Research Article **Open Access**

José A. Flores-Livas, Antonio Sanna, and Stefan Goedecker

Accelerated materials design approaches based on structural classification: application to low enthalpy high pressure phases of SH_3 and SeH_3

DOI 10.1515/nsm-2017-0002
Received November 5, 2016; accepted January 18, 2017

Abstract: We propose a methodology that efficiently asseses major characteristics in the energy landscape for a given space of configurations (crystal structures) under pressure. In this work we study SH_3 and SeH_3 , both of fundamental interest due to their superconducting properties. Starting from the crystal fingerprint, which defines configurational distances between crystalline structures, we introduce an optimal one dimensional metric space that is used to both classify and characterize the structures. Furthermore, this is correlated to the electronic structure. Our analysis highlights the uniqueness of the $Im-3m$ phase of H_3S and $H3_3Se$ for superconductivity. This approach is an useful tool for future material design applications.

The theoretical prediction by Duan et al. [1] and the successful experimental discovery by Drozdov et al. [2] of superconductivity in H_3S at 200 GPa, with the record critical temperature (T_C) of 203 K, rekindled the century-old dream of a room temperature superconductor. A plethora of hydrogen rich materials have been proposed in the quest for novel high-T_C superconductors, as the presence of hydrogen seems to be fundamental to reach the very high phonon frequencies, strong electron-phonon coupling, and therefore high critical temperatures [3–6]. A major emphasis has been given to the class of hydrides [7–19] such as silane [8, 20–23], disilane [24], sulfur hydride [25–33], sodium hydride[34], to name a few.

It is clear that in this research field a key role is played by materials design algorithms [35, 36] both providing predictions [1, 37, 38], prior to experiments, and helping with their interpretation afterwards [39–44]. Several of these computational synthesis tools exist: evolutionary based algorithms [45], particle swarm optimization [46], random sampling methods [47], and hopping based algorithms [48], to name a few.

What all these tools have in common is the use density functional theory (DFT) as the workhorse to evaluate energy and forces due to its reasonable accuracy for a relatively moderate computational cost and overall reproducibility [49]. Within the global optimization codes, what the routines usually do is to rank structures according to their total energy. One usually expects that the absolute enthalpy minimum is the one experimentally realized and many significant predictions have been made in this way. However, due to the limited intrinsic precision of DFT functionals, to the accuracy limit of the computational methods, as well as the non ideal experimental conditions (for instance under high pressure non hydrostaticity can play an important role in the synthesis), all the low enthalpy structures may be experimentally relevant as in the case discussed in Ref. [39].

In this work we consider hydrogen sulfide and hydrogen selenide systems, also studied in Ref. [37], as test case to present a method for the characterization of structures based on fingerprint classification of energetically competitive structures, i.e. in the vicinity of the global minimum structure. We will focus on some key aspects of this classification, in particular on the problem of structural similarity, and the relation with electronic properties. Besides the possible experimental relevance of the set of these low enthalpy structures, as mentioned before, this approach is also useful post-analysis to help in the characterization of the configurational energy landscape for finite systems, and could be used to improve crystal prediction algorithms.

We start from the calculation of the energetically competitive structures in H_3S and H_3Se at high pressure found by the crystal prediction minima hopping method [48] at the level of DFT (sec. 1). We then analyze the predicted structures by means of the structural fingerprint method as introduced by Zhu *et al* [50, 51] (sec. 2). Based on this structural fingerprint, we define an optimal one dimen-

José A. Flores-Livas, Stefan Goedecker: Department of Physics, Universität Basel, Klingelbergstr. 82, 4056 Basel, Switzerland
Antonio Sanna: Max Planck Istitute of Microstructure Physics, Weinberg 2, 06120 Halle, Germany

sional classification (sec. 3.1) that further is correlated to a more general trend, the electronic density of states (DOS).

1 Computational Details

Structural data was generated by using *ab initio* crystal structure prediction, minima hopping method (MHM) [52]. Given only the chemical composition of a system, the MHM aims at finding the global minimum on the enthalpy surface while gradually exploring low-lying structures. This method has been successfully used for global geometry optimization in a large variety of applications [53–57]. Previously we have explored selenium and sulfur hydride under pressure with MHM [37]. We have restricted the study to considered only simulations cells with single formula unit of Se and S -H_3 at different pressures between 50 and 200 GPa. Once identified the global minimum, we allow the algorithm to visit even higher energy structures up to 2 eV. Hundred's of visited structures were relaxed with tight convergence at the density-functional theory level using the Perdew-Burke-Ernzerhof exchange-correlation function [58] and plane-wave basis set with of 900 eV was used to expand the wave function together with the projector augmented wave (PAW) method as implemented in the Vienna ab initio simulation package VASP [59]. Monkhorst-Pack grid meshes with spacing of $2\pi\times0.15$Å were used, resulting in total energy convergence to better than 2 meV/atom and forces on the atoms were less than 2 meV/Å and the resulting stress less than 1 kB.

