Molecularity: a fast and efficient criterion for probing superconductivity

Matías E. di Mauro,[1] Benoît Braïda,[1] Ion Errea,[2,3,4] Trinidad Novoa,[1,*] and Julia Contreras-García[1,†]

[1]*Laboratoire de Chimie Théorique, Sorbonne Université & CNRS, 4 Pl. Jussieu, 75005, Paris, France*
[2]*Fisika Aplikatua Saila, Gipuzkoako Ingeniaritza Eskola, University of the Basque Country (UPV/EHU), Europa Plaza 1, 20018 Donostia/San Sebastián, Spain*
[3]*Centro de Física de Materiales (CSIC-UPV/EHU), Manuel de Lardizabal Pasealekua 5, 20018 Donostia/San Sebastián, Spain*
[4]*Donostia International Physics Center (DIPC), Manuel de Lardizabal Pasealekua 4, 20018 Donostia/San Sebastián, Spain*

(Dated: October 29, 2024)

We present an efficient criterion for doing fast estimations of the critical temperature of hydrogen based superconductors. We start by expanding the applicability of 3D descriptors of electron localization to superconducting states within the framework of superconducting DFT. We first apply this descriptor to a model system, the hydrogen chain, which allows to prove two main concepts: i) that the electron localization changes very little when the transition from the normal to the superconducting state takes place, i.e. that it can be described at the DFT level from the normal state; and ii) that the formation of molecules can be characterized within this theoretical framework, enabling to quickly filter out systems with marked molecular character and hence with low potential to be good superconductors. These two ideas, are then exploited in real binary and ternary systems, showing i) that the bonding type can be characterized automatically; and ii) that this provides a new index which enables to feed machine learning algorithms for a better prediction of critical temperatures. Overall, this sets a grounded theoretical scenario for an automatic and efficient high-throughput screening of potential hydrogen based superconductors.

Keywords: superconductivity, electron localization, molecularity

I. INTRODUCTION

Superconductivity can be considered among the most exciting discoveries in material science of the XXth century due to its implications both at the technological and scientific levels. These implications have led to the discovery of a plethora of superconducting families to which the high pressure hydrides have been added in the last years.

Few hydrate examples are H_3S [1], YH_9 [2], YH_6 [3], and LaH_{10} [4]; all reaching critical temperatures (T_c) well above 200 K at megabar pressures. However, the hard conditions for the synthesis as well as the difficult experimental characterization, make the statement of new high T_c materials difficult from the experimental viewpoint. In this panorama, theory has become a trustworthy diagnosis of hydride superconductivity. For instance, this approach has been successful in describing the important nuclear quantum effects present in hydrogen-rich compounds, which in turn affects its superconducting properties. [5] Nevertheless, this comes at a high computational cost. As an example, the calculation of the T_c of LaH_{10} within the anharmonic approximation takes hundreds of thousands of CPU hours.

Given the high critical temperatures that hydrides have shown to have, the search of high T_c hydride superconductors is still well alive, but it is claiming for a more efficient theoretical approach allowing an efficient scanning of new potential superconductors.

A faster alternative would be to find cheap footprints of high-temperature superconductivity. A full list of them should include vibrational properties as well as electronic, as superconductivity in hydrogen-based systems has an origin in electron-phonon coupling. Here, we focus exclusively on the electronic properties, which have already demonstrated pattern similarities across certain high-T_c structures [6–9]. This allows us to identify promising structures and give first approximations of T_c, without the significant computational expense. If we collect the main characteristics that have been put together over the years, we have some chemical/structural features, e.g. hydrogen-rich systems mixed with s and p elements and highly symmetrical structures favour high T_c. From the electronic structure viewpoint, materials with a high density of states (DOS) at the Fermi level are the best candidates for high-temperature superconductivity [6]. Looking at the normal state-superconductor transition, it has been proposed that the mechanism of superconductivity can be traced to the the formation of electronic pairs be it in the shape of strongly covalent metallic bonds (MgB_2) [10] or lone pairs (Te)[11].

But, even if these features can suggest good trends, they provide necessary but not sufficient conditions. Trying to find a sufficient condition, some of the authors have recently shown that a correlation exists between the critical temperature and the delocalization channels at the density functional theory (DFT) level in binary superconducting hydrides [6]. These channels are determined thanks to the Electron Localization Function (ELF),[12] and a simple quantitative description of them can be done with the *networking value*, a topological descriptor stemming from the ELF.

Nevertheless, this initial proposal was not absent of limi-

* Correspondence email address: trinidad.novoa_aguirre@sorbonne-universite.fr
† Correspondence email address: contrera@lct.jussieu.fr

tations. The fact that DFT calculations can be used for describing the onset of superconductivity needs to be understood. The inherent nature of the networking value, although intuitive, needs to be further explored. This is so much so, if we take into account that for a high throughput exploitation of this index, it is necessary to make sure that it is applicable to more complex systems (ternary, quaternaries, etc) and that no information is missing in the correlation.

These points will be addressed here. Firstly, we will dwell on the use of DFT for describing electron localization in superconductors. With this aim in mind, we will develop a new formulation of the ELF within the superconducting DFT framework and we will apply it to a model system. This will allow us to prove that the normal state DFT analysis of the ELF is sufficient to describe that of the superconducting state. It will also allow us to identify other bonding features, such as the formation of molecules, which quantitatively characterize bad superconductors. With these tools at hand, we will prove in a set of binary and ternary compounds that the new index allows to i) complement the networking value in more complex systems, and ii) improve the fast prediction of T_c; with an special focus in high T_c superconductors (more difficult to predict due to the lack of data).

II. THEORETICAL BACKGROUND

A. The Electron Localization Function (ELF)

The electron localization function was first introduced by Edgecome and Becke to identify regions of localized same-spin electron pairs, or groups of them, in atomic and molecular systems.[13] It is based on the same-spin pair probability as approximated in Hartree-Fock (i.e. considering exchange),

$$P_2^{\sigma\sigma}(r_1, r_2) = \rho_\sigma(r_1)\rho_\sigma(r_2) - \left|\rho_1^\sigma(r_1, r_2)\right|^2 . \quad (1)$$

Here, $\rho_1^\sigma(r_1, r_2)$ is the spin-resolved 1-RDM, and $\rho^\sigma(r_1)$ its diagonal, which corresponds to the electron density for σ-spin. If we assume that there is one σ electron at r_1, we can express the probability of finding another electron with the same spin at r_2 by

$$P_{cond}^{\sigma\sigma}(r_1, r_2) = \frac{P_2^{\sigma\sigma}(r_1, r_2)}{\rho_\sigma(r_1)} = \rho_\sigma(r_2) - \frac{\left|\rho_1^\sigma(r_1, r_2)\right|^2}{\rho_\sigma(r_1)} , \quad (2)$$

which we call the conditional same-spin pair probability. Fixing one of the electrons lets us study the behavior of this probability when $r_2 \to r_1$. Changing to the spherically averaged version of $P_{cond}^{\sigma\sigma}$, that depends on the coordinates (r, s), where r is the reference point and s a distance from it, and doing a Taylor expansion, we can show that

$$P_{cond}^{\sigma\sigma}(r, s) = \frac{1}{3}\left[\sum_i^\sigma |\nabla\phi_i(r)|^2 - \frac{1}{4}\frac{|\nabla\rho_\sigma(r)|}{\rho_\sigma(r)}\right] s^2 + \dots \quad (3)$$

Here, $\phi_i(r)$ are the HF orbitals, and the sum $\sum_i^\sigma$ means that we are only considering the orbitals containing electrons of spin σ.

From (3), we recognize the term in brackets,

$$D_\sigma(r) = \sum_i^\sigma |\nabla\phi_i(r)|^2 - \frac{1}{4}\frac{|\nabla\rho_\sigma(r)|}{\rho_\sigma(r)} , \quad (4)$$

as a measure of localization, as it is the leading term for small distances s between the electrons. This is, when D_σ is small, the probability of finding a σ electron very close to the reference one is also small. This means that the reference electron is very localized, and so is the Fermi hole that comes with it, not allowing a same-spin electron to come near. However, an opposite-spin electron is likely to localize in the same region. For a closed-shell system, we have $\rho(r) = 2\rho_\sigma(r)$, and it is possible to define the spinless quantity

$$D(r) = \frac{1}{2}\sum_i |\nabla\phi_i(r)|^2 - \frac{1}{8}\frac{|\nabla\rho(r)|^2}{\rho(r)} . \quad (5)$$

The function $D(r)$ being opposite to localization, we introduce the Electron Localization Function (ELF),

$$\eta(r) = \left[1 + \left(\frac{D(r)}{D_0(r)}\right)^2\right]^{-1} , \quad (6)$$

where

$$D_0(r) = \frac{3}{10}(3\pi^2)^{2/3}\rho(r)^{5/3} \quad (7)$$

is the term $D(r)$ as evaluated for a uniform electron gas in 3D. Note that in the case of a 1D system -as used in part of this contribution- it becomes $D_0(x) = \pi^2/24\,\rho(x)^3$ [14]. The normalization allows to compare the values of the kernel $D(r)/D_0(r)$ of different systems. Further, the Lorentzian transformation applied on that kernel in the definition of $\eta(r)$ in eq. (6) allows us to have a function that ranges between 0 and 1, that is high in the regions of high localization ($\eta \to 1$), and conserves the topology of the kernel. This, as we shall see, will be very important in the description of electron localization.

A close inspection of eq. (5) allows to identify the first and second terms as the kinetic energy density of the system, $\tau(r)$, and its form in the von-Weizsacker approximation, $\tau_{vW}(r)$ [15], leading to $D(r) = \tau(r) - \tau_{vW}(r)$. This formulation allows to compute the ELF beyond the HF approximation [12]. It also introduces a new interpretation: because the von-Weizsäcker kinetic energy is exact for a bosonic system of the same density $\rho(r)$, the term $D(r)$ is a local measure of the excess kinetic energy due to the fermionic nature of the electrons. If this quantity is high, it means that electron pairs are delocalized in that region, and the ELF value will be small. If the kinetic energy density is not locally increased

as an effect of the Pauli exclusion principle, we say that electrons are localized, which will be reflected on a high value of the ELF.

