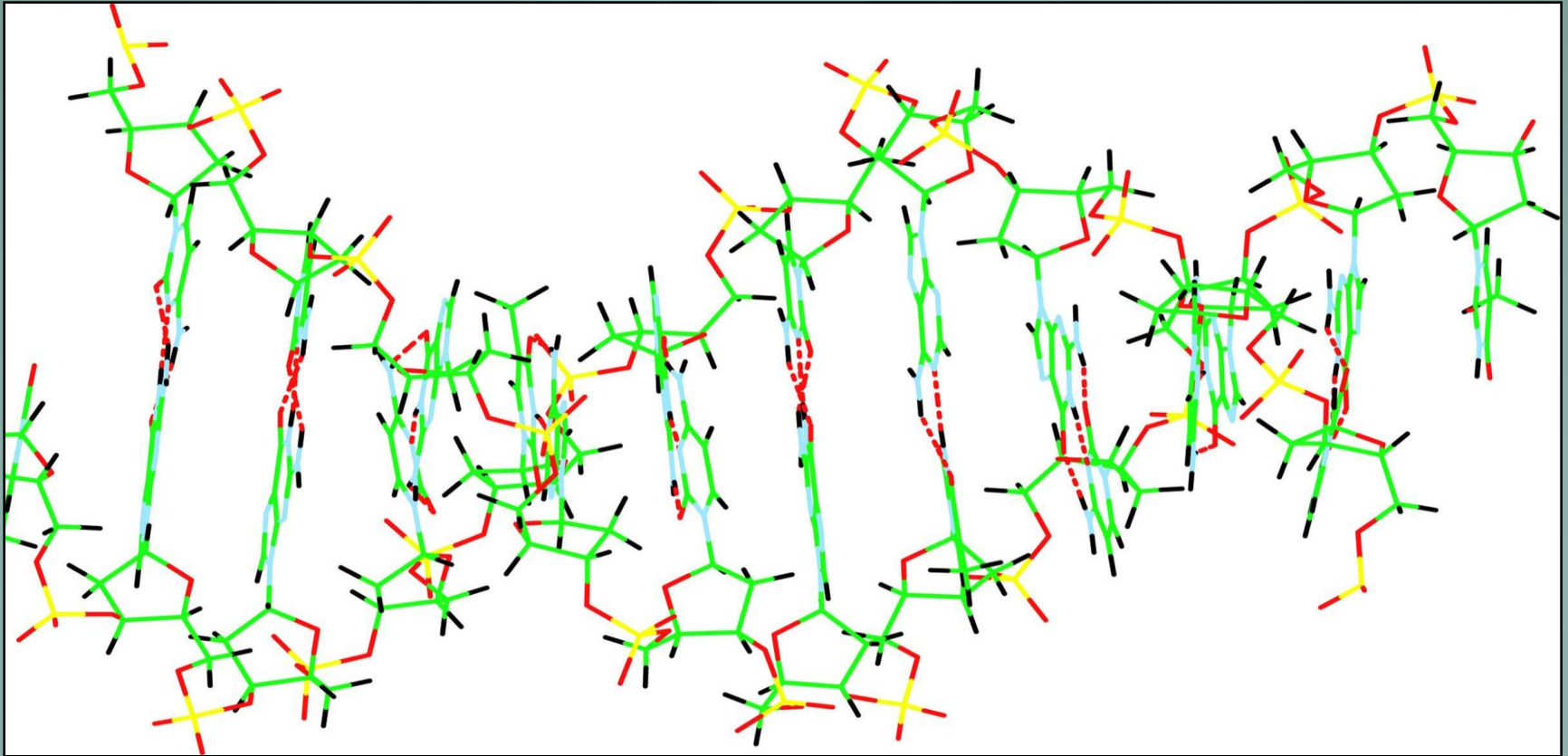


c60@c240

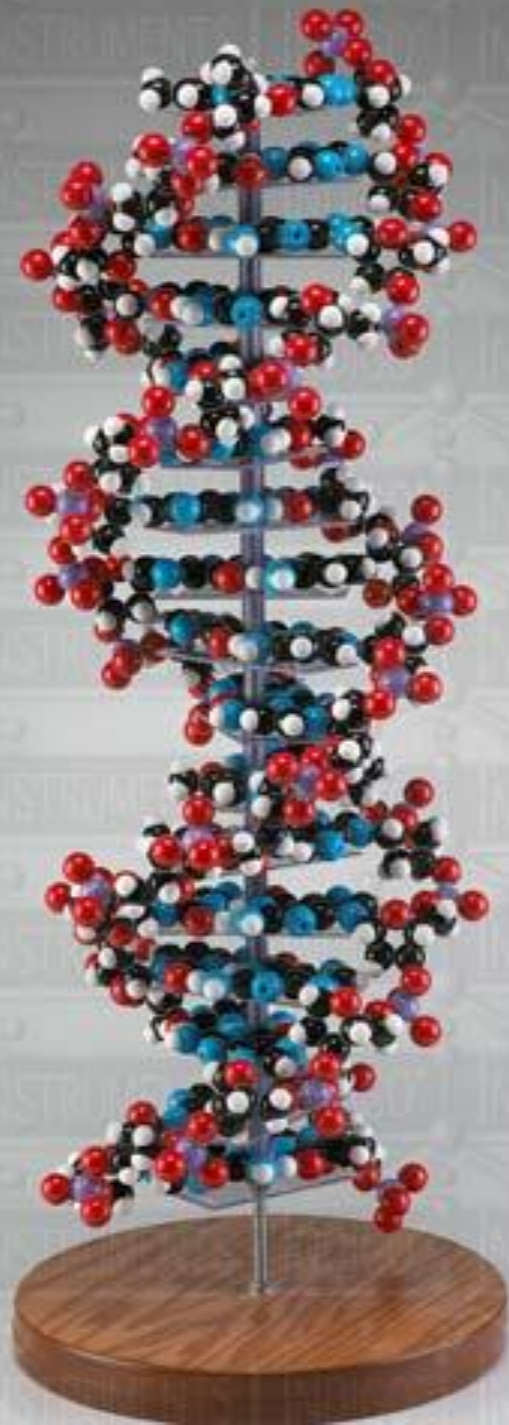
I.V.Kochikov,G.M.Kuramshina,A.V.Stepanova

