Machine Learning and High-Throughput Approaches to Magnetism

S. Sanvito, M. Žic, J. Nelson, T. Archer, C. Oses, and S. Curtarolo

Contents

Abstract

Magnetic materials have underpinned human civilization for at least one millennium and now find applications in the most diverse technologies, ranging from data storage, to energy production and delivery, to sensing. Such great diversity, associated to the fact that only a limited number of elements can sustain a magnetic order, makes magnetism rare and fascinating. The discovery of a new high-performance magnet is often a complex process, where serendipity plays an

S. Sanvito (✉) · M. Žic · J. Nelson · T. Archer
School of Physics and CRANN Institute, Trinity College, Dublin, Ireland
e-mail: sanvitos@tcd.ie; zicm@tcd.ie; archert@tcd.ie

C. Oses · S. Curtarolo
Center for Materials Genomics, Duke University, Durham, NC, USA
Departments of Mechanical Engineering and Materials Science, Physics, and Chemistry, Duke University, Durham, NC, USA
e-mail: stefano@duke.edu

© Springer International Publishing AG, part of Springer Nature 2018

W. Andreoni, S. Yip (eds.), *Handbook of Materials Modeling*,
https://doi.org/10.1007/978-3-319-50257-1_108-1

important role. Here we present a range of novel approaches to the discovery and design of new magnetic materials, which is rooted in high-throughput electronic structure theory and machine learning models. Such combination of methods has already demonstrated the ability of discovering ferromagnets with high Curie temperature at an unprecedented speed.

1 Why New Magnets?

Magnetism is one of the most fascinating macroscopic orders of matter, and it is deeply rooted in quantum mechanics. The Hund's coupling is responsible for the formation of local magnetic moments, m, often localized close to the atomic nuclei, while the exchange interaction, J, makes the moments interacting to each other. This is the m-J paradigm that, together with spin-orbit interaction, generates a multitude of magnetic orders and phenomena.

The formation of the magnetic moment in the solid state is a relatively rare event among the elements of the periodic table. It is possible in $3d$ transition metals and in $4f$ rare-earths, both with an open shell configuration. It is also found, although more rarely, in some $4d$ ions. Light elements, presenting $2p$ valence orbitals, form close shell compounds. Therefore, despite presenting large Hund's coupling, they can sustain a moment only in their radical form, and the possibility of long-range magnetic order remains associated to extremely defective compounds, and it is surrounded by skepticism and debate (Magda et al. 2014). Finally, for heavier elements the valence shell is too delocalized, and m does not form (Janak 1977), with the exception of some low-dimensional structures (Moruzzi and Marcus 1989; Requist et al. 2016).

Once the moments are formed, there are many possible mechanisms to couple them to each other, depending on the details of any given material and in particular on its position with respect to the metal/insulator boundary (Coey 2009). Magnetic moments in metals usually couple by mean of interactions mediated by itinerant electrons, such as the RKKY or the double-exchange ones. These lead to a ferromagnetic order. At the same time, there are mechanisms, such as magnetic super-exchange, which are mostly active in the insulating regime and may produce an antiferromagnetic order. Such rules, however, are not general and the actual magnetic coupling is very sensitive to details. The local chemical environment, the crystal structure, the density of magnetic ions, etc. all may play a critical role so that it is not uncommon to find also ferromagnetic insulators (Wohlfarth 1980) and antiferromagnetic metals (Wadley et al. 2013). In general, however, it is a fact that magnetism is relatively rare in the materials world, with only about 4,000 inorganic compounds among the 100,000 known to date (ICSD 2018) presenting a magnetic order of some kind.

There are several good technological reasons to extend the available menu of magnetic materials. Firstly, there is a growing number of applications that rely on high-performing magnets for which the ideal compound is yet to be discovered or needs to be "rediscovered." Room-temperature applications, of any

kind, require a magnetic ordering temperature in excess of 300°C, a value that ensures stable operation in the temperature range −50°/+120°C. Only about 400 magnets meet such criterion. Furthermore, specific applications dictate several other conditions. For instance, permanent magnets should maximize the magneto-crystalline anisotropy, K_1, and the total magnetization, M_S, so that energy product is large. At the same time, they should be made of elements abundant in the earth crust. In contrast, element criticality is less problematic for the magnets employed by the data storage industry (e.g., Pt is used), but these need to be grown on desired substrates; their grain structure needs to be highly optimized and so does their magnetic hysteresis. Furthermore, changes in technology may generate novel requirements. For instance, it is likely that the advent of heat-assisted magnetic recording may add constrains on the thermal properties of the magnetic recording media.

Sometimes the demand for a new magnet is not dictated by a new technological need. For instance, $Nd_2Fe_{14}B$-based magnets currently cover about 90% of the permanent magnets market, with the remaining share also being largely occupied by rare-earth-containing compounds (e.g., Sm-Co alloys). The average price of a kilogram of neodymium metal was US$8 in 2006, US$56 at the end of 2010, US$450 in 2011, and then fell to US$125 in September 2012. Prize volatility is a significant issue for all those applications requiring large volumes of raw elements (an electric vehicle requires about 3 kg of Nd-Fe-B, a direct-drive wind generator about 250 kg), where steady supply is key. Importantly, the rare-earth prize volatility is not associated to fluctuations in the mining production but to commercial strategies between producers and retailers. Such economical geopolitical constrain is the main driver toward a worldwide effort in developing rare-earth-free permanent magnets.

Finally, in addition of being rare and technologically important, magnetism is complex and often defies our physical/chemical intuition. A good example is given by $SrTcO_3$ (Rodriguez et al. 2011). This is a distorted cubic perovskite displaying a G-type antiferromagnetic order with the remarkably high Néel temperature of $T_N = 750$°C. $SrTcO_3$ is unusual from several points of view. Tc is the magnetic element, one of the few displaying a local moment among the 4*d* transition metal series and highly radioactive. The Néel temperature is exceptionally high with only a handful antiferromagnets having T_N above 500°C. Finally, the chemical substitution of Tc with elements in its neighborhood in the periodic table produces only low-temperature antiferromagnets ($SrCrO_3$, $SrMnO_3$ and $SrFeO_3$), one low-temperature ferromagnet ($SrRuO_3$), and a diamagnetic compound ($SrMoO_3$). As such, the strong antiferromagnetism of $SrTcO_3$, which is due to a subtle interplay between *p*-*d* hybridization and Jahn-Teller distortion (Franchini et al. 2011), appears as a singularity in our magnetic materials landscape.

Given magnetism complexity, the variety of microscopic mechanisms at play, and the subtle sensitivity to details, it is not surprising that the "traditional" trial and error approach to the discovery/design of new magnets has a pretty low throughput. This has been recently challenged by two completely new approaches to materials discovery, namely, high-throughput electronic structure theory (HTEST) and machine

learning (ML) methods. In this new framework, one aims at accelerating the discovery process by performing an extremely large volume of ab initio calculations, which are directed by astute ways for navigating large datasets and identifying patterns among the data. Here we will review such new materials discovery strategy with a particular emphasis on the problems and requirements associated with the magnetic interaction. We will take the discovery of novel magnetic Heusler alloys as prototype, since this is the materials class, where most of the work has focused to date (Sanvito et al. 2017). A large part of the methodology described here is based on the AFLOW project, which is discussed in detail in a chapter of volume 1 (Toher et al. 2018).

2 The High-Throughput Approach to Materials Discovery

2.1 General Principles

In general there are three steps common to all the attempts at HTEST (Curtarolo et al. 2013): (1) the creation of large dataset of computed materials properties, (2) the organization of such calculations in easy-to-access databases, and (3) the screening of the databases in the search of new materials with desired properties or of new rules for designing novel compounds. Let us discuss these three steps in some detail.

Creating large datasets of computed materials properties imposes a number of choices, which somehow determine the further capability of the entire scheme. One strategy consists in producing highly standardized and curated data. Here one decides the level of theory to use and no data are accepted unless they satisfy some stringent criteria. A rather popular choice is that of using density functional theory (DFT) in either the local density or the generalized gradient approximation (LDA and GGA). These ensure enough accuracy, in particular in terms of total energies and elementary linear-response properties, and most importantly the possibility of a high throughput. The convergence parameters (cutoff energies, k-point sampling, force tolerance, etc.) are also highly standardized, so that total energies can be computed across different chemical compositions, structures, and unit cell sizes. Almost all the most popular databases are produced in this way. These include AFLOW.org (Curtarolo et al. 2012b), Materials Project (Jain et al. 2013), OQMD (Kirklin et al. 2015), and 2D (Rasmussen and Thygesen 2015). A second strategy consists in accepting all possible data, regardless of the source (the particular electronic structure theory code used) and the level of theory (e.g., different exchange correlation functionals). The convergence criteria are usually monitored, but stringent tests making different computational platforms compatible with each other are still under development. Typically the data are much more difficult to compare, but the database volume can grow significantly larger than in the case of highly curated data, since it benefits from a large user base. Examples of this strategy are the Harvard Clean Energy (Hachmann et al. 2011) and the NoMaD projects (Ghiringhelli et al. 2017).

