Finding Reaction Pathways with Optimal Atomic Index Mappings

Deb Sankar De, Marco Krummenacher, Bastian Schaefer, and Stefan Goedecker[*]

Department of Physics, Universität Basel, Klingelbergstrasse 82, 4056 Basel, Switzerland

Finding complex reaction and transformation pathways involving many intermediate states is, in general, not possible on the density-functional theory level with existing simulation methods, due to the very large number of required energy and force evaluations. For complex reactions, it is not possible to determine which atom in the reactant is mapped onto which atom in the product. Trying out all possible atomic index mappings is not feasible because of the factorial increase in the number of possible mappings. We use a penalty function that is invariant under index permutations to bias the potential energy surface in such a way that it obtains the characteristics of a structure seeker, whose global minimum is the reaction product. By performing a minima-hopping-based global optimization on this biased potential energy surface, we rapidly find intermediate states that lead into the global minimum and allow us to then extract entire reaction pathways. We first demonstrate for a benchmark system, namely, the Lennard-Jones cluster LJ_{38}, that our method finds intermediate states relevant to the lowest energy reaction pathway, and hence we need to consider much fewer intermediate states than previous methods to find the lowest energy reaction pathway. Finally, we apply the method to two real systems, C_{60} and $C_{20}H_{20}$, and show that the reaction pathways found contain valuable information on how these molecules can be synthesized.

The indistinguishability of atoms manifests itself in various ways. In a chemical reaction, the indistinguishability means that, in principle, any atom of a certain type in the reactant can be mapped onto any atom of the same type in the product. If the system undergoing the chemical reaction consists of M_1 atoms of type 1, M_2 atoms of type 2, etc., then there are consequently $M_1!M_2!,\ldots$ different possible mappings. For simple reactions, it is often obvious which atom in the reactant is mapped onto which atom of the product. In more complicated chemical reactions or similar processes, such as phase or shape transformations in nanoparticles, the best mapping, however, cannot be foreseen. Actually, in such a context, many atomic index mappings may exist that lead to different low-barrier reaction pathways.

Determining the transition states of a reaction or transformation is of great importance in chemistry, physics, and materials sciences. A large number of methods have therefore been proposed to solve this problem, as reviewed for instance by Reiher and co-workers [1]. In the so-called two-sided methods [2–4], one has to provide not only the reactant and the product but also the correct atomic index mapping; i.e., it has to be known beforehand which atom in the reactant is mapped onto which atom in the product. One could of course perform such a transition state and search for all possible atomic index mappings, but this is prohibitive for large systems because of the factorial increase in the number of possibilities.

In the so-called one-sided methods [5,6], one finds a transition state that is, in a certain sense, closest to the initial guess. Once one has located this transition state, one can find the minima that are connected by this saddle point by performing two simple local geometry optimizations: one starting from a point to the left and a point to the right of the saddle point, along the line of negative curvature. In this way, one automatically obtains a correct index mapping, but it is not guaranteed that the two local minima at the end point are the desired reactant and product.

There is a method that would allow us to study, in principle, any chemical reaction or transformation, namely, molecular dynamics (MD). In this approach, there is no index mapping problem since any atom of the reactant is mapped onto its corresponding atom in the product via its trajectory. However, since most chemical reactions are rare events, the number of MD time steps is prohibitive. Metadynamics [7] methods speed up processes that are slow on the timescale of a typical MD time step, by invoking collective variables that "push" the system in a desired direction. The standard collective variables can target certain phases or general structural motifs, but do not target a single configuration. A typical collective variable that is used to drive chemical reactions in a desired direction is, for instance, a certain bond length, whose value in the product is known and different from the value in the reactant. However, it is very likely that there are other stable configurations that have the same bond length, and so the collective variable can drive the system into a configuration that is different from the desired product. Also, very recently proposed collective variables, such as approximate entropies [8], do not target a single configuration.