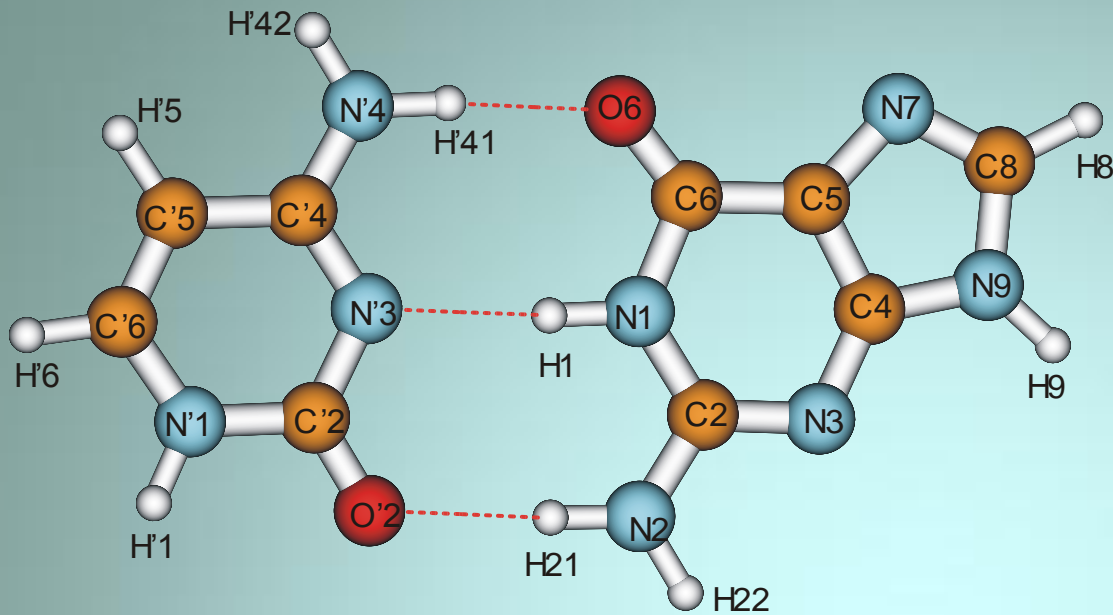
New approach for the correction of Ab initio molecular force fields in Cartesian coordinates.