Common to all the strategies is the necessity to deal with large volumes of calculations produced by several users, often with a rather diverse geographical distribution, and the necessity to store the data in easy-to-access databases. This requires efficient managers to run the calculations, check for convergence, create the appropriate entries for the database, and analyze the data. Again one can adopt different strategies for performing this task. On the one hand, there are examples of a single software platform that performs all these tasks, as, for instance, the AFLOW code (Curtarolo et al. 2012a). On the other hand, one can construct a suit of tools (often written in python), which handle the different tasks, as, for instance, the Aiida platform (Pizzi et al. 2016). The main advantage of the first strategy is the efficiency and the high level of curation, while that of the second is the flexibility of the platform. In both cases the databases are usually accessible through a graphical web interface and a back end suitable for data processing.

Creating large datasets of computational data strictly speaking is not a necessity. In principle one can replace those with experimental data and navigate them with a variety of machine learning tools (see the second part of this chapter). However, although the materials science community produces annually an enormous volume of data, these are communicated via scientific publications in a rather narrative form. As such the experimental data are extremely fragmented and difficult to harvest. There is also a multitude of curated databases, but most of them are proprietary and typically focused only on a limited number of properties (e.g., crystal structure, thermochemistry, etc.). Crystallographic databases are probably the only exception, and projects like ICSD (2018) and the crystallographic open database (Grazulis et al. 2009) currently enjoy a great popularity.

Finally, one needs to develop tools to navigate the databases. Here the central role is played by the *descriptors*. A descriptor is an easy-to-calculate property that defines a material. This can be a direct observable (e.g., the quasiparticle gap, the bulk modulus, etc.) or a quantity that acts as a proxy for a given property, such as the effective mass as a proxy for the mobility of a semiconductor. Descriptors are constructed directly from the elementary electronic properties computed with the chosen electronic structure theory. Their definition usually reflects the understanding of a property from a given community. Examples of descriptors are the *high-throughput topological insulator robustness* for the discovery topological insulators (Yang et al. 2012), the *GIBBS* method for thermal conductivity (Toher et al. 2014), the *spectroscopic limited maximum efficiency* of light-harvesting materials (Yu and Zunger 2012), etc.

2.2 Constructing Magnetic Libraries: Heusler Compounds

In short the HTEST strategy to the discovery of new materials consists in generating large libraries of hypothetical compounds and then in evaluating their properties, including the likelihood of these compounds to be made. Again, the criteria used to generate such initial libraries vary significantly depending on the targeted properties. For instance, one can select a specific crystal structure and try all

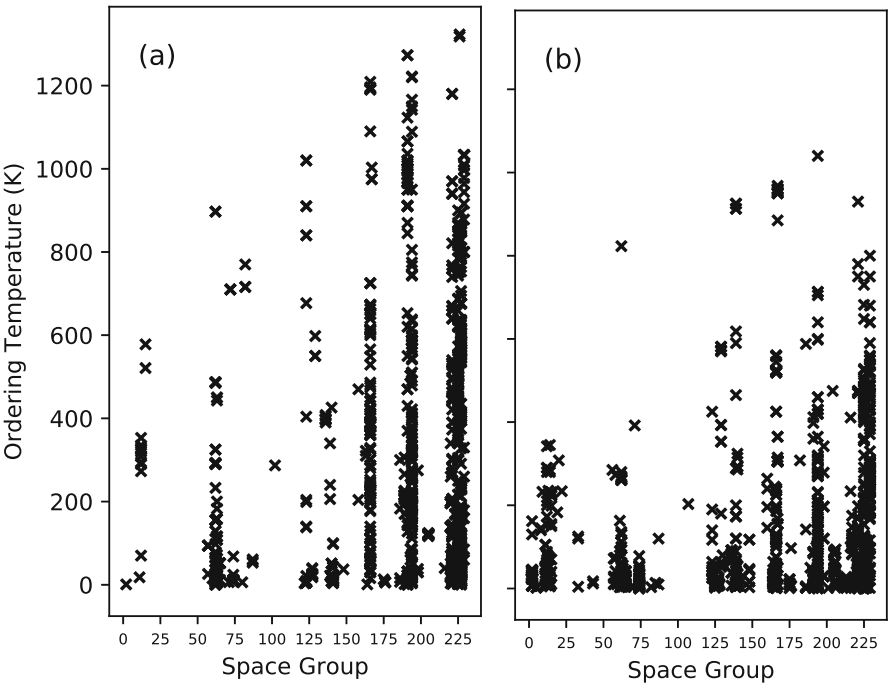

Fig. 1 Critical ordering temperature, either Curie or Néel, for known (**a**) ferromagnets and (**b**) antiferromagnets as a function of the materials space group. Note that magnetic order seems to be more favorable for high space group numbers, namely, for high-symmetry cubic crystals. This is the region in the space group space where we find the magnets with the highest critical temperatures

possible chemical compositions compatible with that structure, sometimes imposing additional constraints such as the total valence electron count (Yan et al. 2015). Magnetic compounds are distributed over a broad range of space groups (see Fig. 1) and, in addition to the limited number of magnetic ions, contain a large variety of nonmagnetic elements. As such establishing a stringent criterion for generating a library of magnetic prototypes is not straightforward. One possible criterion is to look at classes of compounds where the incidence of magnetism is high, as, for instance, in transition metals/rare-earth intermetallics (Dam et al. 2017).

Here we will describe a less selective criterion, namely, a library of Heusler alloys (Sanvito et al. 2017), a prototypical family of ternary compounds populated with several high-performance magnets (Graf et al. 2011). In their most common form, Heusler alloys are cubic compounds consisting of four interpenetrating *fcc* lattices. The so-called *regular* Heusler alloy, X_2YZ (Cu_2MnAl-type), crystallizes in the $Fm\bar{3}m$ cubic space group (No. 225), with the X atoms occupying the 8*c* Wyckoff position (1/4, 1/4, 1/4) and the Y and Z atoms being, respectively, at 4*a* (0, 0, 0) and 4*b* (1/2, 1/2, 1/2). The Y and Z elements form an octahedral-coordinated rock-salt structure, while the X atoms occupy the tetrahedral voids

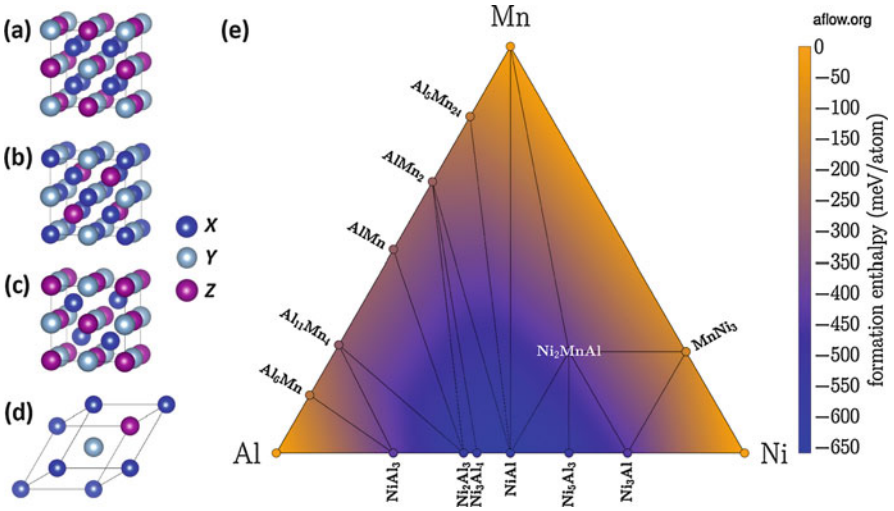

Fig. 2 Possible Heusler alloys: (**a**) regular Heusler, (**b**) inverse Heusler, and (**c**) half Heusler. In panel (**d**) we show the tetrahedral $F\bar{4}3m$ cell used to construct the electronic structure database. (**e**) Ternary convex hull diagram for Al-Mn-Ni. Note the presence of the stable HA, Ni_2MnAl. (Figure from Sanvito et al. 2017)

(see Fig. 2a). Alternatively, in the *inverse* Heusler structure, $(XY)XZ$ (Hg_2CuTi-type), now X and Z form the rock-salt lattice, while the remaining X and Y atoms fill the tetrahedral sites (Fig. 2b), so that one X atom presents sixfold octahedral coordination, while the other has fourfold tetrahedral coordination. Finally, one can also have the so-called *half* Heusler alloy, XYZ (MgCuSb-type). This is obtained by removing one of the X atoms from the regular type, thus leaving a vacancy at one of the tetrahedral sites (Fig. 2c). The Wyckoff positions are now $4a$ (0, 0, 0), $4b$ (1/2, 1/2, 1/2), and $4c$ (1/4, 1/4, 1/4), respectively, for X, Y, and Z. The minimal unit cell describing all three types can be constructed as a tetrahedral $F\bar{4}3m$ cell, containing four (three for the case of the half Heusler) atoms (Fig. 2d). Such a cell allows for a ferromagnetic spin configuration and for a limited number of antiferromagnetic and ferrimagnetic ones.