Temperature accelerated dynamics [9] is another way to speed up rare events that would be impossible to observe with standard MD. This class of methods has the advantage that it does not need collective variables, but on the other hand, it is again not possible to steer the reaction pathways in a certain desired direction. High-temperature MD trajectories are used in a similar spirit in the transition state search using chemical dynamics simulations method [10] to detect saddle points.

The minima hopping guided path search [11] (MHGPS) is another unbiased method used to explore the entire potential energy surface (PES) and to find a large number of local minima and transition states. Essentially the same information can be obtained by combining basin hopping [12] with an eigenvector following approach [13] to find saddle points. So, in principle, all these methods provide the information required to construct any reaction pathway with a correct index mapping, but do not allow for steering. For small systems, exploring the entire PES is feasible and of interest to find a more or less complete set of minima together with the saddle points connecting them. For large systems, such an approach is, however, numerically too expensive and it is advantageous to concentrate on a subset of minima and saddle points that are of interest in a certain context. Transition path sampling [14] is yet another method that allows for the calculation of multiple reaction pathways between given initial and final configurations. However, again, the initial atomic index matching has to be known beforehand.

In this Letter, we will use configurational distances derived from fingerprints that are invariant under atomic index permutations [15] as a driving force toward the desired final configuration. In contrast to the standard collective variables, this fingerprint can uniquely identify a single configuration. The fingerprint distance is zero if and only if the configuration is exactly identical to the desired final configuration up to translations, rotations, and index permutations. Along the reaction pathway, the distance varies continuously until it vanishes at the final configuration. Adding the distance as a penalty $P(\mathbf{R}_1, \ldots, \mathbf{R}_M)$ to the true PES, $E(\mathbf{R}_1, \ldots, \mathbf{R}_M)$, gives a biased PES, $\tilde{E}(\mathbf{R}_1, \ldots, \mathbf{R}_N) = E(\mathbf{R}_1, \ldots, \mathbf{R}_N) + w\ P(\mathbf{R}_1, \ldots, \mathbf{R}_N)$. The parameter w is, in this context, a suitable chosen weight and $\mathbf{R}_1, \ldots, \mathbf{R}_N$ are the Cartesian coordinates of the N atoms in the system. The penalty is based on a fingerprint obtained from the eigenvalues V_i of an overlap matrix S describing the configuration [15]. As described in the Supplemental Material [16], we have used an additional postprocessing step to eliminate discontinuities in the first derivatives of these eigenvalues that arise when eigenvalues cross. Denoting the regularized eigenvalues by $\tilde{V}_i$, we obtain a regularized distance that serves then as the penalty function

$$P = \sqrt{\sum_j (\tilde{V}_j^p - \tilde{V}_j^q)^2}.$$

All derivatives of this function are continuous. Adding the gradient with respect to the atomic positions of this penalty term to the physical forces gives biased forces that drive the system from the present configuration p toward the desired final configuration q. These forces are invariant with respect to index permutations.

