HYBRID TOPOLOGY-BASED COMPUTATIONAL APPROACH FOR CRYSTAL STRUCTURE PREDICTION

V.A. Saleev[1,2], I. A. Baburin[3], D. M. Proserpio[1,4]and A. V. Shipilova[1,2]

[1] Samara Center for Theoretical Materials Science (SCTMS),
Samara State University, Samara, Russia,
[2] Samara State Aerospace University, Samara, Russia
[3]Technische Universität Dresden, Dresden, Germany,
[4] Università degli Studi di Milano, Milano, Italy,

We propose a theoretical approach for the computational design of new crystal structures. Our approach is based on the analysis of the database(s) which contain(s) hypothetical nets and subsequent multi-scale computations at different levels of theory (from classical potentials to *ab initio* methods). To work with large databases, containing ~10^6 nets and to find physically adequate crystal structures, we use a set of computer programs: *GavrogSystre and ToposPro* to manipulate with network topologies and their geometrical relaxation, *GULP and DFTB+* for preliminary structure relaxation, *VASP and CRYSTAL* for accurate *ab initio* calculations of the main physical properties. Additionally, we apply the *CP2K* code to check the dynamical stability of the structures at ambient conditions by performing molecular dynamics simulations. We demonstrate the advantages of the proposed approach when using it to predict a few previously unknown sp^3 carbon allotropes. For our simulations we use state-of-art parallel computing facilities at the supercomputer of *Samara Center for Theoretical Material Science and Supercomputer Center of Samara State Aerospace University*.

Modern methods of crystal structure prediction

Is it possible to predict a crystal structure if we know only its chemical composition? Until recently, the all-known answer for this fundamental question was “No” [1]. This situation strongly differs from the molecular structure prediction, when it has only computing difficulty, but, in principle, the knowledge of the chemical composition is enough to find all possible molecules once the chemical composition is specified. Concerning crystal structure prediction, the situation began to change dramatically after 2000 [2]. Nowadays, many methods of crystal structure prediction have been developed and we can speak in fact about scientific revolution in this important field of science. There are two main approaches for crystal structure prediction: the data mining approach [3] and the approach of computational optimization [4]. In the first case, a large database of experimentally observed structures is required. The data mining approach involves advanced machine learning conceptions, which are based on the empirical laws and biased hypotheses. The second one, the approach of computational optimization, suggests explicit calculations of the landscape, in order to find the most stable structures (*i.e., local minima*). The latter approach looks the most unbiased, non-empirical and more general. However, the main problem of computational optimization is to find the global minimum of the energy landscape, where the number of degrees of freedom (d) (or dimension of energy landscape) can be very large. For the system of N atoms d can be estimated as d=3 N+3-k, where k is the number of correlated dimensions, 3N-3 is the number degrees of freedom, and six dimensions are lattice parameters. Simple random search methods can provide correct solutions only for very small systems, approximately N<10 [5]. For much larger systems different advanced techniques are required, such as Basin-Hopping Method [6], Minima Hopping Method [7], Simulated Annealing Method [8], the Metadynamics Approach [9] and the Evolutionary Approach [10, 11]. All these methods apply random walk technique which needs multiple calculations of the system energy with high accuracy at the Density Functional Theory (DFT) level. These advanced methods are quite demanding and require sufficient supercomputer resources and long time of numerical search but yet do not guarantee that we will arrive at the global minimum of the energy landscape.

Alternatively, the Periodic-Graph Approaches for crystal structure prediction were suggested [12,13]. The crystal structure can be described as an infinite periodic graph (a *net*). Recent theoretical investigations demonstrate the importance of topological graph concept and representation of a crystal structure as a periodic net. Performing topological analysis, we deal with a huge amount of crystallographic data using advanced database-management systems. At present, there are several electronic databases which accumulate the collections of nets, such as RCSR [14], EPINET [15], TTD and TTO [16]. The last ones integrate all collections and contain about 10^6 topological types of nets. GavrogSystre [17] or ToposPro [18] can be used to determine crystal structure topology and to find other topologically similar compounds.