Note that it is not uncommon for Heusler alloys to have a significant amount of site-occupation disorder. In this case there is a special classification depending on the particular sites where the disorder takes place, as illustrated in Table 1 for the complete Heusler alloys and in Table 2 for the inverse ones. Modeling site-occupation disorder is a notoriously challenging problem in ab initio studies due to computational costs, although recent developments look promising for future investigations (Yang et al. 2016). Finally, it is worth mentioning that there exist a number of Heusler alloys presenting significant tetragonal distortion. These are usually associated to Mn-containing compounds, in particular with the Mn_2YZ composition (Faleev et al. 2017), and they are interesting since they can potentially display large magneto-crystalline anisotropy.

Table 1 Site occupancy and general formula for different atomic orders of Heusler compounds. The notations according to the Inorganic Crystal Structure Database (ICSD), the Strukturberichte (SB), the Pearson database, as well the space group are given. (Table adapted from Graf et al. 2011)

Site occupancy	Formula	Type (ICSD)	SB	Pearson	Space group
X, X', Y, Z	$XX'YZ$	LiMgPdSn	Y	cF16	$F\bar{4}3m$ (No. 216)
$X = X', Y, Z$	X_2YZ	Cu_2MnAl	$L2_1$	cF16	$Fm\bar{3}m$ (No. 225)
$X, X' = Y, Z$	XX'_2Z	$CuHg_2Ti$	X	cF16	$F\bar{4}3m$ (No. 216)
$X = X' = Y, Z$	X_3Z	BiF_3	DO_3	cF16	$Fm\bar{3}m$ (No. 225)
$X = X', Y = Z$	X_2Y_2	CsCl	B2	cP2	$Pm\bar{3}m$ (No. 221)
$X = Y, X' = Z$	$X_2X'_2$	NaTl	B32a	cF16	$Fd\bar{3}m$ (No. 227)
$X = X' = Y = Z$	X_4	W	A2	cI2	$Im\bar{3}m$ (No. 229)

Table 2 Site occupancy and general formula for differently ordered half-Heusler compounds. The notations according to the Inorganic Crystal Structure Database (ICSD), the Strukturberichte (SB), the Pearson database, as well the space group are given. Wyckoff position 4*d* (3/4, 3/4, 3/4) denotes the second tetrahedral lattice site, which is void in ordered materials. (Table adapted from Graf et al. 2011)

Site occupancy	Formula	Type (ICSD)	SB	Pearson	Space group
$4a, 4b, 4c$	XYZ	LiAlSi	$C1_b$	cF16	$F\bar{4}3m$ (No. 216)
$4a = 4b, 4c$	XZ_2	CaF_2	C1	cF12	$Fm\bar{3}m$ (No. 225)
$4a, 4b, 4c = 4d$	X_2YZ	Cu_2MnAl	$L2_1$	cF16	$F\bar{4}3m$ (No. 216)
$4a = 4b, 4c = 4d$	XY	CsCl	B2	cP2	$Pm\bar{3}m$ (No. 221)
$4a = 4c, 4b = 4d$	YZ	NaTl	B32a	cF16	$Fd\bar{3}m$ (No. 227)
$4a = 4b = 4c = 4d$	X	W	A2	cI2	$Im\bar{3}m$ (No. 229)

We have constructed a library of potential Heusler alloys by considering all possible combinations of three elements chosen from the 3*d*, 4*d*, and 5*d* periods and from the groups III, IV, V, and VI. These include Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Cl, Co, Cr, Cu, Fe, Ga, Ge, Hf, Hg, In, Ir, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, Os, P, Pb, Pd, Pt, Re, Rh, Ru, Sb, Sc, Se, Si, Sn, Sr, Ta, Tc, Te, Ti, Tl, V, W, Y, Zn, and Zr. We have thus avoided rare-earth elements, since in the last decade their price has been volatile on the market, so that the search for rare-earth-free magnets has become a sought-after target. If one considers all the possible stoichiometry compatible with the elements chosen and all the available crystal structures (regular, inverse, and half Heusler), a total number of 236,115 prototypes will be reached. Their electronic structure has been computed at the AFLOW standard, namely, by DFT with the GGA parameterized by Perdew et al. (1996). The DFT platform is VASP (Kresse and Furthmuller 1996), and the convergence criteria is quite stringent, namely, the convergence tolerance is 1 meV/atom and the typical Brillouin zone sampling is over a dense grid of 6,000–10,000 *k*-points per reciprocal atom (Calderon et al. 2015). The structures are fully relaxed, and in general the electronic structure is initialized in such a way to have the maximum possible spin. This gives a unit-cell ferromagnetic initial configuration whenever the unit cell contains more than one magnetic ion. The calculations are performed

without spin-orbit interaction. More details about the computational method are reported in Calderon et al. (2015).

2.3 The Descriptors

2.3.1 Energy-Related Descriptors

The very first property that one usually wishes to monitor in the HTEST protocol is the likelihood that a particular compound can be grown. The appropriate descriptor is then the Gibbs free energy, $G = H - ST$, where H is the enthalpy, S the entropy, and T the temperature. If G can be calculated, then all thermodynamically stable compounds can be identified, leaving out only those possible in the form of long-living metastable phases. However, the calculation of the Gibbs free energy is too demanding to be carried out in a high-throughput framework, so that in general one chooses to evaluate quantities related only to the DFT total energy.

A very crude but yet selective criterion is given by the enthalpy of formation. For the Heusler X_2YZ one has to check that its enthalpy, H_{X_2YZ}, is lower than the sum of the enthalpies of formation of its elementary constituents, namely, $\Delta H = H_{X_2YZ} - (2H_X + H_Y + H_Z) < 0$. For instance, for Ni_2MnAl, a now well-established magnetic shape memory alloy (Ziebeckt and Webster 1975), one has to check that H_{Ni_2MnAl} is lower than the sum of the enthalpies of formation of *fcc* Ni, Mn, and Al, their energy-lowest structures. As a further approximation, the enthalpy of formation is replaced by the DFT total energy.

A second, much more stringent, stability criterion is obtained by constructing for every three elements the convex hull diagram of the associated ternary compounds (Lukas et al. 2007), *Descriptor 1*. This essentially consists in establishing whether the enthalpy of formation of X_2YZ is lower than that of all the possible binary and elementary decompositions, for instance, XY+XZ, X_2Y+Z, XYZ+X, etc. Again by taking the example of Ni_2MnAl, we find that it is stable against all possible decompositions along the Al-Mn-Ni convex hull diagram, which is presented in Fig. 2e. The most stable structure is the regular Heusler with a formation energy of −404 meV/atom, but there are also three unstable ones with $\Delta H < 0$, namely, Mn_2NiAl (ΔH=−209, 121 meV/atom above tie-plane), NiMnAl (ΔH=−39, 400 meV/atom above tie-plane), and Al_2MnNi (ΔH=−379, 100 meV/atom above tie-plane). Such three unstable structure, all with $\Delta H < 0$, clearly prove that the enthalpy of formation alone says little about the stability of a given compound. It is also worth to mention that even the convex hull, as calculated so far, provides only an incomplete descriptor for the thermodynamical stability. In fact, in principle one has also to check possible decompositions involving alternative ternary compounds. Furthermore, once a ternary material is predicted stable, there is still the possibility that the ground state crystal structure is not a Heusler type. As such, once a stoichiometry has been predicted stable against decomposition, one should perform further crystal structure prediction calculations. Possible methods include random sampling (Pickard and Needs 2011), genetic algorithms (Oganov and Glass 2006), or targeted sampling (d'Avezac et al. 2012).

