The high-throughput highway to computational materials design

Stefano Curtarolo[1,2*], Gus L. W. Hart[2,3], Marco Buongiorno Nardelli[2,4,5], Natalio Mingo[2,6], Stefano Sanvito[2,7] and Ohad Levy[1,2,8]

High-throughput computational materials design is an emerging area of materials science. By combining advanced thermodynamic and electronic-structure methods with intelligent data mining and database construction, and exploiting the power of current supercomputer architectures, scientists generate, manage and analyse enormous data repositories for the discovery of novel materials. In this Review we provide a current snapshot of this rapidly evolving field, and highlight the challenges and opportunities that lie ahead.

Every technology is intimately related to a particular materials set. The steam engines that powered the industrial revolution in the eighteenth century were made of steel and, information and communication technologies are underpinned by silicon. Once a material is chosen for a given technology, it gets locked with it because of the investments associated with establishing large-scale production lines. This means that changing the materials set in an established technology is a rare event and must be considered as a revolution. Moreover, the initial choice of a material is absolutely crucial for the long-lasting success of a technological sector. Importantly, recent times have seen a surge of new technological niches, each one of them potentially looking for a different materials set. Thus, the pressure on the development of new materials is becoming formidable. These should score on many counts. They should be tailored on the specific property that the technology is based on, they often should be compatible with other technologies, should not contain toxic elements, and, if needed in large quantities, should be made of cheap raw materials. As such, searching for materials is a multi-dimensional problem where many boxes should be ticked at the same time.

Although the demand for materials is endlessly growing, experimental discovery is bound by high costs and time-consuming procedures of synthesis. Is there another way? Indeed, this is the burgeoning area of computational materials science called 'high-throughput' (HT) computational materials design. It is based on the marriage between computational quantum-mechanical–thermodynamic approaches[1,2] and a multitude of techniques rooted in database construction and intelligent data mining[3]. The concept is simple yet powerful: create a large database containing the calculated thermodynamic and electronic properties of existing and hypothetical materials, and then intelligently interrogate the database in the search of materials with the desired properties. Clearly, the entire construct should be validated by reality, namely the existing materials must be predicted correctly and the hypothetical ones should eventually be made. Such a reality check feeds back to the theory to construct better databases and increase predictive power.

The HT experimental approach was pioneered over a hundred years ago by Edison[4] and Ciamician[5], but with the advent of efficient and accurate theoretical tools and inexpensive computers, its computational counterpart has become a viable path for tackling materials design. Thus, in the past decade computational HT materials research has emerged[3,6–16] following the impetus of experimental HT approaches[17–19]. In the literature, HT materials research is often confused with the combinatorial evaluation of materials properties. Although a few attempts have been made to clearly define the two concepts[20–22], the distinction is not yet rigorous. Here we define HT as the throughput of data that is way too high to be produced or analysed by the researcher's direct intervention, and must therefore be performed automatically: HT implies an automatic flow from ideas to results. The confusion of HT with combinatorial approaches is thus resolved. The latter, in fact, specifies how the degrees of freedom are investigated, whereas HT strictly defines the overwhelming and automatic flow of the investigations.

The practical implementation of computational HT is highly non-trivial. The method is employed in three strictly connected steps: (i) virtual materials growth: thermodynamic and electronic structure calculations of materials[3,23]; (ii) rational materials storage: systematic storage of the information in database repositories[24,25]; (iii) materials characterization and selection: data analysis aimed at selecting novel materials or gaining new physical insights[15,19,26].

High-throughput is often known for the large databases it generates (for example, the AFLOWLIB.org consortium[24] and the Materials Project[25]). Here we posit that all three HT stages are highly necessary, but that the last one is the most challenging and important. In fact, it is the step that allows one to extract the information and, as such, it requires a deep understanding of the physical problem at hand. The intelligent search of a database is performed by means of 'descriptors'. These are empirical quantities, not necessarily observables, connecting the calculated microscopic parameters (for example, formation and defect energies, atomic environments, band structure, density of states or magnetic moments) to macroscopic properties of the materials (for example, mobility, susceptibility or

Table 1 | Examples of descriptors introduced in the literature.

Problem	Combination of materials properties (gene)	Descriptor
Structure stability: convex hull of an alloy system	Formation enthalpy (H_f) as a function of concentration (x) and the enthalpies (H) of A and B.	$H_f(x) = H(A_{1-x}B_x) - (1-x)H(A) - xH(B)$
Phase stability in off-lattice alloys	Spectral decomposition of alloy vector-energies ($E_{n,p}$, n-rows = species, p-columns = configurations) with principal-component-analysis coefficients (α_i) and truncation error ($\epsilon(d)$) (ref. 3).	$E_{n,p} \cong \alpha_1 E_{n,1} + \cdots + \alpha_{p-1} E_{n,p-1} + \epsilon(d)$
Nanosintered thermoelectrics	Ratio of the average power factor ($\langle P \rangle$) to the grain size (L) (ref. 15).	$\hat{\chi}_{thermo} = \frac{\langle P \rangle}{L}$
Topological insulators (epitaxial growth)	Variational ratio of spin–orbit distortion versus non-spin–orbit derivative strain (E_k^{SOC}, E_k^{noSOC}, spin/no spin–orbit bandgaps at k, a_0 lattice)[16].	$\hat{\chi}_{TI} = -\frac{E_k^{SOC}(a_0)/a_0}{\delta E_k^{noSOC}(a)_0/\delta a_0\big\vert_{a_0}}$
Power conversion efficiency of a solar cell (spectroscopic limited maximum efficiency)	Ratio of the maximum output power density (P_m) to the incident solar energy density (P_{in}) — a function (η) of the radiative electron–hole recombination current (f_r) and the photon absorptivity ($\alpha(E)$) — versus bandgap energy (E_g)[62].	$\eta(\alpha(E), f_r) = P_m/P_{in}; E_g$
Non-proportionality in scintillators	Maximum mismatch between effective masses of electrons (m_e) and holes (m_h)[75].	$\hat{\chi}_{np} = \max\left(\frac{m_e}{m_h}, \frac{m_h}{m_e}\right)$
Morphotropic phase boundary piezoelectrics	Energy proximity between tetragonal, rhombohedra and rotational distortions (ΔE_p). Angular coordinate (α_{AB}) of the energy minimum in the A–B off-centerings energy map for ABO_3 systems[79].	$\Delta E_p \leq 0.5$ eV $\alpha_{AB} \approx 45°$

critical temperatures). In other words, the descriptor is the language with which the researcher speaks to the database, and thus the heart of any effective HT implementation. In Table 1 we illustrate examples of recently introduced descriptors.

Once a good descriptor is identified, the search for better materials within the repository can be performed intrinsically or extrinsically, depending on whether the optimum solutions are already included in the set of calculations or not. Intrinsic searches include just step (iii), require only fast descriptors, and may employ various informatics techniques. Examples of previous such searches include the scanning of better cathode materials[27,28], and the uncovering of unknown compounds[9,29,30], novel topological insulators[16] or thermoelectric materials[15]. Extrinsic searches involve all three steps, because the search for an optimal solution includes iterations leading to an expansion of the repository.

An important component of extrinsic HT computational research is a scheme capable of using the evaluation of descriptors on existing database entries to guide new calculations not yet included in the database. Examples of such schemes published in the literature comprise evolutionary and genetic algorithms[7,8], data mining of spectral decompositions[3] and Bayesian probabilities[10], refinement and optimization by cluster expansion[13,31] and structure map analysis[32–34]. Neural networks[35,36] and support vector machines[37] have also been utilized in a few cases. These methods may sometimes be used to bypass step (iii) of the HT analysis, that is, the formulation of a physically meaningful descriptor, so that a search can still be implemented even with only a superficial understanding of the physical problem.

Areas of current application

Following the general framework outlined above, we describe in this section a few specific examples of computational HT studies reported in the literature, ordered by increasing degree of complexity.

Thermodynamics for the identification of binary and ternary compounds. The identification of stable structures is the first step in the design of materials with various specific functionalities. The proper descriptor of alloy stability, the formation enthalpy, is the simplest example of a parameter used for HT materials development.

Alloys are the workhorse material of many important technological applications. Thus, finding new and improved alloys could be transformative in some areas and would have a substantial economic impact. When improving an existing alloy or designing a new one, scientists rely on databases of alloy thermodynamics and phase diagrams (for example, the Massalski's *Binary Alloy Phase Diagrams*[38] and the Villars's *et al. Pauling File*[39]). Although the utility of these repositories is tremendous, they could be of even greater use if they were more complete. Experimental completeness is difficult to achieve due to the vast combination space and because experimentation is often difficult: it requires high temperatures or pressures, very long equilibration processes, or may involve hazardous, highly reactive, poisonous or radioactive materials. Computational compilation of the properties of materials is more feasible and will lead to much more complete repositories. Examples that demonstrate this are the almost simultaneous prediction and experimental verification of the previously unknown $C11_b$ structure of the Pd_2Ti compound[9,40], the verification by Niu *et al.*[41] of an earlier prediction[42] that the CrB_4 compound, thought for 40 years to have an oI10 structure, is actually more stable in an oP10 structure, and the simultaneous synthesis and solution, by an *ab initio* evolutionary search, of an unexpectedly complex tI56 crystal structure of CaB_6 (ref. 43).