The main idea of the methodology proposed here, that is essentially the topology-based *ab initio* approach, is to exclude random walk approximation in the search for new crystal structures. We suggest to find new structures using the nets collected in the databases. At the first step, we generate new structures in 3-dimensional Euclidean space from the abstract topological nets and select relevant structures by performing geometrical analysis (based on bond lengths and angles). At the second step, we perform a 'quick and dirty' geometrical relaxation of selected compounds and collect the structures with minimal energies. Because the number of computed structures at this step is about $\sim 10^5$, we used fast numerical methods based on the classical force-field approach, as it is implemented in the GULP package [19], or approximate DFT model, the Density-Functional-based Tight-Binding (DFTB) Approach, as it implemented in DFTB+ computer package [20]. After selection of a few structures which have energies in the range of 0.1-0.3 eV (it depends on the type of structure and chemical composition), we perform more accurate relaxations of crystal structures and calculate their properties using full DFT level of calculation (VASP [21] or CRYSTAL [22] packages)

Topology-based computations with Gavrog Systre

The program Systre [23] from the Gavrog software package [17] is designed for the topological analysis of periodic nets. It uses a method of barycentric placement to determine the ideal symmetry of a crystal net. The ideal symmetry means the maximal symmetry in which the given net can be embedded in the 3-dimensional space. Then Systre analyzes its topological structure and generates a unique key for the topological type of a given net and uses this key to look up the structure in its built-in database. Systre also computes a new embedding of the net which shows the ideal symmetry. Systre's built-in database contains 176 zeolite framework types and most of the nets contained in the RCSR [14].

Force-field simulations with GULP

The GULP program is one of most popular computer package for materials simulation using classical empirical potentials [19]. There is an opportunity to obtain unknown parameters of potential functions by using the fit procedure to the experimental data or by the fit of the energy landscape calculated by means of ab initio DFT approaches. GULP makes structure optimization at fixed conditions, comparison of initial and final structures, and mechanical property calculations, including the phonon spectrum. GULP can perform fast calculations on molecules, polymers, surfaces, as well as slabs and solids. There is a parallel version of GULP for multi-processor computing systems.

Density-Functional-based Tight-Binding method with DFTB+

DFTB+ is a fast quantum mechanical simulation package based on the Density Functional Tight Binding (DFTB) method. In this approach, the energy of the system is approximated as a sum of the occupied Kohn-Sham single-particle energy levels and a pairwise repulsive potential. This implies that DFTB calculations are fast enough but (in most cases) keep the accuracy of the full DFT schemes. Using the DFTB+ package one can quickly relax crystal nets and calculate (with a good accuracy) their energies and other relevant physical parameters, such as band gaps etc. The multithread version of DFTB+ guarantees a high computing performance at the modern computer clusters.

Accurate DFT calculations with VASP and CRYSTAL

DFT calculations for crystal structures, based on the self-consistent solution of Kohn-Sham equations [24], are realized in VASP and CRYSTAL program packages. The main features of VASP package are the following: pseudo-potential approximation in which only valence electrons are considered, the plane wave basis set is used to describe the electron wave functions from Kohn-Sham equations. The exchange and correlation functional can be used in different approximation, such as LDA (Local Density Approximation), GGA (Generalized Gradient Approximation), and global and range-separated Hybrid functionals, which include some part of Hartree-Fock exchange. Distinctive features of the CRYSTAL package are full electron approach for atomic numbers Z=1-53 (effective core pseudo-potentials can be used for all Z), and Gaussian type functions basis set for core and valence electrons is used. Both, VASP and CRYSTAL, can be used effectively for *ab initio* modeling of crystal structure and crystal properties. Evidently, VASP is more adequate for description of metals or intermetallic compounds. In case of porous solids such as Covalent-Organic-Frameworks or Metal-Organic-Frameworks, CRYSTAL package is more preferable for calculations. The simultaneous use of both packages provides the cross check of the results obtained in two numerical approaches

DFT-based molecular dynamics simulations with CP2K

CP2K is a program to perform atomistic simulations of solid state, liquid, molecular, and biological systems. To evaluate the energies and/or the forces, it can use either empirical potentials or DFT(B)-based methods. The program package is well parallelized and is widely used for the systems up to 3000 atoms (in DFT(B)-based calculations) or tens of thousands atoms when empirical potentials are employed.