Some of the authors have shown that the ELF can be used to classify bonds in binary supercoductors in six distinct families: molecular systems, covalent systems, systems influenced by weak covalent hydrogen-hydrogen interactions, systems exhibiting electride behavior, ionic systems, and isolated systems. In each instance, the bond nature is identified through analyzing ELF saddle points between different atoms. Moreover, we also found that the value of the ELF at the saddle point which leads to a surface revealing a 3D delocalization through the cell (hereafter called the "networking value") correlates with the critical temperature of superconductors [6].

B. Superconducting Density Functional Theory

The widespread use of DFT for electronic structure calculations, due to its great compromise between accuracy and computational time, has served as a motivation to extend it to a wider variety of systems. The case of superconductors is particular in the sense that it cannot be solved in a perturbative fashion. In fact, this is so because in those systems the phase symmetry is broken, which implies that the number of particles will not be conserved. Superconducting DFT (SC-DFT) successfully treats this problem, [16–20] and accurately reproduces the experimental T_c's of conventional superconductors without introducing any empirical parameters [21].

In SC-DFT, the Hamiltonian comprises an anomalous external potential, $\Delta^*_{ext}(r, r')$, that takes into account the symmetry breaking by allowing Cooper pairs to tunnel in and out of the system

$$\hat{H}_{\Delta_{ext}} = \int \Delta^*_{ext}(r, r')\psi_\uparrow(r)\psi_\downarrow(r')\, dr\, dr' + h.c. \tag{8}$$

where $\psi_\sigma(r)$ are electronic field operators. If we let $\Delta^*_{ext}(r, r')$ go to zero, the Hamiltonian converges to a non-superconducting one, i.e. that of a normal state system. In this way, SCDFT allows to *turn on and off* superconductivity, and to compare the system's properties when it is in the normal or in the superconducting state.

Another feature of SC-DFT is that it considers ionic degrees of freedom explicitly. Therefore, the Hohenber-Kohn theorems in this framework establish a one-to-one mapping between three external potentials and their corresponding densities. Those potentials are: $v_{ext}(r)$, that couples to electrons; $W_{ext}(R)$, that couples to ions; and the aforementioned anomalous potential.

The theory is formulated in the grand-canonical ensemble, as a consequence of the non-particle conserving Hamiltonian in (8). Thus, the variational quantity is the grand-canonical potential, and the electron density is

$$\rho^{SC}(r) = \left\langle \sum_\sigma \psi^\dagger_\sigma(r)\psi_\sigma(r) \right\rangle = Tr\left[\hat{\rho}_0 \sum_\sigma \psi^\dagger_\sigma(r)\psi_\sigma(r)\right], \tag{9}$$

where the *SC* superscript is used to differentiate it from the normal state electron density. We have considered the grand-canonical density matrix,

$$\hat{\rho}_0 = \frac{e^{-\beta(\hat{H}-\mu\hat{N})}}{Tr[e^{-\beta(\hat{H}-\mu\hat{N})}]}, \tag{10}$$

with $\hat{N}$ the number operator.

The other two densities are that of the ions, $\Gamma(\{R_i\})$, and the anomalous density, defined as

$$\chi(r, r') = Tr\left[\hat{\rho}_0\, \psi_\uparrow(r)\psi_\downarrow(r')\right]. \tag{11}$$

This is a two-body object that is responsible of measuring the probability of the appearance of Cooper pairs, and it is the order parameter of the transition.

A Kohn-Sham (KS) scheme leads to electronic equations that yield a Hamiltonian of the form

$$\begin{aligned} \hat{H}_s &= \sum_\sigma \int dr\, \psi^\dagger_\sigma(r)\left[-\frac{\nabla^2}{2} + v_s(r) - \mu\right]\psi_\sigma(r) \\ &+ \int dr dr'\, [\Delta^*_s(r, r')\psi_\uparrow(r)\psi_\downarrow(r') + h.c.], \end{aligned} \tag{12}$$

where $v_s(r)$ is the usual electronic KS potential, and $\Delta^*_s(r, r')$ the mean-field version of the anomalous potential. The KS energies in the superconducting state become

$$E_{n\mathbf{k}} = \sqrt{\xi^2_{n\mathbf{k}} + \Delta^2_{s,n\mathbf{k}}}, \tag{13}$$

with $\xi_{n\mathbf{k}}$ the KS energies of the normal state. We hereby refer to the anomalous potential to the superconducting gap, based on this result.

Diagonalizing the Hamiltonian in (12), the electron density in (9) becomes

$$\rho^{SC}(r) = \sum_{nk}\left[1 - \frac{\xi_{nk}}{|E_{nk}|}\tanh\left(\frac{\beta|E_{nk}|}{2}\right)\right]|\varphi_{nk}(r)|^2, \tag{14}$$

where $\varphi_{nk}(r)$ are the Kohn-Sham orbitals of the normal state. Note that in the normal state limit, where $\Delta_{s,n\mathbf{k}} \to 0$, we recover the normal state density,

$$\rho^{\mathrm{NS}}(r) = 2\sum_{nk} f(\xi_{nk})|\varphi_{nk}(r)|^2, \tag{15}$$

with $f(E_i) = (1 + e^{\beta E_i})^{-1}$ the Fermi-Dirac distribution. This point is crucial for the analysis of the transition.

III. RESULTS

A. Theoretical formulation

The SC-DFT framework allows us to define real-space descriptors other than the electron density in (14). In analogy with the derivation of that density in Ref. [19], one can arrive to an expression for the superconducting one-reduced density matrix (1-RDM):

$$\rho_1^{\mathrm{SC}}(r,r') = \sum_{nk} n_{nk}^{SC} \varphi_{nk}^*(r)\varphi_{nk}(r')\,, \tag{16}$$

which is written in terms of Kohn-Sham orbitals and SC occupations,

$$n_{nk}^{SC} \equiv 1 - \frac{\xi_{nk}}{|E_{nk}|}\tanh\left(\frac{\beta|E_{nk}|}{2}\right). \tag{17}$$

Notice that in the normal state limit, we recover

$$\rho_1^{\mathrm{NS}}(r,r') = 2\sum_{nk} f(\xi_{nk})\varphi_{nk}^*(r)\varphi_{nk}(r')\,. \tag{18}$$

It can be seen that within the SC-DFT framework, the difference of the 1-RDM of both states is only reflected in a change of the occupation numbers.

With the density matrix in (16), we can compute the kinetic energy of the system,

$$T^{SC} = -\frac{1}{2}\int \nabla_{r'}^2 \rho_1^{SC}(r,r')\big|_{r'=r}\, dr\,, \tag{19}$$

$$= \frac{1}{2}\sum_{nk} n_{nk}^{SC}\int |\nabla\varphi_{nk}(r)|^2\, dr\,, \tag{20}$$

letting us define a positive definite kinetic energy density,

$$\tau^{SC}(r) \equiv \frac{1}{2}\sum_{nk} n_{nk}^{SC}|\nabla\varphi_{nk}(r)|^2\,. \tag{21}$$

The von Weizsäcker and Thomas-Fermi KEDs are obtained from the SC density,

$$\tau_{vW}^{SC}(r) = \frac{1}{8}\frac{\left|\nabla\rho^{SC}(r)\right|^2}{\rho^{SC}(r)}\,, \tag{22}$$

$$\tau_{TF}^{SC}(r) = \frac{3}{10}(3\pi^2)^{2/3}\rho^{SC}(r)^{5/3}\,. \tag{23}$$

Note that here $\tau_{TF}^{SC}(r)$ is defined in 3D. Finally, equations (21), (22) and (23) allow us to define the ELF for the superconducting state as:

$$\eta^{SC}(r) = \left[1 + \left(\frac{\tau^{SC}(r) - \tau_{vW}^{SC}(r)}{\tau_{TF}^{SC}(r)}\right)^2\right]^{-1}\,. \tag{24}$$

As all the superconducting quantities defined in this framework, the superconducting ELF (SC-ELF) converges to the (temperature-dependent) normal state ELF when the gap goes to zero.

In order use to the expression in eq. (24) to examine the SC-ELF in a model or in a real system, it is necessary to have an expression for the gap. In SC-DFT, this is done through a connection to many-body perturbation theory, where one can use Green's functions to account for electron-phonon coupling [17]. It is possible, however, to introduce an approximation and represent the dependence of the gap at zero Kelvin with respect to the energies ξ as an isotropic Lorentzian function [22],

$$\Delta_0(\xi) = \frac{\Delta_0}{N_0\pi}\,\frac{\omega/2}{\xi^2 + (\omega/2)^2}\,, \tag{25}$$

where ω is a parameter that adjusts the width of the peak, and N_0 is a normalization such that the height of the peak at $\xi = 0$ is Δ_0. The latter is the constant of the gap at $T = 0$ K in BCS [23], that depends on the critical temperature, $\Delta_0 = 1.76 k_B T_c$, with k_B being Boltzmann's constant. Then, considering the dependence of the gap with respect to the temperature [22, 24], we shall use

$$\Delta(\xi;T) = \Delta_0(\xi)\tanh\left(1.74\sqrt{\frac{T_c - T}{T}}\right). \tag{26}$$

so that in our model, the gap can be obtained for any temperature and energy, for a given T_c. The profile of the gap for different values of those parameters are presented in Figure S5. Note that this approach is also applicable to approximate SC-ELF in DFT calculations of solid systems.[25]

B. One-dimensional model chain

In order to analyze the behavior of SC-ELF we will first apply it to a simple model, the 1D hydrogen chain. This calls for the one-dimensional definition of τ_{TF} presented above. Note that in order to make sure that these results are representative, a careful analysis of the parametric space has been carried out (for a full analysis see [26]).

For a fixed critical temperature, the left panel of Fig. 1 shows how the occupancies of the normal state deviate from the step function as the temperature increases, softening the transition around the Fermi energy. Meanwhile, the superconducting occupation numbers do not suffer big alterations with the temperature, as can be seen on the right panel of Fig. 1. In fact, they tend to resemble the occupations at the critical temperature (in both states, as they are the same), showing the larger correlation of the superconducting state in comparison with the normal state below T_c. This is repeated for other critical temperatures (see Fig. S7). Hence, we will hereby take T_c=300 K for the analysis, and seize the effect of the changes in those functions with respect to the temperature, T.

