REVIEW ARTICLE | INSIGHT
https://doi.org/10.1038/s41563-021-01013-3
nature materials

Electronic-structure methods for materials design

Nicola Marzari[1]✉, Andrea Ferretti[2] and Chris Wolverton[3]

The accuracy and efficiency of electronic-structure methods to understand, predict and design the properties of materials has driven a new paradigm in research. Simulations can greatly accelerate the identification, characterization and optimization of materials, with this acceleration driven by continuous progress in theory, algorithms and hardware, and by adaptation of concepts and tools from computer science. Nevertheless, the capability to identify and characterize materials relies on the predictive accuracy of the underlying physical descriptions, and on the ability to capture the complexity of realistic systems. We provide here an overview of electronic-structure methods, of their application to the prediction of materials properties, and of the different strategies employed towards the broader goals of materials design and discovery.

This Review aims to present the current landscape of electronic-structure methods and of their application to understand, predict and design materials properties. In fact, for the past 20 years, first-principles simulations have outgrown an initial core of developers and practitioners rooted in condensed-matter physics and quantum chemistry to become powerful and widely used tools in many and diverse fields of science and engineering. These approaches have led to remarkable predictions, and their complexity and diversity can benefit from a bird's eye view on capabilities and challenges. Given that simulations are currently run on digital computers, whose throughput capacity[1] in recent decades has doubled every 12–18 months, that machine learning[2] and artificial intelligence[3] are accelerating the capabilities to either search for materials or to predict their properties, and that disrupting paradigms might be leveraged—from neuromorphic[4] to quantum computing[5]—we can safely assume that their footprint (30,000 papers per year already in 2013[6]) and relevance will only keep increasing.

Applications of electronic-structure methods range from nanotechnology to planetary science, from metallurgy to quantum materials, and the current push and excitement to accelerate or complement experiments with simulations makes it even more urgent to highlight not only their capabilities but also their limitations. Simulations do not fail in spectacular ways, but can subtly shift from being invaluable to barely good enough to just useless. The reasons for failure are manifold, from stretching the capabilities of the methods to forsaking the complexity of real materials. But simulations are also irreplaceable: they can assess materials at conditions of pressure and temperature so extreme that no experiment on Earth is able to replicate, they can explore with ever-increasing nimbleness the vast space of materials phases and compositions in the search for that elusive materials breakthrough, and they can directly identify the microscopic causes and origin of a macroscopic property. Last, they share with all branches of computational science a key element of research: they can be made reproducible and open and shareable in ways that no physical infrastructure will ever be.

This Review outlines the framework of density functional theory (DFT) and provides an overview of the many and ever-more-complex approaches that can improve accuracy or extend the scope of simulations (Box 1), with a special focus on Green's function methods and many-body perturbation theory (Box 2). Then, it discusses the capabilities that computational materials science has developed to exploit such a toolbox, and to deliver predictions for the properties of materials under realistic conditions or of ever-increasing complexity. Last, it underscores how physics-driven or data-driven approaches can provide rational, high-throughput or artificial intelligence avenues to materials discovery, and how such efforts are changing the entire research ecosystem.

Electronic-structure methods

Introduction to DFT. DFT[7] is currently the method of choice to compute efficiently and often accurately the ground-state properties of condensed systems: these include the energy of the ground state and its derivatives (for example, forces and stresses, that are used to find equilibrium structures). It is a widely used[8] and mature technology; nevertheless, for materials simulations, codes have only recently started to be verified systematically against each other for numerical accuracy[9–11]. Validation against experiments has a much longer history, and recently, also aided by high-throughput approaches, a number of studies[9,12–14] have appeared that provide comprehensive reference data on the prediction of, for example, elastic and electronic properties. Still, predictive accuracy remains one of the key challenges[15], and progress requires often sound domain-specific knowledge to leverage these calculations.

A key advantage of DFT in its capability to address the complexity of materials is that it shifts the focus from finding the ground-state wavefunction (for N electrons, an exponentially exploding problem of a complex function of $3N$ variables) to finding the ground-state charge density (for any number of electrons, a much more tractable real function of 3 variables). The disadvantage is that the universal functional[7] of DFT is not known, but just proven to exist: *per speculum et in aenigmate* much progress has been made due to a combination of superb physical intuition and some initial serendipity.

The DFT theorems and the Kohn–Sham picture. The first theorem[7] of DFT establishes a one-to-one correspondence between the external potential $v_{ext}(\mathbf{r})$ acting on N interacting electrons and their ground-state charge density $\rho(\mathbf{r})$ (these are scalar fields of the space variable $\mathbf{r}$; typically, and in the Born–Oppenheimer approximation, $v_{ext}(\mathbf{r})$ is the Coulomb attraction from all the atomic nuclei). Owing to this correspondence, it becomes formally equivalent to discuss any physical system either through N and $v_{ext}(\mathbf{r})$, as the Schrödinger equation does, or through its ground-state charge density $\rho(\mathbf{r})$, as DFT does. Thus, $\rho(\mathbf{r})$ is a fundamental quantity, in the sense that it characterizes completely a physical system. The second theorem[7] completes the DFT formulation for the total energy, proving that there exists a functional $F[\rho]$ that provides, for any ground-state $\rho(\mathbf{r})$, the sum of the exact kinetic and electron–electron interaction energies (that is, the expectation values of these operators on the

