TC ESTIME
IS TIME

ROYAL SOCIETY
OF CHEMISTRY

EDGE ARTICLE
Trinidad Novoa, Julia Contreras-García *et al.*
TcESTIME: predicting high-temperature hydrogen-based
superconductors

15 YEARS
ANNIVERSARY

EDGE ARTICLE

Check for updates

TcESTIME: predicting high-temperature hydrogen-based superconductors†

Trinidad Novoa, ⓘ *ab Matías E. di Mauro, ⓘ a Diego Inostroza,a Kaoutar El Haloui,c Nicolas Sisourat,c Yvon Maday ⓘ b and Julia Contreras-García*a

Superconductivity can be considered among the most exciting discoveries in material science of the 20th century. However, the hard conditions for the synthesis and the difficult characterization, make the statement of new high critical temperature (T_c) complex from the experimental viewpoint and have recently led to several hot controversies in the literature. In this panorama, theory has become a trustworthy diagnosis. Nevertheless, this comes at an extremely high computational cost. A faster alternative would be to find cheap footprints of superconductivity from the electronic structure. Some of the authors have recently shown that a correlation exists between T_c, the networking value [Nature Communications, 12, 5381 (2021)], and the molecularity index [arXiv:2403.07584v1 (2024)]. The networking value reflects the metallicity of the parent compound as a measure of its electron delocalization channels, by means of the Electron Localization Function topology (its bifurcation trees). Instead, the molecularity index quantifies the presence of H_2 molecules within the system. All in all, these two quantities characterize bonding features that are related to high T_c: high metallicity and low molecularity boost high T_c states. However, the quantification or these bonding characteristics was initially made by a visual approach, which is not scalable for high throughput screening. We have developed a new code, TcESTIME, which allows to determine the networking value for a given hydrogen-based compound. In this contribution, we present such code and the underlying periodic algorithms we have developed. As a reference, the estimation of T_c for LaH_{10} thanks to this new code amounts to 10 CPU minutes in a computer cluster equipped with Intel Xeon 2.4 GHz processor. Given the new potential for screening, we have applied it to a larger set including ternary hydrogen based superconductors, and have proposed new fits to estimate T_c, leading to errors of *ca.* 33 K. We believe that this contribution settles the bases for an automatic high-throughput screening of hydrogen-based superconductors.

1 Introduction

Superconductivity is one of the most exciting discoveries in materials science of the 20th century with major technological consequences. Indeed, superconductors have perfect diamagnetic properties and zero resistance. These properties have allowed essential technological advances such as the realization of intense magnetic fields used in NMR,[1,2] in high resolution electron microscopy,[3,4] in the development of nuclear fusion reactors,[5,6] in levitation transport,[7,8] in high speed quantum computers,[9,10] *etc.* From the scientific point of view, the great

impact of superconductivity can be quantified not only by the number of publications to which it has given rise, but also by the number of Nobel Prizes that have been awarded in the context of studies of this phenomenon (a total of 5!). Nevertheless, after all these discoveries, an ambient superconductor still remains the Holy Grail, one of the most sought-after objects in solid state physics. In the last decade, the prediction and observation of high-temperature conventional superconductivity in hydrogen-rich systems has proven the exceptional potential of these materials,.[11] Some note-worthy experimental results include H_3S at 203 K,[12] LaH_{10} at 250 K,[13] and YH_9 at 243 K,[14] all of them under very high-pressures, ranging from 90 to 201 GPa. If this class of compounds could be designed for applications in our everyday world, their extraordinary characteristics would introduce a technological revolution through lossless electrical transport and ultra-efficient electric motors or generators. In the field of superconductivity, there is an intense effort between theoretical and experimental work to develop the practical use of hydrogen superconductors. In all of this, the

aLaboratoire de Chimie Théorique (LCT), Sorbonne Université, CNRS, 4 Pl. Jussieu, Paris, 75005, France. E-mail: trinidad.novoa_aguirre@sorbonne-universite.fr

bSorbonne Université, Université Paris Cité, CNRS, Inria, Laboratoire Jacques-Louis Lions (LJLL), Paris, 75005, France

cLaboratoire de Chimie Physique-Matière et Rayonnement (LCPMR), Sorbonne Université, CNRS, 4 Pl. Jussieu, Paris, 75005, France

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4sc04465g

role of theory is to guide the experiments by proposing possible stable compounds and establishing the right footprints for superconductivity.

Important insights on what boosts T_c have been provided in the last years through the analysis of the types of bonds that are present in conventional superconductors.[11,15,16] Starting from the study through first-principles calculations of the structural and electronic properties of more than one hundred compounds, some of us have shown that a high 3D electron delocalization is crucial for understanding the critical temperature of those conventional hydrogen-rich superconductors.[17] At a later stage, we have also described the tendency of the molecularization of hydrogen–hydrogen bonds to hinder the critical temperatures of those compounds, a feature that can be assessed by means of the localization of electrons in H_2-like intramolecular regions.[18] Other works have shown how, for some superconductors, T_c seems to be lowered by saturated covalent bonds,[16] such as those found in H_2 molecular units. In fact, in a one dimensional hydrogen chain, one can see that molecularization hampers the metallic character of the system – crucial precursor to the superconducting transition – giving a hint as to why superconductivity is enhanced without the molecular units. The definition of a function to measure localization in the superconducting state showed that the superconducting transition does not affect the arrangement of electrons greatly, both in the one-dimensional chain as in well-known compounds.[18,19]

Moreover, we have defined a magnitude named as the networking value, which correlates well with the predicted critical temperature, much better than any other descriptor analyzed in the existing literature.[17] This magnitude can be easily calculated for any compound by analyzing isosurfaces of the electron localization function and is applicable to any bonding type. This correlation opens the pathway for a high throughput screening of potential high-temperature superconductors. However, for that aim, the determination of the networking value, which was carried out from visual inspection, needs to be automatized. In this manuscript, we endeavor into such an automatization for the determination of delocalization pathways in periodic systems, and the consequent estimation of T_c. It is worth noting that, since the aforementioned correlation was observed for conventional hydride superconductors, we hereby only refer to materials of that type. Extending the formalism to other types of superconductors is beyond the scope of this contribution.

In the following section, some theoretical background on the Electron Localization Function and the descriptors derived from it are laid out. Then, in Section 3 we present the workflow of TcESTIME, along with the algorithm developed to compute the networking value. In Section 4 we compare our results to those obtained visually in a database of binary compounds. In Section 5, we propose certain improvements on the prediction of the critical temperature by extending the database to contain ternary systems, and presenting new fits for T_c. Finally, in Section 6, we present the open source TcESTIME code and the TcESTIMWEB server, that facilitates the estimation of T_c through an accessible and user-friendly platform.

2 Theoretical background

2.1 The electron localization function

The Electron Localization Function (ELF) was originally designed by Becke and Edgecombe to identify "localized electronic groups in atomic and molecular systems".[20] It relies, through its kernel, $\chi_\sigma(\mathbf{r})$, to $D_\sigma(\mathbf{r})$, the leading term of the Taylor expansion of the spherically averaged conditional same spin pair probability, scaled by the homogeneous electron gas kinetic energy, $D_\sigma^0(\mathbf{r})$,

$$\chi_\sigma(\mathbf{r}) = \frac{D_\sigma(\mathbf{r})}{D_\sigma^0(\mathbf{r})}, \qquad (1)$$

in which

$$D_\sigma(\mathbf{r}) = \tau_\sigma(\mathbf{r}) - \frac{1}{4}\frac{|\nabla\rho_\sigma(\mathbf{r})|^2}{\rho_\sigma(\mathbf{r})}$$

is the difference between the definite positive kinetic energy, $\tau_\sigma(\mathbf{r})$, and the von Weizsäcker kinetic energy functional;[21] while the kinetic energy density of the homogeneous electron gas is

$$D_\sigma^0(\mathbf{r}) = \frac{3}{5}(6\pi^2)^{2/3}\rho_\sigma^{5/3}(\mathbf{r}).$$

This formulation led Savin and co-authors to propose an interpretation of ELF in terms of the local excess kinetic energy due to the Pauli repulsion, enabling its calculation from Kohn–Sham orbitals.[22]

The electron localization function itself is obtained through a lorentzian transformation of χ_σ:

$$\text{ELF}(\mathbf{r}) = \frac{1}{1 + \chi_\sigma^2(\mathbf{r})}. \qquad (2)$$

In this way, the ELF ranges between 0 and 1, and its maxima are located in the regions of higher electron-pair localization, whereas smaller values are found at the boundaries between such regions.