In the following, we denote the normal and superconducting states as NS and SC, respectively. Taking the results for

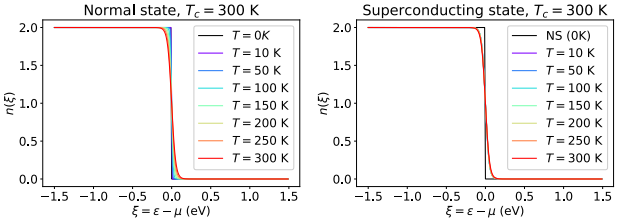


Figure 1. Occupation numbers at different energies. In all panels, that of the normal state at $T = 0$ K is depicted in black. To the left, the occupation numbers of the normal state at different temperatures. To the right, the same is displayed for the superconducting state with T_c=300 K.

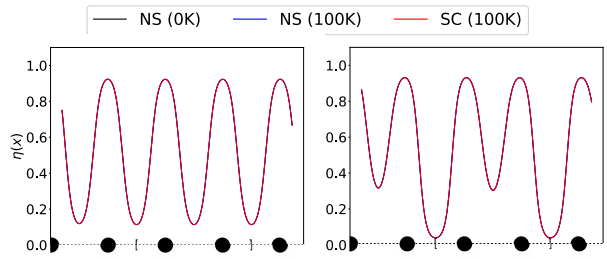


Figure 2. Profile of the ELF in the hydrogen chain for $T_c = 300$ K. Left: metallic, right: dimerized. In both panels, the three overlapping lines represent the three considered models: normal state at $T = 0$ K and $T = 100$ K, and superconducting state at $T = 100$ K. Atomic positions are marked by black circles, and the unit cell is delimited by squared brackets.

the NS and SC occupation numbers into account it is not surprising that when we analyze the SC-ELF in the homogeneous hydrogen chain we see that it does not differ from the localization in the normal state (both T = 0 K and at T = 100 K are analyzed in Fig. 2, for the symmetric and dimerized chains). This is true for the whole range of temperatures and interatomic distances. An analysis on how the negligible differences between the NS and SC real-space descriptors persists in the limit of high-correlation, when the superconducting gap is greatly amplified, is presented in [26]. This shows how the spatial distribution of electrons is resilient below T_c even if the occupation numbers show more sizable changes.

This result has big implications as far as the analysis of pairing in superconductivity is concerned: SC electron localization descriptors can be inferred from the normal state. Hence, with SC-ELF and the hydrogen chain enable us to rationalize the fact that the electron pairing can be obtained for superconductors at the DFT level from the normal state.

C. Introduction of the "Molecularity" index

The 1D model also enables to understand the effect of bonding in metallization. When the distance between the two atoms inside a unit cell is changed, dimerization is simulated (see Fig. 3). The profile of the ELF reveals the molecularization of the system as two of the hydrogens approach each other, with the appearance of two ELF local minima, a higher minimum between the two hydrogen forming a dimer and a lower minimum occurring between two unit cells (at mid-distance of two dimers). The minimum value of the ELF, ϕ, drops to nearly zero when slightly shifting the distances. Note that this ELF value would correspond to the networking value of the 1D-chain, as introduced by the authors in Ref. [6]. Since this value was found to correlate with T_c, this would mean that the molecularity hampers superconductivity. This result agrees with the proposal by Ackland *et al.* from MD simulations [27] and the lower T_c found for molecular systems in [6], in comparison with that of systems presenting other types of bonding.

Simultaneously, the value of the ELF at the higher local minima between the hydrogens in the same unit cell increases with respect to the symmetric case. This is characteristic of a more delocalized behavior of electrons in the intramolecular region. We shall call the value of the ELF at this local minima ϕ^*, or Molecularity index, as it represents the first characterization of the molecularization effect.

Further decreasing the minimum interatomic distance, d_{HH}, accentuates these changes in the topology of the function. In fact, when d_{HH} decreases sufficiently, the atoms in the lattice form units resembling H_2 molecules, characterized by flat ELF profiles within the molecule.

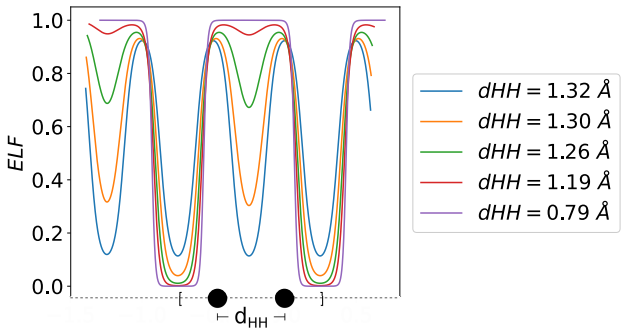


Figure 3. Normal state ELF profile for the dimerized hydrogen chain at $T = 100$ K, considering different minimum interatomic distances, d_{HH}.

As in the Su-Schrieffer-Heeger (SSH) model, a gap opens upon dimerization. Hence, we have also plotted the evolution of the topological descriptors ϕ and ϕ^* with respect to the energy gap (see Fig. S12). It can be seen that a large value for ϕ^* is a feature of an insulating state in the hydrogen chain, where intramolecular distances are shortened. This is

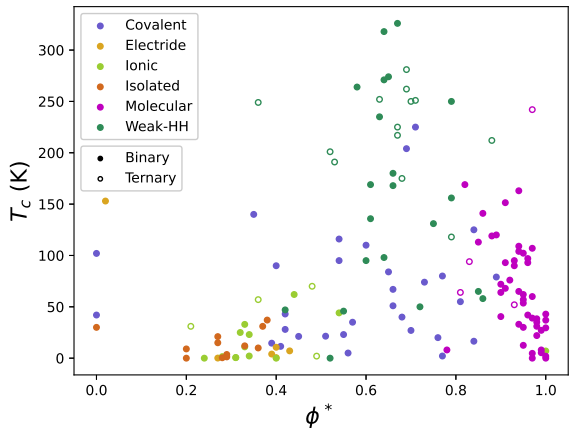


Figure 4. Reference critical temperature T_c (K) with respect to the molecularity index ϕ^* for all binary and ternary data, classified by bonding type families.

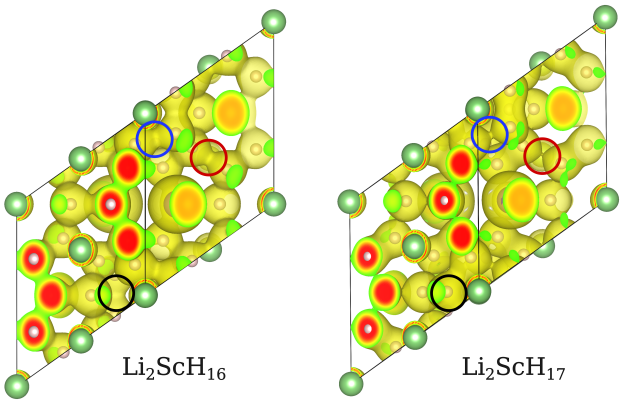


Figure 5. Isosurface of ELF= 0.57 for Li_2ScH_{16} (left, $\phi = 0.63$) and Li_2ScH_{17} (right, $\phi = 0.57$) at 300 GPa. The black circle marks the place where the extra H is added. The blue and red circles show how the values of the ELF at the critical points change with this addition, becoming more prone to form molecular units.

further supported by the increase of the localization in the ELF basins with the decreasing of the intramolecular distance, as it can be seen in Figure 3.

With this quantitative characterization of molecularity in the model, it is then possible to envisage its characterization in real systems at the DFT level. We define the molecularity index in a 3D system, ϕ^*, as the maximum value of the ELF function for which at least two hydrogen atoms become connected. In molecular systems, this will necessarily correspond to the value of the ELF for which molecular units appear, so that the number of atoms inside isosurfaces will be two.

In order to test its use in complex systems, we have calculated it along with the networking value for a set of 129 binary and 21 ternary compounds (see [26] for details). As expected, ϕ^* is high for molecular systems, ranging between 0.8 and 1.0, where no other type of bonding family is present (Fig. 4). We notice that the two bond categories that are most likely to have a high T_c, namely covalent and weak H-H, are dominant in the region where $\phi^* \in [0.45, 0.8]$, meanwhile other types of bonding show generally low critical temperatures. In other words, the molecularity index is a quantitative tool to separate the families of interest. Thus, the molecularity index allows an automated characterization of bonding type, and hence of potential high T_c superconductors.

Moreover, a careful inspection of the new systems clearly shows that the molecularity index is necessary as the complexity of the systems increase, enabling to differentiate systems with similar networking value but very different T_c. This is the case of Li_2ScH_{16} at 300 GPa, with $\phi = 0.63$ and $T_c = 281$ K, and Li_2ScH_{17} at the same pressure, with $\phi = 0.57$ and $T_c = 94$ K. The severe drop of the critical temperature upon inclusion of one extra hydrogen atom is understood when one examines the ELF isosurfaces: the hydrogen atoms rearrange to shorten the minimum distance between them, and as a result they form molecular units (see Fig. 5). This can be measured by an increase of the molecularity index from $\phi^* = 0.69$ in Li_2ScH_{16} to $\phi^* = 0.83$ in Li_2ScH_{17}, the latter being above the threshold of $\phi^* = 0.8$ for molecular systems. As proved in the 1D chain, this is detrimental for high-temperature superconductivity [6, 28]. Hence the new index complements the networking value when going to complex systems, enabling to characterize complex ternary superconductors.

D. Fast estimation of critical temperatures

With the new derived index, it becomes then necessary to propose a new expression to calculate T_c that works both for binary and ternary compounds. To do so, we use Symbolic Regression (SR), implemented in PySR [29]. It corresponds to an evolutionary algorithm where individuals are mathematical expressions that are optimized to minimize the mean squared error of the evaluated expressions with respect to reference data. The output of our SR models are formulas for T_c, thus providing a way to do Machine Learning while retaining the scientific insight that a mathematical formula bears, with new fits that lead to better accuracy and wider applicability.