One has to consider that the probability of finding new ground states per alloy reduces with increasing the number of species. This can be seen by performing a phase stability experiment in the thermodynamic data available in the AFLOW.org repository (Curtarolo et al. 2012b), which is performed by calculating all the available binary, ternary, quaternary, etc. convex hulls (Lukas et al. 2007). The enthalpies can then be analyzed, and the average gain in enthalpy with respect to the number of species N, $\delta H(N) \equiv H(N) \rightarrow (N+1)$, decreases with N. This quantity competes against the disorder state, unavoidably promoted by configurational entropy going as $\approx \log(N)$. If follows that, in our case ternary case, the existence/absence of a ternary compound depends *mostly* on decomposition in binaries than in competition by other ternaries. As such, a first-order stability analysis, based on simply checking binaries, will be an appropriate approach capable to weed out most of the uninteresting compositions.

Additional energy-related descriptors can be defined, depending on the properties and materials class that one wants to investigate. In general, it is useful to have an idea on how robust is the thermodynamical stability of a compounds. This can be estimated by calculating the enthalpy of formation relative to that of the most stable balanced decomposition. If such difference is around $k_{\mathrm{B}}T$ at room temperature (k_{B} is the Boltzmann constant, $k_{\mathrm{B}}T_{\mathrm{RT}} = 25\,\mathrm{meV}$), then one usually defines the material as robustly stable (Sanvito et al. 2017). This descriptor, *Descriptor 2*, is denoted as Δ^{kT}.

Finally, considering that Heusler alloys are often prone to site-occupation misplacement, it is useful to define a descriptor, *Descriptor 3*, which identifies the tendency to disorder. This is the entropic temperature (Curtarolo et al. 2013; Hart et al. 2013; Yong et al. 2014), T_{S}, which is defined for a binary XY alloy as

$$T_{\mathrm{S}} = \max_{i} \left[\frac{\Delta H(X_{x_i} Y_{1-x_i})}{k_{\mathrm{B}}[x_i \log x_i + (1-x_i)\log(1-x_i)]} \right], \tag{1}$$

where i counts all the stable compounds in the XY binary system. In practice T_{S} measures the ability of an ordered phase to resist deterioration into a temperature-driven, entropically promoted, disordered mixture. This is quantified by the concentration-maximized formation enthalpy weighted by the inverse of its ideal entropic contribution (random alloy). By convention we choose the sign of T_{S} so that a positive temperature is needed for competing against the compound stability ($T_{\mathrm{S}} < 0$ if $\Delta H > 0$), and one expects $T_{\mathrm{S}} \rightarrow 0$ for a compound spontaneously decomposing into a disordered mixture.

In Eq. (1), the curvature of the iso-max ideal-latent heat locus can be seen as a descriptor of the (ideal) entropy that a system can absorb upon nucleation in a single phase and internal reorganization containing ideal disorder. As such, it is a viable descriptor to tackle the unavoidable entropy stabilization happening in half Heuslers (entropy of vacancies) or in full Heuslers (entropy of species permutations). In these systems, disorder might overcome enthalpy stabilization and lead to stable solid solutions unobtainable with prediction algorithms mentioned above.

2.3.2 Magnetic Descriptors

A magnet can be defined as "high-performing," when it presents a range of electronic and magnetic properties particularly suitable for a given application. Thus, for instance, a high-performing permanent magnet will need to score high against descriptors different from those of a magnet designed for magnetocaloric applications (Coey 2009). Some magnetic descriptors are observables and can be extracted directly from the DFT calculations. This is, for instance, the case of the magnetic moment, m (*Descriptor 4*), which can be directly compared with the experimental saturation magnetization. In the case the magnet is screened for spintronics applications (for magneto-transport), it is important to assess its spin polarization at the Fermi level. This is defined as (*Descriptor 5*)

$$P_{\mathrm{F}}^{\alpha} = \frac{n_{\mathrm{F}}^{\uparrow}(v_{\mathrm{F}}^{\uparrow})^{\alpha} - n_{\mathrm{F}}^{\downarrow}(v_{\mathrm{F}}^{\downarrow})^{\alpha}}{n_{\mathrm{F}}^{\uparrow}(v_{\mathrm{F}}^{\uparrow})^{\alpha} + n_{\mathrm{F}}^{\downarrow}(v_{\mathrm{F}}^{\downarrow})^{\alpha}}\,, \tag{2}$$

where n_{F}^{σ} and v_{F}^{σ} are, respectively, the density of states (DOS) at the Fermi level and the Fermi velocity for the spin-σ sub-band. In general, P_{F}^{α} describes the degree of spin polarization of the current that the material can sustain in different transport conditions (Mazin 1999). Here $P = 1$ indicates that only one spin species is responsible for the electron transport (Coey and Sanvito 2004). Note that different values of α characterize different transport experiments. For instance, $\alpha = 0$ is for tunneling, $\alpha = 1$ for ballistic transport, and $\alpha = 2$ for diffusive. Note also that P_{F}^{α} is a direct observable only for $\alpha = 0$ (e.g., with a spectroscopical measurement).

There are also magnetic descriptors, whose definition requires additional DFT calculations, possibly combined with other theories. This is, for instance, the case of the magneto-crystalline anisotropy (MCA – *Descriptor 6*), which measures the ability of a magnet to maintain the magnetization direction fixed in space. The MCA contributes to the hysteresis of a magnet, but it is difficult to isolate from other contributions to the total anisotropy (e.g., shape anisotropy). Computing the MCA is in general involved as it requires highly accurate total energy or perturbation theory calculations including spin-orbit interaction (Błoński and Hafner 2009).

Finally, common to all magnets is the critical ordering temperature, T_{C} (Curie or Neél), *Descriptor 7*. This requires knowledge of the compound magnetic excitation spectrum, namely, of the susceptibility, χ. There are several ways to determine such quantity and all require significant additional work. One of such possibilities is that of using time-dependent DFT (Savrasov 1998), which is capable of describing both Stoner excitations and spin waves. In the case the excitation spectrum is dominated by spin waves, a popular choice is that of performing a series of DFT total energy calculations for different magnetic configurations and mapping these onto a Heisenberg model of some kind. There are several different approaches to perform both the mapping and the total energy calculations, including magnetic force theorem (Oswald et al. 1985), spin spirals (Sandratskii 1986), and collinear multi-configurations (Archer et al. 2011). Finally, usually T_{C} is computed either by using mean-field theory or with Monte Carlo simulations.

2.4 Analysis

Out of the 236,115 prototypes contained in our Heusler alloys library, only 35,602 have a DFT energy lower than the sum of that of their elementary phases, namely, they satisfy the condition $\Delta H < 0$. These 35,602 include 6,778 compounds with a magnetic ground state. Such an estimate can be slightly conservative. In fact we are able to describe only those magnetic phases, where the magnetic cell is the same as the primitive one. This clearly excludes many potential magnetic solutions, in particular complex antiferromagnetic and ferrimagnetic phases, which are often found in some of the magnetic Heuslers (Rode et al. 2013). Furthermore, we may expect that the ferromagnetic DFT initialization may force final ferromagnetic solutions to otherwise antiferromagnetic ground states, i.e., the DFT run may converge to a local minimum. In any case 6,778 potentially new magnets is an extraordinary large number, since only 4,000 magnets are known to date.

The application of *Descriptor 1*, however, reduces drastically such number. Since the construction of the ternary convex hull diagrams requires a significantly large number of calculations, we have initially limited our analysis to intermetallic prototypes, namely, to Heusler structures made of $3d$, $4d$, and $5d$ elements only. These are 36,540 in total, but only 248 are on the convex hull; hence only 248 are predicted stable. Most importantly only 20 have a magnetic ground state, meaning that the incidence of stable magnetism in our Heusler library has to be estimated to be around 0.05%. A close look at the newly predicted magnetic Heusler alloys reveals that, with the exception of Rh_2FeZn, only three families are found, namely, Co_2YZ, Mn_2YZ and X_2MnZ.

The Co_2YZ family is perhaps the most interesting; it contains already 25 known compounds all lying on the Slater-Pauling curve (Graf et al. 2011) and houses several half-metals. Our search found four new stable alloys, namely, Co_2VZn, Co_2NbZn, Co_2TaZn, and Co_2MnTi. The first three have the low valence electron count of 25, while for Co_2MnTi this is large, 29. The numerical regression used to evaluate T_C correctly places these four on the Slater-Pauling curve, as shown in Fig. 3, and predicts Co_2MnTi to have the remarkable T_C of 940 K. The other three new compounds have all a T_C around 200 K, but two of them become nonmagnetic upon tetragonal distortion leaving only Co_2VZn magnetic (predicted $T_C \sim 228$ K).