Within the dynamical system framework, pioneered by Bader for the analysis of the electron density,[23] a formal analogy is made between a vector field bounded on a manifold and a velocity field. In the present case, the vector field is the gradient field of the ELF, and the manifold the 3-dimensional geometrical space. This provides a partition of the direct geometrical space into regions called basins, which can be thought of as electronic domains corresponding to the chemical entities of the Lewis picture. Some of those basins will correspond to the valence, and are denoted by V(A, …), where A is the atom of the neighboring atomic core basin, C(A) (which can be constituted by K, L, …shells). In agreement with Lewis's picture, V(A, …) may belong to several atomic shells (see Fig. 1).

2.2 Topological descriptors of the ELF

The properties of the gradient dynamical system are complemented in the context of the ELF topology with the interpretation derived from the f-domain[24] concept, that enables to recover chemical units in the system, as well as to characterize the basins according to common chemical knowledge.
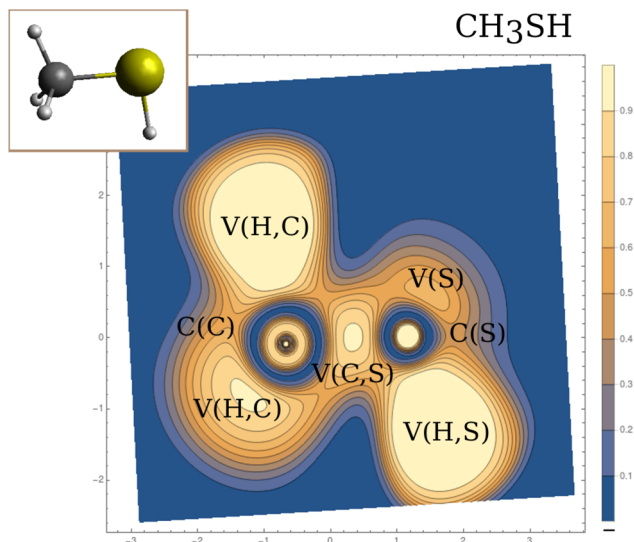
Fig. 1 ELF contour plot for the methanethiol molecule (CH₃SH, drawn in the inset), on the plane containing the C−S and S−H bonds. The core and valence basins, that contain the ELF attractors (maxima, in light colors) and are limited by the saddle points, are labeled and let us identify the cores (C(C) and C(S)), the bonds (V(H,C), V(C,S) and V(S,H)), and the lone pair regions (V(S)).

Introduced by Mezey[25] within the AIM framework, the concept of an *f*-domain accounts for the volume enclosed by an isosurface of a certain value of ELF(**r**) = *f*. As the value *f* increases, successive splitting of the initial domains take place until all of them contain one, and only one, attractor (*i.e.* maximum). These final domains are called irreducible, and the order in which they appear with increasing *f* reveals the nature of the interactions taking place in the system and the relationship between basins. The turning points of the splitting corresponds to that of the highest (3,−1) point of the separatrix connecting the basins. According to the value of the ELF at these nodes, a bifurcation tree can be constructed that reveals the basin hierarchy at a glance. The highest ELF value for which the *f*-domain is tridimensional corresponds to the networking value, ϕ.[17] Fig. 2 illustrates this principle for YH₉. At *f* = 0.79 some basins are connected leading to H₂ molecules (Fig. 2b). As the ELF value diminishes, more connections appear, and at *f* = 0.58 (Fig. 2d) the volume in this isosurface (associated with all points such that *f* ≳ 0.58) is connected. This is the networking value, ϕ = 0.58. In what follows, such a volume inside an *f*-isosurface will be named as "*f*-domain".

In ternary systems, more complex situations arise that required the introduction of another index quantifying the presence of H₂ molecules. This can be done by means of the molecularity index, ϕ^*, which also stems from the analysis of the topology of the ELF.[18] It corresponds to the highest ELF value for which any two hydrogens in the system are placed within the same *f*-domain. In Fig. 2a, where *f* = 0.79, there are two hydrogens (labeled H2 and H3) within the same *f*-domain, while the other hydrogens (H1, for example), remain disconnected. That value corresponds to the molecularity index of the

system, $\phi^* = 0.79$. The isosurface containing H2 and H3 is displayed in 2b.

2.3 Prediction of critical temperatures

The importance of these indexes comes from their relation to the critical temperature of hydrogen-based superconductors. A correlation between T_c and these two indexes has been recently observed,[17,18] providing a much more efficient way of estimating critical temperatures, as the ELF can be obtained practically for free after a SCF calculation is performed.[18] This is true even if the correlation is not very sharp, as it can be used for predicting T_c for large amounts of systems. It can however be improved if other quantities are taken into account: the fraction of hydrogen atoms in the unit cell, H_f, and the fraction of the density of states (DOS) at the Fermi level (ε_F) that correspond to hydrogen, H_{DOS}. Those are defined as

$$H_f = \frac{N_H}{N}, H_{DOS} = \frac{DOS_H(\varepsilon_F)}{DOS(\varepsilon_F)}, \tag{3}$$

where N refers to the number of atoms in the unit cell, N_H to the number of hydrogens in it, and the subscript H of the DOS indicates a projection onto hydrogen orbitals. H_f and H_{DOS} have shown to be necessary, but not sufficient, for high-T_c, and it is typically observed that hydrogen-rich systems with a high DOS at the Fermi level with H-character tend to be better superconductors. In this way, the correlation improves significantly by defining

$$\Phi_{DOS} = \phi H_f \sqrt[3]{H_{DOS}}, \tag{4}$$

allowing for a quick estimation of the critical temperature in binary compounds,

$$T_c = (750 \Phi_{DOS} - 85) \text{ K}. \tag{5}$$

This formalism has already been adopted in the literature for the fast estimation of critical temperatures, *e.g.* to theoretically assess the T_c of different N-doped lutetium hydrides,[26] or of new predicted quaternary hydrides,[27] among others.

Alternative fits to estimate T_c using the same framework have been introduced to extend these relationships to ternary compounds, which required the inclusion of the molecularity index.[18] Those fits were obtained using Symbolic Regression, a Machine Learning technique that yields an analytical expression for, in this case, the estimation of the critical temperatures. The results of ref. 18 show that the inclusion of ϕ^* leads to better results in the estimations. There, the proposed fit

$$T_c^{SR2} = 442.3 \left[1 - (\phi^* - \phi)\right] H_f^3 \sqrt{H_{DOS}}, \tag{6}$$

leads to errors of 36 K in the test set, for which the expression in (5) yields 55 K. Another approach is to filter the systems according to their values of ϕ^*, discarding those outside the range [0.45, 0.8], where T_c is larger. Then, the estimation can be done using