2 The Crystal Fingerprint method

In order to characterize a crystal structure in condensed matter physics it is most convenient to adopt a Bravais matrix and base representation, that allows for a clear description of symmetry properties and therefore is optimal for computational purposes. However, while this description is excellent to characterize a single system, the description becomes inefficient when comparing hundreds of different crystals closely related, simply because comparisons are not straightforward due to the non-unique representation; the same crystal may be described with an infinite number of different, but equivalent ways. This is the typical situation when working with structures generated by crystal structure prediction methods. One therefore has to introduce alternative descriptions that allow to compare structures on an unambiguous ground. The initial aim of the crystal fingerprints was to solve a very specific problem, to classify and remove duplicate crystal structures from the results generated by crystal prediction methods. Early work by Oganov-Valle in fingerprints were developed based on local atom distances and analyzed in diffraction spectra type plots [60]. Another interesting approach to this problem was proposed by Ceriotti et al. [61], in which they analyzed the pairwise distribution distances between frames (molecules) and reduced the high dimensional free-energy surface to a sketch map, see for instance Ref. [62]. Another successful crystal representation inspired by radial distribution functions as used in the physics of X-ray powder diffraction, was introduced by Glawe et al. [63] and used for machine learning algorithms. In this work we will use the *configurational fingerprint approach*, and correlate it to the electronic structure of high temperature superconducting materials, thus we briefly review in this section [50, 51].

For each atom k in a crystal located at $\mathbf{R}_k$, we obtain a cluster of atoms by considering only those contained in a sphere centered at $\mathbf{R}_k$. For this cluster, one calculates the overlap matrix elements $S^k_{i,j}$ as described in Ref. [51] for a non-periodic system, i.e., we construct on each atom one or several Gaussian type orbitals and calculate the resulting overlap integral. The orbitals are indexed by the letters i and j and the index $w(i)$ gives the index of the atom on which the Gaussian $G_i(\mathbf{r})$ is centered, i.e.,

$$S^k_{i,j} = \int d\mathbf{r} G_i(\mathbf{r} - \mathbf{R}_{w(i)}) G_i(\mathbf{r}-\mathbf{R}_{w(j)}) \tag{1}$$

In this first step, the amplitudes of the Gaussians c_{norm} are chosen such that the Gaussians are normalized to one, and the width of each Gaussian $G_i(\mathbf{r})$ is given by the covalent radius of the atom $w(i)$ on which it is centered. To avoid discontinuities in the eigenvalues when an atom enters into or leaves the sphere, a second step constructs a matrix T^k such that

$$T^k_{i,j} = f_c(|\mathbf{R}_{w(i)} - \mathbf{R}_k|) S^k_{i,j} f_c(|\mathbf{R}_{w(j)} - \mathbf{R}_k|). \tag{2}$$

The cutoff function f_c smoothly goes to zero on the surface of the sphere with radius $\sqrt{2n\sigma_c}$,

$$f_c = \left(1 - \frac{r^2}{2n\sigma_c^2}\right)^n. \tag{3}$$

In the limit where n tends to infinity, the cutoff function converges to a Gaussian of width σ_c. This characteristic length scale is typically chosen to be the sum of the two largest covalent radii in the system. The value n determines how many derivatives of the cutoff function are continuous on the surface of the sphere (in our case $n = 3$ was used). One can consider the modified matrix T^k to be the overlap matrix of the cluster where the amplitude of

