Graph Theory Meets *Ab Initio* Molecular Dynamics: Atomic Structures and Transformations at the Nanoscale

Fabio Pietrucci[1,*] and Wanda Andreoni[1,2]

[1]*Centre Européen de Calcul Atomique et Moléculaire (CECAM), Ecole Polytechnique Fédérale de Lausanne, Switzerland*
[2]*Institute of Theoretical Physics, Ecole Polytechnique Fédérale de Lausanne, Switzerland*

(Received 30 May 2011; published 17 August 2011)

Social permutation invariant coordinates are introduced describing the bond network around a given atom. They originate from the largest eigenvalue and the corresponding eigenvector of the contact matrix, are invariant under permutation of identical atoms, and bear a clear signature of an order-disorder transition. Once combined with *ab initio* metadynamics, these coordinates are shown to be a powerful tool for the discovery of low-energy isomers of molecules and nanoclusters as well as for a blind exploration of isomerization, association, and dissociation reactions.

DOI: 10.1103/PhysRevLett.107.085504 PACS numbers: 61.46.Bc, 02.10.Ox, 82.30.−b

Being able to predict the structure of atomic aggregates, from molecules to crystals, is an open (long-standing) challenge for theoretical and computational methods (see, e.g., [1–3]). For nanoscale clusters, this problem is particularly important because their physical properties strongly depend on their structure. Experiments have severe difficulties in providing structural information [4]. Therefore, simulations become crucial in exploring the thermodynamically relevant configurations of atomic systems at the nanoscale. Related to the ''structural problem'' of clusters is that of foreseeing the distinct heteroatomic molecules sharing a given formula unit. Moreover, understanding and predicting isomerization, association, and dissociation reactions are common and crucial issues in the physics and chemistry of both nanoclusters and complex molecules. Computer experiments stand out also in this case as the ideal tool to gain insight. However, a unified and reliable computational approach to the above questions is still lacking.

Numerous computational strategies have been applied to determine low-energy geometries of atomic clusters [2], including simulated annealing [5], genetic algorithms [6], random search [7], basin hopping [8], and minima hopping [9] methods. Studies based on classical potentials have generated large databases; however, exploration of the potential energy surface (PES) with the above techniques is too expensive for nonempirical methods. A first screening is generally made within either classical potential (see, e.g., [10]) or tight-binding schemes (see, e.g., [11]), and only later a small number of candidate structures are selected and optimized at the density functional theory level. In this way, whenever they are not minima of the empirical PES, one risks missing relevant low-energy structures. Advanced first-principles searches for isomers of small molecules are generally made on the basis of static calculations, by using either the ''symmetry adapted stochastic search'' [12] or the ''scaled hypersphere search method'' [13], which also provides interconversion pathways. The former becomes rapidly impractical for low-symmetry cases and an increasing number of atoms ($\gtrsim 10$); the latter is intrinsically restricted to an initial harmonic approximation for the PES.

A more general, efficient, and yet accurate methodology is needed: It should allow the exploration of a large portion of the free-energy surface (not just of the PES) and should not only lead to thermodynamically relevant isomers but also simultaneously disclose the dynamics of structural transformations and chemical reactions. In this Letter, we face this challenge by first developing special structure descriptors (topological coordinates) from spectral graph theory and then combining them with *ab initio* molecular dynamics and accelerated sampling such as metadynamics (MTD) [14]. A special virtue of our coordinates is that they are invariant under permutations of identical atoms, thus avoiding the explosion of equivalent isomers that is typically encountered in other methods. We call them social permutation invariant (SPRINT) coordinates. After introducing their definition and main features, we demonstrate their potential via simulations of Lennard-Jones aggregates and especially silicon clusters and heteroatomic molecules having the formula unit C_4H_5N.

Following spectral graph theory [15], we consider a graph (atomic cluster) and its adjacency (contact) matrix a_{ij} where the ij are all pairs of vertices (atoms): $a_{ij} = 1$ if i and j are connected [there is an edge (bond) between them] and 0 otherwise. a_{ij} is symmetric, non-negative, and also irreducible when it represents a connected graph, i.e., if any pair of vertices is connected through a path. In this case the Perron-Frobenius theorem holds: The largest modulus eigenvalue $\lambda^{\max}$ is real, positive, and nondegenerate, and the corresponding eigenvector $v_i^{\max}$ has all nonzero components with equal sign. We adopt the positive sign convention. In particular, a few very interesting properties can be shown: (i) $\lambda^{\max}$ carries global information on the network: It grows with the number of bonds and lies between the average and the maximum coordination number (CN);

(ii) $v_i^{\max}$ carries information about both the short- and long-range topology of the atomic network surrounding atom i: For any positive integer M,

$$v_i^{\max} = \frac{1}{(\lambda^{\max})^M} \sum_j a_{ij}^M v_j^{\max}, \quad (1)$$

where a_{ij}^M is the number of walks of length M connecting i and j. Equation (1) shows the "social character" of $v_i^{\max}$ [16–18].