$$T_c^{SR4} = 574.7 \phi \sqrt{H_f^3 H_{DOS}}, \tag{7}$$

leading to errors around 39 K in the test set.
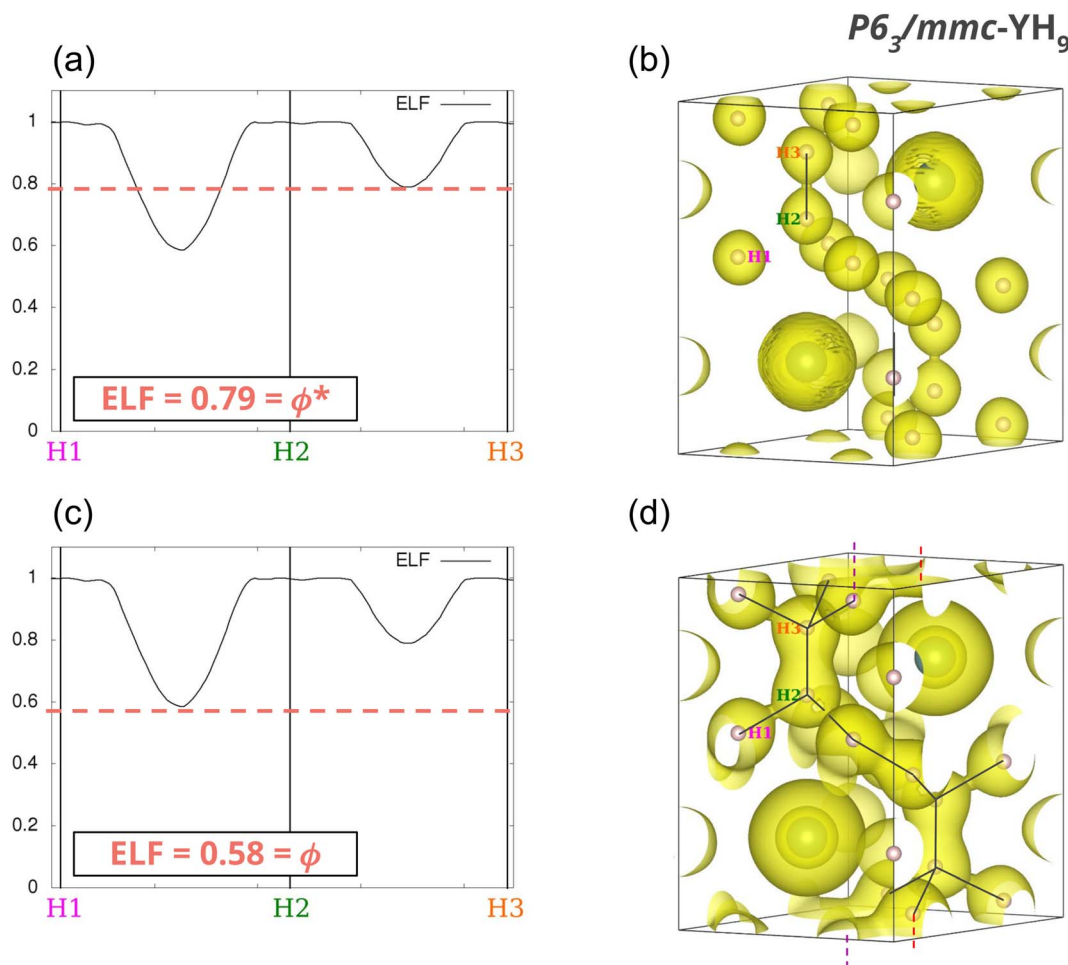
$P6_3/mmc$-YH$_9$

Fig. 2 (a) Profile of the ELF of $P6_3/mmc$-YH$_9$ along the path connecting hydrogens H1, H2 and H3. The dashed line marks the isovalue ELF = 0.79, corresponding to the molecularity index of the system, ϕ^*. (b) Isofurface of ELF = ϕ^* = 0.79. Black line shows how H2 and H3 connect inside the same isosurface. (c) ELF profile of $P6_3/mmc$-YH$_9$ in the same path as in (a), with the isovalue ELF = 0.58 marked by a dashed line, corresponding to the networking value, ϕ. (d) ELF = ϕ = 0.58 isosurface, with black lines representing the three-dimensional periodic isosurface connecting all hydrogens.

3   The TcESTIME algorithm

Automatizing the search of the networking value and of the molecularity index is the missing piece for an efficient computation of the critical temperature. While the estimation of ϕ^* is quite straightforward,‡ that of ϕ is certainly not. The TcESTIME algorithm has been developed for this purpose. The key feature of the networking value is that, for $f = \phi$ the associated ELF f-domain is connected (in the topological sense *i.e.* it has no disjoint components), and thus expands in all three dimensions. To evaluate this connectivity, we build an associated complex network for each candidate networking value, ꬵ, where the nodes are the attractors of the ELF in the unit cell, and the edges represent a gradient path passing through a saddle point and connecting two attractors. Then, we define the periodic connectivity of the network as follows:

1. In 1D: there is at least one path that connects a node to one of its translated analogues in another unit cell.

2. In 3D: there must be three such paths in three different, linearly independent, directions.

Thus, the latter condition is the one that must be fulfilled. In the following sections, we will start by determining the complex network and, then, we will see how to reduce it to one unit cell.

3.1   Constructing the network

The networking value will be determined through the topological analysis of the ELF, performed using the program critic2.[28–30] This code allows reading grid data as the ones obtained from Quantum ESPRESSO[31,32] (used along this application). Using a Newton–Raphson algorithm, it allows the localization of the ELF critical points (CPs): maxima located in the nuclei (nuclear attractors or NUCs), other maxima corresponding to the atomic shell structure and valence (non-nuclear attractors or NNAs), and saddle points corresponding to bonding regions, *i.e.* CPs that are minima in one direction and maxima in the other two (bond critical points or BCPs).[28,29] Each BCP is connected to two attractors, be it nuclear or non-nuclear, inside an isosurface of isovalue that is at most equal to the value of the ELF in the BCP (f-domain). These attractors can be either
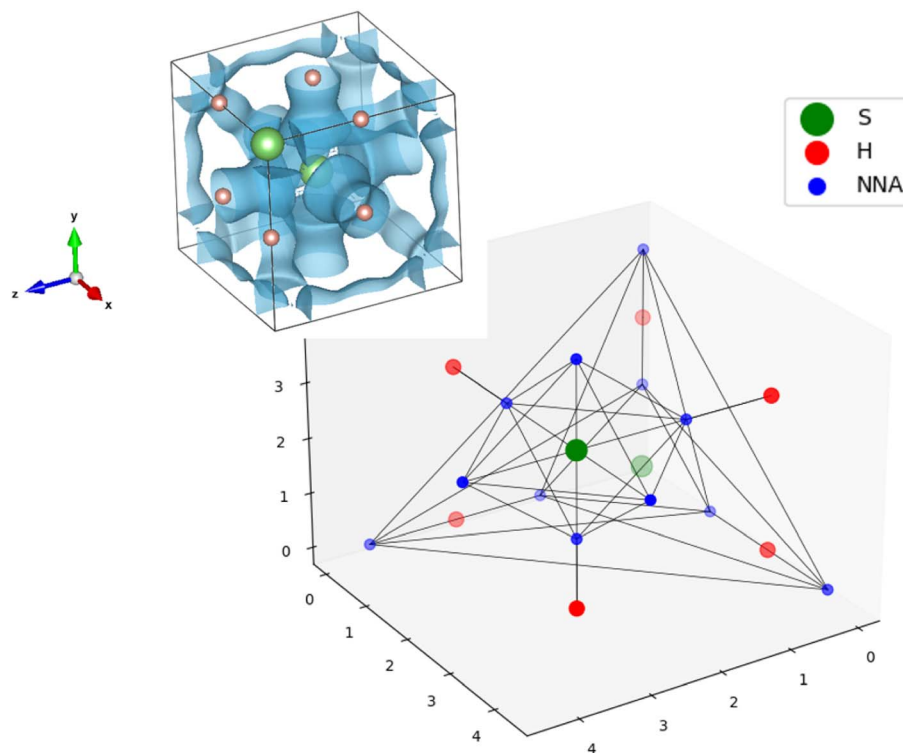
Fig. 3 Unit cell of $(H_2S)2H_2$ in the $Im\bar{3}m$ crystal structure, with S atoms in green and H atoms in red. On the top-left corner, the ELF isosurface for the networking value $\phi = 0.69$ is shown in blue. At the bottom, the representation of the same system in the complex networks created by TcESTIME that connect the atoms and the non-nuclear attractors (NNA) in the unit cell.

in the same unit cell, or in a neighboring one. With this information, it is possible to transform the typically analyzed ELF isosurfaces (Fig. 3-top-left) into graphs (Fig. 3-bottom). Note how the NNAs (in blue) surrounding the S atoms (in green) connect the neighbouring H atoms (in red) for the $f = 0.69$ domain.

