ISSN 1070-3632, Russian Journal of General Chemistry, 2008, Vol. 78, No. 4, pp. 685–695. © Pleiades Publishing, Ltd., 2008.
Original Russian Text © L.A. Gribov, I.V. Mikhailov, 2007, published in Rossiiskii Khimicheskii Zhurnal, 2007, Vol. 51, No. 5, pp. 18–26.

This Virtual Molecular World

L. A. Gribov and I. V. Mikhailov

Vernadskii Institute of Geochemistry and Analytical Chemistry,
Russian Academy of Sciences, ul. Kosygina B-334, Moscow, 119991 Russia
tel: (495)1374147
fax:(495)1376371
e-mail: gribov@geokhi.ru

Received July 26, 2007

Abstract—The existing rules and computing techniques are of little use for directed research dedicated to the real molecular world whose objects are enormous in number and diversity. Any progress is possible only through introducing forward-looking numerical computer experiments into the research practice. This requires changing from analysis of single objects and calculation of individual characteristics and properties to designing a virtual molecular world. Its objects, unlimited in number, can, similarly to real objects, respond to external actions and differ from one another. At the same time, by contrast to real objects, they can grow in complexity through interactions, accumulate energy, transform the initial information, etc. In this context, the prospects of designing such a world were discussed, as well as the methodological requirements on problem statement and solving as regards the theoretical and program construction aspects.

DOI: 10.1134/S1070363208040300

In the last quarter of the XX century the mankind closely approached a novel, infinite in its diversity, world termed the virtual world, which is now actively captured. Though similar in part to the real world, in frequent occasions the virtual world is perfectly different, as visualized by movies and images displayed on PC monitors. The intrusion into the virtual world helps many people "to kill the time," but in a very large number of cases this is absolutely necessary for development of the essential aspects of practical activity. For example, of crucial importance for the progress in modern aviation is the possibility, first, to conduct an enormous number of computer experiments when designing airplanes, and second, to train pilots on complex flight simulators. The latter simulate real situations so that the sensations experienced by trainees are very close to reality. Different, in particular, critical situations are easily changeable. There is every reason to say that, without such experiments, most of airplanes would not have left the runways. Examples of this kind are hundreds in number.

Similar as they are to real objects from the perceptive viewpoint, the mathematical objects "living" and interacting in the virtual world have an essential distinguishing feature: They never behave in a manner absolutely similar to that of material objects. This difference is due to "cutting off" certain particularities by model concepts underlying any version of the virtual world, rather than insufficiently powerful computers or deficient theories and computing algorithms. The Bohr's additivity principle states that attempts to take into account all these particularities and make models more sophisticated eventually preclude comparison of the modeled and experimental data. At the same time, models neglecting these particularities describe the objects and their properties in an inadequately simplified manner and inadmissibly reduce the forecasting capabilities of the approach as a whole. It is between these two extremities where the optimal way lies.

This is not the first time when the mankind deals with the virtual world objects. All mathematical images, which are the fruits of the human brain activity, "live" and function in an abstract space of thinking. Nobody has ever seen a matrix, but such abstract objects are widely used in research activities; they can undergo complex changes, interact, display new properties, be copied, etc.

The theory of knowledge treats building any

Table 1. Number of isomers for the chemical system with the composition $C_{10}H_{16}O$ in relation to various structural restrictions imposed

Type of restriction	Number of possible isomers	Number of isomers rejected
No restrictions	452 458	–
Bredt's rule obeyed	437 360	15 098
Structures with evidently improbable –C≡C–C< moieties absent	417 896	19 464
Acyclic groups absent	387 062	30 834
Structures containing six-member cycles only	13 931	373 131
–C≡C– group absent	13 316	615
>C=C=C< group absent	12 834	482
$–CH_2–OH$ moiety present	1 125	11 709
Two $–CH_3$ groups present	431	694
$–CH_2–CH_3$ moiety present	114	317
$>C=C–CH_2–OH$ moiety present	75	39
$>C=C–CH_2–OH$ moiety present	44	31
$>C=C–CH_3$ moiety absent	20	24
$(C)_2C=C(C)_2$ moiety present	4	16

scientific knowledge as specifically designing some abstract world, in which process the objects of this world are described with all their interactions and transformations. If natural experiments evidence that the objects of the abstract world thus designed behave similarly to real objects, the resultant picture is regarded as adequate to the outworld. In the optimal case, it will be possible to predict how a real process will go under appropriate conditions through analyzing the processes occurring in the virtual world.

Modern computing and graphic facilities were of fundamental importance for interconnecting the real and abstract worlds by means of a major information transfer channel, visual perception.

This brings us to the subject of our study proper.

As known, there is no limit to diversity of the objects of real molecular world. There exist published data concerning over 20 million organic molecules only. A stable system comprised of 30–40 atoms can exist as tens and hundreds of millions of isomorphic structures.