From zeolite nets to new sp^3 carbon allotropes

Recently, there has been a great interest in predicting novel carbon allotropes. A special attention has been paid to the sp^3 allotropes since most studies were conducted in order to elucidate the atomistic structure of the product of the graphite cold compression [25] that is different either from diamond or lonsdaleite phases of carbon. A manifold of computational techniques have been tried to address the problem of crystal structure prediction. However, in many cases the structures predicted by very sophisticated methods appeared to be topologically the same as certain crystal structures known to crystal chemists for many years. To give a couple of examples, we mention (a) the bct-4 carbon [26] that is topologically the same as zeolite BCT and (b) suggested dense phase of carbon with the topology of quartz. This motivated us to have a closer look at the databases of hypothetical zeolite networks compiled by Deem [27] and Treacy [28]. First, we focused only on the nets without 3- and/or 4-rings (in total 5074 candidates) that would normally induce too much strain in the carbon structures. Second, we performed '*geometrical'* relaxation of the nets with the GavrogSystre package. From the set of *geometrically* relaxed structures we extracted 652 nets where the distances to the next-nearest neighbours were by 40% larger than the distances to the nearest neighbours. These structures – that could be considered as stereochemically feasible – were then optimized with the Tersoff force-field as implemented in the GULP package. After this force-field calculation, 257 structures remained 4-coordinated and were subject to further optimizations with the DFTB method as implemented in the DFTB+ package. From the set of the DFTB-optimized structures, we selected 93 representatives that lie within a narrow energetic window (0.40 eV/atom) relative to diamond and performed structural relaxation at the DFT-GGA level with the VASP package, using well-known PBE exchange-correlation functional [29]. Finally, we focus on the six structures that are energetically the lowest ones, within 0.10 eV/atom (or even less) relative to diamond. Their energetic, electronic, vibrational and mechanical properties were calculated at the DFT-GGA(PBE) level of theory as implemented in VASP and CRYSTAL program packages. The calculated elastic moduli suggest that these materials are as hard as diamond. The phonon dispersion curves show no imaginary frequencies throughout the Brillouin zones. More strikingly, we found out that the optical gaps of our structures are by ~1 eV larger than the band gap of diamond. This is quite interesting and unexpected result since – to the best of our knowledge – hypothetical sp^3 carbon allotropes with the gaps larger than diamond are characteristic for their clathrate-like open frameworks [30] that is not the case for our relatively dense structures. Finite temperature molecular dynamics simulations at the DFTB level of theory (NpT ensemble, p=1 bar, T=300 K) have been performed with CP2K program package [31] in order to ensure the dynamical stability of our structures under ambient conditions. The details of calculations and the analysis of obtained results are presented in our recent publication [32].

Table 1. Structural, energetic, electronic and mechanical properties of novel carbon phases

Structure	Space group	ρ, g/cm^3	ΔE (PBE), eV/atom (VASP)	ΔE (PBE), eV/atom (CRYSTAL14)	E_gap, eV PBE/HSE	*B*, GPa (PBE)	*H*, GPa
diamond	$Fd\bar{3}m$	3.509	0.00	0.00	4.2/5.4	441	93.2
#8170628 (*oP*24-I)	*Pbam*	3.409	0.07	0.08	4.7/5.9	418	91.1
#8129388 (*oP*24-II)	*Pnma*	3.408	0.10	0.11	4.9/6.3	412	90.8
#8255250 (*oP*28)	*Pnma*	3.415	0.10	0.12	4.7/6.0	412	90.9
#8155755 (*oP*20)	*Pmma*	3.431	0.09	0.11	4.0/5.1	420	91.4
#8036927 (*mS*32)	*C*2/*m*	3.418	0.11	0.12	4.5/5.7	415	90.8
#8036926 (*mP*16)	*P*2/*m*	3.423	0.10	0.11	4.3/5.5	423	91.0

The work of S.V.A., P.D.M. and S.A.V. from SCTMS was supported by the Megagrant No. 14.25.31.0005 of the Russian Ministry of Science and Education.