International Journal of Quantum Chemistry, Volume 109, Issue 1,
Pages 28-33, 2009

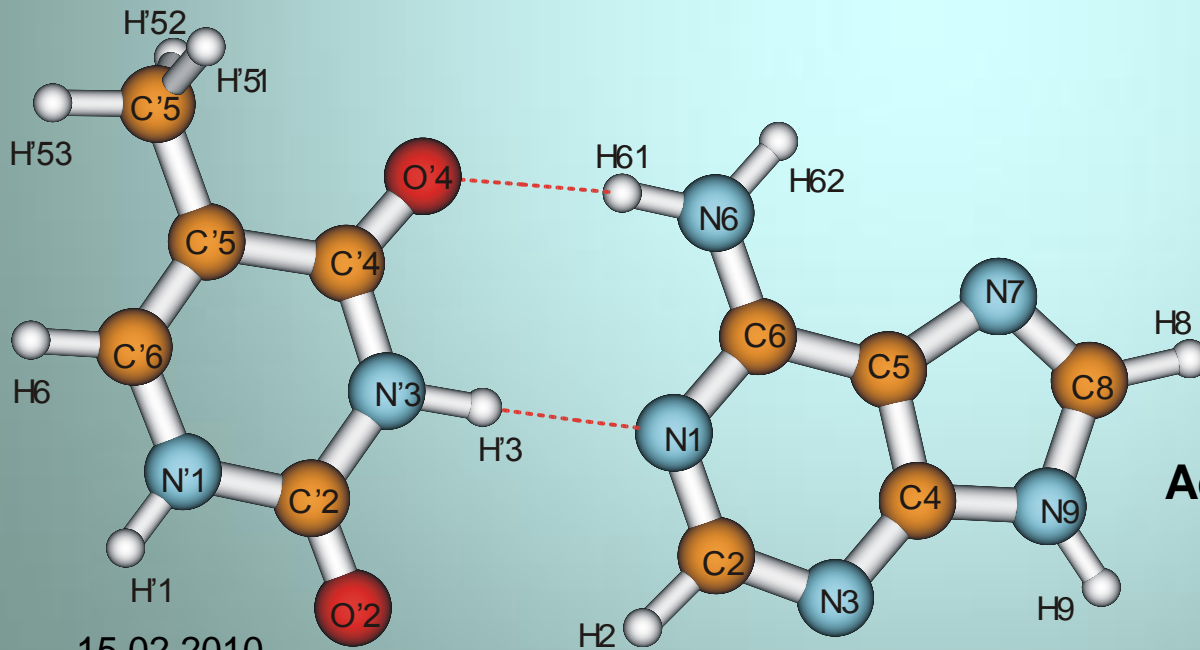


INTERMOLECULAR INTERACTIONS

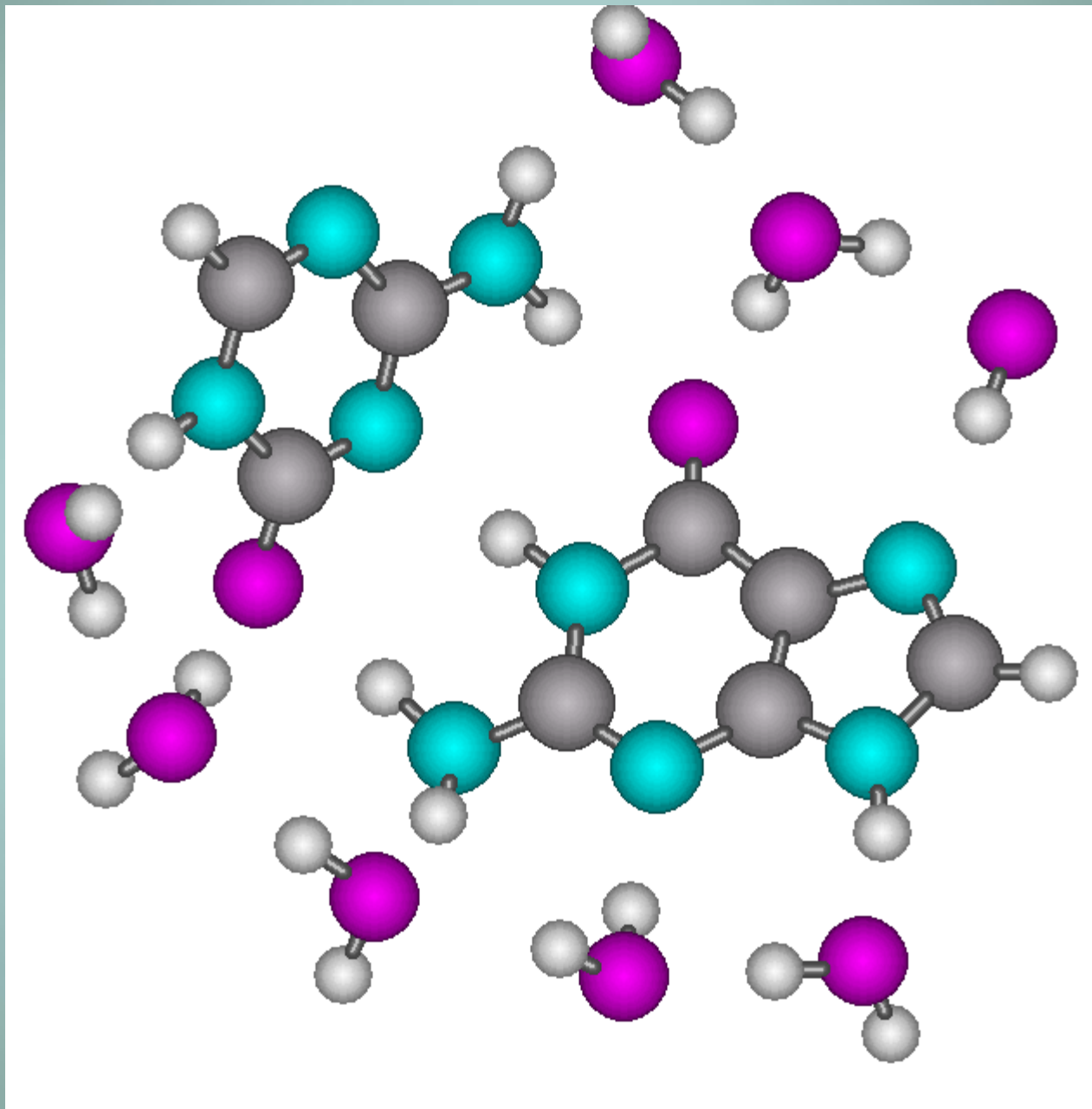




Guanine-Cytosine Base Pair



Adenine-Thymine Base Pair



15.02.2010