In alloy design, the targets of the formation enthalpy descriptor are stable phases. The HT *ab initio* method explores the phase stability landscape of alloys by calculating the descriptor for a large number of possible structures. An HT code must perform these calculations automatically, transform the structures into standard forms that are the easiest to calculate, and automatically set the necessary k-point grid densities, basis-set energy cutoffs and relaxation cycles with a convergence tolerance of the order of a few meV per atom[44]. It should also respond automatically to calculation failures, due to insufficient hardware resources or runtime errors of the *ab initio* calculation itself. These are among the most difficult challenges in HT database generation that have only recently been overcome (ref. 44 gives details about how this automatic data generation is implemented in the AFLOW HT framework). The initial search is performed on a set of known crystal structures, of all lattice types, spanning the entire composition range of the investigated systems[3,9]. In advanced HT studies this set includes hundreds of structures per system[44]. In subsequent steps, the search is often aided by data-mining and optimization techniques that refine and accelerate the structure screening. They include a variety of different approaches: for

example, cluster expansion[2,45] with exhaustive evaluation or genetic search algorithms on fixed-lattice systems[46–48], and evolutionary algorithms for off-lattice structures in mixtures with fixed stoichiometries[30,43,49–52]. These screening and optimization techniques are continuously being improved and adapted for implementation in HT frameworks[13,53]. The search concludes with the automatic construction of the Gibbs free-energy curve for each system from the minimum-energy structures at various component concentrations[44].

As of this writing, the largest computational alloy database, the Binary Alloy Project hosted in the AFLOWLIB.org consortium repository[24], contains the formation enthalpies for hundreds of thousands of intermetallic structures comprising all the transition metal systems and many other intermetallics[24]. The same framework is also being used to generate similar data for ternary alloys. This information overlaps much of the experimental phase-diagram databases and complements them where the data is partial or missing.

By using HT, massive analyses become possible. Figure 1 illustrates the capacity of HT in dealing with all of the 435 *d*-electron binary intermetallics (elements ordered by Pettifor's scale[54]). The top left triangle in Fig. 1 shows the ordering tendency of the mixtures, defined as the maximum temperature at which the entropic term of an ideal solid solution is equal to the formation energy of the mixture, calculated *ab initio*. This is a measure of the strength of a mixture to oppose disorder. Curtarolo *et al.*[24] demonstrated that HT is capable of reproducing the existence of stable ordered structures, or lack thereof, in 80% of the comparisons: in HT, 65% of the existing binaries are found to be compound forming, whereas in only 58% of the systems compounds have been experimentally reported. The bottom right triangle shows this comparison. In cases of agreement between experiments and calculations on the existence of compounds, HT has been found to reproduce the experimental structure ~96.7% of the time (equation (3) of ref. 9). The database has been used to extensively study several alloy classes, for example, platinum group metals[14,55,56], Mg- and Hf-alloys[57,58] and to pinpoint particular missing features of well-known systems (for example, kinetic acceleration in Fe–V alloys[59]). The availability of low-temperature HT data opens new avenues for high-temperature Monte Carlo simulations, and ultimately for the automatic determination of phase diagrams.

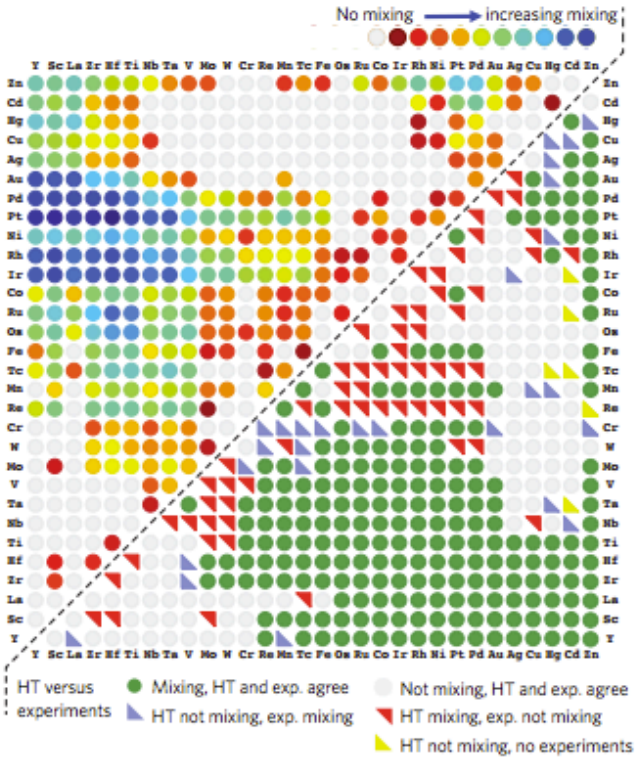

Figure 1 | High-throughput analysis of binary intermetallics[24]. Top left triangle: ordering tendency of the mixtures, as defined in the main text, for elements ordered by Pettifor's chemical scale[54]. Grey circles indicate no ordering, whereas darker blue circles indicate increasing capability to form stable compounds. Bottom right triangle: comparison of HT versus experimental results[24]. Green and grey circles denote agreement between calculation and experimental data on the existence (green) or absence (grey) of compounds. Purple (red) triangles indicate disagreement of HT predictions of compound absence (existence) versus experimental existence (absence). Yellow triangles indicate that data is unavailable for comparison.

Solar materials. Photovoltaic (PV) cells are specialized semiconductor diodes that convert light into direct electric current. Typically, they consist of a transparent conductive oxide layer (such as In-doped SnO_2), an anti-reflection coating (such as Si_3N_4), a p-type doped crystalline Si semiconductor layer with some n-type dopants diffused at the top, and a metallic electrode layer (typically Al), each layer stacked on top of each other. Although such monocrystalline silicon solar cells can convert a useful amount of terrestrial solar energy into electrical energy[60], their construction requires energy-intensive manufacturing at high temperatures (400–1,400 °C) as well as numerous lithographic processes to employ light-trapping techniques. Hence, these conventional cells are too expensive to replace non-renewable energy sources. Indeed, since the introduction of Si solar cells in the 1950s, the search for alternative light-absorbing materials has been an active area of research[61].

Despite the importance of PV materials, the commonly used materials, such as Si, GaAs and $CuInSe_2$, have been discovered accidentally, and been incrementally improved over the years. A systematic analysis of the over 150,000 entries of the inorganic crystal structure database (ICSD) could provide, given the appropriate search parameters, previously undiscovered materials with the appropriate characteristics: semiconductors with strong optical absorption coefficients, a bandgap of ~1.3 eV (the Shockley–Queisser criterion), low cost and, ideally, compatibility with existing technologies. Only recently, the *in silico* screening of PV materials has been attempted with HT techniques: the real difficulty is the characterization of the proper descriptor for the identification of the candidate materials. Yu and Zunger[62] introduced the concept of 'spectroscopic limited maximum efficiency (SLME)', a descriptor combining bandgap, shape of absorption spectra and material-dependent non-radiative recombination losses (all intrinsic materials properties) that were used to tackle the ICSD database[63] (Fig. 2). With their procedure, they were able to identify a set of high SLME materials, including the best already known thin-film solar absorbers, such as $CuInSe_2$, $CuGaSe_2$ and $CuInS_2$, and others that have been found experimentally to be feasible solar absorbers but are much less studied. Interestingly, they were able to identify high SLME materials away from a 1:1:2 stoichiometry (for example, Cu_7TlS_4, Cu_3TlS_2 and Cu_3TlSe_2). Although Tl-containing materials might be unfavourable in practice because of the high toxicity of Tl in the +1 oxidation state, these results suggest that replacing Tl with non-toxic elements could be a viable route to novel high SLME materials.

Water photosplitting. Castelli *et al.*[64] screened a large class of oxide and oxynitride materials (5,400 semiconducting compounds in the cubic perovskite structure covering 52 metals) for optimal solar-light capture in photoelectrochemical cells. Their HT approach, based on the screening of candidate materials by looking at criteria for stability and for the size and position of the bandgap,

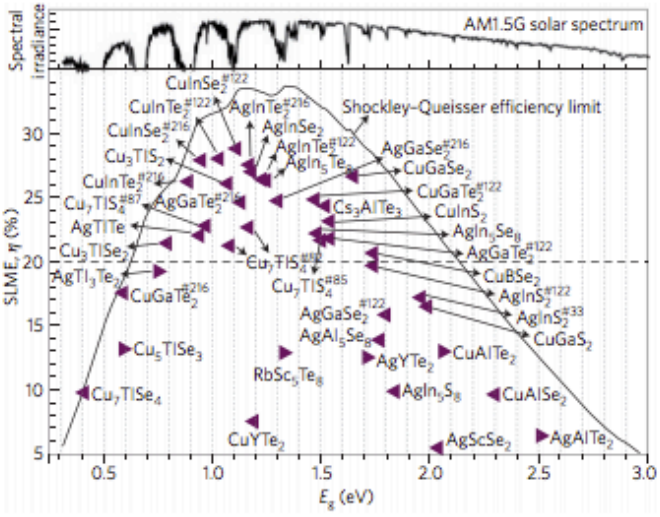

Figure 2 | High-throughput screening of light-absorbing materials for photovoltaic applications. The spectroscopic limited maximum efficiency (SLME) descriptor is plotted versus the minimum gap (E_g) for generalized I–III–VI chalcopyrite materials identified as having a high SLME. In this group, Yu and Zunger found almost all of the currently used PV absorbers as well as those found to be promising in future experiments (with SLME > 20%). Figure adapted with permission from ref. 62, © 2012 APS.

identified ten oxides and five oxynitrides that are well known in the water-splitting community, and predicted nine new combinations for further experimental investigation. In this study, the descriptor was a combination of materials properties (the gene), which include (i) an appropriate bandgap; (ii) well-positioned band edges relative to the water redox levels; (iii) high mobilities, allowing electrons and holes to reach the surface and reduce or oxidize the targets before recombining; and (iv) chemical and structural stability under irradiation. Two of these descriptors (i and iv) are displayed together in Fig. 3, highlighting the process of materials gene construction.