In this Letter, we concentrate on complex reaction pathways, where the system has to cross a substantial number of saddle points. With the right choice of the parameter w, this biased PES has the appearance of a PES of a structure seeker whose global minimum is the desired final state, as can be seen from the disconnectivity graphs in Fig. 1 [17,18]. This means that the downhill barriers (the barriers that one has to overcome when one hops from one minimum into another one with lower energy) are reduced, but have not disappeared. Therefore, a local optimization of the biased PES is not sufficient to find the final configuration. A global optimization is required. However, we exploit the fact that, for a structure seeker, it is easy to find the global minimum with the minima hopping method.
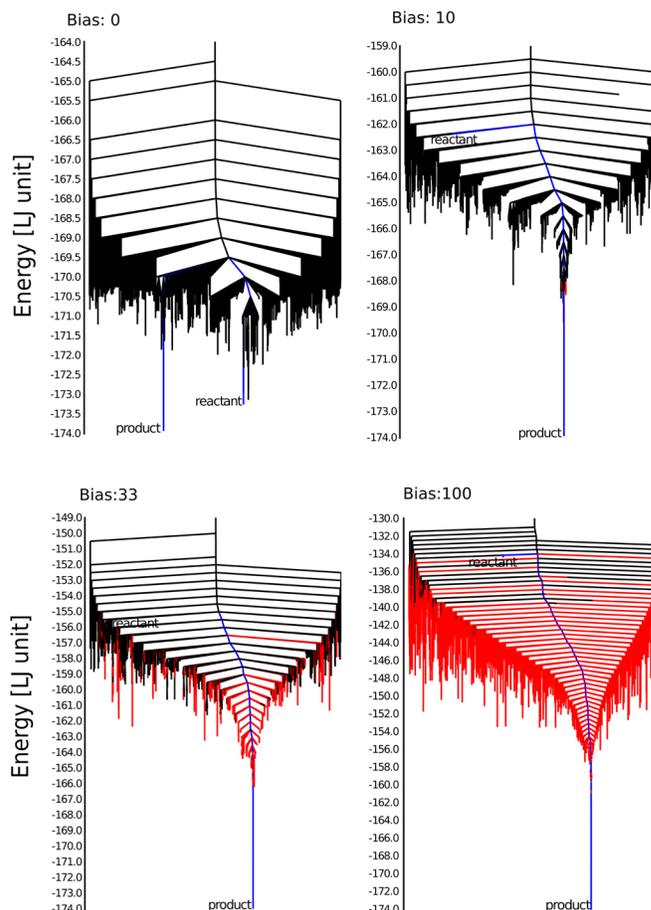

FIG. 1. Disconnectivity graphs of the LJ_{38} cluster with different biased PESs. The red minima indicate fcc-like structures and the blue path shows the transition from the second lowest icosahedral to the global fcc minimum. The graph was constructed with the disconnectionDPS software [18].

We start from an initial configuration (reactant) and then visit with the minima hopping method's consecutive intermediate states until we find the ground state. Because of the stochastic nature of minima hopping, different intermediate states can be visited in different runs. With a reasonable choice for the weight of the penalty, all these intermediate configurations will have low energies and are therefore physically possible intermediate states that are accepted during a minima hopping run. This means that, with this approach, one can find not only one reaction pathway, but all physically relevant low-energy reaction pathways.

Since the hops from one local minimum to the next in minima hopping (MH) are based on MD [19,20] followed by local geometry optimizations [21], the identity of individual atoms of the same type can be traced back for any hop, and one obtains in this way the correct mapping of the atomic indices for the entire complex reaction pathway. The MD trajectories cross barriers, and so one could take some configurations along the MD trajectory as a starting point for a one-sided saddle point search. The MD trajectory can, however, cross over several barriers within one hop of a MH run and, in such a case, it is not clear which point along the MD trajectory should be chosen as the starting point for the saddle point search. We therefore use a recursive variant of the freezing string method [22], as described in Ref. [11] to connect the sequence of accepted local minima by saddle points. The method described so far will be referred to as biased minima hopping guided pathway search (BMHGPS) in the following.

We will first apply our new method to a benchmark system, the Lennard-Jones cluster with 38 atoms (LJ_{38}), for which the lowest energy reaction pathway from the lowest energy icosahedral structure to the fcc ground state is most likely known, since it was studied previously [23–26]. Finding this transformation pathway is quite difficult since it requires a complete overall rearrangement of all the atoms in the system [23].

To find the biased PES that has the strongest structure seeker character, we used four different weights w of the penalty function ($w = 0, 10, 33, 100$). Choosing $w = 0$ gives back the unbiased MHGPS method, which is included for comparison. For each weight, 100 different MH runs were performed. As a reference structure, the fcc global minimum was taken, in order to bias the minimization toward it.