Guanine-Cytosine Base Pair + 8H₂O

27

Compressed

Relaxed

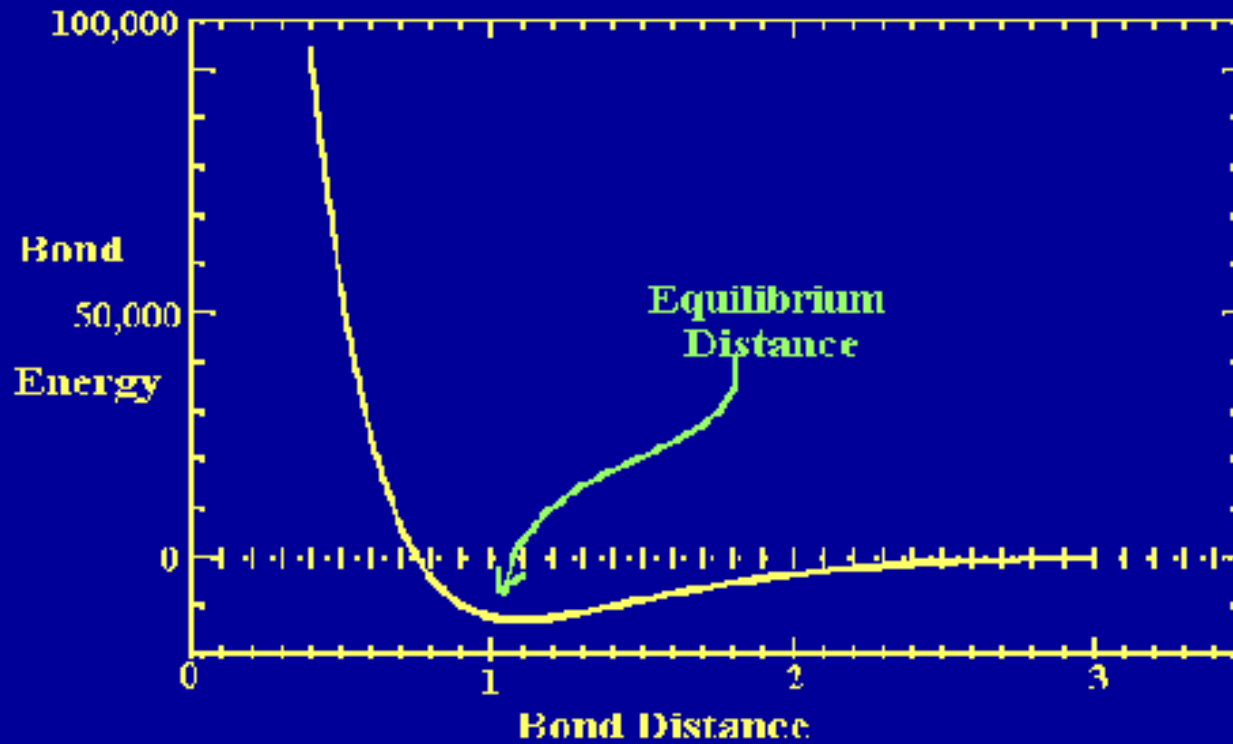
Stretched



Repulsion

Equilibrium

Attraction



INTERMOLECULAR POTENTIAL

For a given spherical potential energy function $U(R)$ the virial expansion up to the second power of density is written as