Carbon capture and gas storage. Lin *et al.*[65] recently proposed another application of HT to energy and environment: the screening of materials for large-scale carbon dioxide capture and sequestration (CCS) in power plants, where carbon dioxide should be captured at its source for subsequent storage in non-atmospheric reservoirs. Capture materials and processes that reduce the parasitic energy imposed by CCS are of extreme interest not only for their industrial impact but also for long-term management of climate change. A complex process such as CCS requires the introduction of composite descriptors that must go beyond single material properties. In their study, Lin *et al.* introduced a complex metric, the 'parasitic energy', to identify the optimal process conditions for each material. This descriptor is predicted by the minimization of the electric load imposed on a power plant by temperature–pressure swing capture processes. Previous investigations[66–68] were limited to only a handful of materials and/or single property descriptors. Lin *et al.* were able to screen hundreds of thousands of zeolite and zeolitic imidazolate framework structures, and identify many different structures that have the potential to reduce the parasitic energy of CCS by 30 to 40% compared with near-term technologies.

Another interesting example is the work by Wilmer *et al.*[69] on the screening of metal–organic frameworks (MOFs) for natural gas storage. In this study, the methane-storage capacity of 137,953 hypothetical MOFs was calculated, and over 300 systems were identified to have a potentially better capacity than any known material. One of such predictions, a methyl-functionalized MOF, was also experimentally verified[69]. In another study, Alapati *et al.*[70] scanned over 100 dehydrogenation reactions on experimentally known compounds, and uncovered several new reactions for potential hydrogen-storage materials. Experimental confirmation of the predictions for the $LiNH_2$:MgH_2 system[71] yielded an 8 wt% capacity and an enthalpy of reaction within about 10% of the predicted value. This particular system had been experimentally studied previously for a different ratio of constituents. The new promising composition was suggested by the large-scale computational scan.

Nuclear detection and scintillators. The design of new scintillator materials for γ-ray nuclear detection has recently attracted much interest due to the potential industrial and security applications. Large-scale experimental studies are already feasible[72] and accentuate the necessity of guidance by HT computations.

Computational predictions were first proposed by Ortiz *et al.*[73], using a highly accurate version of the full potential linear muffin-tin orbital method. They parameterized ~22,000 compounds from the ICSD database[63], and implemented a data-mining reduction, based on electronic structure considerations, to extract 136 potential novel scintillators[73,74]. Using HT methods, Setyawan *et al.*[75,76] later proposed a solution to the 'non-proportionality' puzzle, a long-standing problem in the field of scintillator materials. In ref. 75, they introduced a 'non-proportionality descriptor', the mismatch between effective masses of the carriers, m_e (m_h), near the bottom (top) of the conduction (valence) bands,

$$\hat{\chi}_{np} = \max\left(\frac{m_h}{m_e}, \frac{m_e}{m_h}\right)$$

which correlates with the non-proportionality response for oxides and semiconductors (non-proportionality defined as $NP = Y_{10keV}/Y_{662keV}$ where Y_i denotes the photon light yield resulting from γ radiation with energy E in keV). The explanation of the correlation $\hat{\chi}_{np} \leftrightarrow NP$ led to new scientific understanding of the role of the electron–hole mismatch, free or coupled as an exciton, in the spatial distribution around — or in the migration to — a recombination centre (usually a defect or a dopant). The correlation was experimentally verified with the application of hydrostatic pressure, which modifies the effective masses through the induced strain in the crystalline cell, and increases or decreases the non-proportionality depending on the electro–elastic response of the material. In ref. 76, the descriptor was further optimized by integrating it with theoretical light yield and photo-attenuation length, so that an extensive dataset of compounds could be analysed.

Topological insulators. Quantities directly accessible from the band structure (energy gaps, effective masses and so on) may be used as descriptors or combined in composite descriptors, such as those recently developed to search for novel topological insulator[77] materials[16]. In a first study, Lin *et al.*[78] searched for ternary thermoelectric Heusler compounds and demonstrated that although most of the well-known ones, such as TiNiSn and LuNiBi, are topologically trivial, the distorted *Ln*PtSb-type compounds (such as *Ln*PtBi or *Ln*PdBi, $Ln = f^n$ lanthanides), belonging to the half-Heusler subclass, are topologically non-trivial and could provide a platform for the realization of multifunctional topological devices.

In a more recent study, Yang *et al.*[16] searched the quantum materials repository AFLOWLIB.org and automatically discovered 28 topological insulators (some of them already known) in five different symmetry families by defining a variational descriptor (Table 1) that represents the topological robustness or feasibility of the candidate system. The newly discovered materials included peculiar ternary halides, $Cs(Sn,Pb,Ge)(Cl,Br,I)_3$, which could hardly have been anticipated without HT means. The robustness descriptor combines information on the energy versus strain variations of band structures with and without spin–orbit coupling. This is an example where the definition of the descriptor combines advanced electronic-structure

data with geometrical modifications to obtain the essential gene that is associated with the required property (Table 1).

Piezoelectrics. High-throughput techniques have also been applied to the search of materials with large piezoelectric coefficients, a class of systems with many technological applications. In particular, the discovery of the anomalously large piezoelectric effect in lead zirconate titanate suggests that other materials of the same family could display similar properties. The large piezoelectric response observed in the solid solution between lead titanate and lead zirconate relies on the formation of a morphotropic phase boundary (MPB) between tetragonal and rhombohedral distortions of the perovskite structure. Calculations of the relative stability of those phases is crucial for predicting the formation of the MPB and the potential of materials to exhibit large electromechanical couplings. In a systematic search across all the possible 3,969 ABO_3 compositions, Armiento *et al.*[79] developed simple descriptors (and other criteria) to screen viable phase-diagram end-points (Table 1). These descriptors and criteria were used to extract a set of 49 compositions that can be seen as the fundamental building blocks of isovalent alloys for compounds forming MPBs, suitable for high piezoelectric performance. Among those, they identified three primary composition groups, $(Sn,Pb)(Zr,Hf,Ti)O_3$, $(Ba,Sr,Ca)(Zr,Hf,Ti)O_3$ and $(Li,Na,K,Rb,Cs)(Ta,Nb)O_3$, which coincide with three known materials with MPBs.

More recently, Roy *et al.*[80] screened the ICSD for half-Heusler semiconductor compounds, looking for previously unrecognized piezoelectric systems. In their search, they scanned 987 candidate combinations for insulating character and structural, dielectric and piezoelectric properties, and identified a number of promising systems.

Thermoelectric materials. The efficiency of direct thermoelectric energy conversion devices is directly related to the dimensionless thermoelectric figures of merit (ZT) of their p- and n-type components. ZT is defined as $T\sigma S^2/\kappa$, where σ is the electric conductivity, S is the Seebeck coefficient, κ is the zero current thermal conductivity and T is the temperature[81,82]. The ZT depends on temperature and on doping. Thus, for each operation temperature there is a different set of compounds that display the best ZT, once their doping level has been optimized. In this context ZT would be an ideal candidate for an HT descriptor.

An early example of HT was provided by Madsen[83], who reported an automated search for new thermoelectric materials among 570 Sb-containing compounds in the ICSD database[63]. The study suggested the Zintl compound LiZnSb as a potentially interesting n-type thermoelectric. Subsequent experiments[84] found good agreement between the *ab initio* calculated transport properties for the p-type material and the measured ones, and concluded that p-type LiZnSb is not a good thermoelectric. However, it was not possible to assess the n-type material experimentally. Using a comparable approach, Yang *et al.*[85] screened the thermoelectric properties of 36 half-Heusler compounds.

In addition to HT calculations of bulk crystalline compounds, an immense unexplored territory lies in the field of nanostructured materials. A potential area of development for thermoelectric systems is the case of nano-granular materials. In a recent study, Wang *et al.*[15] compared the thermoelectric power factors (σS^2) of more than 3,000 compounds in the ICSD database. Figure 4 indicates the compounds that are ranked as worthy of further study. Some potentially good solutions captured by HT would be missed by traditional approaches that typically explore compounds similar to known good ones. In addition to predicting new materials, the HT analysis also pinpoints interesting phenomena. Using principal component analysis on the obtained data, the authors established correlations between the power factor and other structural or electronic properties of the compounds. In particular, it was found that, at a given grain size, higher power factors are more likely to occur in sintered compounds with large charge-carrier effective masses and bandgaps, and with a large number of inequivalent atoms in the unit cell. The first two rules were already known from previous models[86], but the HT analysis was able to obtain this trend through an automated procedure. The third finding, relating power factor and unit-cell size, is new for electronic properties. A similar correlation has been known for thermal conductivity, stating that larger unit cells are associated with lower thermal conductivities. Together, these rules for electronic and thermal transport indicate that higher ZTs are to be expected in materials with complex unit cells.

Materials for catalysis. The search for new solid catalysts is probably the most complex problem tackled *ab initio*[87,88]. The complexity stems from the numerous microscopic variables affecting catalysis, such as reactants' energies and transition states on the surface through all the steps of the catalytic process. Although there are several examples of catalytic processes with a complete *ab initio* characterization[89–94], an HT screening of a large number of processes in this manner is at present far beyond current capabilities. A significant advance has been made in medium-throughput computational screening for catalysts with increased activity and improved selectivity[12,87,95–99], representing the foundation on which a future fully HT approach could be developed.

As in the previous examples of HT screening for new materials, the search for improved catalysts relies on simple descriptors.