The details of the parameters used for the SR are presented in the Supplemental Material [26]. Among the several expressions that were obtained to fit the reference T_c, those that yield the lowest errors in the test set, that we call SR1 and SR2, are

$$T_c^{SR1} = 382.5\,(1 - \Delta\phi) H_f H_{DOS}\,, \quad (27)$$

$$T_c^{SR2} = 442.3\,(1 - \Delta\phi) H_f^3 \sqrt{H_{DOS}}\,, \quad (28)$$

where $\Delta\phi = \phi^* - \phi$ takes into account both the networking and the molecularity indexes, H_{DOS} is the fraction of the density of states (DOS) at the Fermi level that correspond to Hydrogen and H_f is the fraction of atoms in the unit cell.

Indeed, this expression is able to differentiate between ternary compounds with similar stoichiometries and different T_c's, leading to MAEs of 38 K and 36 K for eqs. 27 and 28, respectively, in a set of systems that was not used to fit those expressions. Those equations should be compared with the expression from Ref. [6]: $T_c = (750\phi\, H_f \sqrt[3]{H_{DOS}} - 85)K$; with MAE of 55 K in the same dataset. The comparison between the predictions in the test set for SR2 are displayed in Fig. 6-left. The results for the fit SR1 are displayed in Fig. S10. Note that further improvements from these fits can be even achieved from the observation that the estimates also improve in the high T_c region, where predictions are harder due to the little availability of data (see Table S1).

This new approach allows to distinguish between compounds with slightly different hydrogen fractions, as shown for compounds in Fig. 5. This improvement is due to the fact that ϕ^* is able to identify bonding types. Hence, we can also use this ability to filter the data so as to keep the systems with $\phi^* \in [0.45, 0.8]$, which correspond to the bonding families that interest us. Two analytical expressions are proposed here, where the overall errors are much more consistent with those in high-T_c regions, thus being more reliable for our purposes:

$$T_c^{SR3} = 312.0\, H_{DOS}, \tag{29}$$

$$T_c^{SR4} = 574.7\, \phi \sqrt{H_f^3\, H_{DOS}}. \tag{30}$$

The results of the predicted T_c in the test set as obtained with the fit SR4 can be visualized in Fig. 6-right. Results for SR3 are presented in Fig. S11.

For high-throughput analysis, all four models are recommended in order to mutually discard outliers. We believe these new expressions should help in a better prediction and high-throughput analysis of potential high T_c hydrogen based superconductors.

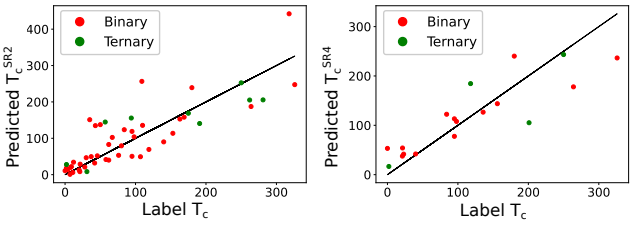


Figure 6. Predicted values of T_c (K) with respect to the reference (label) data, as computes using the fits SR2 (left) and SR4 (right). For the latter, only systems where $\phi^* \in [0.45, 0.8]$ are considered.

IV. CONCLUSIONS

All in all, we have resorted to the SC-DFT approach to develop a new formulation of superconducting ELF in terms of a reorganization of occupation numbers with respect to the normal state. We have first applied these new developments to a model system, leading to two main conclusions. On the one hand, the small changes in occupation numbers lead to small changes in localization from the normal to the superconducting state. Taking into account that most calculations in solid state are carried out within the DFT framework, where 2-body quantities are not usually accessible, having a DFT-based index that only requires Kohn-Sham orbital information enables for a quick screening of the chemistry in potential superconductors. On the other hand, the evolution of ELF upon changes in dimerization prognoses a lowering of the critical temperature upon formation of H_2 molecules. Moreover, the simple picture in the 1D-chain enables to introduce a molecularity index that allows to quantify this process. Building from these results, we have applied the molecularity index to 3D systems and calculated it from a set of binary and ternary systems, showing that i) it allows for the first time to automatically classify the bonding type of these systems and ii) it allows to differentiate tricky situations where molecules appear and previous indexes, developed for binary systems, fail. Hence, this new index goes beyond current proposals (which fail for ternary compounds), allowing for the automatic characterization of complex potential superconducting compounds in a fast manner, with especial emphasis on high T_c prediction. These theoretical advances should help pushing the inverse design of high T_c superconductors to a reliable and cost-efficient limit.

V. ACKNOWLEDGEMENTS

We would like to acknowledge support by ECOS-Sud C17E09 and C21E06, and the Association Nationale de la Recherche under grant ANR-22-CE50-0014. This research was supported by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 810367), project EMC2. I.E. acknowledges funding from ERC under the European Unions Horizon 2020 research and innovation program (Grant Agreements No. 802533 and No. 946629); the Department of Education, Universities and Research of the Eusko Jaurlaritza, and the University of the Basque Country UPV/EHU (Grant No. IT1527-22); and the Spanish Ministerio de Ciencia e Innovación (Grant No. PID2022-142861NA-I00). Computational resources through projects GENCI A0160915069 and A0160815101 are greatly acknowledged.

[1] 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*, 525:73, ags 2015.

[2] Panpan Kong, Vasily S. Minkov, Mikhail A. Kuzovnikov, Alexander P. Drozdov, Stanislav P. Besedin, Shirin Mozaffari, Luis Balicas, Fedor Fedorovich Balakirev, Vitali B. Prakapenka, Eran Greenberg, and Mikhail I. Eremets. Superconductivity up to 243 k in yttrium hydrides under high pressure. *Nature Communications*, 12(5075), 2021.

[3] Ivan A. Troyan, Dmitrii V. Semenok, Alexander G. Kvashnin, Andrey V. Sadakov, Oleg A. Sobolevskiy, Vladimir M. Pudalov, Anna G. Ivanova, Vitali B. Prakapenka, Eran Greenberg, Alexander G. Gavriliuk, Igor S. Lyubutin, Viktor V. Struzhkin, Aitor Bergara, Ion Errea, Raffaello Bianco, Matteo Calandra, Francesco Mauri, Lorenzo Monacelli, Ryosuke Akashi, and Artem R. Oganov. Anomalous high-temperature superconductivity in yh6. *Advanced Materials*, 33(15):2006832, 2021.

[4] AP Drozdov, PP Kong, VS Minkov, SP Besedin, MA Kuzovnikov, S Mozaffari, L Balicas, FF Balakirev, DE Graf, VB Prakapenka, et al. Superconductivity at 250 k in lanthanum hydride under high pressures. *Nature*, 569(7757):528–531, 2019.

[5] Ion Errea, Francesco Belli, Lorenzo Monacelli, Antonio Sanna, Takashi Koretsune, Terumasa Tadano, Raffaello Bianco, Matteo Calandra, Ryotaro Arita, Francesco Mauri, et al. Quantum crystal structure in the 250-kelvin superconducting lanthanum hydride. *Nature*, 578(7793):66–69, 2020.

[6] Francesco Belli, Trinidad Novoa, J. Contreras-García, and Ion Errea. Strong correlation between electronic bonding network and critical temperature in hydrogen-based superconductors. *Nature Communications*, 12:1–11, 2021.

[7] Xing Li, Xiaohua Zhang, Yong Liu, and Guochun Yang. Bonding-unsaturation-dependent superconductivity in P-rich sulfides. *Matter and Radiation at Extremes*, 7(4):048402, 07 2022.

[8] Robert H Lavroff, Julen Munarriz, Claire E Dickerson, Francisco Munoz, and Anastassia N Alexandrova. Chemical bonding dictates drastic critical temperature difference in two seemingly identical superconductors. *Proceedings of the National Academy of Sciences*, 121(14):e2316101121, 2024.

[9] Jiabin Yu, Christopher J Ciccarino, Raffaello Bianco, Ion Errea, Prineha Narang, and B Andrei Bernevig. Non-trivial quantum geometry and the strength of electron–phonon coupling. *Nature Physics*, pages 1–7, 2024.

[10] J. Kortus, I. I. Mazin, K. D. Belashchenko, V. P. Antropov, and L. L. Boyer. Superconductivity of metallic boron in mgb_2. *Phys. Rev. Lett.*, 86:4656–4659, May 2001.

[11] Shuiquan Deng, Jürgen Köhler, and Arndt Simon. Unusual lone pairs in tellurium and their relevance for superconductivity. *Angewandte Chemie International Edition*, 45(4):599–602, 2006.

[12] A. Savin, R. Nesper, S. Wengert, and Thomas F. Fässler. Elf: The electron localization function. *Angew. Chem. Int. Ed. Engl.*, 36:1808–1832, 1997.

[13] K. E. Becke, A. D.and Edgecombe. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.*, 92:5397, 1990.

[14] Martin-Isbjörn Trappe, Jun Hao Hue, and Berthold-Georg Englert. Density-potential functional theory for fermions in one dimension. pages 251–267. World Scientific, 2023.

[15] C. F. von Weizsäcker. Zur theorie de kernmassen. *Z. Phys.*, 96:431, 1935.

[16] L. N. Oliveira, E. K. U. Gross, and W. Kohn. Density-functional theory for superconductors. *Phys. Rev. Lett.*, 60:2430–2433, Jun 1988.

[17] M. Lüders, M. A. L. Marques, N. N. Lathiotakis, A. Floris, G. Profeta, L. Fast, A. Continenza, S. Massidda, and E. K. U. Gross. Ab initio theory of superconductivity. i. density functional formalism and approximate functionals. *Phys. Rev. B*, 72:024545, Jul 2005.

[18] L. N. Oliveira, E. K. U. Gross, and W. Kohn. Density-functional theory for superconductors. *Phys. Rev. Lett.*, 60:2430–2433, Jun 1988.

[19] Antonio Sanna. Introduction to superconducting density functional theory. volume 7 of *The Physics of Correlated Insulators, Metals, and Superconductors*. Forschungszentrum Jülich, 2017.

[20] A. Sanna, C. Pellegrini, and E. K. U. Gross. Combining eliashberg theory with density functional theory for the accurate prediction of superconducting transition temperatures and gap functions. *Phys. Rev. Lett.*, 125:057001, Jul 2020.