As can be observed in the disconnectivity graphs in Fig. 1, the typical double-funnel structure of the PES disappears by adding the penalty function to the energy. For all nonzero weights, the disconnectivity graph has the shape of a structure seeker, for which it is quite easy to find the global minimum. The more the PES is biased, the more fcc-like minima can be observed. On the nonbiased PES, no fcc structure except the global minimum can be seen (shown in this case in blue and not red since it represents the transformation pathway). For large weights, fcc-like structures start to dominate. Choosing even larger weights

would result in a glassy landscape of fcc-like structures, where it would again be more difficult to find a pathway into the global minimum. In addition, very large weights will destroy a faithful mapping between the local minima on the biased PES and the true physical PES. An analysis of the entire transformation obtained by our method shows that it starts by some kind of surface premelting that then propagates toward the center.

All the MH runs were stopped once the global minimum was found. The number of distinct minima that were visited before the global minimum was found was greatly reduced by the bias. Averaging over 100 runs, we visited, on average, 363 minima without any bias: 19 for $w = 10$, 10 for $w = 33$, and 16 for $w = 100$. In order to obtain reaction pathways, the accepted local minima from the MH run on the biased PES were transformed back to the physical PES by a local geometry optimization. Reaction pathways were then generated as described in Ref. [11]. All our pathways shown in Fig. 2 have a highest energy barrier of -169.709 in LJ energy units, which is identical to the lowest highest barrier height found in previous studies [23,27]. The reaction pathway found in these two studies was, however, based on a dataset of 140 000 saddle points, whereas we could obtain this result thanks to the bias with a much smaller dataset. We used all the saddle points obtained in the 100 runs and in this way collected 3238 saddle points for $w = 10$, 2068 for $w = 33$, and 4183 for $w = 100$. Without a bias, we used a dataset of similar
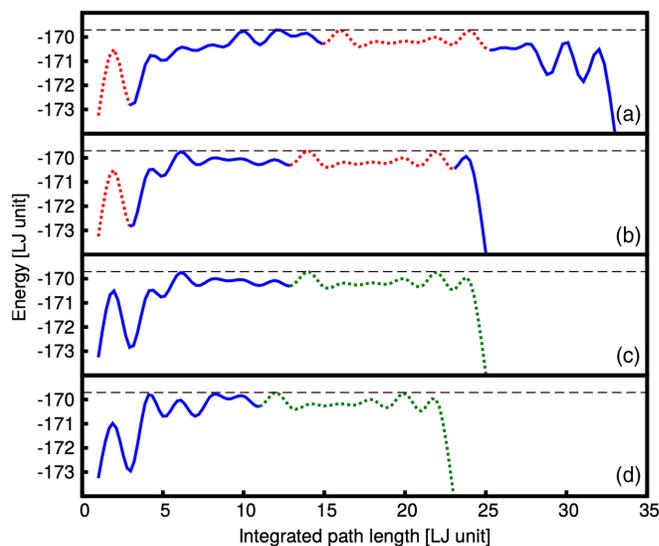

FIG. 2.   BMHGPS pathways found for the LJ_{38} cluster with different bias strengths (a) $w = 100$, (b) $w = 33$, (c) $w = 10$, and (d) $w = 0$ connecting the second lowest icosahedral minimum with the global fcc minimum. The dashed horizontal lines indicate the highest energy along the lowest known pathway [23,27]. The sections indicated by dashed lines are shared by pathways obtained with different weights. The pathways found with a weight of $w = 10$ and 33 are identical.
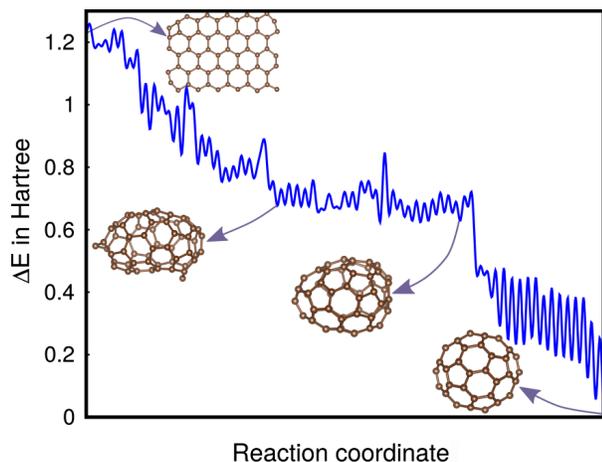
FIG. 3. Typical low-barrier pathway of C_{60} from a graphene flake to a fullerene.