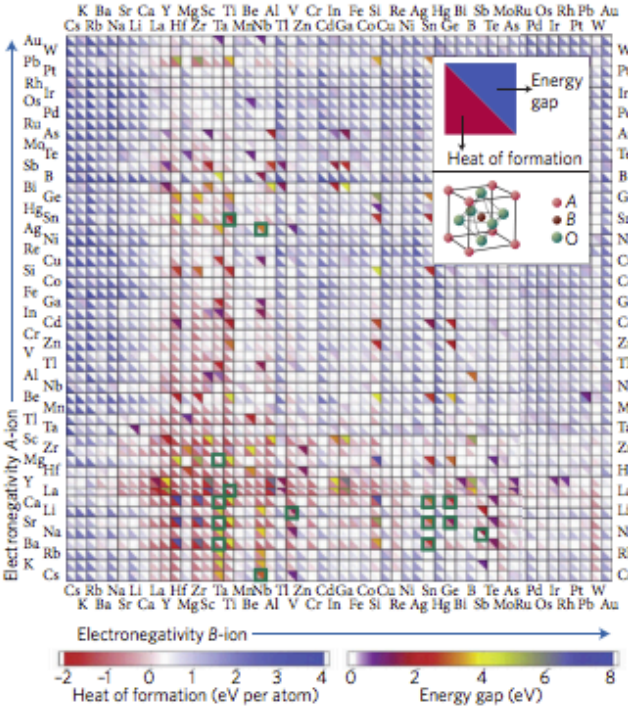

Figure 3 | Example of an efficient HT screening of oxide and oxynitride materials for new photoelectrochemical cells with improved light absorption[64]. Shown are two of the descriptor parameters and their correlation, the calculated heat of formation per atom and the bandgaps, ordered with respect to electronegativity for perovskite binary metal oxynitrides. Each square represents an oxide: in each square the lower-left triangle indicates the formation energy, with red for stability and blue for instability, whereas the upper-right triangle indicates the bandgap, with red for an advantageous bandgap in the range 1.5–3.0 eV. Green squares indicate potential candidate materials. Figure adapted with permission from ref. 64, © 2012 RSC.

However, because the catalysis problem is very complex, these descriptors are reaction specific and their identification requires preliminary analysis of the reaction steps. This is in contrast to the previous examples where a general descriptor of the sought-for functionality is heuristically deduced from the underlying physics and directly employed to screen for candidate materials.

Computational design studies of catalysts carried out in recent years reflect this complexity and include a few principal steps. First, the elementary steps of the reaction, including all the intermediate chemical species, are identified. The energetics of each step, including dissociation and adsorption energies and reaction barriers are evaluated for a number of transition metal surfaces by density functional theory (DFT) calculations, and the active sites, usually a kink or step on the surface at which the activation energies are lowest, are also determined for the specific reaction. Next, correlations are identified between the adsorption energies of the different adsorbates and intermediate species (scaling relations[100]), and between these adsorption energies and their associated transition-state activation energies (Brønsted–Evans–Polanyi relations[87,88,101]). The determination of these relations is of central importance for computational catalyst design, as it allows one to model the catalytic reaction in terms of the minimal number of independent chemical parameters, the descriptors of the process. These descriptors are usually the adsorption energies of the main components of the reactants, which are much easier to evaluate by DFT calculations than activation energies. They can therefore be implemented as practical descriptors of catalytic processes involving dissociation of simple molecules on a metal surface[102]. It is important to note that each active surface site defines a different set of these relations and therefore it is crucial to identify those that lead to the highest reaction rate and the best catalyst.

Once the appropriate descriptor or descriptors have been identified, they are related to measured or calculated total reaction rates for the various metals[87,96], leading to volcano-shaped plots where the optimal catalyst can be identified as the one lying closest to the top of the volcano. These volcano plots emerge due to a fundamental concept in catalysis, the Sabatier principle, stating that for an efficient catalysis the interaction between the surface and reactant should not be too strong nor too weak[103]. When interactions are too weak, reactants will not bind to the catalyst and reaction will not take place. In contrast, if the binding is too strong the catalyst will get blocked, poisoned by reactants or intermediates, or the products will fail to dissociate. Using the volcano plot, the optimal catalyst may be found from a much larger set of candidates than the transition metals used to construct it, because for these candidates the evaluation requires the calculation of only a small number of descriptors. Additional considerations, such as the estimated structural stability and selectivity at the appropriate chemical environment, and cost of the candidates, can be included in the search[87,88,96].

Several descriptor-based searches for new heterogeneous catalysts have been reported in recent years, each screening a few dozen metal surfaces and surface alloys for various chemical reactions[12,87,96,97,99,104–108]. Figure 5 shows the results of one of these studies (Greeley *et al.*[96]): the volcano plot (Fig. 5a) and the optimum active surface alloys for a hydrogen evolution reaction (Fig. 5b). Furthermore, among the predictions made by Studt *et al.*[107], NiZn has been experimentally verified to have better selectivity than the traditional Ag–Pt catalyst at a fraction of the cost, because it contains no precious metals.

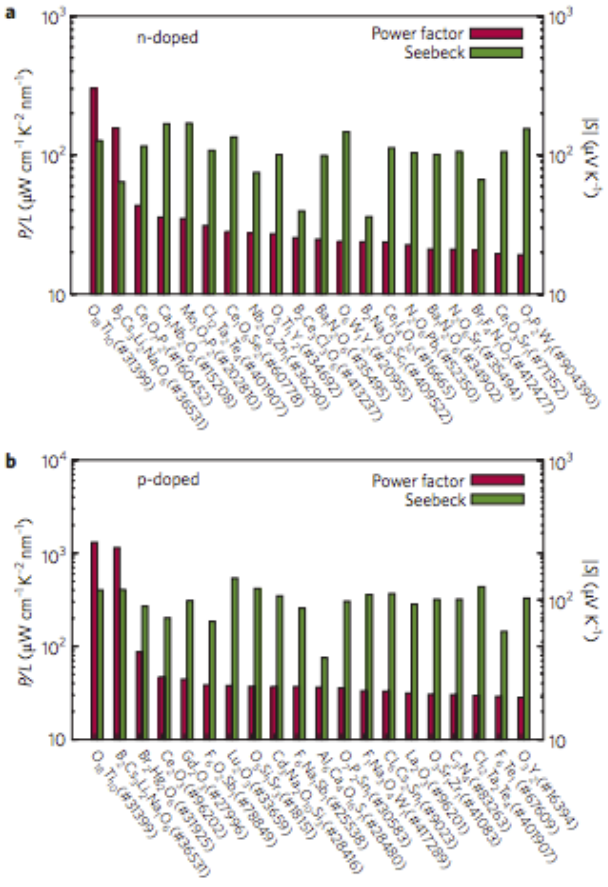

Figure 4 | Example of an HT search for nanosintered thermoelectric materials. a,b, In the limit of very small grains, the direction averaged power factor $\langle P \rangle$ is roughly proportional to the grain size, L. Using the theoretically calculated $\langle P \rangle / L$ at the optimized Fermi level position for n- and p- conduction, 3,000 compounds were analysed. The 20 best compounds in the two lists are shown in **a** (n-doped) and **b** (p-doped), along with the calculated Seebeck coefficient at optimal doping. Figure adapted with permission from ref. 15, © 2011 APS.

Battery materials for energy storage. Efficient energy storage, with high capacity and long cycling lifetime, is one of the main issues in the development of sustainable clean-energy technologies. Lithium-ion batteries are the state of the art in this field. They operate by the transfer of lithium ions from a high chemical potential anode to a low chemical potential cathode through an ion-conducting electron-insulating electrolyte while the electrons flow through an external circuit. Recharging is performed by applying an external potential, forcing the Li ions to migrate back from the cathode to the anode. Almost all anodes currently in use are based on graphite, due to its Li storage capacity, and cycling and safety characteristics[109]. The electrolytes are solutions of Li salts in organic solvents[110]. Several recent attempts have been made to enhance the anode capacity without degrading cycling and safety, using $MoSi_2$-based compounds[111,112]. However, the main route taken by current research to improve Li-ion batteries is to seek new cathode materials with superior properties[27,113]. Finding new battery components experimentally might be impractical due to the very large chemical space and the difficulty of the experiments. High-throughput computational research can make a difference, but as pointed out by Ceder *et al.*[27], a few physical parameters have to be taken into account in addition to structure stability. The search for better materials is a complicated multistep process, with constraints in terms of safety, toxicity, weight, capacity, charge and discharge rates, recyclability and cost[28,114].

The search consists of three steps. First, it starts with a set of candidate chemistries, namely, chemical compositions containing high concentrations of Li ions, a redox active metal, and oxygen or oxide ions (carbonate, borate, phosphate, silicate or arsenate)

that could stabilize a rechargeable Li-containing compound. The chemistry defines the electron activity that can be achieved during the delithiation process and essentially determines the capacity of the battery. Computational screening of the multi-component system is then carried out to identify stable crystallographic structures. This is performed by *ab initio* calculations of the total energies of a large number of structures, chosen with *ad hoc* algorithms[10,115] to parameterize the thermodynamic stability of the systems. This step has been fully implemented in an HT framework (see Fig. 2 of ref. 25).

Second, to achieve good cycling behaviour the structures should facilitate intercalation of Li ions, for example, reversible host–guest chemistry where the ions migrate through layers or tunnels in the host structure with minimal changes in the host itself[113]. The energy density of a battery is determined by the capacity — defined by the chemical process — and the voltage profile, which can be related to the Gibbs free-energy difference between different charged states of the cathode material. Good candidates should exhibit the highest possible voltage, compatible with the limitations of electrolyte stability, usually less than 5 V (ref. 110). Ceder[28] shows an example of the relationship between chemical potential and material resilience to a reducing environment for several classes of compounds (Fig. 6). Considerable advances have been made towards a full HT implementation of this step, but it has not been achieved yet. Standard DFT exchange-correlation functionals — for example, the local density approximation (LDA) and the various forms of generalized gradient approximations (GGA) — may lead to large errors in voltage predictions in redox processes that involve electron transfer between different orbitals, and more advanced and precise hybrid functionals are currently too computationally expensive for HT[116]. The alternative, and less expensive, DFT+U method is usually implemented[117–119]. So far, the most extensive screening of cathode materials through voltage prediction has been carried out by Hautier *et al.*[120] for over 600 phosphate compounds selected out of more than 4,000 structure total-energy calculations. Smaller studies were done on polyanion compounds, with just over 200 different compositions of the sidorenkite structure[121], and on 64 tavorite structured materials[122].