$$\frac{p}{kT} = \rho + B(T)\rho^2$$

where $B(T)$ is the second virial coefficient. It can be obtained using interaction potential function as

$$B(T) = -\frac{2\pi}{3kT} \int_0^{\infty} R^3 \frac{\partial U}{\partial R} \exp\left[-\frac{U(R)}{kT}\right] dR \quad (1)$$

or, in equivalent form,

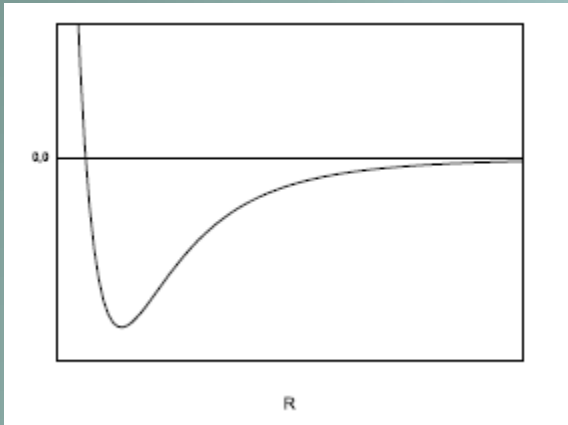
$$B(T) = 2\pi \int_0^{\infty} R^2 \left[1 - \exp\left(-\frac{U(R)}{kT}\right) \right] dR \quad (2)$$

We have a second virial coefficient measured at different temperatures, and calculation of a potential energy function is solving Fredholm integral equation of the first kind, which is an example of typical ill-posed problem

Equation (2) may be directly treated as nonlinear integral equation. If functions $B(T)$ and $U(R)$ belong to $L_2(0,\infty)$, this problem is apparently ill-posed and requires certain regularization technique.

Previous investigations were based on fundamental properties of the potential function that was assumed consisting of repulsive term (monotonically decreasing function for small values of distance R) and attractive term (negative, monotonically increasing to zero function for large R). Well-known representatives of such functions are Lennard-Jones or Morse potentials.

Using restricted set of potential functions may, however, result in the well-posed problem for determining $U(R)$ from Eq. (2). Let us introduce the following set of constraints D :



1) $U(R)$ is convex downwards for $R < R_p$, and convex upwards for $R > R_p$; (R_p - inflection point)

obviously this form of potential generalizes all known empirical potential functions;

2) For $R > R_p$ (actually starting with even smaller distances), $U(R)$ is negative, and monotonically approaches zero as $R \rightarrow \infty$.

Here R_p is some unknown distance. It can be shown that mentioned constraints are sufficient to define a compact set in L_2 space; this guarantees convergence of approximate solutions of Eq. (2) to its exact solution of the problem when data errors in $B(T)$ tend to zero.

Solving Eq. (2) may be implemented as finding minimum of the discrepancy functional

$$\Phi[U] = \int_0^{\infty} [B(T) - \tilde{B}(T)]^2 dT \quad (3)$$

where $B(T)$ is calculated according to Eq. (2), and $\tilde{B}(T)$ are experimentally measured values. Minimization is subject to constraints D .

Technically, potential function $U(R)$ was represented on a discrete grid, and integration in (3) was restricted to the finite interval by introducing R_{\min} , R_{\max} such that input of the omitted intervals was negligible. Discrepancy (3) was minimized for each value of the parameter R_p from a certain interval, and parameter yielding minimum discrepancy was chosen as the solution.

An example: CH₄ (Exp.: Esper et.al. 1995, Fluid Phase Equilibria, 105, 173)

$$\Psi(U, T) = B(U, T) = 2\pi \int_{r_{\min}}^{r_{\max}} R^2 \left[1 - \exp\left(-\frac{U(R)}{kT}\right) \right] dR$$

Use grid: $[x_1, \dots, x_n]$ on interval (r_{\min}, r_{\max}) . Inflection point x_{nn}

$$\Phi[u] = \sum_{j=1}^m [\Psi(u, T_j) - B(T_j)]^2.$$

Limitations $D(r_p)$:

Convexity download

$$u_{i+1}(x_i - x_{i-1}) - u_i(x_{i+1} - x_{i-1}) + u_{i+1}(x_{i+1} - x_i) \geq 0, \quad i = 2, \dots, nn - 1;$$

Convexity upload

$$u_{i+1}(x_i - x_{i-1}) - u_i(x_{i+1} - x_{i-1}) + u_{i+1}(x_{i+1} - x_i) \leq 0, \quad i = nn + 1, nn + 2, \dots, n - 1;$$