According to theoretical predictions (through strongly overestimated), over 10^{20} structural isomers are possible for, e.g., $C_{22}H_{26}O_4$, $C_{20}H_{33}BrO_2$, and $C_{23}H_{37}NO_7$ systems. Clearly, the actual number of isomers is much smaller, with 10^6 being a fairly realistic figure. These structures are able, spontaneously or under external actions, of converting to other structures, breaking down, or combining. They possess the information perception, processing, and storage capabilities and display many other features.

An enormous number of properties of the real molecular world have long been used in practice. It should be noted that a significant proportion of its currently known objects have artificial origin (they are created by researchers). The objects of the real molecular world vigorously grow in number, which suggests that, like in the case of macrocosm, further exploration of the molecular world will inevitably be possible only via combining the natural and forward-looking computer experiments based on the empirical knowledge like "black boxes" and physical theories describing in detail the cause and effect relationships. It will not be an exaggeration to say that further development of chemistry as a science is impossible without equipping chemical laboratories with special "simulators" whose functions will be similar to those of the existing ones.

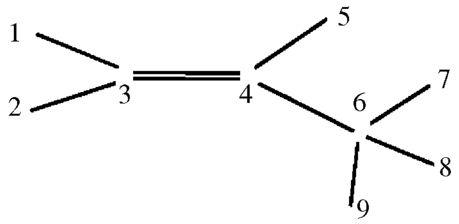
Apparently, P. Dirac was right in declaring that there would be times when hundreds, if not thousands, of chemists would head for computers rather than laboratories.

This raises a question: To what extent is this fantastic and will the problem of the virtual molecular world be solved in a fairly near or very distant future? We seek to answer this question in the discussion below.

Above all, a virtual world evidently needs to be populated with objects, which task can easily be accomplished using the information contained in a large number of the available banks of molecular data. However, this is the first step only: The objects from a bank do not live, in a manner of speaking; they can be manipulated only by taking out or introducing into the bank.

It turns out that a self-growing virtual molecular world can be designed. Indeed, let us presume that only one, more or less complex, heteroatomic virtual molecule was originally introduced into the bank. A computer provided with a structure generator (such generators are now part of all kinds of expert systems [1] used in chemistry) will completely automatically make copies of the initial molecule, thereby designing objects corresponding to its structural isomers. If we will restrict ourselves to only two fundamental axioms of Butlerov's theory, namely, that each chemical element has a certain valence, and carbon atoms can form ordinary, double, and triple bonds, we will obtain an estimate of the number of isomers "from above" and graphic representations of all the structures. As already mentioned, such structures are very large in number. However, this set can be reduced, and in parallel the set of the structures generated can be made closer to reality, via introducing various restrictions, e.g., those concerning their constituent substructures (the necessarily absent/present moieties, presence of cycles, etc.) This is also a smoothly running procedure. Table 1 exemplifies the imposed restrictions in action. Using the formal rules of valence, the structures obtained in the isomer generation stage can be combined or divided into parts, thereby simulating possible synthesis and decomposition reactions.

Notably, a basic precondition to designing the virtual molecular world is as follows: All objects of the virtual world and their transformations should be described in a fairly unambiguous manner using mathematical concepts. For example, the mathematical representa-tion of the structural formula is the so-called contiguity matrix (Fig. 1). This model ranks not very high in the general hierarchy of the models presented in Table 2. It should be noted that Table 2 is incomplete; it demonstrates only the general principles of designing molecular models. More sophisticated models incorporate the parameters of simpler models, but this does not mean that only one model will be sufficient. Application of one complex model combining several simple models would be equivalent to an attempt "to begin again from the beginning" every time. As known, scientific practice always selects those models that are adequate to the problem stated and require the smallest number of concepts for predictions.



	1	2	3	4	5	6	7	8	9
1	0	0	1	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0
3	1	1	0	2	0	0	0	0	0
4	0	0	2	0	1	1	0	0	0
5	0	0	0	1	0	0	0	0	0
6	0	0	0	1	0	0	1	1	1
7	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	1	0	0	0

Fig. 1. Contiguity matrix for the propylene molecule. The matrix lines and columns are numbered identically to atoms; the bond order is indicated in the cells.

It should be mentioned that the need to use different models prevents giving only one information-intensive definition to an object such as molecule. This is fully consistent with the concept advanced by N. Bohr when introducing the additivity principle. N. Bohr emphasized the need to understand that all-sided interpretation of a subject may demand various view-

Table 2. The hierarchy and mathematical representation of molecular models

Model	Mathematical representation
Empirical composition	Atomic numbers and masses of the chemical elements in the molecule
Structural formula	Contiguity matrices
A rigid geometric figure	Bond lengths and bond and torsion angles
An elastic geometric figure	Additionally: an internal coordinate system, elasticity matrix, or a more complex potential function
A geometric figure that tolerates internal rotation and inversion	Additionally: a coordinate system of internal rotation and inversion potentials
Figures for the ground and excited electronic states	The same characteristics for vibronic states
Models with atomic charges, bond orders, and dipole moments of bonds indicated	The values of corresponding parameters

points precluding its unambiguous description and that, strictly speaking, in-depth analysis of any concept and its direct application are mutually excluding [2].