These observations lead us to combine the largest eigenvalue and corresponding eigenvector into the definition of topological (SPRINT) coordinates:

$$S_i = \sqrt{N}\lambda^{\max} v_i^{\max,\text{sorted}}; \qquad i = 1, 2, \ldots, N, \quad (2)$$

where N is the number of atoms and the ith component must be taken after sorting the eigenvector from its smallest to its largest component. It is this sorting operation that makes the set $\{S_i\}$ invariant with respect to the $N!$ permutations of the labeling of N identical atoms (and thus also with respect to point-group symmetries). By systematically avoiding the permutation degeneracy, the SPRINT coordinates are a major improvement over more traditional ones (see, e.g., [19]). An additional advantage is the dimensional reduction from $N(N-1)/2$ elements of the contact matrix to N elements only. Moreover, the constraints imposed on the S_i's by the sorting operation, i.e., $S_1 \leq S_2 \leq \cdots \leq S_N$, strongly reduce their accessible space. The simple example in Fig. 1 illustrates how the SPRINT coordinates work, namely, how they single out the topologically inequivalent atoms and relate to both the CN and the longer-range topology.

Based on the above considerations, we propose to use the S_i's in computer simulations both for postprocessing (or on-the-fly) classification of the explored structures and as reaction coordinates. Here we show their power in combination with MTD. To this purpose, it is convenient to generalize the contact matrix a_{ij} to a smooth and differentiable function of the interatomic distance r_{ij}, still preserving the validity of the Perron-Frobenius theorem. We define

$$a_{ij} = \frac{1-(r_{ij}/r_0)^n}{1-(r_{ij}/r_0)^m}, \quad (3)$$

where r_0, n, and m depend on the typical bond lengths in the cluster [20]. Because of the gradual decay of the function in Eq. (3), the associated S_i's contain information not only on the cluster topology but also on the 3D geometry.

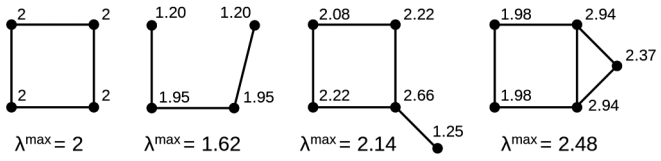

FIG. 1. Examples of simple graphs with each vertex labeled by the topological coordinate S_i in Eq. (2).

Figure 2 shows how the SPRINT coordinates perform as collective variables of MTD simulations, when exploring the structural pattern of Lennard-Jones (LJ) clusters with 13, 38, and 55 atoms [20]. Starting from very improbable configurations (fragments of square or cubic lattices), it takes MTD only ~100 to ~1000 time units to visit the well-known lowest-energy isomers, i.e., icosahedra for $N = 13$ and 55 and a face-centered truncated octahedron for $N = 38$ [Fig. 2(b)]. A crowd of isomers quickly emerges from the MTD trajectory. The spread of the S_i values allows one easily to recognize highly symmetric from low-symmetry structures [see Fig. 2(a)]; more generally, during the simulation, collapse from a wide to a narrow range marks a disorder-order transition.

We next turn to the use of the SPRINT coordinates for the challenging systems and processes for which they were mainly conceived, namely, those requiring an *ab initio* description. First, we tackle the problem of characterizing the thermodynamically relevant configurations of a cluster. We consider silicon clusters that were often taken as a test case for density functional theory-based algorithms designed to "search for the global minimum." Our investigation uses Car-Parrinello [21] molecular dynamics in the local density approximation [20] and does not require simplified interaction models at any step of the simulation. Our specific aim is to demonstrate that the S_i's allow for a fast and efficient exploration of a multitude of low-energy isomers. MTD simulations were run at room temperature, starting from fragments of a simple cubic lattice. In the case of Si_{10}, this method easily yielded the structure