Five Mn_2YZ compounds populate the second family, namely, Mn_2PtRh, Mn_2PtCo, Mn_2PtPd, Mn_2PtV, and Mn_2CoCr. In general Mn_2YZ alloys are found to crystallize in the regular Heusler structure if the atomic number of the Y ion is smaller than that of Mn, $Z(Y) < Z(\mathrm{Mn})$, and the inverse one for $Z(Y) > Z(\mathrm{Mn})$. This, however, is valid when one element from the main groups occupies the $4c$ position. In our case of intermetallic Heuslers, we find that the regular $Fm\bar{3}m$ structure is always the ground state, regardless of the chemical composition. Furthermore, all the new compounds present some degree of antiferromagnetic coupling, which results in either a zero-moment ground state when Mn is the only magnetic ion, and in a ferrimagnetic configuration when other magnetic ions are

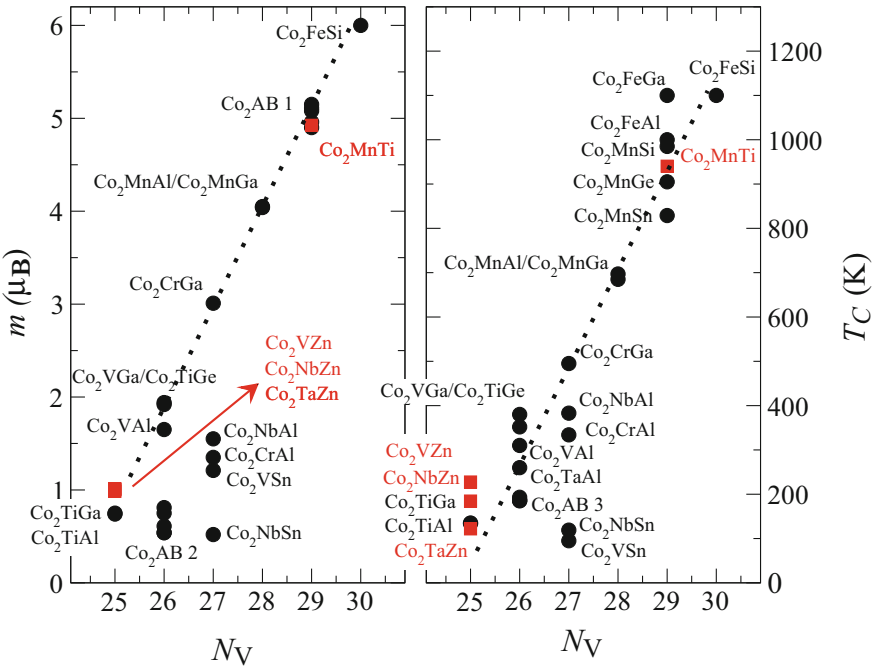

Fig. 3 Slater-Pauling curve for magnetic Heusler alloys of the form Co_2YZ. The magnetic moment per formula unit, m, is plotted against the number of valence electron, N_V, in the left panel, while T_C is displayed on the right. Red symbols correspond to predicted HAs while the black ones to existing materials. For the sake of clarity, several compounds have been named collectively on the picture. Co_2AB 1 (Co_2FeGa, Co_2FeAl, Co_2MnSi, Co_2MnGe, Co_2MnSn), Co_2AB 2 (Co_2TaAl, Co_2ZrAl, Co_2HfGa, Co_2HfAl, Co_2TaGa), Co_2AB 3 (Co_2ZrAl, Co_2HfAl, Co_2HfGa, Co_2TaGa). (Figure from Sanvito et al. 2017)

present. No estimate of T_C is possible in this case since there are no data for regular Mn_2YZ alloys to run the regression against.

Finally, we have found several X_2MnZ compounds (13), most of them including a 4d ion (Ru, Rh, and Pd) in the tetrahedral X position. In general, these compounds have a magnetic moment per formula unit ranging between 4 and 5 μ_B, consistent with the nominal 2+ valence of Mn in octahedral coordination. The regression, run against 18 existing compounds of which 13 are with X=Ru, Rh, or Pd, establishes a correlation between the Mn-Mn nearest neighbor distance, d_{Mn-Mn}, and T_C. This effectively reproduces the empirical Castelliz-Kanomata curves (Castelliz 1955; Kanomata et al. 1987) and returns us the prediction that none of the X_2MnZ alloys will display a T_C exceeding 500 K.

The final and ultimate validation of the high-throughput approach must come from an experimental lab. With this in mind, we have attempted the growth of four newly predicted alloys, namely, Co_2MnTi, Mn_2PtPd, Mn_2PtCo, and Mn_2PtV. Co_2MnTi is chosen because of its high Curie temperature, while among

the Mn_2-based alloys, we have selected two presenting a ferrimagnetic ground state (Mn_2PtCo and Mn_2PtV) and one meeting the stringent robustness criterion (Mn_2PtPd). Details of the growth and magnetic characterization can be found in Sanvito et al. (2017).

Two of the four Heusler alloys have been successfully synthesized, Co_2MnTi and Mn_2PtPd, while the other two, Mn_2PtCo and Mn_2PtV, decompose into binary compounds (notably in both cases, one of the products of the decomposition was an intermetallic with a very large unit cell, and it was not included in our database). Co_2MnTi was found to crystallize in the regular $Fm\bar{3}m$ Heusler structure with a lattice parameter, $a = 5.89$ Å in close agreement to the predicted one, $a = 5.84$ Å. Most strikingly the T_C extrapolated from the zero-field cooled magnetization curve in a field of 1 T is found to be 938 K, essentially identical to that predicted by our regression, 940 K. This is a remarkable result, since to our knowledge it is the first time that a new high-temperature ferromagnet has been discovered by high-throughput means.

Also Mn_2PtPd was found to be a single phase without evidence of decomposition. X-ray diffraction reveals that the structure is that of a tetragonally distorted regular Heusler (*I4/mmm* – $TiAl_3$-type) with lattice parameters $a = 4.03$ Å and $c = 7.24$ Å. Magnetization curves at room temperature and 4 K show no hysteresis or spontaneous magnetization indicating that the compound is antiferromagnetic at low temperature.

3 Machine Learning for Materials Discovery

The general concept underpinning machine learning methods for materials science is relatively simple. One wishes to establish relations between known microscopic features, including chemical and perhaps structural information, with a number of physical/chemical properties. In practice the machine learning model, f_{ML}, defined as

$$f_{\mathrm{ML}} : \vec{v} = (\{Z_i\}, N_{\mathrm{V}}, \ldots) \rightarrow p, \tag{3}$$

associates to a *feature vector*, $\vec{v}$, the property p. Then, $\vec{v}$ contains information such as the atomic numbers, $\{Z_i\}$, of the elements in a compound; the number of valence electrons, N_{V}; etc. The machine learning models are purely numerical and do not require any particular physical knowledge of the property under investigation. There is a multitude of algorithms available, going from simple linear regression to complex neural network schemes (Shalev-Shwartz and Ben-David 2014; Hastie et al. 2013). These can deliver a classification, namely, they can sort materials in classes or the continuous value of a given property (e.g., the Curie temperature of a ferromagnet). The standard procedure in creating a machine learning model is to train the model over a given subset of data, the *training set*, and then to evaluate the quality of the learning by predicting the property of data never seen before, the *test set*.

There are a number of issues related to machine learning applied to materials science, which are relatively general to any algorithm and property. The first concerns the many possible ways to represent the feature vector, which should satisfy a number of criteria. In particular the feature vector should be continuous for small variations in the features, with two materials having the same representation only if their target properties are identical. This criterion essentially establishes a notion of similarity across the materials space. For instance, if one seeks at describing a property that strongly depends on the number of valence electrons, elements are represented by their group in the periodic table better then by their atomic number. In fact, Si and Ge will have the same representation when using the group (both in group IV), while they will appear as rather distant if the atomic number is considered (14 for Si and 32 for Ge). The representation of the crystal structure poses particular problems. The space group, for instance, is difficult to use, since tiny distortions from a high-symmetry structure (high space group number) may result in a space group with a rather low number. Furthermore, whatever representation one chooses, it should be both translational and rotational invariant and should be insensitive on the choice of the unit cell (for solids). Note that these criteria associated to the feature vector are not stringent conditions in general. In principle a solid machine learning model based on an extremely large number of data should be able to establish patterns among the data even if the distance between different feature vectors does not follow an intuitive physical/chemical metric. Unfortunately the materials space is not large, so that using feature vectors with the properties described above is highly desirable.