Clearly, simple models are inadequate to designing a virtual molecular world. At the same time, introducing ever growing number of details can eventually make the model too complex to be of any use. To put it another way, twenty words is insufficient to describe a complex world, but too many words is no good either.

Next, we will emphasize the fact that all these abstract models would be impractical if their characteristics were not reflected in the class of the parameters observed. For example, a model of the empirical composition allows obtaining the emission and X-ray fluorescence spectra, a model organized as a rigid geometric structure, the diffraction patterns and electronic spectra, elastic models, the IR and Raman spectra, etc. With the calculational methods available for unambiguously relating the models and their parameters with those observed in experiments it is possible to interconnect the virtual and real worlds. In this case, the results of experiments carried out in the virtual world will enable understanding and predicting the behavior of objects in the real world. This was made possible by creation of quantum mechanics.

Many brilliant scientists (L. Pauling, R. Mulliken, S. Popple, and other), basing on fundamental concepts, contributed to development of physical theories, computing algorithms, and software to describe the above-listed models and their reflection in the measured parameters at the quantitative level so that numerical data can be derived for all the molecular parameters. The complexity of the structure in question is theoretically unlimited. The results of many-year research activities were summarized in numerous monographs, many of which are well known.

In the real world, the very fact of existence of molecules can be revealed only in experiments that demonstrate how the molecules respond to external actions. In just the same way the virtual world of molecular models of any degree of complexity is also “a thing in itself,” absolutely unsuitable for any predictions. To transform this world into a “thing for us,” it is essential that a computer be able of introducing external actions of different nature and obtaining a response from a model, that can be compared with the experimental data. For example, a virtual exposure of a molecular model to electromagnetic radiation allows obtaining the response in the form of a spectral distribution curve for the absorption coefficient, i.e., in that corresponding to the absorption spectrum recorded in a real experiment (Figs. 2, 3). Figure 4 demonstrates the possibilities of calculating the dynamics (changes over time) of the optical spectra of the molecules excited by supershort pulses. Essential for describing the molecular spectra dynamics are the theory and methods for calculation of the optical spectra, NMR spectra, X-ray diffraction on three-dimensional atomic lattices and electrons on molecules, knowledge of chemical transformation laws, etc.

Finding solutions to methodological problems relevant to exploration of the virtual molecular world took over half a century (although the agreement between the calculated and experimental data, and

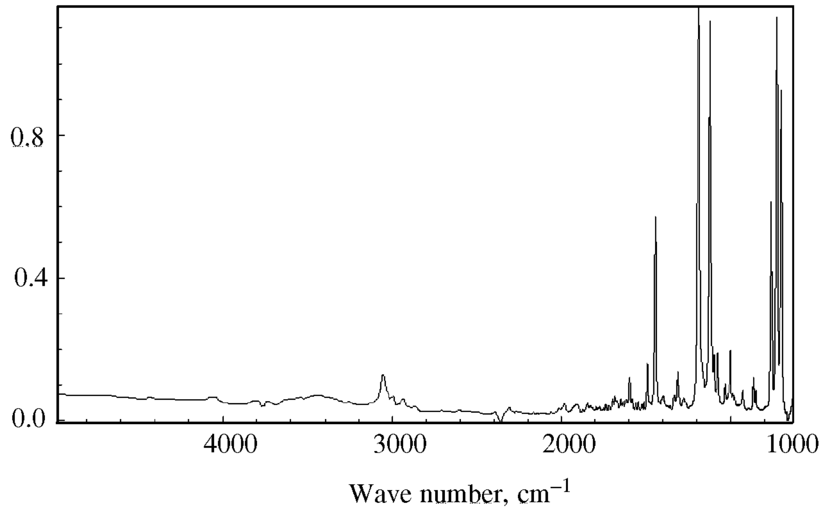


Fig. 2. Response of a virtual structure to the black body (a heated rod) emission. An IR absorption spectrum is recorded.