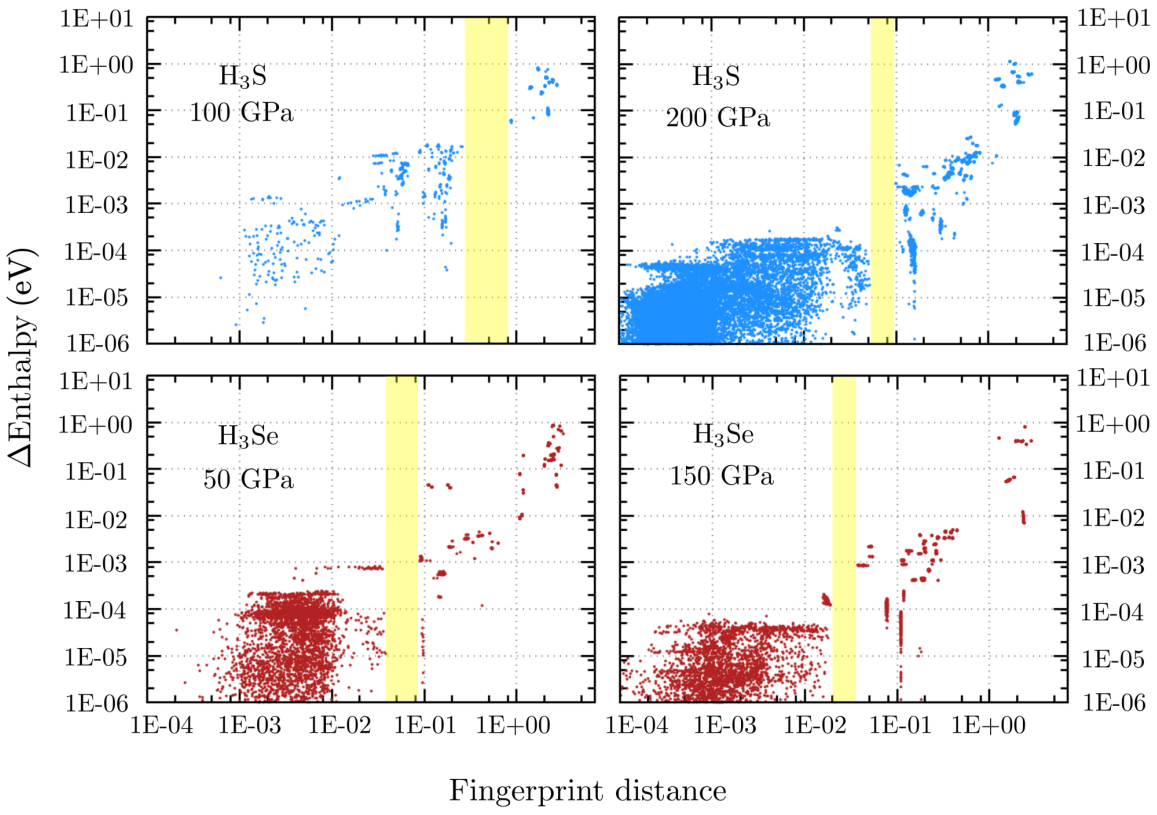

Figure 1: (Top panels) show the correlation between the enthalpy difference (eV/unit cell) and the fingerprint distance (arbitrary units) for H_3S at 100 GPa and 200 GPa and (Bottom panels) for H_3Se at 50 GPa and 150 GPa. Each point in the plot correlates energy difference and fingerprint distance between two structures. For each set of structures, the enthalpy difference is with respect to the global minimum. Based on energetics, different structures are distinguishable above 10 meV (1E-02 eV/u.c.); below this threshold the structures are consider identical. Indicated in yellow areas is the fingerprint "gap", in which the function differentiates between structures.

the Gaussian at atom i is determined by $c_{\text{norm}} f_c(|\mathbf{R}_i - \mathbf{R}_k|)$. In this way atoms close to the surface of the sphere give rise to very small eigenvalues of T^k and are weighted less than the atoms closer to the center.

The eigenvalues of this matrix T^k are sorted in descending order and form the atomic fingerprint vector $\mathbf{V}_k$. [1] The Euclidean norm $|\mathbf{V}_k - \mathbf{V}_l|$ measures the dissimilarity between the atomic environments of atoms k and l.

The atomic fingerprints $\mathbf{V}_k^p$ and $\mathbf{V}_k^q$ of all the N_{at} atoms in two crystalline configurations p and q can now be used to define a *fingerprint distance* (FD) $d(p, q)$ between two crystals,

$$d(p,q) = \min_{\mathbf{P}} \left(\sum_{k}^{N_{at}} |\mathbf{V}_k^p - \mathbf{V}_{\mathrm{P}(k)}^q|^2 \right)^{1/2} \tag{4}$$

where P is a permutation function which matches a certain atom k in crystal p with atom $P(k)$ in crystal q. The optimal permutation function which minimizes $d(p, q)$ can be found with the Hungarian algorithm [64] in polynomial time. If the two crystals p and q are identical, the Hungarian algorithm will in this way assign corresponding atoms to each other. The Hungarian algorithm needs as its input only the cost matrix C given by

$$C_{k,l} = |\mathbf{V}_k^p - \mathbf{V}_l^q|^2. \tag{5}$$

A key concept is that $d(p, q)$ satisfies the properties of a metric, positiveness, symmetry, coincidence axiom and

1 Since one cannot predict exactly how many atoms will be in the sphere, we estimate a maximum length for the atomic fingerprint vector. If the number of atoms is too small to generate enough eigenvalues to fill up the entire vector, the entries at the end of the fingerprint vector are filled up with zeros. This also guarantees that the fingerprint is a continuous function with respect to the motion of the atoms when atoms might enter or leave the sphere. If an atom enters into the sphere, some zeros towards the end of the fingerprint vector are transformed in a continuous way into some very small entries which only contribute little to the overall fingerprint.

the triangle inequality, and have been demonstrated in the work by Zhu *et al.*[50]. We restrict this study only to the use of $\mathbf{R}_k$ typical s-type Gaussian orbitals with centered spheres (with about 80-120 atoms).

3 Results

In Fig 1 we plot the FD versus enthalpy difference. For the low pressure case of H_3S we consider 63 structures, for the high pressure (200 GPa) 423 structures. H_3Se contains 227 for the low pressure and 241 structures for the high pressure. Each point in these plot correlates energy difference and fingerprint distance between two structures. For each set of pressure, the energy difference is with respect the global minimum. An ideal fingerprint produces, in these type of plots, a clear gap along the FD axis, clearly separating structures that are classified as identical from those that are different. In both materials, independently of pressure there is a clear distinguishable gap. After visual inspection of the structures, we observe that many of the structures are fairly similar and hardly distinguishable by eye, many of them only having small distortions of hydrogen and different vector cell description. Based on energetics, and after visual inspection, different structures are distinguishable above 10 meV (1E-02 eV/u.c.), below this threshold the structure is consider identical. Fingerprint distances (FD) > 0.1 can differentiate between structures for the case of SH_3 at low pressure. We confirm that at low pressure, both systems still have high molecular nature, this also means that many energetically degenerated structures are possible withing little change in their energies (we will come back to this point in the next section). However, at high pressure the situation changes and in both systems the fingerprint distance is shifted toward lower FD values. For instance in H_3S at 200 GPa FD is 4E-02, while for selenium-hydride we found the gap for > 2E-02 FD values. We found that there is a clear correlation in both systems between pressure and fingerprint distance, i.e. the gap is shifted towards lower FD values as increasing pressure. This is due to the changes in the potential energy landscape; as increasing pressure in both systems there are much less structures with small variations in the hydrogen positions.