Two main issues were observed when constructing the network:

1. Due to the rapid variations of the ELF in the core shells, some connections between NNA's in the cores and the corresponding nuclear attractor are missed by critic2.

2. The edges of the graphs are very sensitive to the tuning of the critic2 parameters, namely the distance thresholds for which two CPs are considered to be the same one.

3.1.1 Effective core radii. Although shell ELF attractors are spherically degenerated, they become non degenerated when the atoms are coordinated.[33] An example is shown in Fig. 4a, where the maxima around the Cr are highly concentrated in the direction of the Hydrogens. This great variety and amount of critical points per unit cell sets a big algorithmic difference between solids and molecules. Whereas the valence region has been found to be exhaustively tracked by algorithms analogous to those employed for the analysis of the topology induced by the electron density, missing connections between NUC's and core NNA's are a common feature in the networks that are built from the critic2 output data. Hence, the complete characterization of the ELF topology in the solid requires a hybrid method, with a combination of core-valence approaches.

We use chemical intuition to avoid this problem and presuppose that all ELF maxima that correspond to the same core are connected, as it is in fact the core the one that needs to be part of the three-dimensional lattice. This requires the definition of an effective core radius, which can be done from the ELF of isolated atoms, as it has been shown that the values in the core region are quite rigid, staying practically unchanged upon bonding and induced pressure.[34] Core radii values, r_A, are tabulated for elements A with $Z \in [3, 38] \cup \{48\}$ in ref. 35 (also available in Table S1 of ESI†). Hence, all critical points around a nuclear maximum of atom A and within a distance r_A of it will be linked together, as shown by the black circles in Fig. 4b.

As it is usual in the case of ELF, it is complex to analyze its topology for hydrogen atoms. Having only one electron – no Pauli repulsion – the ELF surface is very flat. This can sometimes lead to having many NNAs close to the hydrogens, and the connections between them and the bond CPs are easier to miss. Therefore, it is useful to assign a core radius to H, since it allows to group all those CPs that are found very close to the nucleus. We choose $r_H = 0.7a_0$, which corresponds to half the bonding distance of the H_2 molecule.

The definition of this radius is particularly convenient for systems where molecular hydrogen is present, as the profile of the ELF along the bond axis is very flat and it is more common that the algorithm fails to find the BCPs or to properly connect them to the NUCs and NNAs.

3.1.2 Tuning critic2 parameters. The algorithm implemented in critic2 for the search of the critical points of the ELF
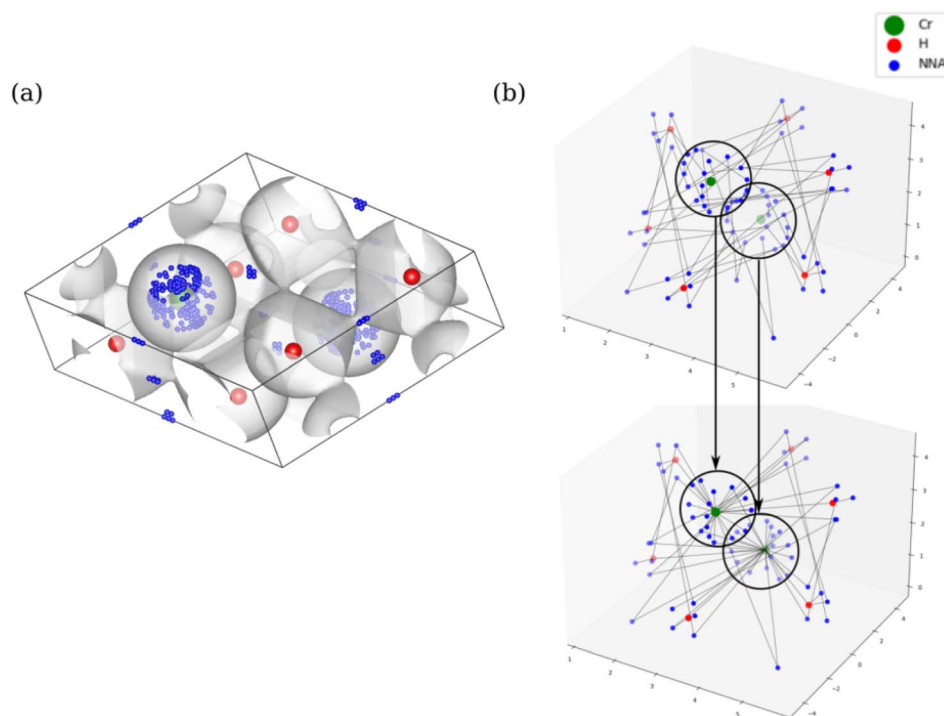
Fig. 4 (a) ELF isosurface at $\phi_{ref} = 0.35$ of the CrH$_3$ $P6_3/mmc$ crystal at 81 GPa. ELF non-nuclear attractors as found with the default critic2 parameters are marked in blue, while the atomic positions of Cr and H are in green and red, respectively. (b) Graphs of the same system as represented by TcESTIME, with Cr and H nuclear attractors in green and red, respectively, and NNA's depicted in blue (computed with optimized parameters). The original graph on top is transformed into the one on the bottom when all NNA's inside an effective atomic radius are considered to be connected.

requires a set of thresholds that let the program identify two critical points as being the same or not. The default minimum distance to consider two CPs to be equivalent is $0.2a_0$. We have noticed that networking value is more accurately estimated when this threshold is raised to $0.3a_0$, which we have set as default in TcESTIME. This bypasses the need to get a finer grid when the algorithm oscillates and does not converge properly, and it avoids the unnecessary computation of too many equivalent CPs in highly symmetric systems (*e.g.* spherically symmetric maxima and saddle points in the cores).

Another parameter that can be tuned is the minimum distance between the nuclear coordinates of hydrogen and an ELF attractor for the latter to be considered nuclear. To avoid a large amount of NNAs to form around the hydrogens, we have raised this quantity to $0.6a_0$, which has shown to optimize the search of the networking value in the reference set.

These parameters can be changed by the user, as well as other more specific ones, as TcESTIME also allows to run critic2 externally with the preferred options, using the resulting output file as input.

3.2 Reducing the analysis to the unit cell

Considering the definition of periodic three-dimensional connectivity given before, to determine ϕ we must search for paths that connect a node to itself in another unit cell. However, for each trial f this means using supercells composed of at least 8 cells of unit size, in order to check for all possible connectivities outside the unit cell. We will try to simplify this picture thanks to periodicity. This is done in two steps: translating the graph to the unit cell, and then checking for closed paths.

To reduce the network to the unit cell, we associate a direction and weight to the edges, that will correspond to the translation vector of the connection. For example, if node 1 is directed towards node 2 with translation $\mathbf{t}_{12} = (-1, 0, 0)$ it means that, in every unit cell, 1 is connected to 2 in a neighboring cell to its left (see Fig. 5). It is worth noting that these directions are simply an attribute of the connection, and it does not mean that these are directed graphs in the complex network sense. (In the previous example, we do consider that node 2 is connected to 1, but the translation would change sign. This would not be true in a classic directed graph.)

With this "labelled" graph, we can translate all the information of the periodic connectivity to the unit cell while keeping the translational information.