The second issue is related to the so-called curse of dimensionality, which establishes that as the dimension of the feature space gets larger, the number of possible distinct configurations increases exponentially. In other words the density of points needed to extrapolate the data in a multidimensional space grows exponentially with the dimension of the space itself. This essentially means that in order to obtain the same quality of learning, the number of data needed grows exponentially with the dimension of the feature vector, i.e., adding more information to the feature vector may decrease the change to establish patterns in the data. In materials science the curse of dimensionality is a serious issue for two main reasons. Firstly, the chemical materials space is large but not huge. Secondly, the availability of experimental data is limited by the fragmentation of the various databases, and the fact that results are communicated in scientific publications, which are little data oriented. This has two main consequences. Firstly, most of the machine learning studies to date train on theoretical data (as the ones contained in AFLOWLIB.org) (Isayev et al. 2017) and not experimental ones, though predictions show to be consistent with the experimental results. Secondly, one has to design machine learning algorithms, where the dimension of the feature vector remains small (Ghiringhelli et al. 2015).

When applied to magnetism, the use of machine learning models remains very limited. Here we will present two illustrative examples, looking, respectively, at predicting the magnetic moment of Fe-containing Heusler alloys and at sorting tetragonally distorted ones into hard and soft magnets. Other examples of machine

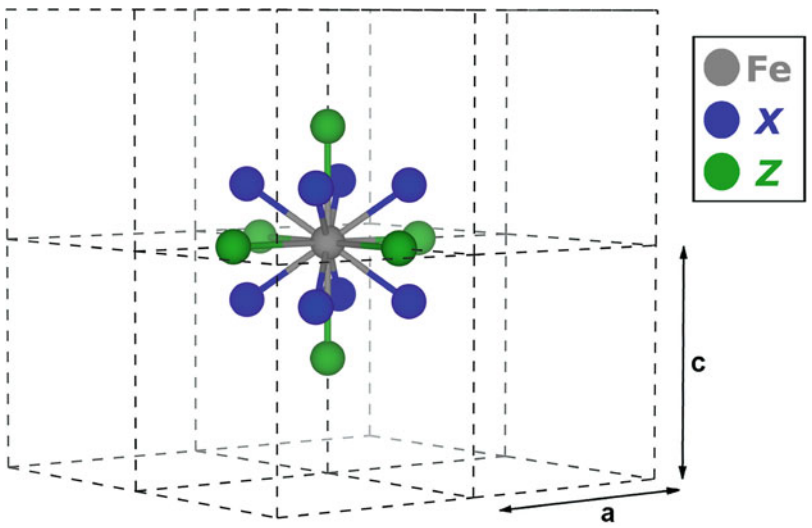

Fig. 4 The local coordination of the atomic sites in a Heusler alloy. The neighbors of the central Fe atom form two shells of different symmetry. Atoms belonging to the nearest neighbor shell, shown in blue, coordinate the central atom tetrahedrally. The next nearest neighbor shell is made out of six (green) atoms and has octahedral symmetry. (Figure adapted from Žic et al. 2017)

learning approaches to magnetism include the identification of novel magnetic phases with enhanced magnetic anisotropy (Kusne et al. 2014), the discovery of intermetallic compounds (Oliynyk and Mar 2018), and the prediction of the Curie temperature of transition-metal rare-earth compounds (Dam et al. 2017).

3.1 Magnetic Moment Predictions

We will now present a machine learning strategy to predict the magnetic moment of $L2_1$ Fe-containing Heusler alloys (Žic et al. 2017) of the $X_2\mathrm{Fe}Z$ type. In general the magnetic moment of $3d$ transition metals is highly localized and can be described as atomic-like. This may get modified by the surrounding ions, which should be then included into the definition of a suitable feature vector. We represent the local structure by constructing clusters centered around the Fe ions as illustrated in Fig. 4. The center of the cluster is occupied by Fe, the first coordination shell is populated by ions X with atomic number Z_1, while the second nearest neighbors are ions Z of atomic number Z_2.

A feature vector can then be defined as the following:

$$\vec{v} = (\{Z_i\}, \{r_{0i}\}, \{N_i\}, a_{\mathrm{lat}}), \tag{4}$$

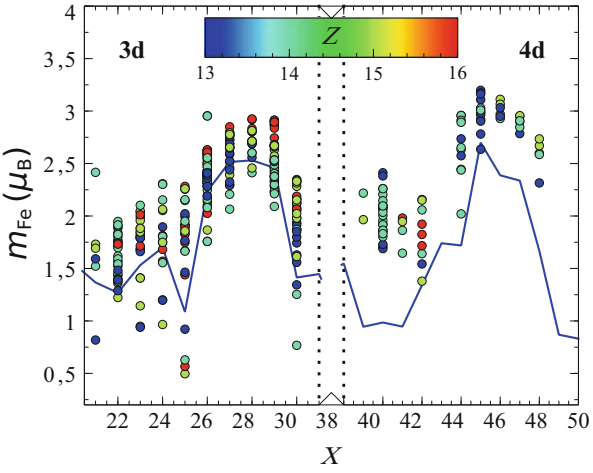

Fig. 5 *Left panel* – Magnetic moment of Fe, m_{Fe} (in μ_{B}), for a wide range of nearest neighbors at a constant Wigner-Seitz volume ($R_{\mathrm{WS}} = 2.7$ bohr). The atomic number of the next nearest neighbor, Z_3, is color coded, while here we plot data as a function of the first nearest neighbor atomic number, Z_1. We can notice a linear increase of the magnetic moment across the transition metal series which does not depend on Z_3. The symbols are the DFT data, while the corresponding machine learning trend is shown with the blue line. *Right panel* – a data sample containing a wider range of main group elements. The data elucidates the origin of the oscillation in the machine learning trend throughout the main group series. (Figure adapted from Žic et al. 2017)

where $\{Z_i\}$ are the atomic number of the ions in the cluster ($i = 1, 2$), $\{N_i\}$ is their valence, r_{0i} is the distance between the central Fe ion and the i-th one, while a_{lat} is the lattice constant of the relative Heusler compound.

We have then used such feature vector in a random forest regression algorithm trained over a large number of Heusler alloys, whose electronic structure was computed at the level of GGA-DFT (Žic et al. 2017). In fact, we have constructed a number of machine learning models using feature vectors similar to that presented here and capable of predicting the magnetic moment of an atom at the center of the cluster of any type (not just Fe). After having noticed that the Fe magnetic moment is little affected by the choice of Z_2, as long as this crystal position is occupied by an element belonging to one of the main groups, we have focused our attention on the dependence of the Fe moment, m_{Fe}, over the Z_1. In general we have discovered that the valence of X determines m_{Fe}. In fact, if this is smaller than 8, then the moment will decrease; otherwise it will increase. This trend is true for all the elements in the $3d$ and $4d$ periods, as shown in Fig. 5, so that the largest Fe moment is found when $X =$ Ni and Pd.

3.2 Anisotropy Analysis: Saving Computational Time

As a second example of machine learning methods applied to magnetism, we describe here the construction of a classification scheme to sort out hard from soft magnets (Žic 2017). The magnetic hardness of a compound is determined, at the microscopic level, by the magneto-crystalline anisotropy (Bloński and Hafner 2009). This is a well-defined quantity, which depends on the spin-orbit coupling strength, the crystal structure, and the magnetic ion responsible for the magnetic moment. In general it can be computed either by using perturbation theory or by total energy differences. In both cases the calculation is delicate as it required an extremely accurate evaluation of the Fermi surface. In fact, the anisotropy is very sensitive to the occupation of the various d orbitals so that one needs to determine the position of the Fermi energy at a high degree of accuracy. As a result sampling the k-points over a very fine mesh is mandatory, and the calculations end up to be quite time-consuming. Clearly, if one can construct a machine learning strategy to the calculation of the magneto-crystalline anisotropy, the high-throughput analysis of magnetic materials will become more feasible.