3.1 Optimal 1D classification of crystal structures

As seen in Sec. 2, the structural fingerprint maps the structures into a metric space. This allows for further analysis but still lacks the possibility for a simple (visual) representation since this space still is a high dimensional one. For a simply representation it would be best to map it into an Euclidean space of low dimensionality (degree two or three). This means to associate, to each element in the set (crystal), coordinates in an Euclidean space in such a way that the distance is as close as possible to the fingerprint distance. Clearly the higher the dimension, the more accurate would be the mapping, however at the cost of being less intuitive the graphical representation.

Since the case of the (low enthalpy) SH_3 / SeH_3 structures is relatively simple (and we known the global minimum), we can consider for this work the case of a one dimensional representation, for that we introduce a rather simple mapping algorithm and yet easily expandable to higher dimensions: we associate to each crystal in the set a random position in the one dimensional space, such a way the length scale of the distribution is of the order of the average fingerprint distance. Then we define the following force field:

$$F(i,j) = \mathrm{sgn}(x_i - x_j)||(x_i - x_j)| - d(i,j)| \tag{6}$$

for all pairs of systems i, and j, where x_i and x_j are the position in the 1D axis. And we run dynamics. Since many local minima exist, the operation is repeated for different starting order (randomized) to minimize the quantity

$$S = \sum_{i,j} ||(x_i - x_j)| - d(i,j)|. \tag{7}$$

This stochastic approach may require many iterations to reach the optimal configuration especially in multidimensional spaces and when large number of systems are involved. In the systems considered in this work and for the number of structures, the algorithm finds a minimum within a few thousand iterations (a couple of minutes of computing time). Other alternative to use instead, while dealing for larger data-sets, is to resort genetic algorithms. This have been proved to efficiently solve similar mathematical problems, as in Ref. [65].

The resulting 1D optimization is shown in the bottom panels of Fig. 2 and Fig. 3, respectively for SH_3 and SeH_3. First we can distinguish that structures are more clustered at high pressure than at low pressure, this indicates that at high pressure the crystal data-set covers all the well defined minima, while at lower pressure a larger number

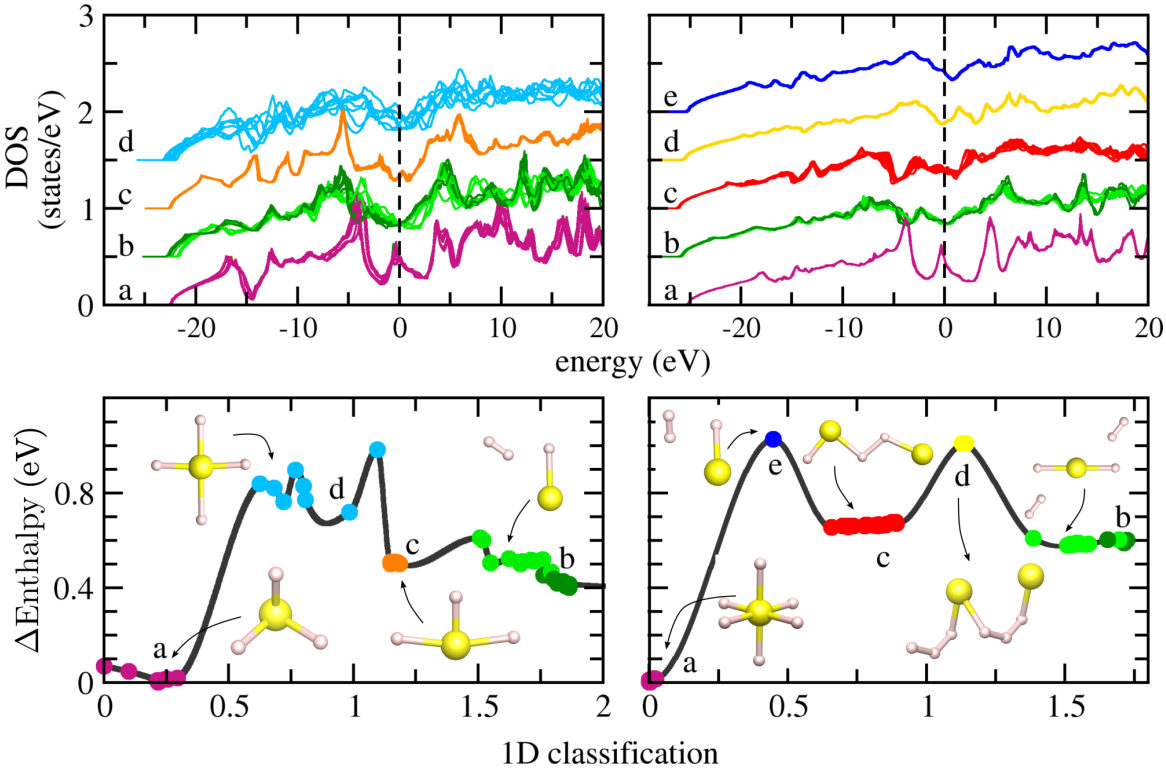

Figure 2: (Color online) Density of states (top panels) and one dimensional classification (bottom –see text) of SH_3 structures generated by the minima hopping algorithm, at 100 GPa of pressure (left) and at 200 GPa of pressure (right). The color code is used to match a structure group with its corresponding density of states (and alphabetically labeled). The solid black lines that connect the points (bottom panels) serve only as guide to the eye. In the structures, hydrogen is represented in white and sulfur in yellow. For each family we shown a representative structure for the data set, depicting variations in structure motif, coordination number and bond type.