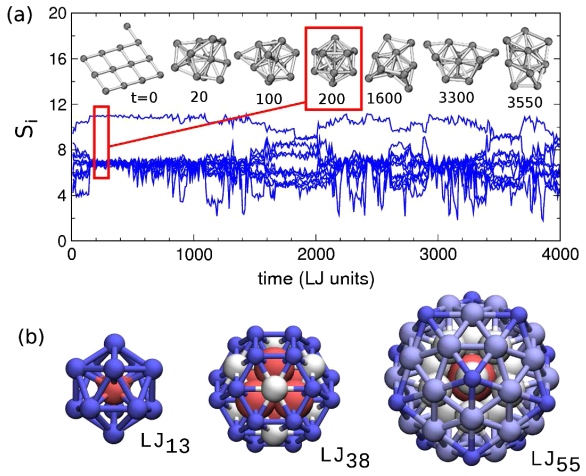

FIG. 2 (color online). MTD of Lennard-Jones clusters at $T = 0.1$ (LJ units). (a) LJ_{13}: time evolution of the SPRINT coordinates. Some of the explored isomers are shown. (b) Lowest-energy minima of LJ_{13}, LJ_{38}, and LJ_{55}. Colors vary from red to white to blue with decreasing values of S_i and highlight the equivalent atoms.

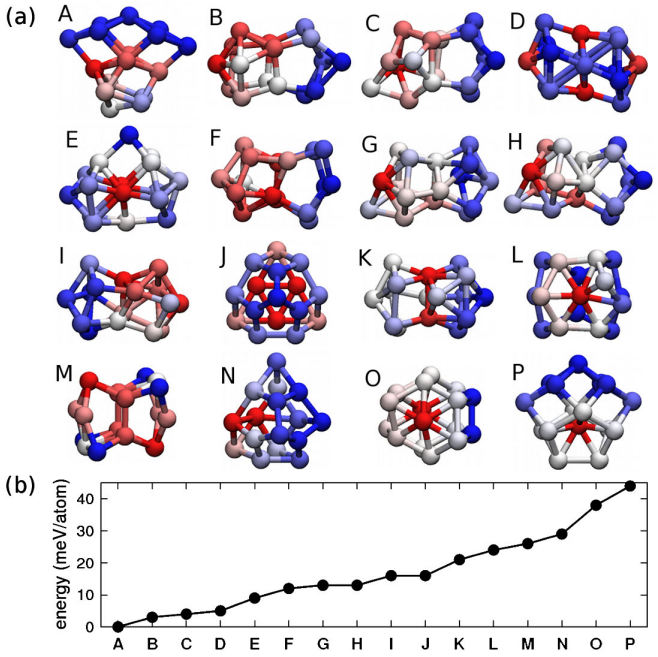

FIG. 3 (color online). Si_{16}: Optimized geometries and relative energies per atom. Colors vary as in Fig. 2.

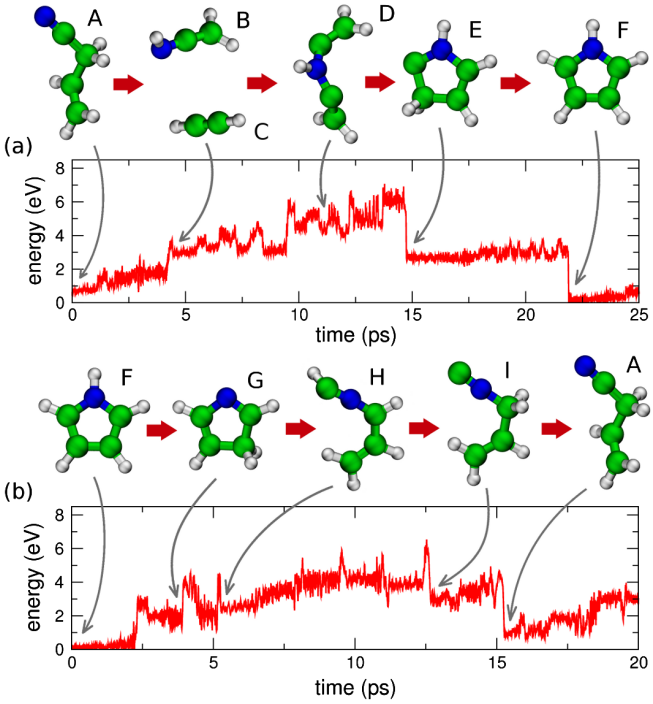

FIG. 4 (color online). C_4H_5N: Two independent *ab initio* MTD simulations driven by the SPRINT coordinates, starting from (a) allyl cyanide and (b) pyrrole.