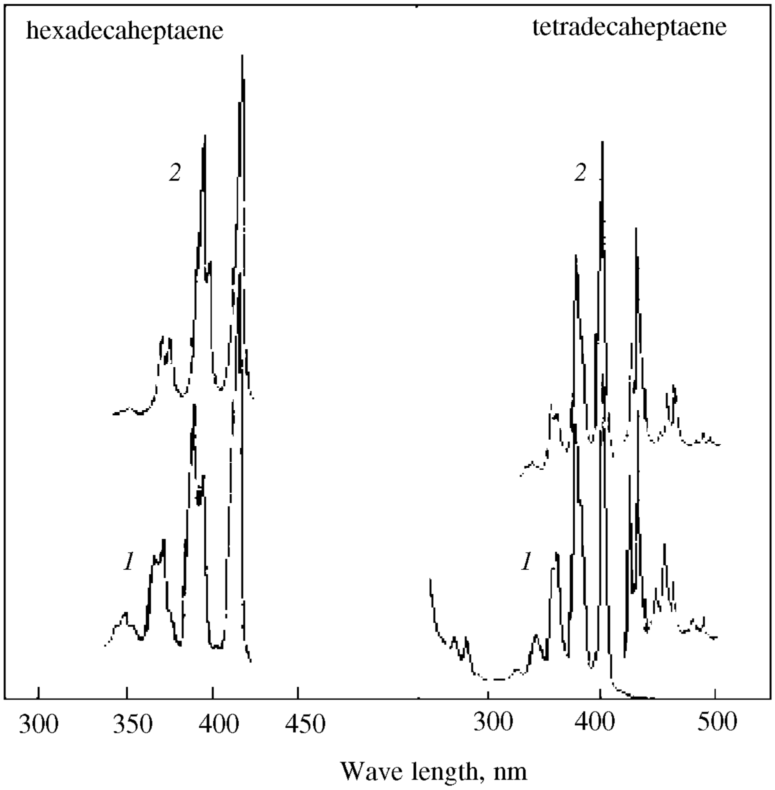


Fig. 3. (*1*) Experimental and (*2*) calculated absorption spectrum of hexadecaheptaene and tetradecaheptaene molecules.

hence the forecasting capability of the models are still not always at the desired level). The progress thus achieved was due, above all, to happy empirical discoveries that underlie the models, like representing the electron functions as linear combinations of atomic orbitals (LCAOs), Gaussian basis, etc., as well as combining the operator and matrix methods in quantum problem statement and solving, rather than simply to active development of mathematical approaches and perfect computer facilities.

Clearly, the whole body of experience in development of the theory of the molecular structure and properties indicated unequivocally what kind of general methodological basis should underlie bridging the basic concepts of quantum theory with the

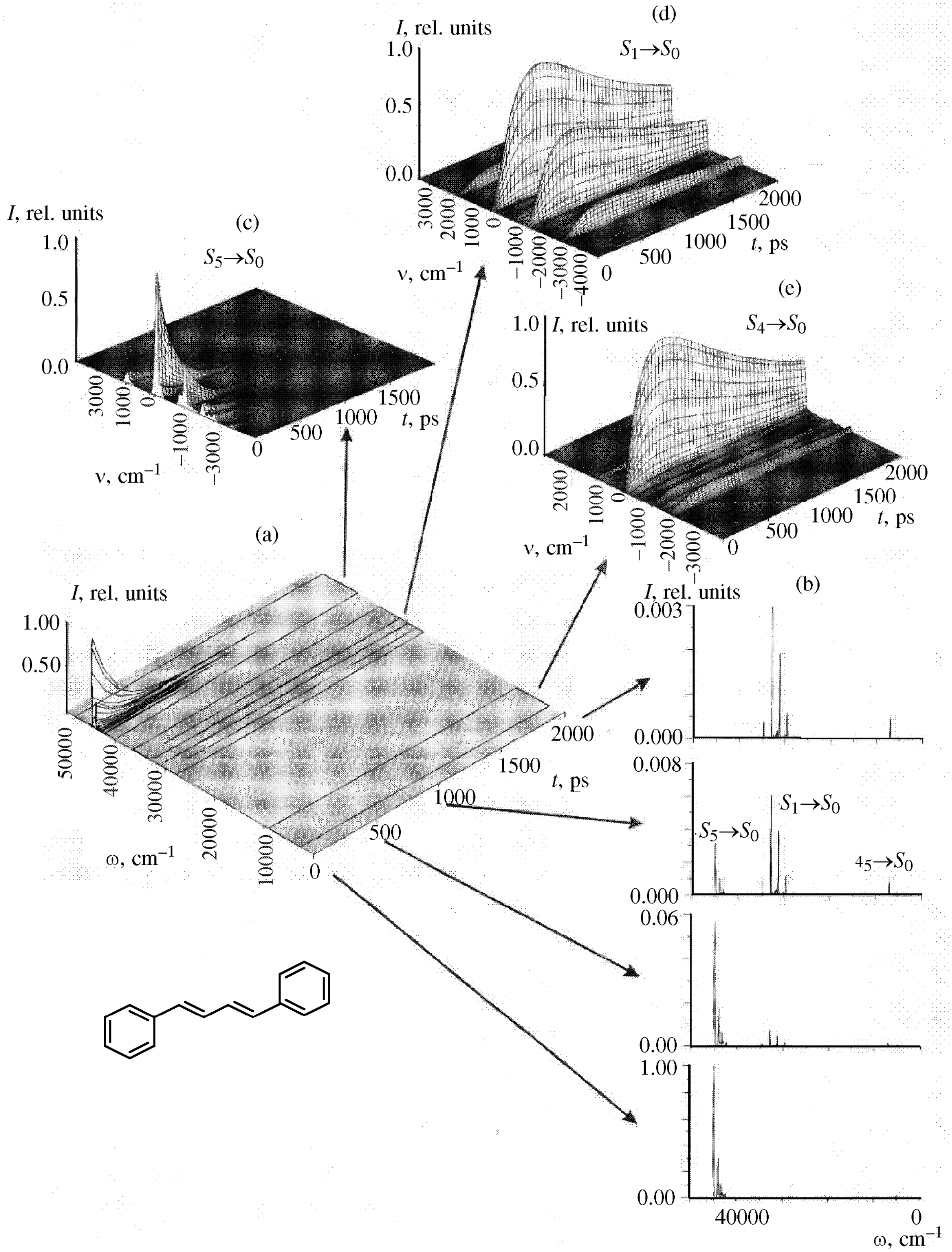


Fig. 4. Electronic spectrum of the model of the diphenylbutadiene molecule upon excitation of all vibrational sublevels of the S_5 state: (a) full-scale, (b) sections at $t = 0$, 500, 1000, and 2000 ps, and (c, d, e) regions corresponding to different electronic transitions.