of different structures are found. These can be local minima but also saddle points, since the only criterion for their convergence is a threshold on the inter-atomic forces. Note also that all the structures (including the lowest energy one) could be unstable towards phonon displacement with finite **q**. The full dynamic stability has to always be verified (independently) to guarantee that a given prediction is physically relevant. At lower pressures (particularly evident for SH_3 at 100 GPa) we observe a large spread of structures that are distributed in a more homogeneous way on the 1D axis. This indicates that minima in the potential energy surface are very shallow, indeed of a highly molecular nature. We can also see that whenever the structure contains internal blocks of detached H_2 molecules, the more the FD, and consequently the 1D classification, is spread out, regardless of their energetics, see for example set b and c for SH_3 at 200 GPa.

One of the most interesting features that result from this type of analysis is the identification of low-enthalpy structures, only 1D distance separated from the global minimum. In short this points to the existence of energetically plausible structures that are structure-wise very different from the ground state and therefore could in principle be accessible to synthesis because they belong to a different funnel. A funnel is the basin of attraction where the global minimum or local minima lives, different funnels are typically separated by high energy barriers and feature distinct structures. However, there are no flagrant cases that point to this situation in these systems, apart from the structure group of SeH_3 at 50 GPa labeled c and b (see different motifs and similar energetics). Structure b is the pre-cubic structure ($Im-3m$) that becomes stable at high pressure and is superconducting. Although the 1D structure distance from c to b is not large, they are separated enough to be classified as two different funnels, and under further compression both, will evolve to the cubic structure (main attractor funnel at high pressure). We can also point out that the highly metallic cubic structure is always localized by the classification at the (left) edge of the set (violet circles in all classification plots) and in all cases is well separated from the rest of the structures, indicat-

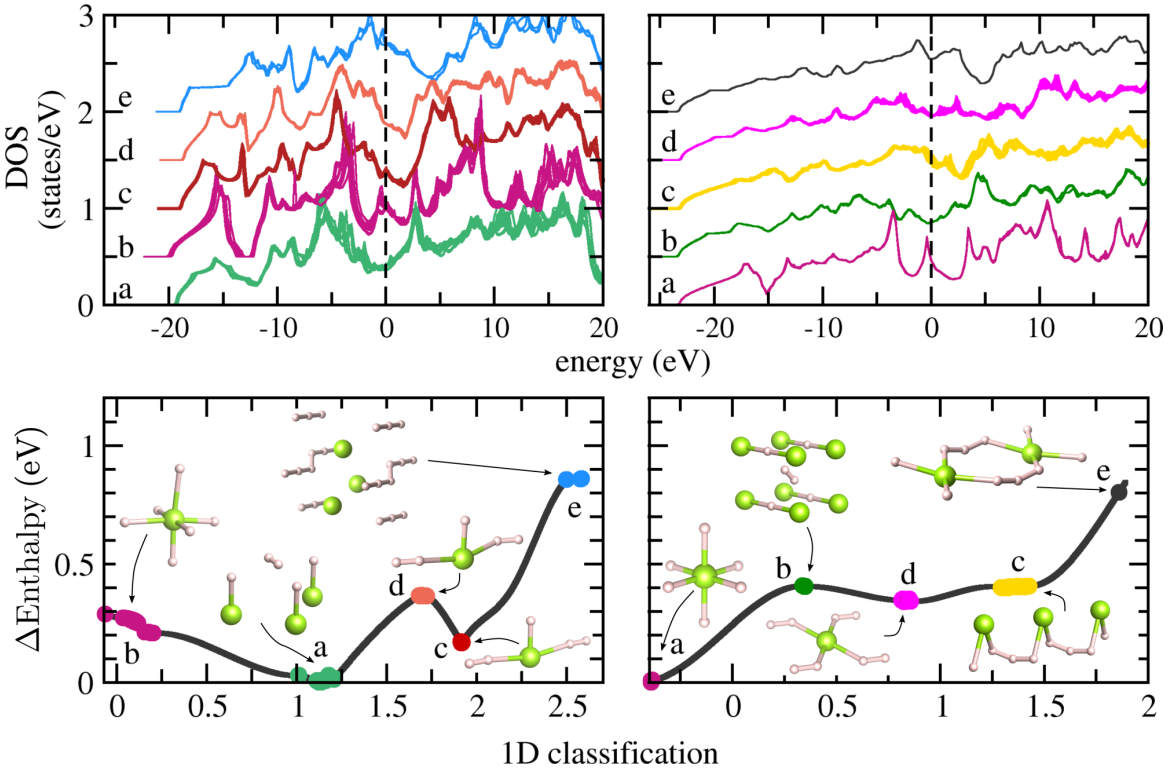

Figure 3: (Color online) Density of states (top panels) and one dimensional classification (bottom –see text) of SeH_3 structures generated by the minima hopping algorithm. At 50 GPa of pressure (left) and at 150 GPa of pressure (right). The color code is use to match a structure group with their corresponding density of states (as well alphabetically labeled). The solid black line that connects the points (bottom panels) serves only as guide to the eye. Hydrogen is represented in white and selenium in green spheres. For each family we shown a representative structure for the data-set, depicting variation in structure motif, coordination number and bond type.

ing the uniqueness of the $Im-3m$ phase, in both SH_3 and SeH_3 at high pressure.