Third, following the screening for voltage and cycling potential, the remaining candidates are evaluated for charge and discharge rates. As illustrated by Ceder *et al.*[27], the rate-limiting processes, which are the ionic and electronic conductivities, are evaluated through *ab initio* calculations of the diffusion barriers for ion migration and of the availability of charge carriers and their mobility in the polaronic intercalation compounds of the electrodes (the most promising new cathode materials are electrical insulators such as borates, phosphates and silicates). This step of evaluation, the most computationally expensive, is carried out at the end of the screening process and it has not yet been implemented in HT. Indeed, Mueller *et al.*[122] calculated lithium-diffusion activation energies for just three structures out of the 64 tavorite compounds proposed, and no lithium-diffusion or polaron-migration calculations were reported in the larger studies[120,121]. A future implementation of an HT framework for this step would require a definition of an efficient descriptor for these transport properties, which is currently non-existent.

Outlook

The computational HT approach, although at a good level of maturity, is still far from being a magic box for materials discovery. Several important properties and classes of materials have not been addressed in this framework, and further algorithm implementations, repositories and data-mining interfaces, are necessary. Here we give important examples of materials and property challenges to be addressed, and an overview of the possible evolution of the framework.

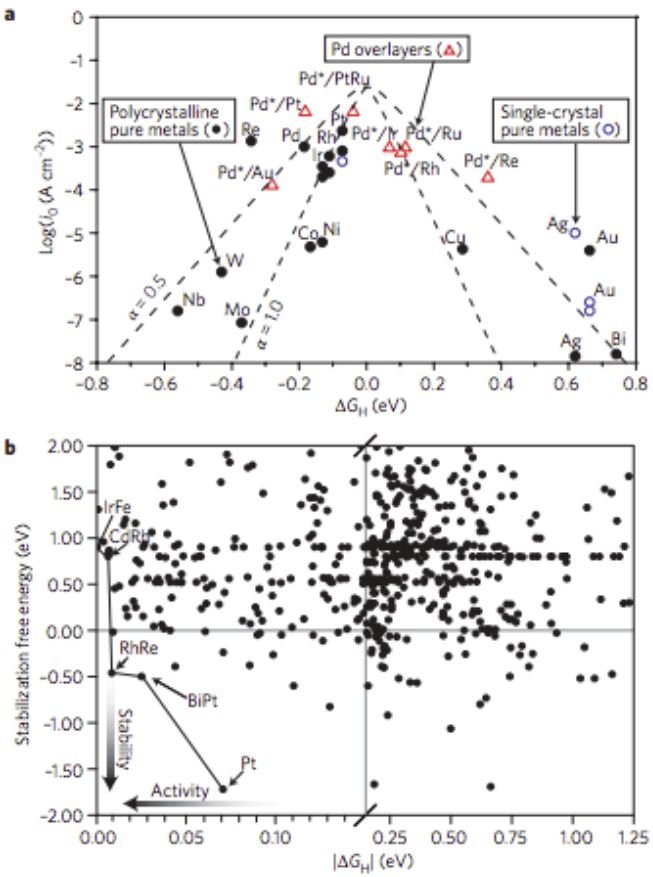

Figure 5 | High-throughput screening of electrocatalytic materials for a hydrogen evolution reaction. a, Catalytic activity (experimentally measured exchange current density) versus free energy of hydrogen adsorption for metals and metal overlayers. The dashed lines denote activity predictions of simple mean-field models, with α as transfer coefficients. The * indicates that the free energies were calculated including corrections due to the coverage for Pd overlayers. **b**, Free energy of formation versus hydrogen evolution reaction activity of surface alloys: each point denotes a different surface alloy. Stable alloys with optimal activities are linked by the solid line in the lower left quadrant. Figure adapted from ref. 96, © 2006 NPG.

Stability at finite temperatures. *Ab initio* calculations of metallic systems often predict ordered compounds that may be difficult to fabricate due to entropy and kinetic effects. The assessment of the thermal stability of structures and the practicality of their realization requires accurate estimation of the various entropic contributions (for instance, configurational, vibrational and magnetic) of the competing phases predicted by HT calculations. For specific sets of structures sharing a common parent lattice, this can be accomplished by high-temperature Monte Carlo simulations. However, a general HT *ab initio* framework for this problem is still lacking.

Thermoelectrics. Design of better thermoelectrics requires a better estimation of the carrier lifetimes. The standard approximation of a constant relaxation time is not satisfactory[83,85]. However, recent work has shown that it is possible to predict carrier lifetimes and electronic mobility with good accuracy, fully from first principles[123,124]. Also, the calculation of the lattice thermal conductivity can now be accurately calculated *ab initio*[125–131]. The definition of a simpler descriptor that can speed up massive HT computation remains a considerable challenge. For bulk systems, a possible approach is to relate the lattice thermal conductivity to the Debye temperature and Grüneisen parameter that can be estimated by

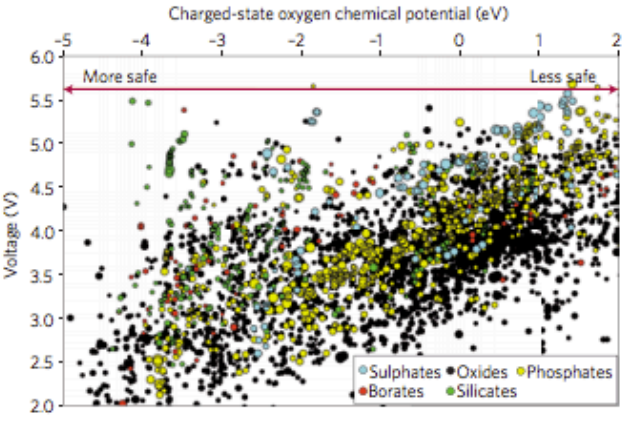

Figure 6 | High-throughput study of safety versus voltage in lithium batteries. Shown are the oxygen chemical potentials at which the charged state decomposes versus their Li intercalation voltage. The more negative the chemical potential, the higher the material resilience to a reducing environment (safety). Figure adapted with permission from ref. 28, © 2010 Cambridge Univ. Press.

the 'GIBBS' isothermal–isobaric approach[132]. For nanostructured materials, an additional challenge is the evaluation of conductivities at interfaces, where the transport problem couples with the difficulty of predicting nanoscale geometries. Here the solution lies in a quantum-transport treatment that has only recently been considered for HT applications[133].

Magnetic materials. Although magnetism can be found in a multitude of materials (several thousands), the choice of magnets available for mainstream applications is much more limited (around two dozen)[134]. There are two main reasons for such limited diversity. First, any standard application, regardless of the particular technology it concerns, needs to operate in the temperature range between −50 °C and +120 °C, which requires the magnet to have a critical ordering temperature, T_C, of at least 550 K. Unfortunately, there are only a few hundred magnets with such a high critical temperature. The second reason is that magnetic materials need to satisfy additional physical constraints for each specific application. Thus, for instance, energy-related technologies (electric turbines, electric motors and so on) require large magnetic energy densities, whereas magnetic sensors often need sensitive magnetic–electric responses.

Only by exploiting HT techniques can one explore the possibility of synthesizing new high-performance magnetic materials, and search into large materials classes that are known to be populated by high-temperature magnets. Particular classes of interest are the intermetallic ternary materials, such as the Heusler compounds[135]. A simple combinatorial calculation gives an upper limit for the number of possible Heusler compounds (including half-Heusler) of about 230,000. Among these, about 1,500 are known and have been synthesized in the past. However, there is still a significant number for which a synthetic strategy has not been designed. Particularly important would be the development of permanent magnets without critical elements (so as to counteract the current 'rare-earth crisis'[136]), or of magnets whose properties are specifically targeted to electronics applications, such as magnetic random access memories. Finally, HT technologies can be used to design entire magnetic heterostructures, and the case of tunnel-magnetoresistance devices seem particularly attractive[137].

Heusler alloys. This Review includes a few references to research on Heusler alloys. In addition to the reviewed topics (thermoelectricity[85], topological insulators[78,138], piezoelectricity[80]), such systems have drawn general attention in the computational materials community[135]: chemical stability was investigated by the Zunger group[139], and bandgap and lattice constants were computed, for optoelectronic applications, by Gruhn[140]. A repository of Heusler alloys' calculations would thus be useful for scientific and industrial applications. This is an undergoing task of the AFLOWLIB.org consortium[24]. By combining the wealth of binary intermetallics and the parameterization of all the possible Heusler alloy combinations (full-[$AlCu_2Mn$], half-[AgAsMg] and anti-[$CuLi_2Sn$]), one can rapidly determine the thermodynamic stability and the appropriate electronic-structure features.

Alloy theory at the nanoscale. The extension of the HT framework to predict alloy stability at the nanoscale has great technological implications, especially for catalysis. Many phenomena are chemically dependent on the stable surface of the catalyst[96,98], and the proper parameterization of the surface stability (energy) and surface tension (stress)[141] will greatly help the development of new catalysts (for example, the size-dependent phase transitions in Fe–C mixtures have been shown to be responsible for deactivation in nanotube growth[142]).

Catalysis. Considerable improvements are necessary to advance to a more comprehensive HT framework for computational catalyst design. One direction is related to the current limitations of DFT calculations in treating non-metallic surfaces, electronic bandgaps and excited electronic states, and the chemistry of atoms and molecules on such surfaces in various environments[143]. Improvements in this field would be needed to extend the current framework from transition metals to other useful materials families, such as oxides, sulphides, nitrides and zeolites. Furthermore, for the development of catalysts with long lifetimes, the thermodynamics of the catalyst–product mixture must be elucidated, especially at the nanoscale where the quest for more active surface dictates size reduction and, as a consequence, possible size-induced thermokinetic deactivation[142,144].

On the HT conceptual level, developing a systematic methodology to determine appropriate descriptors for an as extensive as possible variety of catalytic properties would be of crucial importance. It is not yet apparent how this challenge could be met, but it is clear that it is the key for the implementation of truly HT computational catalyst discovery and design, concomitant with the necessary experimental validation[22].