[21] A. Sanna, C. Pellegrini, and E. K. U. Gross. Combining eliashberg theory with density functional theory for the accurate prediction of superconducting transition temperatures and gap functions. *Phys. Rev. Lett.*, 125:057001, Jul 2020.

[22] M. Tinkham. *Introduction to Superconductivity*. International series in pure and applied physics. McGraw Hill, 1996.

[23] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Phys. Rev.*, 108:1175–1204, Dec 1957.

[24] Koji Kajimura and Hisao Hayakawa, editors. *Advances in Superconductivity III*. Springer Tokyo, 1991.

[25] Wilver A. Muriel, Trinidad Novoa, Carlos Cardenas, and Julia Contreras-García. Introducing electron correlation in solid state calculations for superconducting states. *Faraday Discussions*, 2024.

[26] See Supplemental Material at [URL will be inserted by publisher] for more details about the models, calculation details, figures and tables. The Supplemental Materials also contains references [6, 14, 28–42].

[27] S. van de Bund and G. J. Ackland. Competition between superconductivity and molecularization in the quantum nuclear behavior of lanthanum hydride. *Phys. Rev. B*, 108:184102, Nov 2023.

[28] Santanu Saha, Simone Di Cataldo, Federico Giannessi, Alessio Cucciari, Wolfgang von der Linden, and Lilia Boeri. mapping superconductivity in high-pressure hydrides: The superhydra project. *Physical Review Materials*, 7(5):054806, 2023.

[29] Miles Cranmer. Interpretable machine learning for science with pysr and symbolicregression.jl, 2023. arXiv:2305.01582.

[30] K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical Methods for Physics and Engineering: A Comprehensive Guide*. Cambridge University Press, 3 edition, 2006.

[31] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[32] Sason S Shaik and Philippe C Hiberty. *A chemist's guide to valence bond theory*. John Wiley & Sons, 2007.

[33] R. F. W. Bader and M. E. Stephens. Spatial localization of the electronic pair and number distributions in molecules. *Journal of the American Chemical Society*, 97:7391–7399, 1975.

[34] Alexey I. Baranov and Miroslav Kohout. Electron localization and delocalization indices for solids. *Journal of Computational Chemistry*, 32:2064–2076, 2011.

[35] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. Fabris, G. Fratesi, S. de Gironcoli, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. Quantum ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21:395502, 2009.

[36] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio Jr, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M .Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero de-la Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni. Advanced capabilities for materials modelling with quantum ESPRESSO. *J. Phys.: Condens. Matter*, 29:465901, 2017.

[37] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[38] Ying Sun, Yanchao Wang, Xin Zhong, Yu Xie, and Hanyu Liu. High-temperature superconducting ternary Li-R-H superhydrides at high pressures (R=Sc, Y, La). *Phys. Rev. B*, page 24519, 2022.

[39] Wendi Zhao, Defang Duan, Xianxiang Yao, Zihao Huo, Qiwen Jiang, and Tian Cui. Pressure-induced high-t_c superconductivity in the ternary clathrate system Y-Ca-H. *Phys. Rev. B*, page 14521, 2022.

[40] Lan-Ting Shi, Yong-Kai Wei, A-Kun Liang, Robin Turnbull, Cai Cheng, Xiang-Rong Chen, and Guang-Fu Ji. Prediction of pressure-induced superconductivity in the novel ternary system $ScCah_{2n}$ (n = 1˜6). *J. Mater. Chem. C*, pages 7284–7288, 2021.

[41] Zihan Zhang, Tian Cui, Michael J. Hutcheon, Alice M. Shipley, Hao Song, Mingyang Du, Vladimir Z. Kresin, Defang Duan, Chris J. Pickard, and Yansun Yao. Design principles for high-temperature superconductors with a hydrogen-based alloy backbone at moderate pressure. *Phys. Rev. Lett.*, page 47001, 2022.

[42] Yao Sun, Shuai Sun, Xin Zhong, and Hanyu Liu. Prediction for high superconducting ternary hydrides below megabar pressure. *J. Phys. Condens. Matter*, page 505404, 2022.

Supplemental Material
Molecularity: a fast and efficient criterion for probing superconductivity

Matías E. di Mauro, Benoît Braïda, Trinidad Novoa,* and Julia Contreras-García†
Laboratoire de Chimie Théorique, Sorbonne Université & CNRS, 4 Pl. Jussieu, 75005, Paris, France

Ion Errea
Fisika Aplikatua Saila, Gipuzkoako Ingeniaritza Eskola,
University of the Basque Country (UPV/EHU), Europa Plaza 1, 20018 Donostia/San Sebastián, Spain
Centro de Física de Materiales (CSIC-UPV/EHU), Manuel de Lardizabal Pasealekua 5, 20018 Donostia/San Sebastián, Spain and
Donostia International Physics Center (DIPC), Manuel de Lardizabal Pasealekua 4, 20018 Donostia/San Sebastián, Spain
(Dated: October 31, 2024)

I. THE TIGHT-BINDING MODEL IN REAL SPACE

In order to make use of the tight-binding formalism to evaluate the real-space properties of the hydrogen chain, we begin by building a basis set of Gaussian atomic functions:

$$\chi_{lA}(x) = e^{-\alpha(x-la)^2}\,, \tag{1}$$

$$\chi_{lB}(x) = e^{-\alpha(x-(R_{AB}+la))^2}\,, \tag{2}$$

that depends on the unit cell index, l. In this way, every unit cell contains two basis functions. Fig. S1 is a spatial representation of the two atomic orbitals in the unit cell. In general, we consider the length of the unit cell, a, to be equal to $1.0\,u$, with u some arbitrary units. In the next section, we shall see which values of a properly represent a hydrogen chain. The same is done for the Gaussian exponent α.

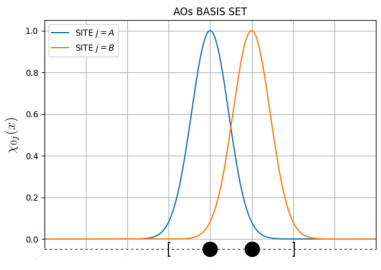


Figure S1. Atomic orbitals used in the unit cell of the hydrogen chain tight-binding model, represented by two identical Gaussian functions. The black spheres mark the positions of the two nuclei inside the unit cell, delimited by squared brackets. Here, the unit cell length is taken as $1\,u$, and the Gaussian exponent as $\alpha = 10\,u^{-2}$

With this, the atomic orbitals fulfilling Bloch's theorem, or Bloch sums, are

$$\varphi_{Ak}(x) = \frac{1}{\sqrt{N_l}}\sum_l e^{ikla}e^{-\alpha(x-la)^2}\,, \tag{3}$$

$$\varphi_{Bk}(x) = \frac{1}{\sqrt{N_l}}\sum_l e^{ik(R_{AB}+la)}e^{-\alpha(x-(R_{AB}+la))^2}\,, \tag{4}$$

with N_l the number of unit cells considered in the model. Here, k is selected to take values inside the first BZ, $k \in \left[-\frac{\pi}{a}, \frac{\pi}{a}\right]$, with $k = \frac{2\pi}{N_l a}l$ and l taking integer values in the range $\left[-\frac{N_l}{2}, \frac{N_l}{2}\right]$. For simplicity, we will generally refer to k in units of $\frac{2\pi}{a}$. A graphical representation of these orbitals for $k = 0.02$ (i.e. $k = 0.02 \times \frac{2\pi}{a}$) is offered in Figure S2, for a total number of $N_l = 101$.

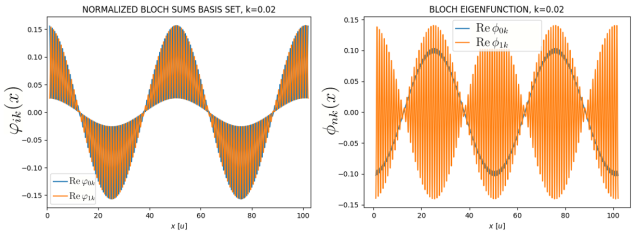


Figure S2. Left: Real part of the Bloch sums of each band index, $i = 0$ and $i = 1$, for $k = 0.02 \times \frac{2\pi}{a}$. Right: Real part of the Bloch wavefunctions at the same k point for the symmetric chain.

Finally, the Bloch wavefunctions that diagonalize the Hamiltonian of the hydrogen chain can be written as a linear combination of the Bloch sums in (3) and (4),

$$\phi_{nk}(x) = \frac{1}{\sqrt{N_l}}\sum_i\sum_{R_i} c_{ik}^n e^{ikR_i}e^{-\alpha(x-R_i)^2}\,. \tag{5}$$

In this case, the band index n can take the values 0 and 1. The diagonalization determines the coefficients c_{ik}^n, ant it is performed using the Gaussian elimination method [1], as implemented in NumPy [2]. The result of the diagonalization will necessarily depend on the hopping parameters. Let us recall that the hopping parameter between site A and the site B to its right is v, whereas between the same site A and the atom

* Correspondence email address: trinidad.novoa_aguirre@sorbonne-universite.fr
† Correspondence email address: contrera@lct.jussieu.fr

B on its left, the hopping parameter is w. In the special case in which $v = w$, the resulting coefficients are $c^0_{Ak} = c^0_{Bk} = \frac{1}{\sqrt{2}}$ and $c^1_{Ak} = -c^1_{Bk} = \frac{1}{\sqrt{2}}$, corresponding to a bonding and antibonding orbitals, respectively. Figure S2 shows an example of such a Bloch wavefunction.

Once the Bloch wavefunctions have been defined and computed, the next step is to evaluate the real-space functions of interest. In order to do so, we will generally assume that the Bloch wavefunctions are orthonormal, which allows expressions such as

$$\rho(x) = \sum_{nk} n_{nk} |\phi_{nk}(x)|^2 \,, \tag{6}$$

for the electron density, and similar ones for the other functions.