3.2 Correlation between classification and electronic properties

The structure fingerprint distance and the one dimensional classification can be correlated to a more general trend: the electronic density of states (DOS). For this analysis we proceed in the following way: we compute the electronic DOS for all the systems in the sets and group them in classes according to the similarity. This classification can be somehow arbitrary and there are cases (structures) in which it is not uniquely define. For that cases we compute the occupation at the Fermi level (per structure) and build an histogram type plot. As shown in the top panels of Fig. 2 and Fig. 3 a clear separation in families is achieved for all the cases. The Fermi level is set to 0 eV in all the plots. Families are depicted within different colors and correspond to the crystals depicted in the bottom panels.

The 1D classification correctly clustered structures that are labeled as similar and that have similar DOS, after deep investigation we did not find situations in which different classes (families of structures or DOS) overlap. This means that despite of its simplicity, by mapping into the 1D Euclidean space we are able to classify different existent funnels in both compositions. It is also important to notice that spread (structures avoid clustering, as in the low pressure case) in the 1D classification corresponds as well a spread in the shape of the DOS function. This becomes clear and perhaps considered as obvious, but it is not trivial; we are not directly comparing the DOS with the crystal structure, we are stepping through several non trivial steps as the conversion of crystal representation into distances and reducing it to 1D-mapping. Therefore, our mapping is actually allowing a direct comparison between structures and electronic properties without the complexity of the Bravais lattice representation.

From a physical point of view there are some patterns that is worth to point out. The superconducting phase (cubic/rhombohedral) $Im-3m/R3m$ respectively at high and

low pressure - see Ref. [37] for details) has always a characteristic sharp peak near the Fermi level, highly metallic in comparison with the rest of structures living in other funnels, that instead tend to show a dip at the Fermi level as in indication of a remaining molecular character in their bonding. This mirrors the fact that the classification locates cubic/rhombohedral phases at the extreme side (in this case left) of the configurational space.

4 Conclusions

In this work we presented a new type of analysis and classification for materials design algorithms and we applied it for the first time to the case of high temperature, high pressure, superconducting families SH_3 and SeH_3. Materials design algorithms are intended to compute low enthalphy structures of a system at a given chemical composition. However due to both the limited accuracy of calculations and the complexity of the physical conditions it is not always the case that the experimentally synthesized system has the lowest predicted enthalphy. It is important to develop methods to characterize the full set of predicted structures. Based on the structural fingerprint, we define a metric distance between two given structures. The distances between structures are minimized with a simple algorithm; by doing this we are able to reduce the free-energy landscape to 1D-dimensional representation. With this 1D-dimensional structural classification it is easy to observe patterns of clustering of stable/metastable structures and their grouping into families. We showed that the corresponding structural grouping also corresponds to that of their electronic structure. From this description it appears that the covalent bonded cubic/rhombohedral phase responsible for the measured high $T_C$is unique. Finally, our post 1D-dimensional analysis could be useful in genetic-based algorithms for crystal prediction: identifying outsiders at each generation and keeping a good asset of novelty in crystal motifs, thus accelerating materials design.

Acknowledgement: J.A.F-L. gratefully acknowledges support from the Universität Basel and the computer resources provided by the CSCS in Lugano. This work was done partially within the NCCR MARVEL project.

References

[1] Defang Duan, Yunxian Liu, Fubo Tian, Da Li, Xiaoli Huang, Zhonglong Zhao, Hongyu Yu, Bingbing Liu, Wenjing Tian, and Tian Cui. *Sci. Rep.*, 4, Nov 2014.

[2] A. P. Drozdov, M. I. Eremets, I. A. Troyan, V. Ksenofontov, and S. I. Shylin. *Nature*, 525, 73–76, 2015.

[3] N. Ashcroft. *Phys. Rev. Lett.*, 21:1748–1749, Dec 1968.

[4] C. F. Richardson and N. W. Ashcroft. *Phys. Rev. Lett.*, 78:118–121, Jan 1997.

[5] P. Cudazzo, G. Profeta, A. Sanna, A. Floris, A. Continenza, S. Massidda, and E. Gross. *Phys. Rev. Lett.*, 100:257001, Jun 2008.

[6] Jeffrey M. McMahon and David M. Ceperley. *Phys. Rev. B*, 84:144515, Oct 2011.

[7] N. Ashcroft. *Phys. Rev. Lett.*, 92:187002, May 2004.

[8] Xiao-Jia Chen, Viktor V. Struzhkin, Yang Song, Alexander F. Goncharov, Muhtar Ahart, Zhenxian Liu, Ho-kwang Mao, and Russell J. Hemley. *Proceedings of the National Academy of Sciences*, 105(1):20–23, 2008.

[9] D. Y. Kim, R. H. Scheicher, S. Lebègue, J. Prasongkit, B. Arnaud, M. Alouani, and R. Ahuja. *Proceedings of the National Academy of Sciences*, 105(43):16454–16459, 2008.