FIG. 4. Typical low-barrier pathway of $C_{20}H_{20}$ from a pagodane to a dodecahedral configuration.

size as the previous studies, namely, 83 000 saddle points. This reduction in the number of required saddle points is due to the fact that our biasing allows us to preferentially find the intermediate states and saddle points in the region of configurational space that is in between the reactant and the product. Those states are the relevant ones for our wanted reaction pathways from the reactant to the product.

Because of the significant efficiency gains obtained by BMHGPS, it can also be applied on the density-functional theory (DFT) level and we will use it to find highly complex reaction pathways for C_{60} and $C_{20}H_{20}$. The DFT calculations were performed with the BigDFT code [28] using dual space Gaussian pseudopotentials [29]. An initial pathway was obtained with the self-consistent charge density functional tight-binding method as implemented in DFTB+ [30] with s and p atomic orbitals for carbon. These pathways were then further refined with BigDFT calculations.

Several MD studies of C_{60} formation failed to find a route from high-energy structures to the buckyball ground state, presumably because the necessary timescales are not accessible to MD simulations [31–37]. Only a semiclosed C_{60} pseudocage was found after a simulation time of 250 ns using a classical Brenner potential and an adaptive temperature MD method [38].

Using our new method, we were able to find a reaction pathway from an initial structure that was a planar graphene flake to the buckyball ground state. Good weights w are in the range of 0.2–0.3, which means that the energy difference between the ground state and the first Stone-Wales defect is increased from 0.06 to 0.08 Ha, and the energy difference between the initial flake and the ground state is increased from 1.21 to 1.97 Ha. A dataset of 9000 saddle points was used for all reaction pathway calculations. It was obtained from 4 MH runs that were stopped once the global minimum on the biased PES was found.
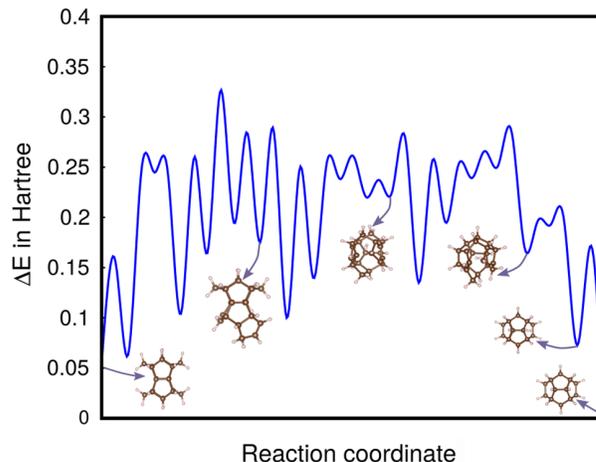
The Dijkstra algorithm [39] can then be used to find the reaction pathway with the lowest highest barrier along the path. Such a path is shown in Fig. 3. As in the case of LJ_{38}, even for a given lowest highest barrier, multiple pathways can be found. Even more pathways can be found if one allows for slightly higher barriers at the very beginning of the reaction pathway. A selection of such pathways is shown in the Supplemental Material [16]. Even though the barriers along the pathways are quite high, there is a clear tendency to lower the energy as one approaches the ground state. This is also true for all the pathways that we have found and so many pathways will contribute to the formation of the C_{60}. The dynamics of the synthesis process can be understood from our database by using a master equation [40] or kinetic Monte Carlo [41] approach. This bias of all the pathways explains why C_{60} can be synthesized in a one-step procedure by graphite vaporization at high temperatures [42].