Our dataset was constructed from a Heusler alloys library, from which we have selected about 300 compounds presenting a magnetization exceeding 0.5 μ_B/fu and tetragonal distortion $c/a \neq 1$. In this case the thermodynamical stability was not considered so that the dataset includes both stable and unstable structures. The magnetic anisotropy was then calculated at the level of DFT-GGA using magnetic force theorem. These initial calculations have returned us magneto-crystalline anisotropy values ranging between 0 and 15 MJm^{-3}, although for most of the compounds, the anisotropy is smaller than 8 MJm^{-3}.

We have then constructed a machine learning model based on a ridge classifier (Pedregosa et al. 2011) and different choices of feature vector. In particular the best model uses the following vector:

$$\vec{v} = (\{Z_i\}, \{N_i\}, \{\zeta_i\}, \{M_i\}, V), \tag{5}$$

where Z_i, N_i, and M_i are the atomic number, the number of valence electron, and the local magnetic moment of the different species in X_2YZ, respectively. In addition V is the unit cell volume and ζ_i is a quantity associated to the spin-orbit coupling strength (Dunn 1961). Importantly, the magnetic moments are estimated using a generalization of the machine learning model described in the previous section (the central atom does no longer need to be Fe) (Žic et al. 2017), so that the feature vector does not comprise any quantities computed ab initio. We have then set the magneto-crystalline anisotropy value of 0.8 MJm^{-3} as the boundary between hard and soft magnets. This choice ensures that the population of hard and soft magnets in our dataset is identical.

Although hard magnets are our target, we have decided to train the machine learning model so to identify the soft ones. The reason behind such decision is primarily related to the fact that hard magnets are scarce and the cost of

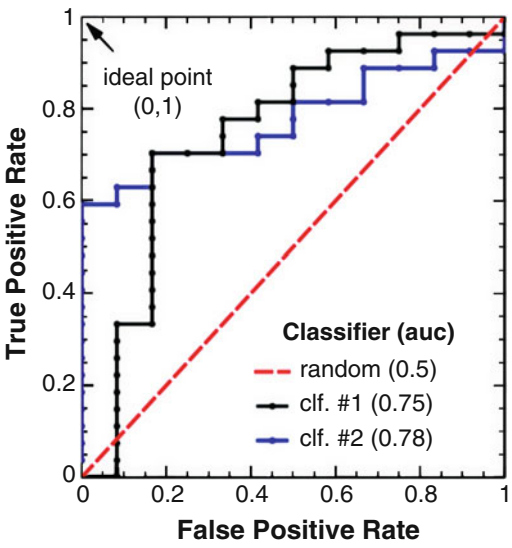

Fig. 6 Receiver operating curve (ROC) for the two hard magnet classifiers discussed in the text (left). The area under the curve (auc) is indicated in the brackets. The machine learning model described in the text is clf 2, while clf 1 represents a second model with a different choice of feature vector. The dashed red line is for a random classifier

unintentionally dismissing a viable candidate is high. This means that a very selective machine learning model will be able to predict reliably hard magnets but will likely discharge many of them. We therefore choose a strategy where we maximize the number of correctly identified soft magnets, namely, the true positive rate (TPR), under the constraint that no hard magnet is mistaken for a soft magnet, i.e., we aim to keep the false positive rate (FPR) as small as possible. This will select the soft compounds and discharge both soft and hard.

In Fig. 6 we present the receiver operating curve (ROC) for the machine learning model described in the text (clf 2), for a second one constructed with a different feature vector (clf 1) and for a random classifier (dashed red line). One can move along the ROC by changing the decision value cutoff. Ideally one wishes the model to be at the top left corner of the ROC, namely, to have a true positive rate of 100% (all compounds with the desired property are selected) and a false positive rate of 0 (no compounds are misclassified). In a realistic machine learning model, this is not possible and one has to take a compromise. In our case the best solution is that of having a false positive rate of 0 and a true positive rate of around 60%. When applied to soft magnets, this means that 60% of the soft magnets will be found without an error, meaning that 30% of the total compounds in the dataset will be correctly classified as soft (we consider a sample where the soft/hard ratio is 1). For those one will not need to perform any further calculation. Clearly such result becomes important if the dataset is very extended. Let us imagine, for the sake of the argument, that one wishes to compute the magneto-crystalline anisotropy for all the 236,000 Heusler alloys contained in the AFLOWLIB.org database. Our machine learning scheme will then save about 79,000 calculations.

4 Conclusion

In this chapter we have described how high-throughput electronic structure theory and machine learning strategies can accelerate the discovery and the design of new magnets. This indeed is an area of magnetism, which is still in its infancy, but that has already demonstrated a strong potential. At a first look, it appears that the efficiency of the approach is not high and the most skeptical reader may argue that a more conventional trial and error approach can be more fruitful. This is true only in part. There are, in fact, a few key points to be noted.

Firstly, one has to expect that as the databases grow, the need for further calculations will be reduced. The mapping of the Heusler alloys thermodynamical stability has certainly been a challenging computational undertaking. However, we were able to conduct a first rapid screening of the intermetallic subset simply because the data for binary compounds were already in the AFLOWLIB.org database. It is particularly important to remark that many of these data were generated for research projects completely different from the one described here, meaning that data sharing is a unique asset of any high-throughput initiative.

The opposite argument is also valid, namely, our data have been useful not only to discover new magnets but also for screening materials according to other properties. For instance, the half Heusler library was investigated in the search of compounds presenting low thermal conductivity (Carrete et al. 2014). This concept of "recycling data" is indeed an important one, and certainly the success of the method needs to be evaluated over the entire range of discoveries that a dataset enables.

Finally, even data that do not contribute to a direct discovery (materials that are not thermodynamically stable) still can be used to construct machine learning algorithms to rapidly navigate the data themselves. Most importantly such machine learning models will help us in directing further calculations, so to improve the discovery throughput and identify gaps and opportunities in the materials space.

Acknowledgements This work is supported by Science Foundation Ireland (Grants No. 14/IA/2624). JN thank the Irish Research Council for financial support. SC and CO acknowledge support by DOD-ONR (N00014-13-1-0635, N00014-15-1-2863, N00014-16-1-2326) and the consortium AFLOW.org – Duke University – for computational assistance. SC acknowledges the Alexander von Humboldt Foundation for financial support.

References

Archer T, Pemmaraju C, Sanvito S, Franchini C, He J, Filippetti A, Delugas P, Puggioni D, Fiorentini V, Tiwari R, Majumdar P (2011) Exchange interactions and magnetic phases of transition metal oxides: benchmarking advanced ab initio methods. Phys Rev B 84:115114

Błoński P, Hafner J (2009) Density-functional theory of the magnetic anisotropy of nanostructures: an assessment of different approximations. J Phys Condens Matter 21:426001

Calderon C, Plata J, Toher C, Oses C, Levy O, Fornari M, Natan A, Mehl M, Hart G, Nardelli M, Curtarolo S (2015) The AFLOW standard for high-throughput materials science calculations diagrams. Comput Mat Sci 108:233–238

Carrete J, Li W, Mingo N, Wang S, Curtarolo S (2014) Finding unprecedentedly low-thermal-conductivity half-heusler semiconductors via high-throughput materials modeling. Phys Rev X 4:011019

Castelliz L (1955) Beitrag zum ferromagnetismus von legierungen der ubergangsmetalle mit elementen der b-gruppe. Z Metallk 46:198–203

Coey J (2009) Magnetism and magnetic materials. Oxford University Press, Oxford

Coey J, Sanvito S (2004) Magnetic semiconductors and half-metals. J Phys D Appl Phys 37: 988–993

Curtarolo S, Setyawan W, Hart G, Jahnatek M, Chepulskii R, Taylor R, Wang S, Xue J, Yang K, Levy O, Mehl M, Morgan D (2012a) AFLOW: an automatic framework for high-throughput materials discovery. Comput Mat Sci 58:218–226

Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor R, Nelson L, Hart G, Sanvito S, Nardelli M, Mingo N, Levy O (2012b) AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. Comput Mat Sci 58:227–235

Curtarolo S, Hart G, Nardelli M, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12:191–201

Dam HC, Nguyen VC, Pham TL, Nguyen AT, Kino H, Terakura K, Miyake T (2017) A regression-based feature selection study of the curie temperature of transition-metal rare-earth compounds: prediction and understanding. arXiV:cond-mat

d'Avezac M, Luo JW, Chanier T, Zunger A (2012) Genetic-algorithm discovery of a direct-gap and optically allowed superstructure from indirect-gap Si and Ge semiconductors. Phys Rev Lett 108:027401

Dunn TM (1961) Spin-orbit coupling in the first and second transition series. Trans Farad Soc 57:1441