3.2.1 Finding closed paths: tree algorithm. Having defined the network in the unit cell, the connectivity along a direction is reduced to having closed paths P_i connecting a node i to itself within the unit cell. Considering all edges $k \in P_i$, the net translation of P_i is defined as:

$$\mathbf{t}_{ii} = \sum_k s_k \mathbf{t}_k, \tag{8}$$

with $\mathbf{t}_k$ the translation vector of k and s_k the sign assigned that the edge within P_i, *i.e.* $s_k = 1$ if the path goes in the original
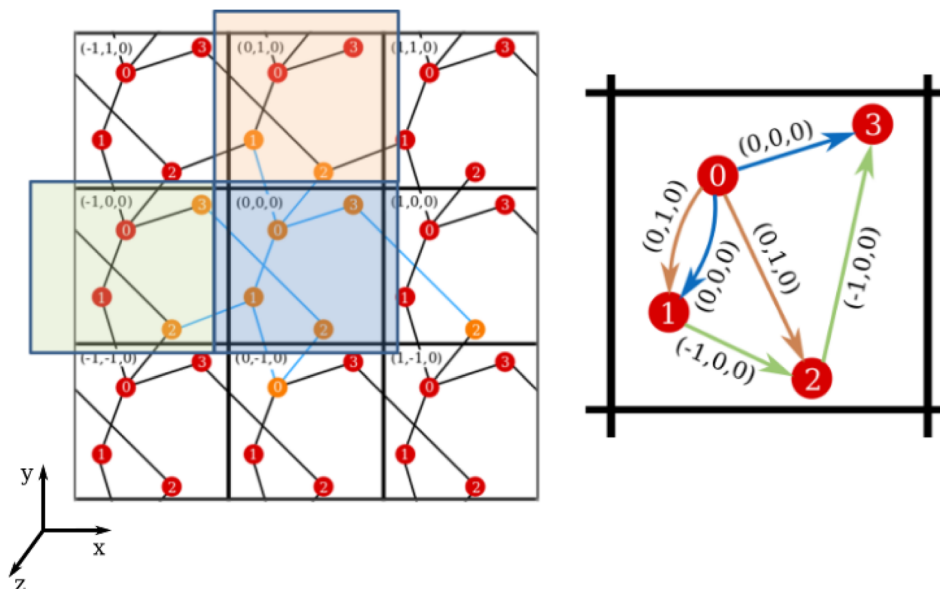
Fig. 5 Left: connections between nodes in a single unit cell, (0,0,0), and those surrounding it. The nodes in orange are those in or connected to the (0,0,0) unit cell, forming the minimal network that captures periodicity. Right: simplified representation of all the connections in the lattice in one unit cell. The network edges are labeled with the translation between two nodes. The color of the arrows represents the translation to the respective neighboring cell on the left panel.

direction of the edge, and $s_k = -1$ otherwise. Note that the first and the last edges must contain i: $k^1 = i, j$ and $k^{n_i} = l, i$, with n_i the number of edges in P_i. For the system in Fig. 5, we can identify a closed path $3 \rightarrow 2 \rightarrow 0 \rightarrow 3$, with a net translation of $\mathbf{t}_{33} = (1, -1, 0)$.

While it is rather easy to visually identify all possible closed paths for small systems, it is not straightforward to do so for larger ones. We propose an algorithm that is based in the breadth-first search algorithm,[36] that builds the tree data structure of connections in a network starting from one node, that we call n_0. In the original algorithm, this is done by checking all the connections of the initial node, and creating a new branch for each one of them, which are put in a queue, while the initial node is marked as visited. The process is repeated for each of the nodes in the queue, which are then taken off of it when they are visited, and finishes when it is empty. Because in our networks there could be more than one connection between the nodes (see for example the double connection between nodes 0 and 1 in Fig. 5), we use the adjacency matrix, A, instead of the visited flag, where $A_{ij} = A_{ji}$ is equal to the number of edges connecting nodes i and j. The matrix is updated by subtracting 1 to A_{ij} (and A_{ji}) when a connection between said nodes is added to the tree. The translations of these connections are also taken into account and stored, taking care of its sign, as can be seen in the example in Fig. 6.

Once the tree starting from a node n_0 is found, the nodes n_r that appear more than once are identified. For those there will be a closed path in the network, and its net translation can be found by subtracting the net translation of each of the branches, up to n_r (see Fig. 6). If there is more than one connection between two nodes, we consider a different path for each of them, with different net translations.
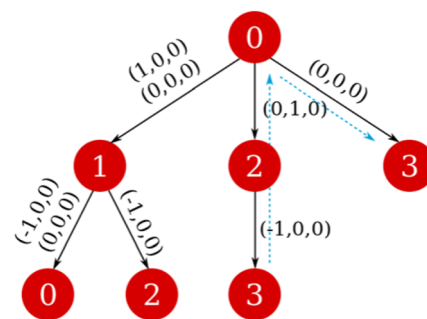

Fig. 6 Tree data structure of the network starting from node 0. The blue dashed arrows in each pannel mark the closed path $3 \rightarrow 2 \rightarrow 0 \rightarrow 3$.

4 Results for reference systems

The networking value was computed for an ensemble of 129 systems, corresponding to the hydrogen-based superconductors of ref. 17, from where the reference values ϕ_{ref} were taken.§ These values were originally visually determined, analyzing different ELF isosurfaces and finding the one for which a 3D network formed.

Table 1 Systems for which TcESTIME fails to compute a networking value, as indicated by their chemical formula and space group; and their expected networking value, ϕ_{ref}, and critical temperature, T_c

Chemical formula	Space group	ϕ_{ref}	T_c (K)	Reference
LaH_8	$C2/m$	0.52	131	37
PoH	$P6_3/mmc$	0.34	0.65	38
H_2I	$Pnma$	0.36	5.3	39

The program was able to find a networking value for 126 systems, with a mean absolute error (MAE) of 0.04 with respect to the reference values ϕ_{ref}, and a standard deviation of 0.08, and an average computation time of *ca.* 2 s per system. A distribution of the difference $\phi_{\mathrm{ref}} - \phi$ is reported in Fig. 8, where we compare the estimations made with and without the considerations that were made concerning the cores: the core radius for non-H and H atoms, and the nuclear distance parameter for hydrogen, too. For simplicity, we call all these considerations the effective core radius r_{eff}. The $Im\bar{3}m$–$(H_2S)2H_2$ structure of Fig. 3 is a successful example for which TcESTIME found the exact expected networking value, $\phi = \phi_{\mathrm{ref}} = 0.69$. TcESTIME was not able to find a networking value for a total of 3 systems, as reported in Table 1. A more detailed look into the

Fig. 7 ELF isosurface at the reference netowrking value, $\phi_{\mathrm{ref}} = 0.52$, of the LaH$_8$ $C2/m$ crystal (top-left corner) and its graph representation as constructed by TcESTIME, with the La and H atoms in green and red, respectively, and the NNA's in blue. The black circle marks one of the regions where critical points fail to get connected.

Fig. 8 Histogram of the error in the estimation of the networking value ϕ with respect to the reference value ϕ_{ref}. Results with (orange) and without (blue) the inclusion of effective core radii are presented.

ELF graph for the $C2/m$–LaH$_8$ crystal at the reference value $\phi_{\mathrm{ref}} = 0.52$ (see Fig. 7), shows that it is likely that two unconnected bond critical points should have been considered as equivalent. A hand tuning of critic2's parameters solves the issue for these problematic systems in most cases. From Fig. 9b one can note that both approaches, TcESTIME and visual search of ϕ, yield the most similar results for the structures with higher predicted T_c, which interest us the most.

Using the same values of H_{DOS} as in ref. 17, it is straightforward to compute Φ_{DOS}. The correlation between Φ_{DOS}, as computed by TcESTIME, and the reported critical temperature of the reference systems is recovered (see Fig. 9a).

5 Improving the estimation of T_c

5.1 Better fits for ternary hydrides

It has become evident that the search for high-temperature hydrogen-based superconductors has to expand towards ternary systems, a much broader space that has not yet been explored exhaustively. Here, we will use the new automated code to simultaneously analyze the set of 129 binary[17] and 21 ternary[18] compounds. The ternary compounds are composed of X-RE-H, with X and RE *s*-block and rare-earth elements, respectively. A list of those systems is provided in Table 3 of the ESI.† For the computational details concerning the computation of the ELF and DOS in those systems, see ref. 18.