Monotonicity

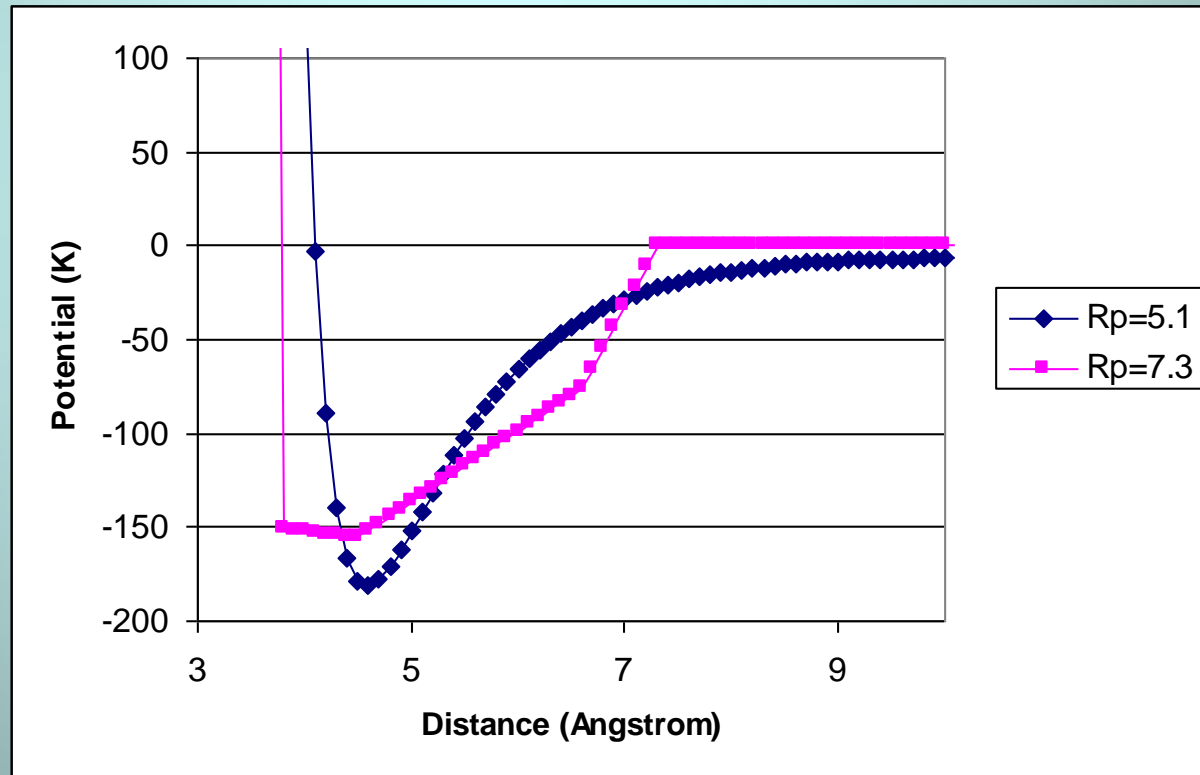
$$u_{i+1} - u_i \geq 0, \quad i = nn - 1, nn, \dots, n - 1;$$

$$u_i \leq 0, \quad i = nn, nn + 1, \dots, n.$$

$$\min_{r_p} \min_{u \in D(r_p)} \Phi[u],$$

Fig.1 represents solutions of the inverse problem obtained for two cases:

- 1) When R_ρ parameter was chosen close to its exact value (5.1 Å);
- 2) When R_ρ parameter was seriously in error (7.3 Å).