techniques and methods to handle practically important complex systems. However, even at the present time all the achievements are reduced to routine calculations of three-dimensional structures of isolated molecules, as well as of their vibrational spectra, dipole moments, and distributions of electron density and external electrostatic fields, with all this concerning the ground states of the molecules solely. When applied in analysis of high-excited states and, e.g., in prediction of the appearance of the fine-structured electronic-vibrational spectra, these computing techniques do not provide reasonable agreement between the experimental and calculated data.

Even less favorable is the situation with the theory of chemical transformations initiated by various external actions. The basic model is still that of "barrier overcoming," proposed several tens of years ago. The main drawback suffered by this model consists in that it does not consider the time factor, although all chemical transformations obviously develop in time. Also, this theory does not describe photochemical reactions and many other molecular processes.

In this context, further step in development of the theory of chemical transformations had to consist, above all, in designing physical models that, based on several first fundamental principles and concepts, will allow arriving at conclusions consistent with the maximal number of the major general regular features of the molecular processes of different nature (spectral, structural transformations, etc.) without summarizing the empirical results. Also, computer simulation required that the ways of quantitative description of the proceeding of such processes be specified.

We listed below some of the well-known general chemical regularities whose description in terms of a simple physical model is the most desired.

(1) There exist structural isomerization, decomposition, and synthesis reactions.

(2) All reactions can be subdivided into fast (fractions of second to several hours) and slow (days, months, years) reactions.

(3) There can be exo- and endothermic reactions.

(4) Reactions are initiated by heating and photoexcitation.

(5) The kinetic relationships manifested in spectra and many reactions are described by exponential functions of matrices.

(6) A large number of reactions adhere to the Arrhenius equation. The reaction induction effect is observed fairly often.

(7) Reactions involving complex molecules proceed only at local reaction centers. In the case of very large molecules, the activation energy of the reaction decreases.

(8) "No-energy" (without exposure to severe external impacts) and vibrational processes can proceed during reactions.

(9) Radical rearrangements of the structures of complex molecules are possible only as a result of consecutive structural isomerization. The same processes are responsible for energy and information transfer in the molecular space.

(10) Molecules can process fuzzy information.

(11) Monitoring of fast processes reveals quantum beats.

(12) All chemical transformations satisfy the short-range principle.

In further discussion of the problem of virtual molecular world designing we will restrict ourselves to these concepts.

It turned out that, to arrive at similar conclusions "from the first principles," it is sufficient to proceed from the following few basic concepts.

(1) A set of atoms, localized within a limited space, can occur in the stationary states corresponding to the structural isomeric forms (long-lived states) or to weakly bound complexes. These states are characterized by points in space, represented in normal coordinates.

(2) Conventional optical processes do not involve transitions between such points (e.g., isomer-isomer transitions).

(3) Every type of chemical transformation is brought into correspondence with a transition from one point to another in the space of states. The probability of such a transition is at a maximum under resonance of the energy levels of the substructures being combined and is determined by the corresponding matrix element of their interaction.

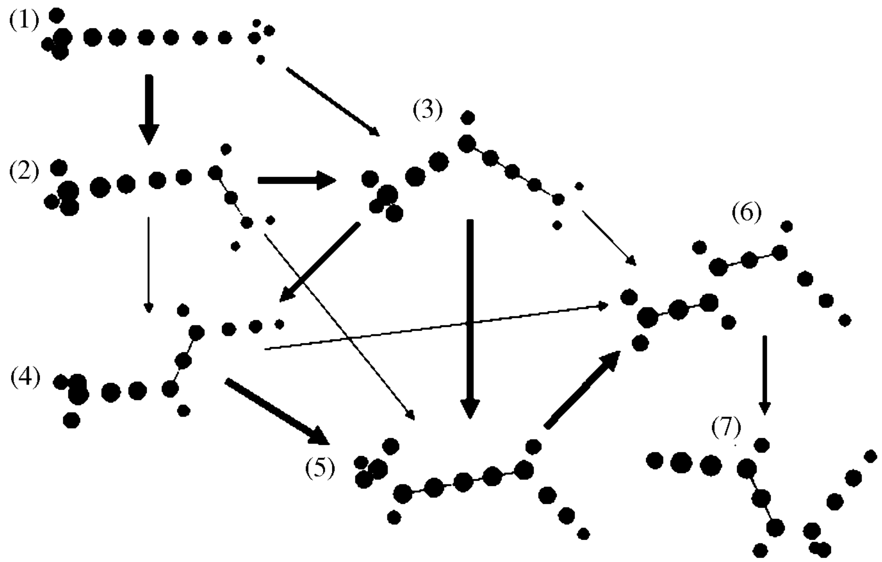


Fig. 5. Mutual transformation routes for isomeric forms of the C_8H_6 system. The arrow thickness correlates with the probability of the structural transition.