The values of the networking value according to TcESTIME have a MAE of 0.02 with respect to the ones estimated visually, showing similar trends to those observed for the binary compounds. As can be seen in Fig. 9 the correlation between Φ_{DOS} and T_c still holds for the ternaries, although some of them do seem to further deviate from the tendency.

A new linear fit can be obtained considering Φ_{DOS} of both binary and ternary compounds, as computed by TcESTIME, and using the least squares method:

$$T_c^{\mathrm{LS}} = (456.34\Phi_{\mathrm{DOS}} - 9.46) \text{ K}. \tag{9}$$

This estimation of T_c yields a MAE of 38.71 K in the whole dataset, and a standard deviation of 53.13 K.

5.2 Gradient boosting regression

The availability of these data opens the possibility of proposing new models to estimate T_c. We present here a Machine Learning (ML) model obtained using ϕ, ϕ^*, H_f and H_{DOS} as descriptors to estimate the target quantity, T_c. The introduction of this model permits to consider both binary and ternary systems during training. Moreover, their performance can be evaluated in a test set that was not used during training, which is crucial to assess the transferability of the model, and was not done in the original linear fit.

Gradient boosting regression (GBR) is an approach first proposed by Friedman[40] to extend the boosting algorithm to regression problems. GBR consists of integrated models with high performance and good stability, combining several weak prediction models (decision trees) to create a more powerful and accurate one. In practice, one constructs an additive model
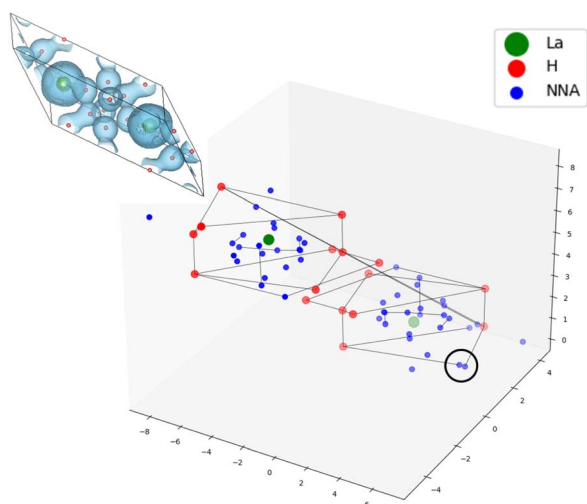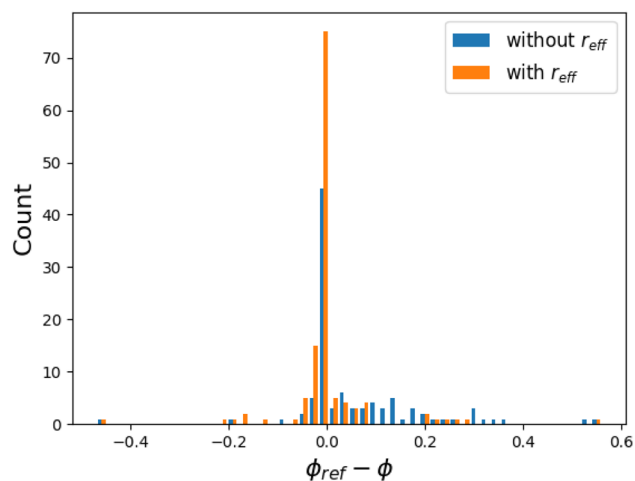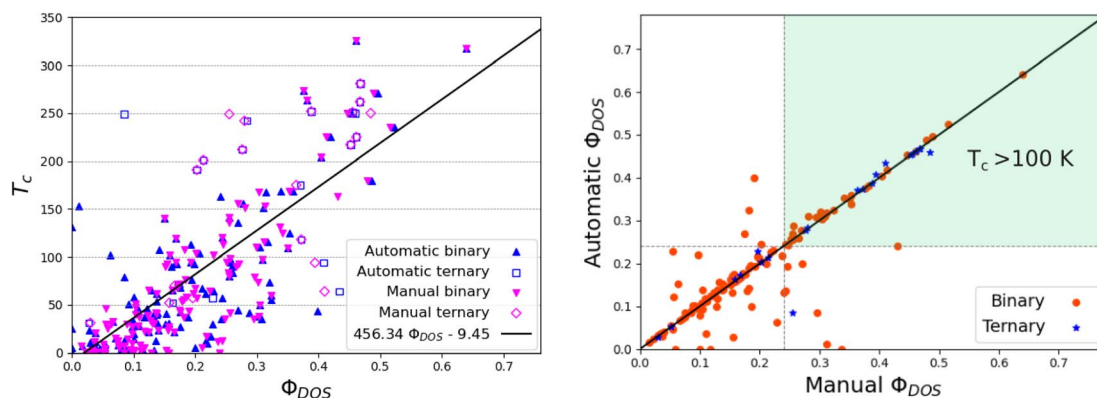
Fig. 9 Left: Correlation between Φ_{DOS} and the superconducting critical temperature, T_c for the reference values of the networking value estimated manually and those computed automatically with TcESTIME, for binary and ternary compounds. The black line represents the linear dependence of T_c on Φ_{DOS} according to TcESTIME. Right: comparison between manual and automatic (TcESTIME) estimation of Φ_{DOS}. The green shaded region corresponds to $T_c > 100$ K, according to the linear fit w.r.t. Φ_{DOS}.
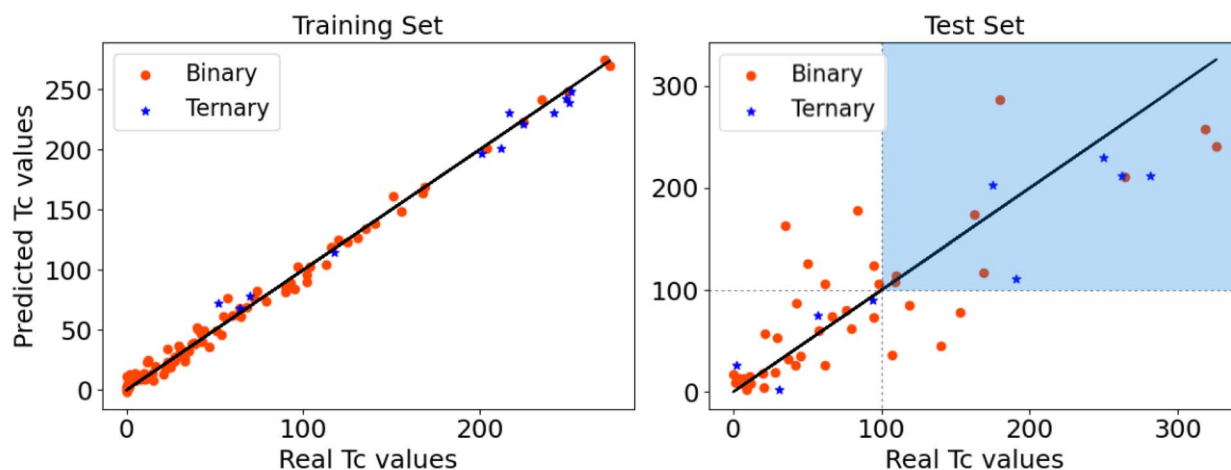
Fig. 10 Predicted T_c (K) values with respect to its reference values in train (left) and test (right) sets, as predicted by the GBR model. The shaded blue rectangle marks the region where $T_c > 100$ K for the test set systems.

in a step-by-step manner, enabling optimization of any differentiable loss function. During each step, a regression tree is trained using the negative gradient of the specified loss function.