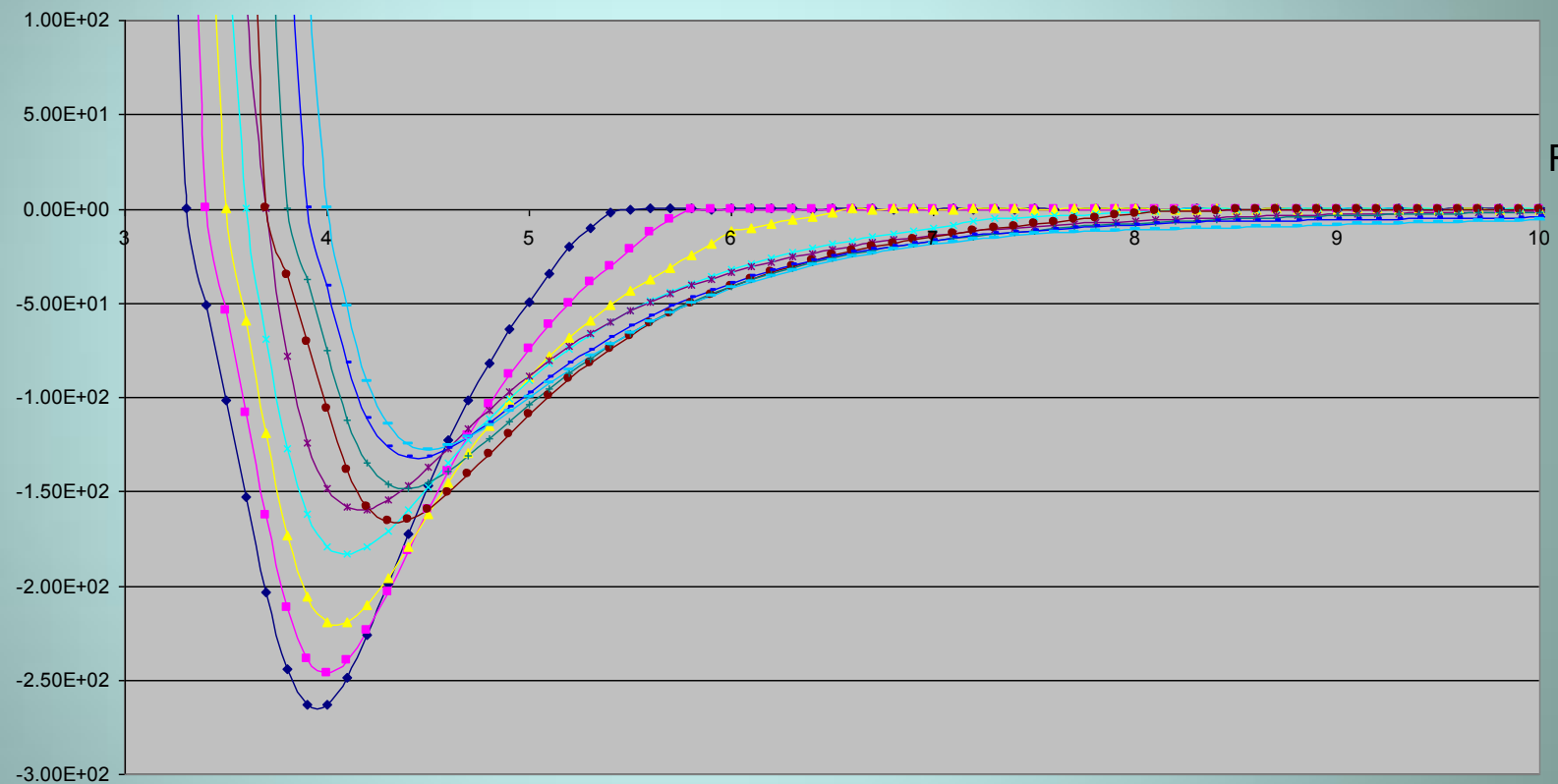


Note that the first solution (smooth curve with inflection point at $R = 5.1 \text{ \AA}$) practically coincides with the exact solution, because no additional noise was introduced into input data. However, the second solution (piecewise linear curve) obtained under assumption that inflection point is located at $R = 7.3 \text{ \AA}$, yields practically the same discrepancy value (within 0.1 per cent of the relative error in the values of $B(T)$). The exact model data and simulated virial coefficients corresponding to both potentials (Fig. 1) are compared in the following Table 1.

$T(K)$	$B(\text{model})$	$B(5.1)$	$B(7.3)$
232.095	-77.61	-77.55	-77.62
250.000	-67.73	-67.76	-67.73
270.42	-58.41	-58.47	-58.40
290.016	-50.96	-51.01	-50.95
309.718	-44.60	-44.62	-44.59
331.357	-38.64	-38.62	-38.65
350.366	-34.13	-34.05	-34.13

V,

Methane potentials for different inflection points



Determination of a methane intermolecular potential model for use in molecular simulations from *ab initio* calculations