Battery technologies. Additional advancements would be needed to expand energy-storage materials research beyond the current scope of Li-ion batteries[27,28]. For example, a variety of conversion electrodes, where transition metal binary phases react with lithium, potentially possess much higher energy density than intercalation devices. The wealth of such compounds with different degrees of covalence and transition metal oxidation states could enable the tuning of operation voltages, for positive or negative electrodes, and the possibility of selecting low-cost and environmentally friendly materials[145]. Likewise, batteries employing a higher valence cation such as magnesium or aluminum, which could have considerably increased capacity with reduced weight and volume[109,146], have not yet been considered in computational studies.

Algorithms and repositories. To be effective, the wealth of calculations produced by HT needs to be open-domain, shared in online repositories and equipped with effective search capabilities. Examples are the AFLOWLIB.org consortium[24], the Materials Project[25], the Computational Materials Repository[147], The Electronic Structure Project[74] and the Carnegie Mellon's Alloy Database[148]. For efficiency, the repositories should be integrated so that they can share information through standardized calculation and

communication protocols. This will bring the HT field beyond the monolithic, undistributed approach of each single research group, and allow a better use of the results generated by the entire community. Furthermore, another computational frontier will be the implementation of 'materials daemons', *ad hoc* artificial-intelligence codes implementing the descriptors and autonomously crawling across the various linked databases, scanning and directing calculations until optimum materials are found.

Summary

In this Review, we described the milestones that have been reached with HT computational materials research. We proposed a comprehensive framework, in which they can be conceptually classified, to address the increasing demands of modern technology. For the many other achievements winking at the horizon, crucial components still need to be put in place: efficient HT codes, open and distributed networks of repositories, fast and effective descriptors, and strategies to transfer knowledge to practical implementations. This is an adventurous journey where players will race on a highway without speed limits.

Received 17 August 2012; accepted 9 January 2013; published 20 February 2013

References

1. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge Univ. Press, 2004).
2. De Fontaine, D. in *Solid State Physics* (eds Ehrenreich, H. & Turnbull, D) Vol. 47, 33–176 (Wiley, 1994).
3. Curtarolo, S., Morgan, D., Persson, K., Rodgers, J. & Ceder, G. Predicting crystal structures with data mining of quantum calculations. *Phys. Rev. Lett.* **91,** 135503 (2003).
4. Bryan, G. S. *Edison, the Man and His Work* (Knopf, 1930).
5. Ciamician, G. The photochemistry of the future. *Science* **36,** 385–394 (1912).
6. Ceder, G. *et al.* Identification of cathode materials for lithium batteries guided by first-principles calculations. *Nature* **392,** 694–696 (1998).
7. Jóhannesson, G. H. *et al.* Combined electronic structure and evolutionary search approach to materials design. *Phys. Rev. Lett.* **88,** 255506 (2002).
8. Stucke, D. P. & Crespi, V. H. Predictions of new crystalline states for assemblies of nanoparticles: perovskite analogues and 3-D arrays of self-assembled nanowires. *Nano Lett.* **3,** 1183–1186 (2003).
9. Curtarolo, S., Morgan, D. & Ceder, G. Accuracy of *ab initio* methods in predicting the crystal structures of metals: review of 80 binary alloys. *Calphad* **29,** 163–211 (2005).
10. Fischer, C. C., Tibbetts, K. J., Morgan, D. & Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nature Mater.* **5,** 641–646 (2006).
11. Bligaard, T. *et al.* Parteo-optimal alloys. *Appl. Phys. Lett.* **83,** 4527–4529 (2003).
12. Andersson, M. P. *et al.* Toward computational screening in heterogeneous catalysis: Pareto-optimal methanation catalysts, *J. Catal.* **239,** 501–506 (2006).
13. Levy, O., Hart, G. L. W. & Curtarolo, S. Uncovering compounds by synergy of cluster expansion and high-throughput methods. *J. Am. Chem. Soc.* **132,** 4830–4833 (2010).
14. Levy, O., Chepulskii, R. V., Hart, G. L. W. & Curtarolo, S. New face of rhodium alloys: revealing ordered structures from first principles. *J. Am. Chem. Soc.* **132,** 833–837 (2010).
15. Wang, S., Wang, Z., Setyawan, W., Mingo, N. & Curtarolo, S. Assessing the thermoelectric properties of sintered compounds via high-throughput *ab-initio* calculations. *Phys. Rev. X* **1,** 021012 (2011).
16. Yang, K., Setyawan, W., Wang, S., Buongiorno Nardelli, M. & Curtarolo, S. A search model for topological insulators with high-throughput robustness descriptors. *Nature Mater.* **11,** 614–619 (2012).
17. Xiang, X. D. *et al.* A combinatorial approach to materials discovery. *Science* **268,** 1738–1740 (1995).
18. Takeuchi, I. *et al.* Identification of novel compositions of ferromagnetic shape-memory alloys using composition spreads. *Nature Mater.* **2,** 180–184 (2003).
19. Koinuma, H., & Takeuchi, I. Combinatorial solid-state chemistry of inorganic materials. *Nature Mater.* **3,** 429–438 (2004).
20. Maclean, D. *et al.* Glossary of terms used in combinatorial chemistry. *Pure Appl. Chem.* **71,** 2349–2365 (1999).
21. Schüth, F. in *Hochdurchsatz-Untersuchungen*, in *Winnacker-Küchler Chemische Technik* 5th edn (eds Dittmeyer, R., Keim, W., Kreysa, G. & Oberholz, A.) Ch. 5, 549–585 (Wiley, 2004).
22. Maier, W. F., Stöwe, K. & Sieg, S. Combinatorial and high-throughput materials science. *Angew. Chem. Int. Ed.* **46,** 6016–6067 (2007).
23. Setyawan, W. & Curtarolo, S. High-throughput electronic band structure calculations: challenges and tools. *Comp. Mater. Sci.* **49,** 299–312 (2010).
24. Curtarolo, S. *et al.* AFLOWLIB.ORG: A distributed materials properties repository from high-throughput *ab initio* calculations. *Comp. Mater. Sci.* **58,** 227–235 (2012).
25. Jain, A. *et al.* A high-throughput infrastructure for density functional theory calculations. *Comp. Mater. Sci.* **50,** 2295–2310 (2011).
26. Spivack, J. L. *et al.* Combinatorial discovery of metal co-catalysts for the carbonylation of phenol. *Appl. Catal. A* **254,** 5–25 (2003).
27. Ceder, G., Hauthier, G., Jain, A. & Ong, S. P. Recharging lithium battery research with first-principles methods. *Mater. Res. Soc. Bull.* **36,** 185–191 (2011).
28. Ceder, G. Opportunities and challenges for first-principles materials design and applications to Li battery materials. *Mater. Res. Soc. Bull.* **35,** 693–701 (2010).
29. Kolmogorov, A. N. & Curtarolo, S. Prediction of new crystal structure phases in metal borides: a lithium monoboride analog to MgB_2. *Phys. Rev. B* **73,** 180501(R) (2006).
30. Kolmogorov, A. N. *et al.* New superconducting and semiconducting Fe–B compounds predicted with an *ab initio* evolutionary search. *Phys. Rev. Lett.* **105,** 217003 (2010).
31. Hart, G. L. W., Blum, V., Walorski, M. J. & Zunger, A. Evolutionary approach for determining first-principles hamiltonians. *Nature Mater.* **4,** 391–394 (2005).
32. Pettifor, D. G. Structure maps revisited. *J. Phys. Condens. Matter* **15,** V13–V16 (2003).
33. Morgan, D., Rodgers, J. & Ceder, G. Automatic construction, implementation and assessment of Pettifor maps. *J. Phys. Condens. Matter* **15,** 4361–4369 (2003).
34. Levy, O., Hart, G. L. W. & Curtarolo, S. Structure maps for hcp metals from first-principles calculations. *Phys. Rev. B* **81,** 174106 (2010).
35. Bhadeshia, H. K. D. H., Dimitriu, R. C., Forsik, S., Pak, J. H. & Ryu, J. H. Performance of neural networks in materials science. *Mater. Sci. Technol.* **25,** 504–510 (2009).
36. Sumpter, B. G. & Noid, D. W. On the design, analysis, and characterization of materials using computational neural networks. *Annu. Rev. Mater. Sci.* **26,** 223–277 (1996).
37. Serra, J. M., Baumes, L. A., Serna, M. M, P. & Corma, A. Zeolite synthesis modelling with support vector machines: a combinatorial approach. *Comb. Chem. High Throughput Screen.* **10,** 13–24 (2007).
38. Massalski, T. B., Okamoto, H., Subramanian, P. R. & Kacprzak, L. (eds) *Binary Alloy Phase Diagrams* (ASM, 1990).
39. Villars, P. *et al. Crystal Impact, Pauling File. Inorganic Materials Database and Design System, Binaries Edition* (ASM, 2003).
40. Wodniecki, P., Wodniecka, B., Kulińska, A., Uhrmacher, M. & Lieb, K. P. The $TiPd_2$ compound studied by PAC with ^{181}Ta and ^{111}Cd probes. *J. Alloys Compound* **385,** 53–58 (2004).
41. Niu, H. *et al.* Structure, bonding, and possible superhardness of CrB_4. *Phys. Rev. B* **85,** 144116 (2012).
42. Bialon, A. F. *et al.* Possible routes for synthesis of new boron-rich FeB and $Fe_{1-x}Cr_xB_4$ compounds. *Appl. Phys. Lett.* **98,** 081901 (2011).
43. Kolmogorov, A. N., Shah, S., Margine, E. R., Kleppe, A. K. & Jephcoat, A. P. Pressure-driven evolution of the covalent network in CaB_6. *Phys. Rev. Lett.* **109,** 075501 (2012).
44. Curtarolo, S. *et al.* AFLOW: an automatic framework for high-throughput materials discovery. *Comp. Mater. Sci.* **58,** 218–226 (2012).
45. Sanchez, J. M. & de Fontaine, D. Ising model phase-diagram calculations in the fcc lattice with first- and second-neighbor interactions. *Phys. Rev. B* **25,** 1759–1765 (1982).
46. Drautz, R., Díaz-Ortiz, A., Fähnle, M. & Dosch, H. Ordering and magnetism in Fe-Co: dense sequence of ground-state structures. *Phys. Rev. Lett.* **93,** 067202 (2004).
47. D'Avenzac, M. & Zunger, A. Identifying the minimum-energy atomic configuration on a lattice: Lamarckian twist on Darwinian evolution. *Phys. Rev. B* **78,** 064102 (2008).
48. Hart, G. L. W. & Forcade, R. W. Generating derivative structures: Algorithm and applications. *Phys. Rev. B* **77,** 224115 (2008).
49. Abraham, N. L. & Probert, M. I. J. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Phys. Rev. B* **73,** 224104 (2006).
50. Oganov, A. R. & Glass, C. W. Crystal structure prediction using *ab initio* evolutionary techniques: Principles and applications. *J. Chem. Phys.* **124,** 244704 (2006).
51. Yao, Y., Tse, J. S. & Tanaka, K. Metastable high-pressure single-bonded phases of nitrogen predicted via genetic algorithm. *Phys. Rev. B* **77,** 052103 (2008).