We used the GBR algorithm as implemented in the sklearn Python package (version 1.3.0).[41] The squared error was used as a loss function for optimization. The ensemble of the data was split into a training and test set, corresponding to 67% and 33% of the dataset, respectively. Some parameters of the GBR model were optimized to minimize the root mean squared error on the test set. With this, the learning rate was set to 0.2, the number of estimators to 50, the maximum depth to 4, and the subsample size to 0.5 (leading to a stochastic gradient boosting model).

The results of our model are reported in Fig. 10. No difference is observed in the accuracy of the predictions between the binary and ternary systems. The MAE in the test set is of 33.08 K, lower but of similar magnitude to what was obtained using Symbolic Regression in ref. 18. The GBR method allows to characterize the importance of the descriptors as a percentage, showing that it is H_{DOS} that has the largest influence in the

estimation of T_c, with 45% importance. It is followed by 22% and 21% importance for ϕ^* and ϕ, respectively, whereas for H_f it is only of 11%. Despite the large importance of H_{DOS} in the prediction of T_c within this model, the other descriptors are crucial in the estimation, as there is no clear correlation between H_{DOS} and T_c in the database, as it can be seen in Fig. S4 of the ESI.†

A table containing the reference and predicted critical temperatures for all the systems in the dataset using the different fits is provided in the ESI.†

6   Access to TcESTIME

6.1   TcESTIME: open source code to download

The TcESTIME code is available for download at **https://github.com/juliacontrerasgarcia/Tcestime**. To use it, the ELF cube file and the DOS Quantum ESPRESSO output files are needed as input. A version of TcESTIME that supports VASP output files is currently under development. TcESTIME provides three different estimations of T_c, corresponding to
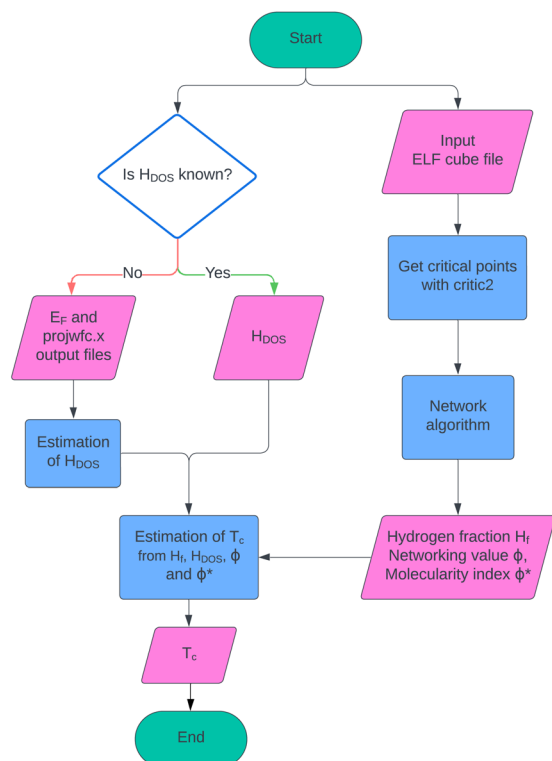
Fig. 11 Workflow of TcESTIME, from the input of the ELF cube file and projected DOS information, to the estimation of T_c using one of the four proposed fits.

the predictions obtained using: (i) the linear fit with respect to Φ_{DOS} (see eqn (9)); (ii) two analytical expressions obtained with Symbolic Regression in ref. 18 (see eqn (6) and (7)); and (iii) the fit obtained with the GBR algorithm. A summary of the

workflow of TcESTIME including input and output is presented in Fig. 11.

6.2 TcESTIMWEB: web server

In order to facilitate the estimation of T_c from a normal DFT calculation, we have also given access to the code *via* a web interface. The TcESTIME web server is an efficient tool for calculating the critical temperature of hydrogen-based systems, supporting Quantum Espresso (QE) and Electron Localization Function data. It can be accessible online at **https://lct-webtools.sorbonne-universite.fr/tcestime/**.

To use web server, the user must prepare the following files:

● A QE compilation archive (.tar.gz) containing the pdos_wfc files and the output file of a QE calculation containing the Fermi energy (*e.g.* nscf.out).

● An ELF file in.cube format.

Once in the platform, the user should navigate to the upload section, provide their name and email address, and upload the ELF and QE files. Upon completion, the user should press the "Go" button to submit the calculations (see Fig. 12a).

After submission, the user will be directed to a waiting page displaying their calculation ID. Processing time may vary depending on file size and complexity, but results are typically generated within seconds. The results page will display four T_c estimates using different methods (Least Squares, GBR, SR2, and SR4) (see Fig. 12b). For example, for the H_3S crystal, where the reference is $T_c = 225$ K, the estimated values are:

$$\text{Least squares: } T_c^{LS} = 181.99 \text{ K} \tag{10}$$

$$\text{GBR: } T_c^{GBR} = 222.75 \text{ K} \tag{11}$$

$$\text{SR2: } T_c^{SR2} = 130.49 \text{ K} \tag{12}$$

$$\text{SR4: } T_c^{SR4} = 185.33 \text{ K} \tag{13}$$

(a) Upload screen

(b) Results screen

Fig. 12 Example calculation on the web server: (a) the initial upload screen is displayed, where the user must provide the files needed for the calculation; (b) the results screen shows one of the four estimations (the other three are omitted in this example).

© 2025 The Author(s). Published by the Royal Society of Chemistry

Finally, a link to the results page is provided *via* email for future access. It should be noted that this content is only accessible for 24 hours, after which the data is deleted from the server.

7 Conclusions

We have presented TcESTIME and its web access, TcES-TIMWEB, a program that does a fast prediction of the critical temperature of potential hydrogen-rich superconductors, by exploiting the relationship between T_c and four key quantities: H_f, H_{DOS}, ϕ, and ϕ^*. The main focus of this work has been the automatization of the evaluation of the networking value, which is done by interfacing the search of critical points and gradient paths of the ELF using critic2, with a newly-developed tree algorithm. The use of complex networks in the latter allows to reduce the analysis to only one unit cell, considerably saving computational time.

The performance of the new computational tool in the original dataset of binary hydrides with respect to the reference values (obtained by visual inspection) was satisfactory, obtaining almost identical results, and only failing to provide an output in 3 systems out of 129 (roughly 2% of them). The new computational tool was also tested on the set of 21 ternary compounds of ref. 18. A new linear fit between T_c and the values of Φ_{DOS} obtained with TcESTIME was proposed, in order to provide a consistent framework for the prediction of T_c simultaneously for binary and ternary compounds.

Moreover, thanks to the new code, a new fit for T_c was proposed using the GBR machine learning technique. Its performance in the test set, corresponding to 33% of the (binary and ternary) hydride dataset presented a MAEs of *ca.* 33 K. This fit allows for a more transferable prediction of T_c, and has the best predictions obtained from the ELF and DOS descriptors up to date.

All in all, the TcESTIME program is a tool that was highly needed to pursue the high-throughput prediction of new hydrogen-rich superconducting materials through the analysis of the ELF. Noticeably, it can certainly highly speed-up the search for good superconductors by experimentalists or theoretical physicists that would like to perform experiments/ expensive calculations only on promising systems.

Data availability

The TcESTIME and the web server can be found in the following links: **https://github.com/juliacontrerasgarcia/Tcestime**, **https://lct-webtools.sorbonne-universite.fr/tcestime/**.

Author contributions

T. N. developed the main code, was in charge of formal analysis and writing. M. d. M. took part in the methodology, doing most calculations on ternary systems. D. I. developed the web application. K. E. H. and N. S. were in charge of methodology, developing the machine learning model. Y. M. contributed in the conception of the algorithm, the formal analysis and

participated in the review and edition of manuscript. J. C. G was in charge of the conceptualization, the formal analysis and writing the manuscript.

Conflicts of interest

There are no conflicts to declare.

Notes and references

‡ The molecularity index is the highest value of the $(3, -1)$ points of the ELF that split all hydrogens into separate f-domains.