To evaluate the orthogonality, we first resort to the nearest-neighbors approximation, in which the only non-zero overlaps between the basis functions are

$$\int \chi^*_{lA}(x)\chi_{lA}(x)dx = \int \chi^*_{lB}(x)\chi_{lB}(x)dx = C \,, \tag{7}$$

and that between two adjacent atomic sites,

$$\begin{aligned} \int \chi^*_{lA}(x)\chi_{lB}(x)\,dx &= \int \chi^*_{l-1B}(x)\chi_{lA}(x)\,dx \\ &= \int \chi^*_{lB}(x)\chi_{l+1A}(x)\,dx = \mathcal{S} \end{aligned} \tag{8}$$

because we consider PBC, we must also include

$$\int \chi^*_{N_l B}(x)\chi_{1A}(x)\,dx = \mathcal{S} \,. \tag{9}$$

Equation (7) allows to normalize the basis set. Once this is done, the Bloch overlap will yield

$$\int \varphi^*_{ik}(x)\varphi_{jk'}(x)\,dx = \delta_{k,k'} c^{k,k'}_{i,j} \,, \tag{10}$$

where it can be shown that $c^{k,k}_{A,A} = c^{k,k}_{B,B} = 1$, and $c^{k,k}_{i,j}$ for $i \neq j$ is dependent on $\mathcal{S}$. In this way, equation (10) shows that the Bloch sums are orthogonal with respect to the wave vector index k, but not with respect to the site. This changes when we diagonalize the Hamiltonian and find the Bloch eigenfunctions, which are orthonormal

$$\int \phi^*_{nk}(x)\phi_{n'k'}(x)\,dx = \delta_{n,n'}\delta_{k,k'} \,. \tag{11}$$

Equation (11) is true analytically, but in practice there is a numerical error that comes from the integration of the Bloch sums in (10): the overlap of two Bloch sums of different k is not exactly zero. Given the importance of the orthogonality condition in the computation of the density and the other real-space descriptors, we define the numerical error associated to those computations as

$$ERROR = \max_{k \neq k'} \left\{ \int \varphi^*_{ik}(x)\varphi_{jk'}(x)\,dx \right\} \,. \tag{12}$$

This quantity sets the precision of the results presented in the following, and it will be of special interest to minimize it when choosing the parameters of the model.

As a final remark concerning the computation of the real-space quantities in this model, it is important to note that the expression of the ELF is not valid for one-dimensional systems, as the expression of the TF KED changes in those cases to [3]

$$\tau_{TF}(x) = \frac{\pi^2}{24}\rho(x)^3 \,. \tag{13}$$

It is this one-dimensional version of the TF KED and the corresponding ELF, that we consider for all of the calculations of this Section.

II. THE MODEL PARAMETERS

The construction of the hydrogen chain using the tight-binding formalism relies on a set of parameters that need to be chosen carefully for the model to properly represent the physical system. Firstly, we choose the lattice parameter, a, for which we hereby adopt the value of $a = 2.64$ Å. This determines the value of the arbitrary units, $1\,u = 2.64$ Å. This value is in accordance with what has been observed in the literature for high-temperature hydrogen-based superconductors of interest [4, 5]. In this way, for the symmetric chain we shall set the hydrogen-hydrogen distance to $d_{HH} = 0.5\,u = 1.32$ Å.

The Gaussian exponent, α, is optimized using VB theory on a similar hydrogen chain, using a 6-31G* basis, and fitting the proposed Gaussian orbitals to those results [6]. The outcome of such a calculation indicates that α should live in the range of $[8, 15]\,u^{-2}$ or, equivalently, $\alpha \in [1.15, 2.15]$ Å^{-2}. In general, we will use $\alpha = 10\,u^{-2} = 1.43$ Å^{-2}.

In theory, the hopping parameters, v and w, could be evaluated analytically. In practice this is not possible, as divergences arise from the Coulomb potential of hydrogen in one dimension,

$$V_0(x) = \frac{1}{4\pi\epsilon_0}\frac{1}{x} \,, \tag{14}$$

with ϵ_0 the permittivity of vacuum. This issue is avoided by using the WHA, that has been proposed for the Hückel model, that is, in principle, physically equivalent to the tight-binding formalism. For one-dimensional systems, the WHA allows to establish the relationship

$$t = \kappa \cdot \epsilon \cdot \mathcal{S} \,, \tag{15}$$

for the hopping t, with respect to the onsite energy ϵ and the overlap $\mathcal{S}$ of eq. 8. There, κ is a constant that is set to 0.787 [1] and the onsite energy for the hydrogen's $1s$ orbital is considered, $\epsilon = -13.6$ eV. With this, the hopping parameters are restrained to the range $t \in [-3.0, -1.5]$ eV. Moreover, equation (15) is particularly useful when atomic distances are variable, because it allows to scale the hopping parameters, simply by assessing the new overlap integrals, $\mathcal{S}$.

In the case of the symmetric chain, when only one hopping parameter is present, we will use the value $t = v = w = -1.5$ eV. It is interesting to note that, actually, in that case the only effect of changing t is that the energy scale varies, resulting in flatter bands for a higher $|t|$, affecting the DOS at the Fermi energy.

Another two parameters that arise from the discretization of the problem in real and reciprocal space must be determined, namely the real-space grid step, Δx, and the number of k points, N_k (that is equal to the number of unit cells, N_l). To ensure the reliability of our results, we perform a convergence test on those parameters, in order to minimize the numerical error arising from such an approximation. First, we fix $N_k = 101$ and vary the grid step Δx from $10^{-1}\,u$ to $10^{-4}\,u$. We evaluate the error of the overlap with respect to that distance, as defined in equation (12), as it is displayed in Fig. S3. In order to check the convergence of the localization properties, we also evaluate progress with respect to Δx of the minimum value of the ELF, baptized here as ϕ, for different temperatures and models (NS and SC). Because temperature does not seem to have much effect on the convergence, only the results for $T = 10$ K are shown in Fig S4. Setting $\Delta x = 10^{-3}\,u$ to ensure an error below 10^{-5} the same analysis is performed by varying the number of k-points. Only odd numbers are considered to ensure a proper sampling of the BZ, that incorporates the Γ-point, i.e. $x = 0\,u$. Considering the results obtained in this convergence test, the value of $N_k = 101$ is chosen for the rest of the calculations.

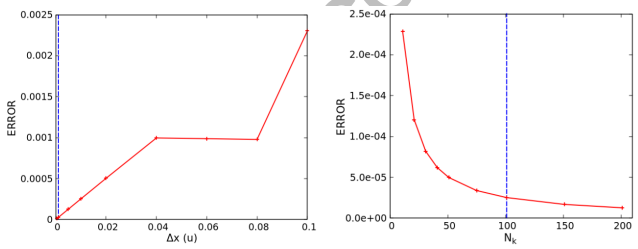


Figure S3. Error in the orthogonality of the Bloch sums with respect to: the grid step considered in the discretization of the spatial functions, $\Delta x\,(u)$ (left); and the number of k-points considered in the discretization of reciprocal space, N_k. The dashed vertical line marks the chosen value of Δx abd N_k.

[1] Here, the value of κ is scaled from the typical value of $\kappa = 1.75$ for covalent bonds, to yield the correct values for the hopping when the on site energy is shifted from $\epsilon = 0$ eV to $\epsilon = -13.6$ eV.

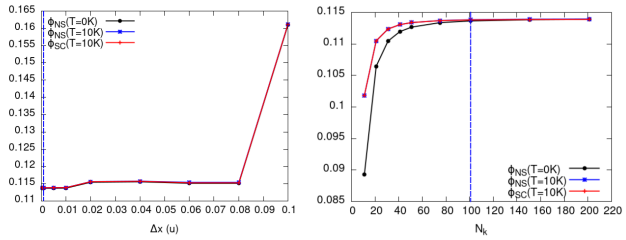


Figure S4. Value of the minima of the ELF of the NS at zero temperature (red), the NS at $T = 10$ K (green) and the SC state at $T = 10$ (K) (blue), for $T_c = 300$ K; with respect to: the grid step considered in the discretization of the spatial functions, $\Delta x\,(u)$ (left); and the number of k-points considered in the discretization of reciprocal space, N_k. The dashed vertical line marks the chosen value of Δx abd N_k.

III. APPROXIMATION OF THE SUPERCONDUCTING GAP

Figure S5 offers the profile of the gap function for $\omega = 0.2$ eV, considering different critical temperatures, T_c, and at different temperatures, T. This functions represent the width of the window around the Fermi energy where electrons will form Cooper pairs. One can see how the gap goes to zero everywhere when the temperature reaches $T = T_c$, where the transition occurs. A similar thing occurs to the anomalous density,

$$\chi(\xi;T) = \frac{\Delta(\xi;T)}{2\sqrt{\xi^2+\Delta(\xi;T)^2}} \tanh\left(\frac{\sqrt{\xi^2+\Delta(\xi;T)^2}}{2k_BT}\right), \quad (16)$$

as displayed in Fig. S6.

Taking this into account, one can compare the occupation numbers for the metallic and the superconducting state, as can be seen in Figure S7. For a fixed critical temperature, it can be seen that the occupancies of the normal state deviate from the step function as the temperature increases, softening the transition around the Fermi energy, as expected. On the other hand, the superconducting occupation numbers do not suffer great alterations with the temperature.

For the highest critical temperature, $T_c = 300$ K, Fig. S7 shows that a larger range of occupancies is spanned.

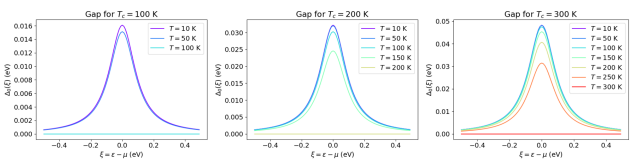


Figure S5. Gap function around the Fermi energy, considering different critical temperatures: $T_c = 100$ K (left), $T_c = 200$ K (middle), and $T_c = 300$ K (right). In each case, the dependence on the temperature T is shown by different colored lines.

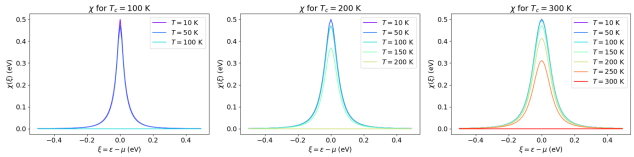


Figure S6. Anomalous density as a function of the energy, for three different critical temperatures: T_c = 100 K (left), T_c = 200 K (middle), and T_c = 300 K (right). In each case, the dependence on the temperature T is shown by different colored lines.