(tetracapped trigonal prism) that finds consensus as the most thermodynamically favored [5,6,22]. Increasing the cluster size, a rich variety of low-energy structures was observed. For example, in the case of Si_{16}, five independent simulations generated hundreds of different geometries over a cumulative time of 400 ps. Out of them, a set of 80 was selected and their atomic positions were optimized: Sixteen were found within only 45 meV/atom [23] as shown in Fig. 3. Our result for the lowest-energy minimum agrees with previous studies [11,24]. However, the latter identified only very few isomers: This is a common outcome of *ab initio* calculations and risks giving a misleading view of silicon nanoclusters. Indeed, our findings show, on the contrary, that already at this small size (7–9 Å) an impressive diversity of structural motifs lies within a narrow energy range, from fused units (e.g., *B*, *C*, *D*, and *M*) to capped (deformed) cores like the tri- or tetracapped trigonal prism (*A*, *E*, *F*, *G*, *I*, and *K*), from symmetric configurations to quasiamorphous. All geometries (Fig. 3) are "inhomogeneous"; namely, "highly social" (high-connected) atoms (red) coexist with "asocial" (low-connected) atoms (blue) [20]. By differentiating the atomic environment over the whole extension of the cluster and easily identifying topologically equivalent positions, the SPRINT coordinates qualify as precise structural fingerprints.

As a second demonstration of the power of our scheme, we considered a heteroatomic system of given composition and attempted a blind exploration of the possible reactions it might undergo at a given temperature, namely, without relying on any information about either mechanisms or products. We chose the formula C_4H_5N, used a density functional theory description with a generalized gradient approximation functional, and applied room temperature Car-Parrinello MTD starting from a few different molecules [20]. Because of the presence of different elements, sorting of the principal eigenvector in Eq. (2) was made within sets of alike atoms.

In Fig. 4, we present part of two of our computer experiments. One [in Fig. 4(a)] started from allyl cyanide (*A*). A complex and rich reaction pathway is observed already in the first 30 ps, with free-energy barriers of the order of the eV. (*A*) dissociates into ethenimine (*B*) and acetylene (*C*) through the simultaneous breaking of the $NCCH_2$—$CHCH_2$ bond and hydrogen transfer from the tail carbon to nitrogen. Reassociation readily follows and leads to pyrrole (*F*) passing through two intermediates: a linear (*D*) and a cyclic (*E*) carbene. Note that the SPRINT coordinates preserve their validity (the Perron-Frobenius theorem continues to hold) even when the system dissociates into two or more fragments, because the long tail of the function in Eq. (3) maintains a weak connection in the contact matrix. The other simulation [in Fig. 4(b)] initiated from pyrrole (*F*). The ring opening leading to linear carbene (*H*) proceeds via isomerization to 3H-pyrrole (*G*). The latter is driven by direct hydrogen transfer. The $F \rightarrow G \rightarrow H$ pathway represents the same reaction mechanism proposed in Ref. [26] on the basis of laborious static quantum-chemistry calculations. Linear carbene (*H*) converts then to allyl isocyanide (*I*), which easily transforms to allyl cyanide (*A*) by a 180° rotation of the CN group.

Evolution over an additional 200 ps revealed many other chemical species, both C_4H_5N isomers (e.g., methacrylonitrile, cyanocyclopropane, and isocyanocyclopropane) and dissociation products (e.g., methylacetylene, cyanomethane and isocyanomethane, acrylonitrile, propadiene, cyclopropene, allyl radical, ethene, methane, hydrogen cyanide, and molecular hydrogen). This scenario is in agreement with experimental evidence of the abundance of products of C_4H_5N thermal pyrolysis [27].

A few comments can be added on possible extensions of the scheme here proposed: (a) In pathological cases, e.g., for structures having all atoms with identical CN, one could include other eigenvectors of the contact matrix beside the one corresponding to the maximum eigenvalue [28]. (b) The use of the SPRINT coordinates is by no means restricted to MTD. They can be analogously combined with the Monte Carlo method and also with any other accelerated sampling or global optimization procedure. Note that we implemented these collective variables in the freely available plug-in PLUMED [29].