Faleev SV, Ferrante Y, Jeong J, Samant MG, Jones B, Parkin SS (2017) Origin of the tetragonal ground state of Heusler compounds. Phys Rev Appl 7:034022

Franchini C, Archer T, He J, Chen XQ, Filippetti A, Sanvito S (2011) Exceptionally strong magnetism in the 4*d* perovskites $R\mathrm{TcO}_3$ (R = Ca, Sr, Ba). Phys Rev B 83:220402

Ghiringhelli L, Vybiral J, Levchenko S, Draxl C, Scheffler M (2015) Big data of materials science: critical role of the descriptor. Phys Rev Lett 114:105503

Ghiringhelli L, Carbogno C, Levchenko S, Mohamed F, Huhs G, Lueders M, Oliveira M, Scheffler M (2017) Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. NPJ Comput Mater 3:46

Graf T, Felser C, Parkin S (2011) Simple rules for the understanding of Heusler compounds. Prog Solid State Chem 39:1–50

Grazulis S, Chateigner D, Downs RT, Yokochi AT, Quiros M, Lutterotti L, Manakova E, Butkus J, Moeck P, Le Bail A (2009) Crystallography open database – an open-access collection of crystal structures. J Appl Crystallogr 42:726–729

Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Snchez-Carrera RS, Gold-Parker A, Vogt L, Brockway AM, Aspuru-Guzik A (2011) The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. J Phys Chem Lett 2(17):2241–2251

Hart G, Curtarolo S, Massalski T, Levy O (2013) Comprehensive search for new phases and compounds in binary alloy systems based on platinum-group metals, using a computational first-principles approach. Phys Rev X 3:041035

Hastie T, Tibshirani R, Friedman J (2013) The elements of statistical learning: data mining, inference, and prediction. Springer, New York

ICSD (2018) FIZ Karlsruhe and NIST, inorganic crystal structure database. http://icsdfiz-karlsruhede/icsd/

Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A (2017) Universal fragment descriptors for predicting properties of inorganic crystals. Nat Comm 8:15679

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater 1(1):011002

Janak J (1977) Uniform susceptibilities of metallic elements. Phys Rev B 16:255–262
Kanomata T, Shirakawa K, Kaneko T (1987) Effect of hydrostatic pressure on the curie temperature of the Heusler alloys Ni_2MnZ (Z = Al, Ga, In, Sn and Sb). J Magn Magn Mater 65:76
Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, Aykol M, Rühl S, Wolverton C (2015) The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. npj Comput Mat 1:15010
Kresse G, Furthmuller J (1996) Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15–50
Kusne AG, Gao T, Mehta A, Ke L, Nguyen MC, Ho KM, Antropov V, Wang CZ, Kramer MJ, Long C, Takeuchi I (2014) On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. Sci Rep 4:6367
Lukas H, Fries S, Sundman B (2007) Computational thermodynamics, the Calphad method. Cambridge University Press, Cambridge
Magda G, Jin X, Hagymási I, Vancsó P, Osváth Z, Nemes-Incze P, Hwang C, Biró L, Tapasztó L (2014) Room-temperature magnetic order on zigzag edges of narrow graphene nanoribbons. Nature 514:608–611
Mazin I (1999) How to define and calculate the degree of spin polarization in ferromagnets. Phys Rev Lett 83:1427–1430
Moruzzi VL, Marcus PM (1989) Magnetism in FCC rhodium and palladium. Phys Rev B 39: 471–474
Oganov A, Glass C (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. J Chem Phys 124:244704
Oliynyk AO, Mar A (2018) Discovery of intermetallic compounds from traditional to machine-learning approaches. Acc Chem Res 51:59–68
Oswald A, Zeller R, Braspenning P, Dederichs P (1985) Interaction of magnetic impurities in Cu and Ag. J Phys F 15:193
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825
Perdew J, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868
Pickard CJ, Needs R (2011) Ab initio random structure searching. J Phys Condens Matter 23:053201
Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) Aiida: automated interactive infrastructure and database for computational science. Comput Mat Sci 111:218–230
Rasmussen FA, Thygesen KS (2015) Computational 2D materials database: electronic structure of transition-metal dichalcogenides and oxides. J Phys Chem C 119(23):13169–13183
Requist R, Baruselli P, Smogunov A, Fabrizio M, Modesti S, Tosatti E (2016) Metallic, magnetic and molecular nanocontacts. Nat Nanotech 11:499–508
Rode K, Baadji N, Betto D, Lau YC, Kurt H, Venkatesan M, Stamenov P, Sanvito S, Coey J, Fonda E, Otero E, Choueikani F, Ohresser P, Porcher F, André G (2013) Site-specific order and magnetism in tetragonal Mn_3Ga thin films. Phys Rev B 87:184429
Rodriguez E, Poineau F, Llobet A, Kennedy B, Avdeev M, Thorogood G, Carter M, Seshadri R, Singh D, Cheetham A (2011) High temperature magnetic ordering in the 4*d* perovskite $SrTcO_3$. Phys Rev Lett 106:067201
Sandratskii L (1986) Energy band structure calculations for crystals with spiral magnetic structure. Phys Status Solidi B 136:167
Sanvito S, Oses C, Xue J, Tiwari A, Zic M, Archer T, Tozman P, Venkatesan M, Coey M, Curtarolo S (2017) Accelerated discovery of new magnets in the Heusler alloy family. Sci Adv 3:e1602241
Savrasov S (1998) Linear response calculations of spin fluctuations. Phys Rev Lett 81:2570–2573
Shalev-Shwartz S, Ben-David S (2014) Understanding machine learning. Cambridge University Press, Cambridge

Toher C, Plata J, Levy O, de Jong M, Asta M, Nardelli MB, Curtarolo S (2014) High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasi-harmonic Debye model. Phys Rev B 90:174107

Toher C, Oses C, Hicks D, Gossett E, Rose F, Nath P, Usanmaz D, Perim DCFE, Calderon CE, JPlata J, Lederer Y, MichalJahnátek, Setyawan W, Wang S, Xue J, Chepulskii KRRV, Taylor RH, Gomez G, Shi H, Supka AR, Orabi RARA, Gopal P, Cerasoli FT, Liyanage L, Wang H, Siloi I, Agapito LA, Nyshadham C, Hart GLW, Carrete J, Legrain F, Mingo N, Zurek E, Isayev O, Tropsha A, Sanvito S, Hanson RM, Takeuchi I, Mehl MJ, Kolmogorov AN, Yang K, D'Amico P, Calzolari A, Costa M, Gennaro RD, Nardelli MB, Fornari M, Levy O, Curtarolo S (2018) The AFLOW fleet for materials discovery. In: Handbook of materials modeling. Methods: theory and modeling, vol 1. Springer

Žic M (2017) Towards data-driven magnetic materials discovery. Ph.D Thesis, Trinity College Dublin

Žic M, Archer T, Sanvito S (2017) Designing magnetism in Fe-based Heusler alloys: a machine learning approach. arXiv p 1706.01840

Wadley P, Novák V, Campion R, Rinaldi C, Martí X, Reichlová H, Železný J, Gazquez J, Roldan M, Varela M, Khalyavin D, Langridge S, Kriegner D, Máca F, Mašek J, Bertacco R, Holý V, Rushforth A, Edmonds K, Gallagher B, Foxon C, Wunderlich J, Jungwirth T (2013) Tetragonal phase of epitaxial room-temperature antiferromagnet cumnas. Nat Commun 4:2322

Wohlfarth EP (1980) Ferromagnetic materials: a handbook on the properties of magnetically ordered substances. Elsevier, New York

Yan F, Zhang X, Yu Y, Yu L, Nagaraja A, Mason T, Zunger A (2015) Design and discovery of a novel half-Heusler transparent hole conductor made of all-metallic heavy elements. Nat Commun 6:7308

Yang K, Setyawan W, Wang S, Nardelli MB, Curtarolo S (2012) A search model for topological insulators with high-throughput robustness descriptors. Nat Mater 11:614–619

Yang K, Oses C, Curtarolo S (2016) Modeling off-stoichiometry materials with a high-throughput ab-initio approach. Chem Mater 28:6484–6492

Yong J, Jiang Y, Usanmaz D, Curtarolo S, Zhang X, Shin J, Li L, Pan X, Tachuchi I, Greene R (2014) Composition-spread growth and the robust topological surface state of Kondo insulator SmB_6 thin films. Appl Phys Lett 105:222403

Yu L, Zunger A (2012) Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. Phys Rev Lett 108:068701

Ziebeckt K, Webster P (1975) Helical magnetic order in Ni_2MnAl. J Phys F Met Phys 5:1756–1766