Using the same approach as for C_{60} we also obtained the reaction pathways of $C_{20}H_{20}$ from its pagodane configuration to its dodecahedron ground state. A representative reaction pathway, obtained from a database of 900 saddle points, is shown in Fig. 4. In contrast to C_{60} there is no systematic lowering of the energy of the intermediate structures as one approaches the ground state. This reflects the fact that there is no one-step synthesis recipe to obtain the dodecahedron from the pagodane structure [43]. The experimental synthesis procedure consists of 11 single steps, each of which involves other ingredients that modify the PES and which therefore give the necessary driving force toward the ground state [44–46].

In summary, we introduced a bias that is invariant under atomic index permutations and that can target a single well-defined configuration as the final configuration of a chemical reaction or physical transformation. In this way, we overcome the index mapping problem that would

require an exponentially large number of applications of standard two-sided saddle point search methods. The forces arising from the bias, by construction, do not depend on the indexing of the atoms. We have thus reduced the combinatorial atomic indexing problem, which has an exponential scaling, to a global minimization problem on a biased PES involving an indexing invariant penalty function. For suitably chosen weights, the biased PES has the form of a structure seeker and the optimization problem is rapidly solvable, in practice, by global optimization methods such as minima hopping. In contrast to standard and order parameters that can drive the system only into certain regions of a low-dimensional order parameter space, our method pulls the system toward a single configuration in the full $3N$-dimensional configuration space. Collective variables are, in general, system specific and finding them can be nontrivial. The penalty function we proposed is universal and can be applied to any reaction or transformation in molecules or clusters.

We expect that our method will give atomistic insight into complex reaction pathways found, for instance, in catalysis, as well as complex phase and shape transformations in nanoparticles. It may also help in finding new organic-chemistry synthesis pathways. Permutationally invariant distances could also be useful in other reaction path search methods in which the atomic index mapping is a limiting factor.

[*]stefan.goedecker@unibas.ch

[1] G. N. Simm, A. C. Vaucher, and M. Reiher, J. Phys. Chem. A **123**, 385 (2019).

[2] G. Henkelman, B. P. Uberuaga, and H. Jnsson, J. Chem. Phys. **113**, 9901 (2000).

[3] R. Granot and R. Baer, J. Chem. Phys. **128**, 184111 (2008).

[4] S. A. Ghasemi and S. Goedecker, J. Chem. Phys. **135**, 014108 (2011).

[5] G. Henkelman and H. Jnsson, J. Chem. Phys. **111**, 7010 (1999).

[6] Y. Abashkin and N. Russo, J. Chem. Phys. **100**, 4477 (1994).

[7] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).

[8] P. M. Piaggi and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **115**, 10251 (2018).

[9] M. R. Sørensen and A. F. Voter, J. Chem. Phys. **112**, 9599 (2000).

[10] S. Vzquez, X. Otero, and E. Martinez-Nunez, Molecules **23**, 3156 (2018).

[11] B. Schaefer, S. Mohr, M. Amsler, and S. Goedecker, J. Chem. Phys. **140**, 214102 (2014).

[12] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).

[13] J. Doye and D. Wales, Z. Phys. D **40**, 194 (1997).

[14] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).

[15] A. Sadeghi, S. A. Ghasemi, B. Schaefer, S. Mohr, M. A. Lill, and S. Goedecker, J. Chem. Phys. **139**, 184118 (2013).

[16] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.123.206102 for more details on the method and fingerprint distances, influence of the bias on the equilibrium configurations and different reaction pathways from graphene flake to C_{60}.

[17] O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).

[18] M. Miller, D. Wales, and V. de Souza, disconnectionDPS, http://www-wales.ch.cam.ac.uk/software.html.