(4) Every process occurring in ensembles of molecular systems develops in time and is described by the system of differential equations

$$dn/dt = P(t)n,$$

where n is the column matrix of the population of states (energy levels) of subsystems, and $P(t)$, transition probability matrix, whose elements can take into account the time factors of the resonant energy or population transfer between the subsystems. The kinetics of the process is determined by the initial conditions.

It is seen that all the basic concepts are fairly simple and clear. In particular, the resonance is remarkable in that, even at a very small coupling coefficient, the energy can be exhaustively transferred from one subsystem to another.

The above-described concepts were represented in the mathematical form, which allows developing a logically consistent sequence of computing algorithms [3].

The theory of chemical transformations, underlain by the resonance tunneling concept, allows the probabilities of the structural isomerization, decomposition, and synthesis reactions to be fairly easily, though roughly, estimated “from above”. Such estimates can be obtained by calculating the $\exp(-b^2)$ parameters, where b^2 is the norm of vector connecting the points in the multidimensional space of normal coordinates, corresponding to, e.g., two isomeric structures. This allows constructing the corresponding contiguity matrices to not only derive the possible routes of chemical transformations (at least as “yes/no”) but also predict how the processes will develop in time (see an example in Fig. 5).

Figures 6 and 7 exemplify the time dependence of the isomer-isomer transformations (dynamic spectra and ground level populations) under initial photoexcitation.

Along with radiation, external actions such as temperature, density of the medium, electric and magnetic fields, etc. are suitable for the virtual molecular objects in model experiments.

Molecular objects are able of not only perceiving but also memorizing and processing in a complex way the external actions. It was shown that all the operations of formal logics in the so-called Boolean algebra correspond unambiguously to real processes in the molecules and their transformations. Thus, a new scientific direction is being developed which is in some cases termed molecular logics. Virtually no empirical experience was accumulated for this area, but it shows much promise: For example, the processes

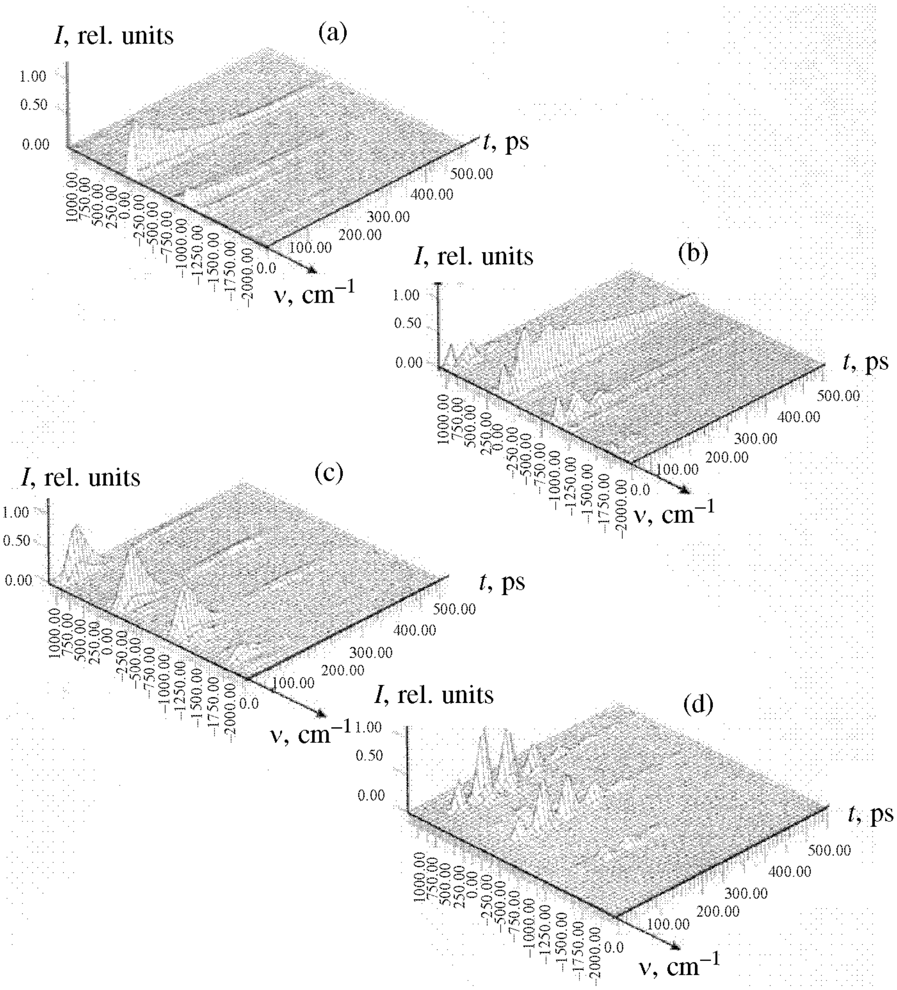


Fig. 6. Calculated luminescence spectra of (a) hepta-1,6-dienylbenzene, (b) hepta-1,5-dienylbenzene, (c) hepta-1,4-dienylbenzene, and (d) hepta-1,3-dienylbenzene molecules, generated under wide-band excitation of hepta-1,6-dienylbenzene.