§ In the original reference, 132 binary systems are computed, but three of those have been taken out due to irreproducibility of the original results: $C2/c$-KH$_6$ at 166 GPa, $Pmmn$-SbH$_3$ at 300 GPa, and $P6_3/mmc$-HTe at 300 GPa.

1 A. Raza and S. Ali, Superconductors for Magnetic Imaging Resonance Applications, *Mater. Res. Found.*, 2022, **132**, 230–255.

2 Y. S. Greenberg, Application of superconducting quantum interference devices to nuclear magnetic resonance, *Rev. Mod. Phys.*, 1998, **70**, 175.

3 J.-P. Adriaanse, Advances in Imaging and Electron Physics, *Elsevier*, 2019, **209**, 215–338.

4 D. Williams and C. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science; Cambridge library collection*; Springer, vol. 1, 2009.

5 P. Bruzzone, W. H. Fietz, J. V. Minervini, M. Novikov, N. Yanagi, Y. Zhai and J. Zheng, High temperature superconductors for fusion magnets, *Nucl. Fusion*, 2018, **58**, 103001.

6 N. Mitchell, *et al.*, Superconductors for fusion: a roadmap, *Supercond. Sci. Technol.*, 2021, **34**, 103001.

7 F. N. Werfel, U. Floegel-Delor, R. Rothfeld, T. Riedel, B. Goebel, D. Wippich and P. Schirrmeister, Superconductor bearings, flywheels and transportation, *Supercond. Sci. Technol.*, 2011, **25**, 014007.

8 P. Bernstein and J. Noudem, Superconducting magnetic levitation: principle, materials, physics and models, *Supercond. Sci. Technol.*, 2020, **33**, 033001.

9 J. Clarke and F. K. Wilhelm, Superconducting quantum bits, *Nature*, 2008, **453**, 1031–1042.

10 H.-L. Huang, D. Wu, D. Fan and X. Zhu, Superconducting quantum computing: a review, *Sci. China Inf. Sci.*, 2020, **63**, 1–32.

11 J. A. Flores-Livas, L. Boeri, A. Sanna, G. Profeta, R. Arita and M. Eremets, A perspective on conventional high-temperature superconductors at high pressure: Methods and materials, *Phys. Rep.*, 2020, **856**, 1–78.

12 A. P. Drozdov, M. I. Eremets, I. A. Troyan, V. Ksenofontov and S. I. Shylin, Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system, *Nature*, 2015, **525**, 73.

13 A. Drozdov, P. Kong, V. Minkov, S. Besedin, M. Kuzovnikov, S. Mozaffari, L. Balicas, F. Balakirev, D. Graf and V. Prakapenka, others Superconductivity at 250 K in lanthanum hydride under high pressures, *Nature*, 2019, **569**, 528–531.

14 P. Kong, V. S. Minkov, M. A. Kuzovnikov, A. P. Drozdov, S. P. Besedin, S. Mozaffari, L. Balicas, F. F. Balakirev, V. B. Prakapenka, E. Greenberg and M. I. Eremets, Superconductivity up to 243 K in yttrium hydrides under high pressure, *Nat. Commun.*, 2021, **12**, 5075.

15 R. H. Lavroff, J. Munarriz, C. E. Dickerson, F. Munoz and A. N. Alexandrova, Chemical bonding dictates drastic critical temperature difference in two seemingly identical superconductors, *Proceedings of the National Academy of Sciences*, 2024, **121**, e2316101121.

16 X. Li, X. Zhang, Y. Liu and G. Yang, Bonding-unsaturation-dependent superconductivity in P-rich sulfides, *Matter Radiat. Extremes*, 2022, **7**, 048402.

17 F. Belli, T. Novoa, J. Contreras-García and I. Errea, Strong correlation between electronic bonding network and critical temperature in hydrogen-based superconductors, *Nat. Commun.*, 2021, **12**, 5381.

18 M. E. di Mauro, B. Braïda, I. Errea, T. Novoa and J. Contreras-García, *Molecularity: a fast and efficient criterion for probing superconductivity*, 2024.

19 A. Wilver, C. C. Muriel, N. Trinidad and J. Contreras-García, Introducing electron correlation in solid-state calculations for superconducting states, *Faraday Discuss.*, 2024, 1–14.

20 A. D. Becke and K. E. Edgecombe, A simple measure of electron localization in atomic and molecular systems, *J. Chem. Phys.*, 1990, **92**, 5397–5403.

21 C. F. von Weizsäcker, Zur Theorie der Kernmassen, *Z. Phys.*, 1935, **96**, 431–458.

22 A. Savin, A. D. Becke, J. Flad, R. Nesper, H. Preuss and H. G. von Schnering, A New Look at Electron Localization, *Angew Chem. Int. Ed. Engl.*, 1991, **30**, 409–412.

23 R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford Univ. Press, 1990.

24 M. Calatayud, J. Andrés, A. Beltrán and B. Silvi, *Theor. Chem. Acc.*, 2004, **112**, 453.

25 P. G. Mezey, Quantum chemical shape: new density domain relations for the topology of molecular bodies, functional groups, and chemical bonding, *Can. J. Chem.*, 1994, **72**, 928–935.

26 A. Denchfield, H. Park and R. J. Hemley, Electronic structure of nitrogen-doped lutetium hydrides, *Phys. Rev. Mater.*, 2024, **8**, L021801.

27 A. Denchfield; H. Park; R. J. Hemley, Designing Quaternary Hydrides with Potential High T_c Superconductivity, *arXiv*, preprint, arXiv:2403.01688, 2024,.

28 A. Otero-de-la-Roza, M. A. Blanco, Á. M. Pendás and V. Luaña, Critic: a new program for the topological analysis of solid-state electron densities, *Comput. Phys. Commun.*, 2009, **180**, 157–166.

29 A. Otero-de-la-Roza, E. R. Johnson and V. Luaña, Critic2: A program for real-space analysis of quantum chemical interactions in solids, *Comput. Phys. Commun.*, 2014, **185**, 1007–1018.

30 J. Contreras-Garcia, A. M. Pendas, B. Silvi and J. M. Recio, Computation of local and global properties of the ELF topology in crystals, *J. Chem. Theory Comput.*, 2009, **113**, 1068.

31 P. Giannozzi, *et al.*, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter*, 2009, **21**, 395502.

32 P. Giannozzi, *et al.*, Advanced capabilities for materials modelling with Quantum ESPRESSO, *J. Phys.: Condens. Matter*, 2017, **29**, 465901.

33 B. Silvi and A. Savin, Classification of chemical bonds based on topological analysis of electron localization functions, *Nature*, 1994, **371**, 683.

34 J. Contreras-García, P. Mori-Sánchez, B. Silvi and J. M. Recio, A Quantum Chemical Interpretation of Compressibility in Solids, *J. Chem. Theory Comput.*, 2009, **5**, 2108–2114.

35 M. Kohout and A. Savin, *Int. J. Quantum Chem.*, 1996, **60**, 875–882.

36 T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms*, The MIT Press, Cambridge, Massachusetts, 3rd edn, 2009, p. 594.

37 H. Liu, I. I. Naumov, R. Hoffmann, N. W. Ashcroft and R. J. Hemley, Potential high-Tc superconducting lanthanum and yttrium hydrides at high pressure, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 6990–6995.

38 G. Wu, Y. Ma, H. Yu, D. Li, B. Liu and T. Cui, Prediction of stoichiometric poh n compounds: crystal structures and properties, *RSC Adv.*, 2015, **5**, 103445–103450.

39 D. Duan, F. Tian, Y. Liu, X. Huang, D. Li, H. Yu, Y. Ma, B. Liu and T. Cui, Enhancement of T_c in the atomic phase of iodine-doped hydrogen at high pressures, *Phys. Chem. Chem. Phys.*, 2015, **17**, 32335–32340.

40 J. H. Friedman, Greedy function approximation: a gradient boosting machine, *Ann. Math. Stat.*, 2001, 1189–1232.

41 F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.