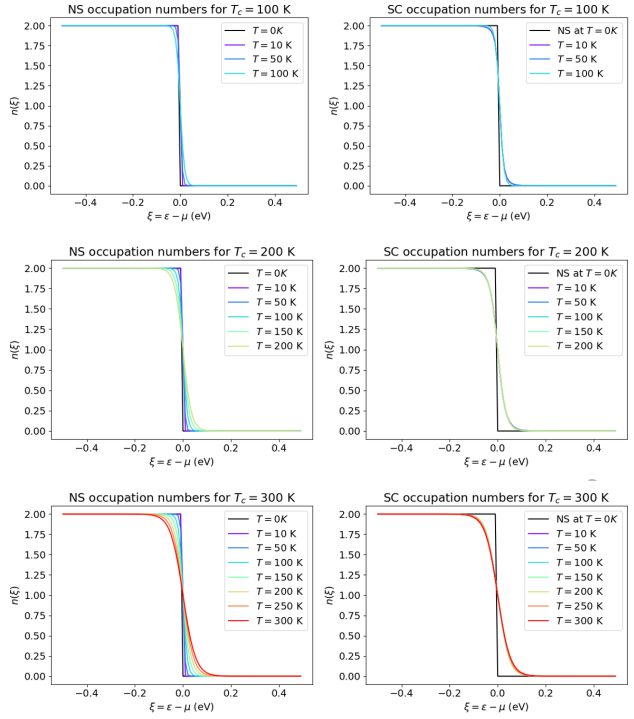


Figure S7. Occupation numbers at different energies. In all panels, that of the normal state at T = 0 K is depicted in black. To the left, each figure shows the occupation numbers of the normal state at different temperatures, for a give T_c. To the right, the same is displayed for the superconducting state.

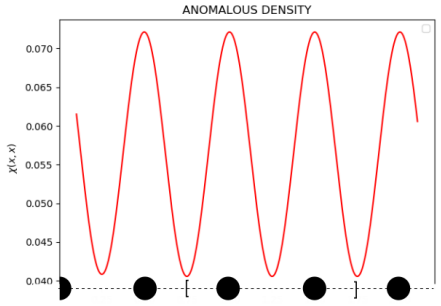


Figure S8. Anomalous SC density, $\chi(x)$, along the hydrogen chain for the amplified gap (T_c = 5000 K). The atomic positions are marked by the black circles, with the squared brackets delimiting the unit cell.

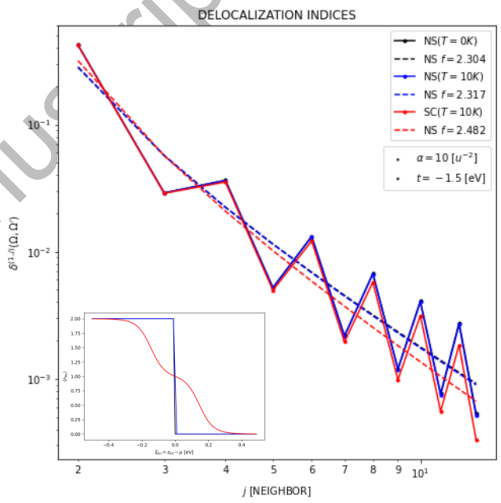


Figure S9. Delocalization indices in the NS at 0 K (black) and at 10 K (blue), and for the SC state at 10 K (red). The dashed lines shows the best algebraic fit for each curve, and the inset the occupation numbers of each state, in the same color coding.

IV. HIGH-CORRELATION LIMIT

The limit case of highest correlation in the superconducting state corresponds to when $\Delta \to \infty$ and $T \to 0$. The first condition can also be reformulated as $T_c \to \infty$. From the occupations derived in the main text, one can already infer that the occupancies will go to 1 in those places where $\Delta(\xi) \neq 0$. Taking the very large value of T_c = 5000 K and a low temperature of T = 10 K, we obtain the occupation numbers shown in the inset of Figure S9.

Surprisingly, even at this limiting case the profile of the real-space functions in the symmetric chain in the SC state, namely the electron density, the KED, and the ELF; seem indistinguishable from those of the normal state. The anomalous density does increase in this limiting case, but its magnitude remains far below that of the electron density, see Fig. S8.

The analysis of the localization (LI) and delocalization indices (DI),[7, 8] on the other hand, shows a more clear tendency. For the LI, we obtain the values of $\lambda_{NS}(T = 0K)$ = 0.45654, $\lambda_{NS}(T = 10K)$ = 0.45653, and $\lambda_{SC}(T = 10K)$ = 0.45645. Although the differences remain very small, there seems to be a tendency for higher localization in the superconducting state. This is further supported by the higher decay rate of the SC DI, as it becomes clear from Figure S9.

V. CALCULATION DETAILS

We have carried out the calculations for the ELF and DOS of these systems using DFT within the Kohn-Sham framework and the Plane-Waves Pseudo Potentials method as implemented in Quantum ESPRESSO [9, 10]. We chose the PBE scheme for the exchange-correlation functional [11], including scalar relativistic effects. Binary compounds were taken from our previous contribution, Ref. [4]. Ternary compounds were taking from autrui (see Table I in S.I. and references therein) and cutoffs were set to 80 Ry and 800 Ry for the wavefunction and density energy cut-offs, respectively. The Brillouin Zone was sampled with an unshifted 12x12x12 regular grid.

VI. MACHINE LEARNED MODEL

The new fits to estimate T_c have been obtained using Symbolic Regression, as implemented by PySR [12]. We have chosen four input quantities to be considered to compute T_c, for binary and ternary systems: ϕ, H_f, H_{DOS}, and $1-\Delta\phi$, where $\Delta\phi = \phi^* - \phi$, as they have all shown to be high in high-critical temperature compounds. The output expressions have different accuracy and complexity, the latter being a measure of how many nested operators are present. Eligible unary operators were x^2, x^3, $\sqrt{x}$, and $\sqrt[3]{x}$; while only multiplication was allowed as a binary operator. In every case, the models were trained for 200 iterations, and the dataset was divided training and test sets, corresponding to 2/3 and 1/3 of the systems, respectively. The results in the test set for the best two fits, SR1 and SR2, are displayed in Figure S10. The performance of both of them can be assessed by the MAE in the test set, which is ca. 36-38K (see Table S1). This value goes up to nearly 50K for systems with $T_c \geq 77$K, where errors are expected to raise due to the lower amount of systems in the dataset living in that high-T_c region.

One thing that can be done to avoid this problem, is to filter the data to include a similar amount of high and low-T_c systems. Instead of doing it in a random way, we have chosen to filter keeping only the systems with $\phi^* \in [0.45, 0.8]$, which correspond to the bonding families that interest us. We thus trained a model using this data and the magnitudes ϕ, H_f, and H_{DOS}; and keeping the same method and parameters for the training. The overall errors in the two best fits, SR3 and SR4, are of 48.0 and 35.9 K, the latter being much suitable for the estimation of T_c. However, in both cases the MAEs are more consistent with those in high-T_c regions, compared to SR1 and SR2 (see Table S1).

VII. FIGURES AND TABLES

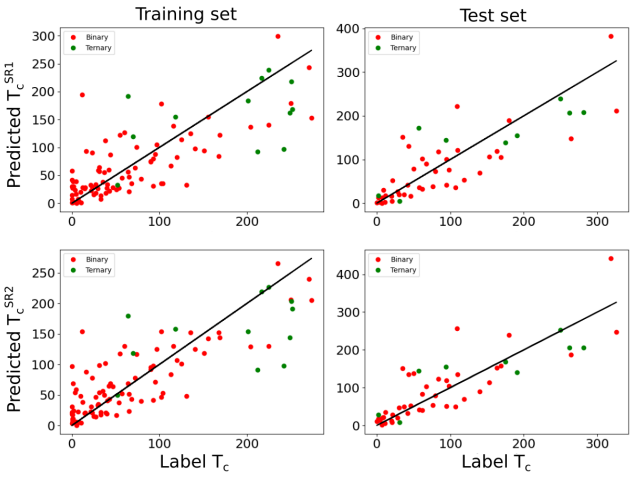


Figure S10. Reference (label) T_c (K) values with respect to its predicted values in the training (left) and test (right) sets, according to the two fits: SR1 (top) and SR2 (bottom); as obtained with the input quantities $1-\Delta\phi$, H_f, anf H_{DOS}.

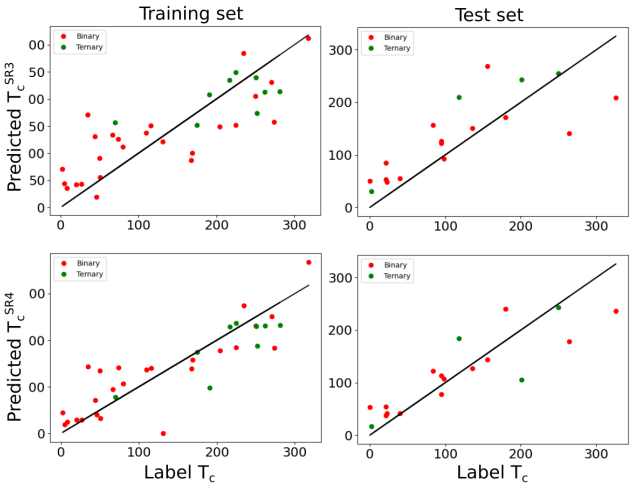


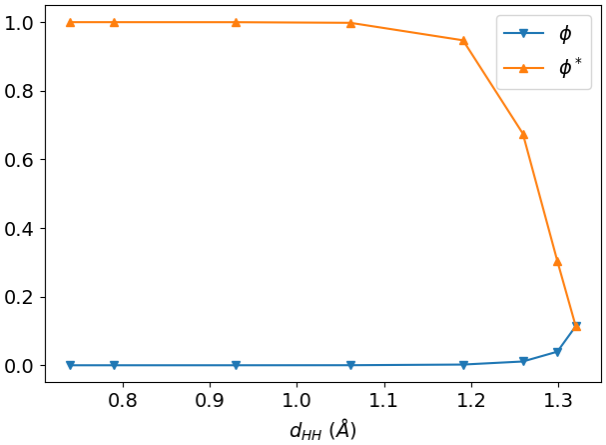
Figure S11. Reference (label) T_c (K) values with respect to its predicted values in the training (left) and test (right) sets, according to the two fits: SR3 (top) and SR4 (bottom); as obtained with a smaller dataset containint only systems where $\phi^* \in [0.45, 0.8]$, and using the input quantities ϕ, H_f, anf H_{DOS}.