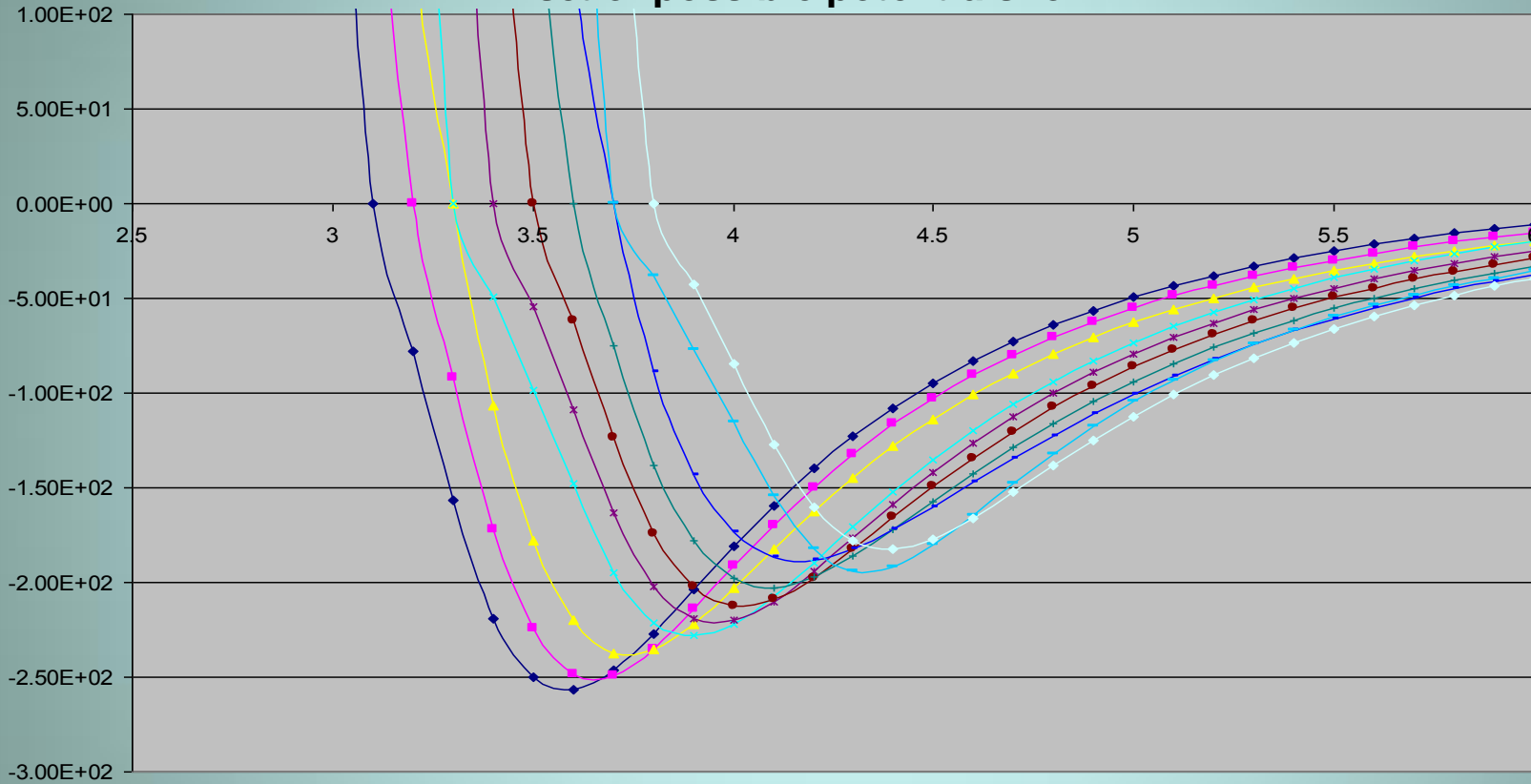
J. Chem. Phys. 110, 3368 (1999) [Richard L. Rowley](#), [Tapani Pakkanen](#)

Interaction energies of van der Waals and hydrogen bonded systems calculated using density functional theory: Assessing the PW91 model

J. Chem. Phys. 114, 3949 (2001); [Seiji Tsuzuki](#), [Hans P. Lüthi](#)

The *intermolecular* interaction *potential* for the *methane* dimer has its minimum at an *intermolecular distance* of 4.0 Å

A set of possible potentials for Kr



Acknowledgements

Prof. Anatoly Yagola, MSU, Faculty of Physics

Dr. Igor Kochikov, MSU, Scientific Research Computer Center

M.S. Nadezhda Anikeeva, MSU, Faculty of Physics

\$\$: RFBR grant No 07-03-00145a

CONCLUSIONS

A scheme of modeling vibrational spectra of the large size molecular systems based on stable numerical methods is considered. A new regularizing algorithm for solving inverse problem of intermolecular potential restoration from experimental thermodynamic data on second virial coefficient dependent on temperature, using a priori constraints on convexity and monotonicity of intermolecular potential, is presented.

Obvious advantage of new approach is a generalization of all known empirical finite parameter potentials. Mathematical formulation of the problem is reduced to minimization of a non-quadratic functional on a set of linear constraints defining a concave-convex potential function with the fixed inflection point. Position of inflection point is determined using enumerative technique; functional is minimized by the projection of conjugate gradients method.