52. Lyakhov, A. O. & Oganov, A. R. Evolutionary search for superhard materials: Methodology and applications to forms of carbon and TiO_2. *Phys. Rev. B* **84,** 092103 (2011).
53. Nelson, L. J., Zhou, F., Hart, G. L. W. & Ozoliņš, V. Compressive sensing as a new paradigm in model building. *Phys. Rev. B* **87,** 035125 (2013).
54. Pettifor, D. G. A chemical scale for crystal-structure maps. *Solid State Commun.* **51,** 31–34 (1984).
55. Levy, O., Jahnatek, M., Chepulskii, R. V., Hart, G. L. W. & Curtarolo, S. Ordered structures in rhenium binary alloys from first-principles calculations. *J. Am. Chem. Soc.* **133,** 158–163 (2011).
56. Jahnatek, M. *et al.* Ordered structures and vibrational stabilization in rhutenium alloys from first principles calculations. *Phys. Rev. B* **84,** 214110 (2011).
57. Taylor, R. H., Curtarolo, S. & Hart, G. L. W. Guiding the experimental discovery of magnesium alloys. *Phys. Rev. B* **84,** 084101 (2011).
58. Levy, O., Hart, G. L. W. & Curtarolo, S. Hafnium binary alloys from experiments and first principles. *Acta Mater.* **58,** 2887–2897 (2010).
59. Bloch, J. *et al.* Prediction and hydrogen-acceleration of ordering in iron-vanadium alloys. *Phys. Rev. Lett.* **108,** 215503 (2012).
60. Green, M. A., Emery, K., Bücher, K., King, D. L. & Igari, S. Solar cell efficiency tables (version 9). *Prog. Photovoltaics: Res. Applications* **5,** 51–54 (1997).
61. Wadia, C., Alivisatos, A. P. & Kammen, D. M. Materials availability expands the opportunity for large-scale photovoltaics deployment. *Environ. Sci. Technol.* **43,** 2072–2077 (2009).
62. Yu, L. & Zunger, A. Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. *Phys. Rev. Lett.* **108,** 068701 (2012).
63. Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. The inorganic crystal structure data base. *J. Chem. Inf. Comput. Sci.* **23,** 66–69 (1983).
64. Castelli, I. E. *et al.* Computational screening of perovskite metal oxides for optimal solar light capture. *Energy Environ. Sci.* **5,** 5814–5819 (2012).
65. Lin, L-C. *et al. In silico* screening of carbon-capture materials. *Nature Mater.* **11,** 633–641 (2012).
66. Krishna, R. & van Baten, J. M. *In silico* screening of metal-organic frameworks in separation applications. *Phys. Chem. Chem. Phys.* **13,** 10593–10616 (2011).
67. Krishna, R. & Long, J. R. Screening metal-organic frameworks by analysis of transient breakthrough of gas mixtures in a fixed bed adsorber. *J. Phys. Chem. C* **115,** 12941–12950 (2011).
68. Yazaydin, A. *et al.* Screening of metal-organic frameworks for carbon dioxide capture from flue gas using a combined experimental and modeling approach. *J. Am. Chem. Soc.* **131,** 18198–18199 (2009).
69. Wilmer, C. E. *et al.* Large-scale screening of hypothetical metal–organic frameworks. *Nature Chem.* **4,** 83–89 (2012).
70. Alapati, S. V., Johnson, J. K. & Sholl, D. S. Large-scale screening of metal hydride mixtures for high-capacity hydrogen storage from first-principles calculations. *J. Phys. Chem. C* **112,** 5258–5262 (2008).
71. Lu, J., Fang, Z. Z., Choi, Y. J. & Sohn, H. Y. Potential of binary lithium magnesium nitride for hydrogen storage applications. *J. Phys. Chem. C* **111,** 12129–12134 (2007).
72. Derenzo, S. *et al.* New scintillators discovered by high-throughput screening. *Nuclear Inst. Methods Phys. Res. A* **652,** 247–250 (2011).
73. Ortiz, C., Eriksson, O. & Klintenberg, M. Data mining and accelerated electronic structure theory as a tool in the search for new functional materials. *Comp. Mater. Sci.* **44,** 1042–1049 (2009).
74. Klintenberg, M. *The Electronic Structure Project*; http://gurka.fysik.uu.se/esp-fs/
75. Setyawan, W., Gaume, R. M., Feigelson, R. S. & Curtarolo, S. Comparative study of nonproportionality and electronic band structures features in scintillator materials. *IEEE Trans. Nucl. Sci.* **56,** 2989–2996 (2009).
76. Setyawan, W., Gaume, R. M., Lam, S., Feigelson, R. S. & Curtarolo, S. High-throughput combinatorial database of electronic band structures for inorganic scintillator materials. *ACS Comb. Sci.* **13,** 382–390 (2011).
77. Hasan, M. Z. & Kane, C. L. Colloquium: Topological insulators. *Rev. Mod. Phys.* **82,** 3045–3067 (2010).
78. Lin, H. *et al.* Half-Heusler ternary compounds as new multifunctional experimental platforms for topological quantum phenomena. *Nature Mater.* **9,** 546–549 (2010).
79. Armiento, R., Kozinsky, B., Fornari, M. & Ceder, G. Screening for high-performance piezoelectrics using high-throughput density functional theory. *Phys. Rev. B* **84,** 014103 (2011).
80. Roy, A., Bennett, J. W., Rabe, K. M. & Vanderbilt, D. Half-Heusler semiconductors as piezoelectrics. *Phys. Rev. Lett.* **109,** 037602 (2012).
81. Mahan, G. D. & Sofo, J. O. The best thermoelectric. *Proc. Natl Acad. Sci. USA* **93,** 7436–7439 (1996).
82. Snyder, G. J. & Toberer, E. S. Complex thermoelectric materials. *Nature Mater.* **7,** 105–114 (2008).
83. Madsen, G. K. H. Automated search for new thermoelectric materials: the case of LiZnSb. *J. Am. Chem. Soc.* **128,** 12140–12146 (2006).
84. Toberer, E. S., May, A. F., Scanlon, C. J. & Snyder, G. J. Thermoelectric properties of p-type LiZnSb: Assessment of *ab initio* calculations. *J. Appl. Phys.* **105,** 063701 (2009).
85. Yang, J. *et al.* Evaluation of half-Heusler compounds as thermoelectric materials based on the calculated electrical transport properties. *Adv. Func. Mater.* **18,** 2880–2888 (2008).
86. Mahan, G. D. in *Solid State Physics* Vol. 51 (ed. Ehrenreich, F. S. H.) 81–158 (Academic, 1998).
87. Nørskov, J. K., Bligaard, T., Rossmeisl, J. & Christensen, C. H. Towards the computational design of solid catalysts. *Nature Chem.* **1,** 37–46 (2009).
88. Nørskov, J. K., Abild-Pedersen, F., Studt, F. & Bligaard, T. Density functional theory insurface chemistry and catalysis. *Proc. Natl Acad. Sci. USA* **108,** 937–943 (2011).
89. Hansen, E. W. & Neurock, M. First-principles-based Monte Carlo simulation of ethylene hydrogenation kinetics on Pd. *J. Catal.* **196,** 241–252 (2000).
90. Linic, S. & Barteau, M. A. Construction of a reaction coordinate and a microkinetic model for ethylene epoxidation on silver from DFT calculations and surface science experiments. *J. Catal.* **214,** 200–212 (2003).
91. Reuter, K., Frenkel, D. & Scheffler, M. The steady state of heterogeneous catalysis studied by first-principles statistical mechanics. *Phys. Rev. Lett.* **93,** 116105 (2004).
92. Honkala, K. Ammonia synthesis from first-principles calculations. *Science* **307,** 555–558 (2005).
93. Kandoi, S. Prediction of experimental methanol decomposition rates on platinum from first principles. *Top. Catal.* **37,** 17–28 (2006).
94. Ferguson, G. A. *et al.* Exploring computational design of size-specific subnanometer clusters catalysts. *Top. Catal.* **55,** 353–365 (2012).
95. Nørskov, J. K. & Christensen, C. H. Toward efficient hydrogen production at surfaces. *Science* **312,** 1322–1323 (2006).
96. Greeley, J., Jaramillo, T. F., Bonde, J., Chorkendorff, I. & Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature Mater.* **5,** 909–913 (2006).
97. Sehested, J. *et al.* Discovery of technical methanation catalysts based on computational screening. *Top. Catal.* **45,** 9–13 (2007).
98. Greeley, J. & Nørskov, J. K. Electrochemical dissolution of surface alloys in acids: Thermodynamic trends from first-principles calculations. *Electrochimica Acta* **52,** 5829–5836 (2007).
99. Nørskov, J. K. & Abild-Pedersen, F. Catalysis: Bond control in surface reactions. *Nature* **461,** 1223–1225 (2009).
100. Abild-Pedersen, F. *et al.* Scaling properties of adsorption energies for hydrogen-containing molecules on transition metal surfaces. *Phys. Rev. Lett.* **99,** 016105 (2007).
101. Nørskov, J. K. *et al.* Universality in heterogeneous catalysis. *J. Catal.* **209,** 275–278 (2002).
102. Bligaard, T. *et al.* The Brønsted-Evans-Polanyi relation and the volcano curve in heterogeneous catalysis. *J. Catal.* **224,** 206–217 (2004).
103. Deutschmann, O., Knözinger, H., Kochloefl, K. & Turek, T. *Heterogeneous Catalysis and Solid Catalysts. 1. Fundamentals* Ch. 1, 457–481 (Wiley, 2011).
104. Strasser, P. *et al.* High throughput experimental and theoretical predictive screening of materials — A comparative study of search strategies for new fuel cell anode catalysts. *J. Phys. Chem. B* **107,** 11013–11021 (2003).
105. Linic, S., Jankowiak, J. & Barteau, M. A. Selectivity driven design of bimetallic ethylene epoxidation catalysts from first principles. *J. Catal.* **224,** 489–493 (2004).
106. Greeley, J. & Mavrikakis, M. Alloy catalysts designed from first principles. *Nature Mater.* **3,** 810–815 (2004).
107. Studt, F. *et al.* Identification of non-precious metal alloy catalysts for selective hydrogenation of acetylene. *Science* **320,** 1320–1322 (2008).
108. Greeley, J. *et al.* Alloys of platinum and early transition metals as oxygen reduction electrocatalysts. *Nature Chem.* **1,** 552–556 (2009).
109. Whittingham, M. S. Materials challenges facing electrical energy storage. *Mater. Res. Soc. Bull.* **33,** 411–419 (2008).
110. Xu, K. Nonaqueous liquid electrolytes for lithium-based rechargeable batteries. *Chem. Rev.* **104,** 4303–4417 (2004).
111. Xiao, J. *et al.* Exfoliated MoS_2 nanocomposite as an anode material for lithium ion batteries. *Chem. Mater.* **22,** 4522–4524 (2010).
112. Chen, X., He, J., Srivastava, D. & Li, J. Electrochemical cycling reversibility of $LiMoS_2$ using first-principles calculations. *Appl. Phys. Lett.* **100,** 263901 (2012).
113. Whittingham, M. S. Lithium batteries and cathode materials. *Chem. Rev.* **104,** 4271–4301 (2004).
114. Hautier, G., Jain, A. & Ong, S. P. From the computer to the laboratory: materials discovery and design using first-principles calculations. *J. Mater. Sci.* **47,** 7317–7340 (2012).
115. Hautier, G., Fischer, C. C., Jain, A., Mueller, T. & Ceder, G. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **22,** 3762–3767 (2010).