Table S1. Mean absolute errors (MAEs) in the train and test sets using the three different approximations for the estimation of T_c presented in this work. The errors estimating reference T_c's above 77 K are labeled as $MAE_{\geq 77K}$.

T_c approx.	MAE train (K)	MAE test (K)	$MAE_{\geq 77K}$ train (K)	$MAE_{\geq 77K}$ test (K)
SR1	34.7	37.7	50.3	50.9
SR2	31.9	36.4	41.9	47.7
SR3	47.1	48.0	44.1	54.2
SR4	36.7	35.9	38.9	42.5

Table S2. List of X-RE-H systems (X is *s*-block element, RE is rare earth) in the dataset, including their chemical formula, pressure (GPa), space group, and superconducting critical temperature T_c(K), networking value ϕ, H_{DOS}, bonding type, and reference from where it was taken. The systems classified as Ionic and Molecular with an asterisk use a different definition to that of Belli *et al.* (2021).

Chem. formula	Pressure (GPa)	Space group	T_c (K)	ϕ	H_{DOS}	Bonding type	Ref.
$LiScH_{10}$	300	$R\bar{3}m$	52	0.30	0.28044	Molecular	[13]
Li_2ScH_{20}	300	$Immm$	242	0.36	0.74533	Molecular	[13]
Li_2ScH_{16}	300	$Fd\bar{3}m$	262	0.63	0.6838	Weak H-H	[13]
Li_2ScH_{16}	230	$Fd\bar{3}m$	281	0.63	0.68616	Weak H-H	[13]
Li_2ScH_{17}	300	$Fd\bar{3}m$	94	0.57	0.59984	Molecular*	[13]
Li_2LaH_{17}	300	$Fd\bar{3}m$	118	0.50	0.67135	Weak H-H	[13]
Li_2YH_{16}	300	$Fd\bar{3}m$	251	0.59	0.76919	Weak H-H	[13]
Li_2YH_{17}	300	$Fd\bar{3}m$	64	0.55	0.79807	Molecular*	[13]
$CaYH_{20}$	600	$P4/mmm$	250	0.62	0.81869	Weak H-H	[14]
Ca_2YH_{18}	200	$P\bar{3}m1$	217	0.59	0.75185	Weak H-H	[14]
Ca_3YH_{24}	200	$Fm\bar{3}m$	225	0.58	0.80000	Weak H-H	[14]
CaY_3H_{24}	200	$Fm\bar{3}m$	252	0.55	0.55800	Weak H-H	[14]
$CaScH_2$	250	$Fm\bar{3}m$	31	0.19	0.0237	Ionic	[15]
$CaScH_4$	200	$P6/mmm$	2	0.17	0.0978	Ionic*	[15]
$CaScH_6$	200	$P4/mmm$	57	0.36	0.60139	Ionic*	[15]
$CaScH_8$	200	$P4/mmm$	212	0.42	0.5569	Weak H-H	[15]
$CaScH_{12}$	160	$Pm\bar{3}m$	175	0.55	0.48617	Weak H-H	[15]
$BeLaH_8$	50	$Fm\bar{3}m$	191	0.29	0.66781	Weak H-H	[16]
$BeYH_8$	100	$Fm\bar{3}m$	249	0.12	0.69714	Weak H-H	[16]
$BeCeH_8$	30	$Fm\bar{3}m$	201	0.29	0.77857	Weak H-H	[17]
$BeThH_7$	20	$P6/mmc$	70	0.28	0.50172	Ionic	[17]



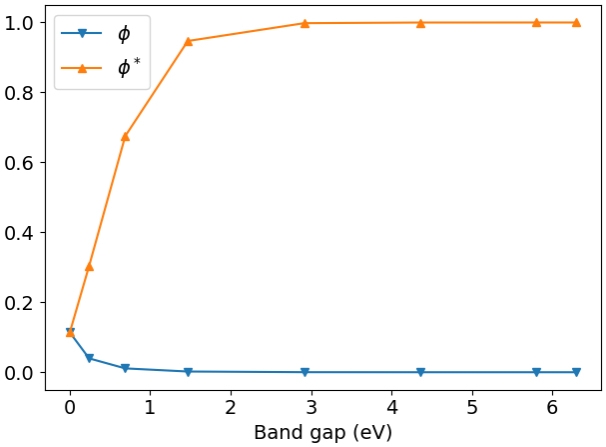
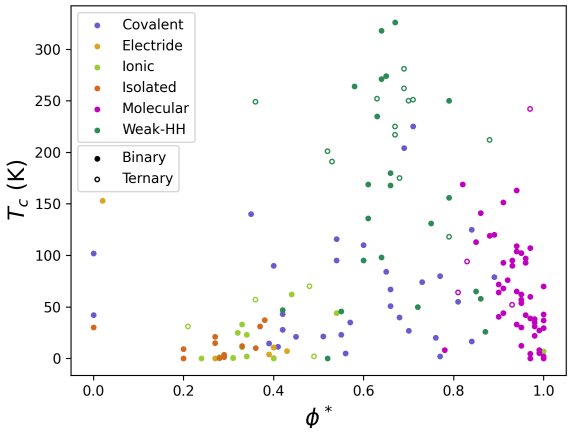


Figure S12. Evolution of the topological descriptors ϕ and ϕ^* with respect to the intramolecular distance (left) and the band gap (right).



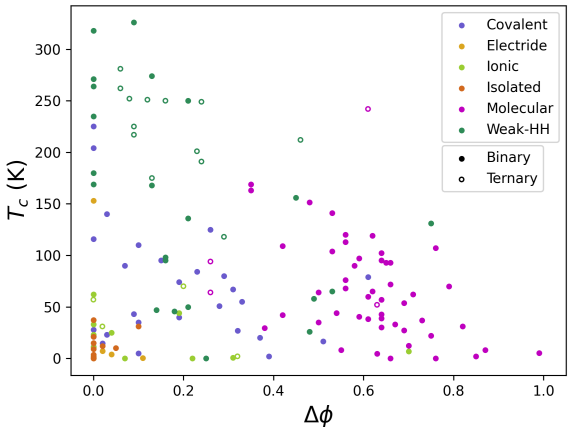


Figure S13. Reference critical temperature T_c (K) with respect to the molecularity index ϕ^* (left), and to the different between that and the networking value, $\Delta\phi = \phi^* - \phi$ (center), for all binary and ternary data.

[1] K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical methods for physics and engineering: a comprehensive guide*. Cambridge University Press, 3 edition, 2006.

[2] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[3] Martin-Isbjörn Trappe, Jun Hao Hue, and Berthold-Georg Englert. Density-potential functional theory for fermions in one dimension. pages 251–267. World Scientific, 2023.

[4] Francesco Belli, Trinidad Novoa, J. Contreras-García, and Ion Errea. Strong correlation between electronic bonding network and critical temperature in hydrogen-based superconductors. *Nature Communications*, 12:1–11, 2021.

[5] Santanu Saha, Simone Di Cataldo, Federico Giannessi, Alessio Cucciari, Wolfgang Von Der Linden, and Lilia Boeri. mapping superconductivity in high-pressure hydrides: The superhydra project. *Physical Review Materials*, 7(5):054806, 2023.

[6] Sason S Shaik and Philippe C Hiberty. *A chemist's guide to valence bond theory*. John Wiley & Sons, 2007.

[7] R. F. W. Bader and M. E. Stephens. Spatial localization of the electronic pair and number distributions in molecules. *Journal of the American Chemical Society*, 97:7391–7399, 1975.

[8] Alexey I. Baranov and Miroslav Kohout. Electron localization and delocalization indices for solids. *Journal of Computational Chemistry*, 32:2064–2076, 2011.

[9] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. Fabris, G. Fratesi, S. de Gironcoli, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. Quantum ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21:395502, 2009.

[10] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio Jr, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M .Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero de-la Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni. Advanced capabilities for materials modelling with quantum ESPRESSO. *J. Phys.: Condens. Matter*, 29:465901, 2017.

[11] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[12] Miles Cranmer. Interpretable machine learning for science with pysr and symbolicregression.jl, 2023. arXiv:2305.01582.

[13] Ying Sun, Yanchao Wang, Xin Zhong, Yu Xie, and Hanyu Liu. High-temperature superconducting ternary Li-R-H superhydrides at high pressures (R=Sc, Y, La). *Phys. Rev. B*, page 24519, 2022.

[14] Wendi Zhao, Defang Duan, Xianxiang Yao, Zihao Huo, Qiwen Jiang, and Tian Cui. Pressure-induced high-t_c superconductivity in the ternary clathrate system Y-Ca-H. *Phys. Rev. B*, page 14521, 2022.

[15] Lan-Ting Shi, Yong-Kai Wei, A-Kun Liang, Robin Turnbull, Cai Cheng, Xiang-Rong Chen, and Guang-Fu Ji. Prediction of pressure-induced superconductivity in the novel ternary system $ScCah_{2n}$ (n = 1˜6). *J. Mater. Chem. C*, pages 7284–7288, 2021.

[16] Zihan Zhang, Tian Cui, Michael J. Hutcheon, Alice M. Shipley, Hao Song, Mingyang Du, Vladimir Z. Kresin, Defang Duan, Chris J. Pickard, and Yansun Yao. Design principles for high-temperature superconductors with a hydrogen-based alloy backbone at moderate pressure. *Phys. Rev. Lett.*, page 47001, 2022.

[17] Yao Sun, Shuai Sun, Xin Zhong, and Hanyu Liu. Prediction for high superconducting ternary hydrides below megabar pressure. *J. Phys. Condens. Matter*, page 505404, 2022.