occurring in molecules can serve for recognition of images based on fuzzy (which is extremely important!) input information (see [3]).

As already mentioned, the states and changes of a very large number of objects in the virtual world can be described by setting the populations (energy levels) of the subsystems containing identical structures. This allows introducing the concept of a multidimensional molecular space and characterizing it with the population coordinates. All processes in this space, manifested both in spectra and chemical reactions, can be described in terms of the uniform system of the differential equations, presented above. Through solving them it is possible to examine various generalized characteristics, such as the routes of spectral and chemical transformations, energy and information fluxes, etc. (see Figs. 8, 9).

The coordinates are the populations of the ground energy levels of isomers A, B, and C. At $t = 0$, the upper level of isomer A is populated. These routes were calculated for different parameters of the transition probabilities matrix $P(t)$.

The above-said will suffice to realize, on the one hand, that designing a virtual molecular world whose objects, for variously stated problems (demands), behave in a manner fairly close to that in real experiments, is a complex task and, on the other, that the case in point is a scientific forecast based on a very serious work done in advance, rather than a fantasy.

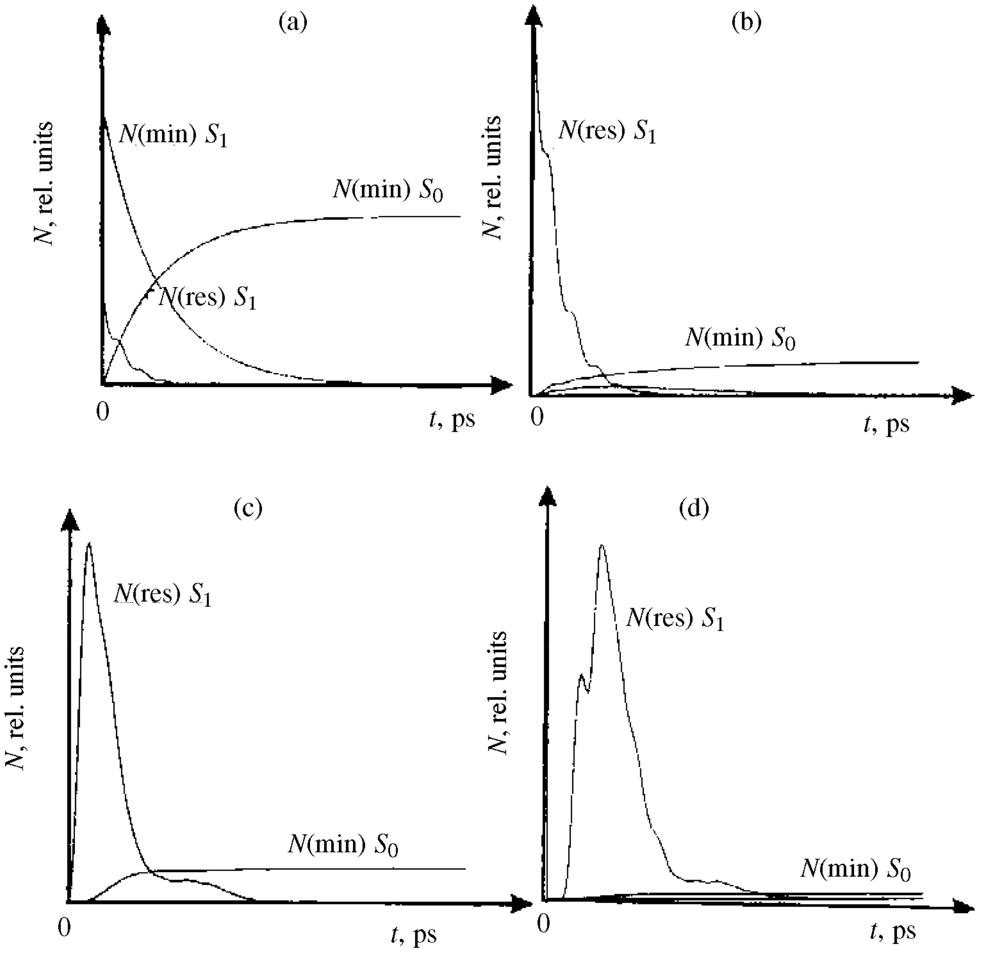


Fig. 7. Variation of the levels populations for hepta-1,6-dienylbenzene isomers under wide-band excitation: (a) hepta-1,6-dienylbenzene, (b) hepta-1,5-dienylbenzene, (c) hepta-1,4-dienylbenzene, and (d) hepta-1,3-dienylbenzene.