In conclusion, we remark that spectral graph theory is at the heart of many important and diverse applications: just to name some, the tremendously successful page-rank algorithm of Google [17], the centrality concept in social networks [16], and the compact description of protein structures [18]. By introducing the SPRINT topological descriptors and proving their validity and unique advantages as reaction coordinates, our work pioneers the application of spectral graph theory to the broad area of dynamic atomistic simulations and, in particular, to *ab initio* approaches. Here we have focused on isolated molecules and nanoclusters. However, it is straightforward to extend this scheme to condensed phases; applications are currently ongoing to phase transitions in solids and chemical reactions in solution.

The authors are indebted to Alessandro Laio for insightful discussions and Alessandro De Vita for a critical reading of the manuscript.

*fabio.pietrucci@epfl.ch

[1] J. Pannetier *et al.*, Nature (London) **346**, 343 (1990).
[2] S. Woodley and R. Catlow, Nature Mater. **7**, 937 (2008).
[3] A. R. Oganov, A. O. Lyakhov, and M. Valle, Acc. Chem. Res. **44**, 227 (2011).
[4] S. J. L. Billinge and I. Levin, Science **316**, 561 (2007).
[5] P. Ballone *et al.*, Phys. Rev. Lett. **60**, 271 (1988); N. Binggeli, J. L. Martins, and J. R. Chelikowsky, *ibid.* **68**, 2956 (1992); U. Rothlisberger, W. Andreoni, and M. Parrinello, *ibid.* **72**, 665 (1994); U. Rothlisberger and W. Andreoni, J. Chem. Phys. **94**, 8129 (1991).
[6] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995); K. Ho *et al.*, Nature (London) **392**, 582 (1998).
[7] L. D. Lloyd and R. L. Johnston, Chem. Phys. **236**, 107 (1998); M. Saunders, J. Comput. Chem. **25**, 621 (2004).
[8] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).
[9] S. Goedecker, J. Chem. Phys. **120**, 9911 (2004); S. R. S. Schoenborn, S. Goedecker, and A. Oganov, *ibid.* **130**, 144108 (2009).
[10] G. Rossi *et al.*, Phys. Rev. Lett. **93**, 105503 (2004).
[11] S. Goedecker, W. Hellmann, and T. Lenosky, Phys. Rev. Lett. **95**, 055501 (2005).
[12] S. E. Wheeler, P. V. R. Schleyer, and H. F. Schaefer, J. Chem. Phys. **126**, 104104 (2007).
[13] K. Ohno and S. Maeda, Chem. Phys. Lett. **384**, 277 (2004).
[14] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 562 (2002).
[15] B. Bollobas, *Modern Graph Theory* (Springer-Verlag, Berlin, 1998).
[16] P. Bonacich, Am. J. Sociology **92**, 1170 (1987).
[17] K. Bryan and T. Leise, SIAM Rev. **48**, 569 (2006).
[18] M. Porto *et al.*, Phys. Rev. Lett. **92**, 218101 (2004).
[19] M. Randić, J. Chem. Inf. Comput. Sci. **15**, 105 (1975); A. Balaban, D. Ciubotariu, and M. Medeleanu, J. Chem. Inf. Comput. Sci. **31**, 517 (1991); P. P. Bera *et al.*, J. Phys. Chem. A **110**, 4287 (2006); J. M. Vasquez-Perez *et al.*, J. Chem. Phys. **131**, 124126 (2009).
[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.107.085504 for computational details.
[21] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[22] K. Raghavachari and C. M. Rohlfing, J. Chem. Phys. **89**, 2219 (1988); N. Binggeli and J. R. Chelikowsky, Phys. Rev. B **50**, 11 764 (1994); J. T. Lyon *et al.*, J. Am. Chem. Soc. **131**, 1115 (2009).
[23] The detailed ranking will depend on the exchange-correlation functional (see, e.g., [24,25]); a critical assessment may require comparison of different levels of theory, which is outside the scope of this Letter.
[24] S. Yoo and X. C. Zeng, J. Chem. Phys. **123**, 164303 (2005).
[25] X. Zhu *et al.*, J. Chem. Phys. **120**, 8985 (2004).
[26] M. Martoprawiro, G. B. Bacskay, and J. C. Mackie, J. Phys. Chem. A **103**, 3923 (1999).
[27] J. C. Mackie *et al.*, Int. J. Chem. Kinet. **23**, 733 (1991); A. Doughty and J. C. Mackie, J. Phys. Chem. **96**, 272 (1992).
[28] U. Bastolla *et al.*, Proteins **73**, 872 (2008).
[29] M. Bonomi *et al.*, Comput. Phys. Commun. **180**, 1961 (2009).