[10] Ji Feng, Richard G. Hennig, N. W. Ashcroft, and Roald Hoffmann. *Nature*, 451, 445–448, 2008.

[11] Shibing Wang, Ho-kwang Mao, Xiao-Jia Chen, and Wendy L. Mao. *Proceedings of the National Academy of Sciences*, 106(35):14763–14767, 2009.

[12] Yansun Yao and Dennis D. Klug. *Proceedings of the National Academy of Sciences*, 107(49):20893–20898, 2010.

[13] Guoying Gao, Artem R. Oganov, Peifang Li, Zhenwei Li, Hui Wang, Tian Cui, Yanming Ma, Aitor Bergara, Andriy O. Lyakhov, Toshiaki Iitaka, and Guangtian Zou. *Proceedings of the National Academy of Sciences*, 107(4):1317–1320, 2010.

[14] Duck Young Kim, Ralph H. Scheicher, Ho-kwang Mao, Tae W. Kang, and Rajeev Ahuja. *Proceedings of the National Academy of Sciences*, 107(7):2793–2796, 2010.

[15] Yinwei Li, Guoying Gao, Yu Xie, Yanming Ma, Tian Cui, and Guangtian Zou. *Proceedings of the National Academy of Sciences*, 107(36):15708–15711, 2010.

[16] James Hooper, Tyson Terpstra, Andrew Shamp, and Eva Zurek. *The Journal of Physical Chemistry C*, 118(12):6433–6447, 2014.

[17] Takaki Muramatsu, Wilson K. Wanene, Maddury Somayazulu, Eugene Vinitsky, Dhanesh Chandra, Timothy A. Strobel, Viktor V. Struzhkin, and Russell J. Hemley. *J. Phys. Chem. C*, 119(32):18007–18013, 2015.

[18] Thomas Jarlborg and Antonio Bianconi. *Scientific reports*, 6, Article number: 24816 (2016).

[19] M Mahdi Davari Esfahani, Zhenhai Wang, Artem R Oganov, Huafeng Dong, Qiang Zhu, Shengnan Wang, Maksim S Rakitin, and Xiang-Feng Zhou. *Scientific reports*, 6, 2016.

[20] M. I. Eremets, I. A. Trojan, S. A. Medvedev, J. S. Tse, and Y. Yao. *Science*, 319(5869):1506–1509, March 2008.

[21] Olga Degtyareva, John E. Proctor, Christophe L. Guillaume, Eugene Gregoryanz, and Michael Hanfland. *Solid State Communications*, 149(39-40):1583–1586, October 2009.

[22] Michael Hanfland, John E. Proctor, Christophe L. Guillaume, Olga Degtyareva, and Eugene Gregoryanz. *Phy. Rev. Lett.*, 106(9):095503, March 2011.

[23] Timothy Strobel, P. Ganesh, Maddury Somayazulu, P. Kent, and Russell Hemley. *Phys. Rev. Lett.*, 107:255503, Dec 2011.

[24] José A. Flores-Livas, Maximilian Amsler, Thomas J. Lenosky, Lauri Lehtovaara, Silvana Botti, Miguel A. L. Marques, and Stefan Goedecker. *Phys. Rev. Lett.*, 108:117004, Mar 2012.

[25] Ion Errea, Matteo Calandra, Chris J Pickard, Joseph R Nelson, Richard J Needs, Yinwei Li, Hanyu Liu, Yunwei Zhang, Yanming Ma, and Francesco Mauri. *Nature*, 532:81–84, 2016.

[26] N. Bernstein, C. Stephen Hellberg, M. D. Johannes, I. I. Mazin, and M. J. Mehl. *Phys. Rev. B*, 91:060511, Feb 2015.

[27] Ion Errea, Matteo Calandra, Chris J. Pickard, Joseph Nelson, Richard J. Needs, Yinwei Li, Hanyu Liu, Yunwei Zhang, Yanming Ma, and Francesco Mauri. *Phys. Rev. Lett.*, 114:157004, Apr 2015.

[28] Defang Duan, Xiaoli Huang, Fubo Tian, Da Li, Hongyu Yu, Yunxian Liu, Yanbin Ma, Bingbing Liu, and Tian Cui. *Phys. Rev. B*, 91:180502, May 2015.

[29] Ryosuke Akashi, Mitsuaki Kawamura, Shinji Tsuneyuki, Yusuke Nomura, and Ryotaro Arita. *Phys. Rev. B*, 91:224513, Jun 2015.

[30] Christoph Heil and Lilia Boeri. *Phys. Rev. B*, 92:060508, Aug 2015.

[31] Yu Xie, Quan Li, Artem R Oganov, and Hui Wang. *Acta Crystallographica Section C: Structural Chemistry*, 70(2):104–111, 2014.

[32] Luciano Ortenzi, Emmanuele Cappelluti, and Luciano Pietronero. *Phys. Rev. B 94*, 064507 – Published 10 August 2016

[33] Ryosuke Akashi, Wataru Sano, Ryotaro Arita, and Shinji Tsuneyuki. *Phys. Rev. Lett.*, 117:075503, Aug 2016.