Certainly, implementation of this forecast will imply, above all, a kind of “great consolidation” of expertise around designing a virtual molecular world. Naturally, this cannot be done by simply collecting theoretical and program developments in an unsystematic manner. It is essential to design a logically consistent architecture and a system of algorithms that will allow “taking different routes and covering different distances when floating freely in the virtual molecular sea.”

There is a need in formulating the man–computer interaction principles underlain by a maximal friendliness of its participants, as well as in designing advanced computer graphic facilities.

Certainly, much effort is needed to accomplish these complex tasks. The initial step can consist in development of a new-generation powerful dialog expert system with a novel destination, capable of providing a breakthrough into the “operations space” of the microcosm. Otherwise, any progress in the field of chemistry, nanotechnologies, and molecular devices and machines of various designations is hardly possible. Such expert systems can prove to be irreplaceable in educational and professional training activities.

Of principal importance is also the fact that, at the present time, the expertise is represented primarily by computer program products. It is virtually impossible to “get” into the available complex programs and extract something helpful in development of new ideas. Without appropriate program base this will automatically make all the works done in the field of

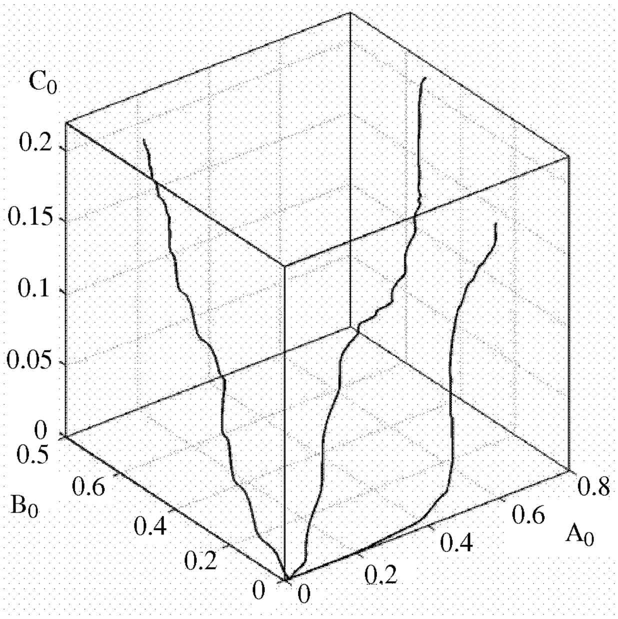


Fig. 8. Examples of the routes for photochemical reactions.

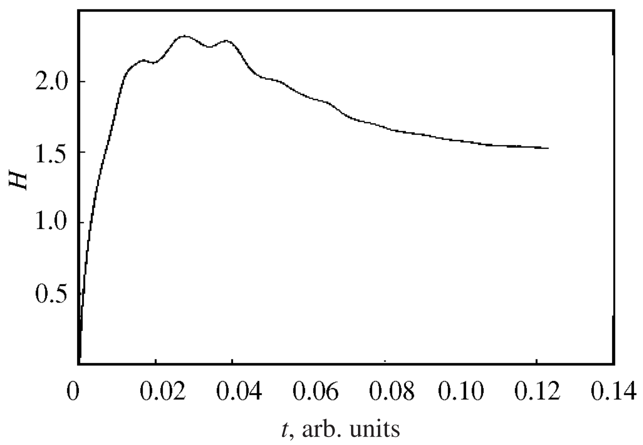


Fig. 9. A model example of the time dependence of the information entropy for the case when the upper energy level of the initial isomer is “discharged.”

the microcosm accessible for those countries where such products are available.

Naturally, small laboratories cannot adequately respond to the challenge posed. However, it is essential, above all, to realize the need to move in the desired direction and to start this movement. We will remind in this connection that, in a not so distant past, the task of high-precision guidance of intercontinental-range missiles was successfully accomplished, which also implied pooling efforts by many theoreticians and mathematicians. That problem is close in complexity to the problem of interest, with the latter being of even greater practical significance. In this context, it should be noted that development of nanotechnologies, which undoubtedly implies designing the appropriate theoretical provision, was declared as one of the high-priority problems by the Russian Federation Government.

REFERENCES

1. Elyashberg, M.E, *Usp. Khim.*, 1999, vol. 68, no. 7, p. 579.
2. Bohr, N., *Izbrannye nauchnye trudy* (Selected Scientific Works), Moscow: Nauka, 1971б vol. 2.
3. Gribov, L.A. and Baranov, V.I., *Teoriya i metody rascheta molekulyarnykh protsessov*: *spektry, khimicheskie prevrashcheniya i molekulyarnaya logika* (Theory and Methods of Calculation of Molecular Processes: Spectra, Chemical Transformations, and Molecular Logics), Moscow: KomKniga, 2006.