[19] S. Roy, S. Goedecker, and V. Hellmann, Phys. Rev. E **77**, 056707 (2008).

[20] M. Sicher, S. Mohr, and S. Goedecker, J. Chem. Phys. **134**, 044106 (2011).

[21] B. Schaefer, S. A. Ghasemi, S. Roy, and S. Goedecker, J. Chem. Phys. **142**, 034112 (2015).

[22] A. Behn, P. M. Zimmerman, A. T. Bell, and M. Head-Gordon, J. Chem. Phys. **135**, 224108 (2011).

[23] J. P. K. Doye, M. A. Miller, and D. J. Wales, J. Chem. Phys. **111**, 8417 (1999).

[24] J. Neirotti, F. Calvo, D. L. Freeman, and J. Doll, J. Chem. Phys. **112**, 10340 (2000).

[25] V. A. Mandelshtam and P. A. Frantsuzov, J. Chem. Phys. **124**, 204511 (2006).

[26] R. M. Sehgal, D. Maroudas, and D. M. Ford, J. Chem. Phys. **140**, 104312 (2014).

[27] J. P. K. Doye, M. A. Miller, and D. J. Wales, J. Chem. Phys. **110**, 6896 (1999).

[28] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, and R. Schneider, J. Chem. Phys. **129**, 014109 (2008).

[29] A. Willand, Y. O. Kvashnin, L. Genovese, A. Vazquez Mayagoitia, A. K. Deb, A. Sadeghi, T. Deutsch, and S. Goedecker, J. Chem. Phys. **138**, 104109 (2013).

[30] B. Aradi, B. Hourahine, and T. Frauenheim, J. Phys. Chem. A **111**, 5678 (2007).

[31] P. Ballone and P. Milani, Phys. Rev. B **42**, 3201 (1990).

[32] J. R. Chelikowsky, Phys. Rev. Lett. **67**, 2970 (1991).

[33] J. R. Chelikowsky, Phys. Rev. B **45**, 12062 (1992).

[34] X. Jing and J. R. Chelikowsky, Phys. Rev. B **46**, 5028 (1992).

[35] J.-Y. Yi and J. Bernholc, Phys. Rev. B **48**, 5724 (1993).

[36] C. Xu and G. E. Scuseria, Phys. Rev. Lett. **72**, 669 (1994).

[37] P. Marcos, M. López, A. Rubio, and J. Alonso, Chem. Phys. Lett. **273**, 367 (1997).

[38] S. Maruyama and Y. Yamaguchi, Chem. Phys. Lett. **286**, 343 (1998).

[39] E. W. Dijkstra, Numer. Math. **1**, 269 (1959).

[40] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).

[41] A. F. Voter, in *Radiation Effects in Solids*, edited by K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga (Springer, Netherlands, Dordrecht, 2007), pp. 1–23.

[42] W. Krätschmer, L. D. Lamb, K. Fostiropoulos, and D. R. Huffman, Nature (London) **347**, 354 (1990).

[43] H. Prinzbach, F. Wahl, A. Weiler, P. Landenberger, J. Wrth, L. T. Scott, M. Gelmont, D. Olevano, F. Sommer, and B. von Issendorff, Chem. Eur. J. **12**, 6268 (2006).

[44] W.-D. Fessner, B. A. Murty, and H. Prinzbach, Angew. Chem., Int. Ed. Engl. **26**, 451 (1987).

[45] G. S. Prakash, V. Krishnamurthy, R. Herges, R. Bau, H. Yuan, G. A. Olah, W. D. Fessner, and H. Prinzbach, J. Am. Chem. Soc. **110**, 7764 (1988).

[46] W.-D. Fessner, B. A. Murty, J. Wörth, D. Hunkler, H. Fritz, H. Prinzbach, W. D. Roth, P. V. R. Schleyer, A. B. McEwen, and W. F. Maier, Angew. Chem., Int. Ed. Engl. **26**, 452 (1987).