[34] Viktor V Struzhkin, DuckYoung Kim, Elissaios Stavrou, Takaki Muramatsu, Hokwang Mao, Chris J Pickard, Richard J Needs, Vitali B Prakapenka, and Alexander F Goncharov. *Nature Communications* 7, Article number: 12267 (2016) doi:10.1038/ncomms12267

[35] Tiago FT Cerqueira, Rafael Sarmiento-Pérez, Maximilian Amsler, F Nogueira, Silvana Botti, and Miguel AL Marques. *Journal of chemical theory and computation*, 11(8):3955–3960, 2015.

[36] Richard J Needs and Chris J Pickard. *APL Materials*, 4(5):053210, 2016.

[37] A. José Flores-Livas, Antonio Sanna, and E.K.U. Gross. *Eur. Phys. J. B*, 89(3):1–6, 2016.

[38] Christian Kokail, Christoph Heil, and Lilia Boeri. *Phys. Rev. B*, 94:060502, Aug 2016.

[39] José A. Flores-Livas, Maximilian Amsler, Christoph Heil, Antonio Sanna, Lilia Boeri, Gianni Profeta, Chris Wolverton, Stefan Goedecker, and E. K. U. Gross. *Phys. Rev. B*, 93:020508, Jan 2016.

[40] Andrew Shamp, Tyson Terpstra, Tiange Bi, Zackary Falls, Patrick Avery, and Eva Zurek. *Journal of the American Chemical Society*, 138(6):1884–1892, 2016.

[41] Hanyu Liu, Yinwei Li, Guoying Gao, John S Tse, and Ivan I Naumov. *The Journal of Physical Chemistry C*, 120(6):3458–3461, 2016.

[42] A Bianconi and T Jarlborg. *Novel Superconducting Materials*, 1:37–49, 2015.

[43] Eva Zurek. *Comments on Inorganic Chemistry*, (just-accepted), 2016, http://dx.doi.org/10.1080/02603594.2016.1196679

[44] Radosław Szcześniak and Artur P Durajski. *The European Physical Journal B*, 88(12):1–6, 2015.

[45] Colin W Glass, Artem R Oganov, and Nikolaus Hansen. *Computer Physics Communications*, 175(11):713–720, 2006.

[46] Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. *Computer Physics Communications*, 183(10):2063–2070, 2012.

[47] Chris J Pickard and RJ Needs. *Journal of Physics: Condensed Matter*, 23(5):053201, 2011.

[48] Stefan Goedecker. *The Journal of Chemical Physics*, 120(21):9911, 2004.

[49] Kurt Lejaeghere, Gustav Bihlmayer, Torbjörn Björkman, Peter Blaha, Stefan Blügel, Volker Blum, Damien Caliste, Ivano E Castelli, Stewart J Clark, Andrea Dal Corso, et al. *Science*, 351(6280), 2016.

[50] Li Zhu, Maximilian Amsler, Tobias Fuhrer, Bastian Schaefer, Somayeh Faraji, Samare Rostami, S Alireza Ghasemi, Ali Sadeghi, Migle Grauzinyte, Chris Wolverton, et al. *The Journal of chemical physics*, 144(3):034203, 2016.

[51] Ali Sadeghi, S. Alireza Ghasemi, Bastian Schaefer, Stephan Mohr, Markus A. Lill, and Stefan Goedecker. *The Journal of Chemical Physics*, 139(18), 2013.

[52] M. Amsler and S. Goedecker. *The Journal of Chemical Physics*, 133(22):224104, 2010.

[53] Maximilian Amsler, José A Flores-Livas, Lauri Lehtovaara, Felix Balima, S Alireza Ghasemi, Denis Machon, Stéphane Pailhes, Alexander Willand, Damien Caliste, Silvana Botti, et al. *Physical review letters*, 108(6):065501, 2012.

[54] Maximilian Amsler, José A Flores-Livas, Tran Doan Huan, Silvana Botti, Miguel AL Marques, and Stefan Goedecker. *Physical review letters*, 108(20):205505, 2012.

[55] Silvana Botti, José A Flores-Livas, Maximilian Amsler, Stefan Goedecker, and Miguel AL Marques. *Physical Review B*, 86(12):121204, 2012.

[56] Silvana Botti, Maximilian Amsler, José A Flores-Livas, Paul Ceria, Stefan Goedecker, and Miguel AL Marques. *Physical Review B*, 88(1):014102, 2013.

[57] Maximilian Amsler, José A Flores-Livas, Miguel AL Marques, Silvana Botti, and Stefan Goedecker. *The European Physical Journal B*, 86(9):1–3, 2013.

[58] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[59] G. Kresse and J. Furthmüller. *Comput. Mat. Sci.*, 6:15–50, 1996.

[60] Artem R. Oganov and Mario Valle. *The Journal of Chemical Physics*, 130(10), 2009.

[61] Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. *Proceedings of the National Academy of Sciences*, 108(32):13023–13028, 2011.

[62] Sandip De, Albert P Bartók, Gábor Csányi, and Michele Ceriotti. *Physical Chemistry Chemical Physics*, 18(20):13754–13769, 2016.

[63] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross. *Phys. Rev. B*, 89:205118, May 2014.

[64] Harold W Kuhn. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[65] Henning Glawe, Antonio Sanna, E K U Gross, and Miguel A L Marques. *New Journal of Physics*, 18(9):093011, 2016.