116. Chevrier, V. L., Ong, S. P., Armiento, R., Chan, M. K. Y. & Ceder, G. Hybrid density functional calculations of redox potentials and formation energies of transition metal compounds. *Phys. Rev. B* **82,** 075122 (2010).
117. Zhou, F., Cococcioni, M., Marianetti, C. A., Morgan, D. & Ceder, G. First-principles prediction of redox potentials in transition-metal compounds with LDA + *U*. *Phys. Rev. B* **70,** 235121 (2004).
118. Cococcioni, M. & de Gironcoli, S. Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. *Phys. Rev. B* **71,** 035105 (2005).
119. Wang, L., Maxisch, T. & Ceder, G. Oxidation energies of transition metal oxides within the GGA+U framework. *Phys. Rev. B* **73,** 195107 (2006).
120. Hautier, G. *et al.* Phosphates as lithium-ion battery cathodes: an evaluation based on high-throughput *ab initio* calculations. *Chem. Mater.* **23,** 3495–3508 (2011).
121. Hautier, G. *et al.* Novel mixed polyanions lithium-ion batery cathode materials predicted by high-throughput *ab initio* computations. *J. Mater. Chem.* **21,** 17147–17153 (2011).
122. Mueller, T., Hautier, G., Jain, A. & Ceder, G. Evaluation of tavorite-structured cathode materials for lithium-ion batteries using high-throughput computing. *Chem. Mater.* **23,** 3854–3862 (2011).
123. Restrepo, O. D., Varga, K. & Pantelides, S. T. First-principles calculations of electron mobilities in silicon: phonon and Coulomb scattering. *Appl. Phys. Lett.* **94,** 212103 (2009).
124. Murphy-Armando, F. & Fahy, S. First-principles calculation of carrier-phonon scattering in n-type $Si_{1-x}Ge_x$ alloys. *Phys. Rev. B* **78,** 035202 (2008).
125. Broido, D. A., Malorny, M., Birner, G., Mingo, N. & Stewart, D. A. Intrinsic lattice thermal conductivity of semiconductors from first principles. *Appl. Phys. Lett.* **91,** 231922 (2007).
126. Ward, A., Broido, D. A., Stewart, D. A. & Deinzer, G. *Ab initio* theory of the lattice thermal conductivity in diamond. *Phys. Rev. B* **80,** 125203 (2009).
127. Ward, A. & Broido, D. A. Intrinsic phonon relaxation times from first-principles studies of the thermal conductivities of Si and Ge. *Phys. Rev. B* **81,** 085205 (2010).
128. Kundu, A., Mingo, N., Broido, D. A. & Stewart, D. A. Role of light and heavy embedded nanoparticles on the thermal conductivity of SiGe alloys. *Phys. Rev. B* **84,** 125426 (2011).
129. Shiomi, J., Esfarjani, K. & Chen, G. Thermal conductivity of half-Heusler compounds from first-principles calculations. *Phys. Rev. B* **84,** 104302 (2011).
130. Garg, J., Bonini, N., Kozinsky, B. & Marzari, N. Role of disorder and anharmonicity in the thermal conductivity of silicon-germanium alloys: A first-principles study. *Phys. Rev. Lett.* **106,** 045901 (2011).
131. Lindsay, L., Broido, D. A. & Reinecke, T. L. Thermal conductivity and large isotope effect in GaN from first principles. *Phys. Rev. Lett.* **109,** 095901 (2012).
132. Blanco, M. A., Francisco, E. & Luaña, V. GIBBS: isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debye model. *Computer Phys. Commun.* **158,** 57–72 (2004).
133. Calzolari, A., Marzari, N., Souza, I. & Buongiorno Nardelli, M. *Ab initio* transport properties of nanostructures from maximally localized Wannier functions. *Phys. Rev. B* **69,** 035108 (2004).
134. Coey, J. M. D. *Magnetism and Magnetic Materials* Ch. 1, 10–23 (Oxford Univ. Press, 2009).
135. Graf, T., Felser, C. & Parkin, S. S. P. Simple rules for the understanding of Heusler compounds. *Prog. Solid State Chem.* **39,** 1–50 (2011).
136. Nakamura, E. & Sato, K. Managing the scarcity of chemical elements. *Nature Mater.* **10,** 158–161 (2011).
137. Yuasa, S. & Djayaprawira, D. D. Giant tunnel magnetoresistance in magnetic tunnel junctions with a crystalline MgO(001) barrier. *J. Phys. D* **40,** R337–R354 (2007).
138. Chadov, S. *et al.* Tunable multifunctional topological insulators in ternary Heusler compounds. *Nature Mater.* **9,** 541–545 (2010).
139. Zhang, X., Yu, L., Zakutayev, A. & Zunger, A. Sorting stable versus unstable hypothetical compounds: The case of multi-functional ABX half-Heusler filled tetrahedral structures. *Adv. Func. Mater.* **22,** 1425–1435 (2012).
140. Gruhn, T. Comparative *ab initio* study of half-Heusler compounds for optoelectronic applications. *Phys. Rev. B* **82,** 125210 (2010).
141. Cahn, J. W. Surface stress and the chemical equilibrium of small crystals—I. The case of the isotropic surface. *Acta Metall.* **28,** 1333–1338 (1980).
142. Harutyunyan, A. R. *et al.* Reduced carbon solubility in Fe nano-clusters and implications for the growth of single-walled carbon nanotubes. *Phys. Rev. Lett.* **100,** 195502 (2008).
143. Wellendorf, J. *et al.* Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Phys. Rev. B* **85,** 235149 (2012).
144. Cervantes-Sodi, F. *et al.* Viscous state effect on the activity of Fe nano-catalysts. *ACS Nano* **4,** 6950–6956 (2010).
145. Cabana, J., Monconduit, L., Larcher, D. & Palacin, M. R. Beyond intercalation-based Li-ion batteries: The state of the art and challenges of electrode materials reacting through conversion reactions. *Adv. Mater.* **22,** E170–E192 (2010).
146. Aurbach, D. *et al.* Prototype systems for rechargeable magnesium batteries. *Nature* **407,** 724–727 (2000).
147. Landis, D. D. *et al.* The computational materials repository. *Computing Sci. Eng.* **14,** 51–57 (2012).
148. Mihalkovič, M. & Widom, M. *Ab initio* calculations of cohesive energies of Fe-based glass-forming alloys. *Phys. Rev. B* **70,** 144107 (2004).

Acknowledgements

We thank Marco Fornari, Greg Rohrer, Shidong Wang, Kesong Yang, Junkai Xue, Richard Taylor, Camilo Calderon, Cheng-Ing Chia, Omar Knio, Ichiro Takeuchi, Mike Mehl, Harold Stokes, Rodney Forcade, Gerbrand Ceder, Alex Zunger, Wahyu Setyawan and Aleksey Kolmogorov for useful comments. This work was supported in part by DOD-ONR (N00014-11-1-0136, N00014-09-1-0921) and by the Duke University—Center for Materials Genomics. S.S. thanks financial support from CRANN.

Competing financial interests

The authors declare no competing financial interests.