References

1. Maddox J. Nature, **335**, 201 (1988).
2. Oganov A.R. (Editor). *Modern methods of crystal structure prediction* //Wiley-VCH, Berlin, 2011.
3. Curtarolo S., Morgan D., Persson K, Rodgers J, Ceder G. Phys. Rev. **91**, 135505 (2003).
4. Oganov A.R., Valle M. J. Chem. Phys.**130**, 104504 (2009).
5. Souvatzis P., Eriksson O., Katsnelson M, and Rudin S.P. Phys. Rev. Lett. **100,** 095901 (2008).
6. Wales D.J. Energy Landscapes. Cambridge University Press, Cambridge, 2003.
7. Goedecker S. J. Chem. Phys. 123, 9911 (2004); Godecker S., Hellmann W., Lenosky T. Phys. Rev. Lett. **95**, 055501 (2005).
8. Pannetier J., Bassasalsia J., Rodriguez-Carvajal J.,Caignaert V. Nature, **346**, 343 (1990).
9. Martonak R., Laio A., and Parrinello M. Phys. Rev. Lett. 90, 075503 (2003).
10. Oganov A.R.,and Glass C.W. J. Chem. Phys. **124**, 244704 (2006).
11. Zhu, A. R. Oganov, M. A. Salvadó, P. Pertierra and A. O. Lyakhov, Phys. Rev. B,**83**, 193410 (2011).
12. Winkler B., Pickard C.J., Milman V., Thimm G. Chem. Phys. Lett. 337, 36 (2001).
13. Blatov A.V. and Proserpio D.M., in: *Modern Methods of Crystal Structure Prediction* (ed. Oganov A.R.), Wiley-VCH, Berlin, 2011, pp. 1–28.
14. RCSR, *Reticular Chemistry Structure Resource*, http://rcsr.anu.edu.au
15. EPINET, *Euclidean Patterns in Non-Euclidean Tilings*, http://epinet.anu.edu.au
16. TTD&TTO, *Topological Databases and Topological Types Observed*, http://topospro.com
17. Gavrog, *Generation, Analysis, and Visualization of Reticular Ornaments*, http://www.gavrog.org
18. Blatov V.A., Shevchenko A.P., Proserpio D.M., Cryst. Growth Des., **14**, 3576 (2014).
19. Gale J.D. and Rohl A.L., Mol. Simul., **29**, 291 (2003).
20. Seifert G.,Porezag D. and Frauenheim T., Int. J. Quantum Chem., **58**, 185 (1996).
21. Kresse G. and Furthmüller J., Phys. Rev. B, **54**, 11169 (1996).
22. Dovesi R, Orlando R, Erba A, et al., Int. J. Quantum Chem., **114**, 1287 (2014).
23. Delgado-Friedrichs O. and O'Keeffe M., Acta Crystallogr.,Sect. A: Found. Crystallogr., **59**, 351 (2003)
24. Kohn W.,and Sham L.J.Phys. Rev. **140**, A1133(1965).
25. Karttunen A.J., Fässler T.F., Linnolahti M. and Pakkanen T.A., Inorg. Chem., **50**, 1733 (2011).
26. Umemoto K., R. Wentzcovitch R.M., Saito S. and Miyake T., Phys. Rev. Lett., **104**, 125504 (2010).
27. Deem M.W., Pophale R.,Cheeseman P.A. and Earl D.J., J. Phys. Chem. C, **113**, 21353 (2009).
28. Treacy M.J., Randall K.H., Rao S., Perry J.A. and Chadi D., Z. Kristallogr., **212**, 768 (1997).
29. Perdew J.P., Burke K., and Ernzerhof M.Phys. Rev. Lett., 77, 3865 (1996)
30. Karttunen A.J., Fassler T.F., Linnolahti M. and Pakkanen T.A., Inorg. Chem., **50**, 1733 (2011).
31. J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, Wiley Interdisciplinary Reviews: Computational Molecular Science, **4**, 15, (2014).
32. Baburin I.A., Proserpio D.M., Saleev V.A., and Shipilova A.V. Phys. Chem. Chem. Phys. **17**, 1332 (2015).