[1]Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [2]Centro S3, CNR-Istituto Nanoscienze, Modena, Italy. [3]Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA. ✉e-mail: nicola.marzari@epfl.ch

Box 1 | Hierarchies of electronic-structure methods

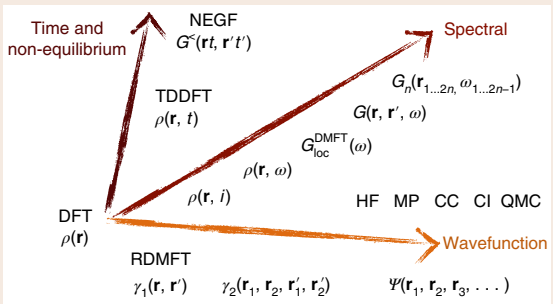
The wavefunction domain. Greatly developed in quantum chemistry, the wavefunction domain starts from Hartree–Fock (HF) and post-Hartree–Fock approaches (a broad class of methods[184] ranging from Møller–Plesset (MP) perturbation theory to coupled cluster (CC) and configuration interaction (CI)) to deliver improved accuracy. Some of these approaches have been extended to treat solid-state systems, also in combination with stochastic sampling[185]—quantum Monte Carlo (QMC) having a long history of delivering accurate results for materials[186,187]. Remarkably, the total energy can also be written as an explicit functional of the second-order reduced-density-matrix γ_2, but the conditions for which γ_2 is the contraction of a proper wavefunction are not known, precluding variational searches. These are known for the first-order γ_1, but the correlation functional of reduced-density-matrix functional theory (RDMFT)[188] is unknown and has to be approximated.

The spectral domain. The spectral domain targets spectral properties with many-body Green's function methods; these introduce diagrammatic approximations for the non-local (that is, function of two space variables $\mathbf{r}$ and $\mathbf{r}'$) and dynamical (that is, function of the frequency ω) self-energy $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$, allowing one to obtain the one-particle Green's function and thus the spectral function and the total energy. Beyond, there is a hierarchy of equations of motion for the n-particle Green's function G_n in terms of the G_{n+1}. Dynamical mean-field theory (DMFT)[57,58] builds a self-energy for a localized manifold (typically for some d or f electrons) through the infinite-dimensional limit of the Anderson impurity model, obtaining an orbital-local or short-ranged $G_{\mathrm{loc}}^{\mathrm{DMFT}}(\omega)$ that can describe strong correlations. Extensions include dynamical cluster approximations or the combination of DMFT and the GW approximation[59].

In spectral formulations, the quasiparticle weights (determined from the derivatives of the self-energy with respect to frequency) do not have to be integers, and complex features such as satellites and side bands can emerge. Functionals dependent on orbital densities $\rho(\mathbf{r}, i)$, where i labels the different orbitals, can give rise to local and orbital-dependent self-energies[189] aimed at reproducing the spectral properties[190] of the interacting system. The resulting approaches are then flexible enough to address both total energies and spectral properties[25,191].

The time-dependent and non-equilibrium domain. This domain is under very active development, also driven by the experimental capabilities to probe electron dynamics at ultrashort timescales. In time-dependent DFT (TDDFT)[64,192], a correspondence is established between the time-dependent charge density and the external time-dependent potential and initial wavefunction; a KS picture then leads to a time-dependent exchange-correlation potential that depends on the entire history of the system. TDDFT can target the real-time evolution of a system of interacting electrons, but can also describe neutral excitations, since in exact TDDFT these are given by the poles of the density-response function[193].

While practical TDDFT is still approximate, it asks the right question with respect to excitations (where are the poles?), and already adiabatic functionals (that are not history dependent) provide very good results in molecules. In solids simple xc approximations do not bind excitons, but progress can be made by directly approximating (in linear response) the f_{xc} kernel[64], or by using potentials from hybrid functionals[49] or DFT+U[194]. These latter lead to time-dependent Hamiltonians with an effective non-local screened exchange, mirroring real-time approaches[66,175] to the solution of the Bethe–Salpeter equation (see Box 2). Last, non-equilibrium Green's functions (NEGF) and the Kadanoff–Baym equations[65] allow for time-dependent generalizations of many-body perturbation theory that can, for example, address transients and time-resolved spectroscopies[63].



Electronic-structure methods. The complex landscape of electronic-structure methods is captured here grouping hierarchies of methods that progressively extend scope and accuracy while increasing cost and complexity.

exact ground-state wavefunction). $F[\rho]$ is universal, since it does not depend on the external potential; adding to it $\int v_{\mathrm{ext}}\rho\mathrm{d}\mathbf{r}$ provides a total-energy functional $E_{v_{\mathrm{ext}}}[\rho]$ that is minimized by the exact ground-state density and gives the exact ground-state energy[7]. The original requirements of having non-degenerate ground states and of v–representability (that is, that the functional be defined only on charge densities that are actual ground states of local external potentials) have then been relaxed by the Levy and Lieb formulations[16].

While $F[\rho]$ is formally well defined, it is not known. To make progress, Kohn and Sham (KS) suggested to decompose it in the sum of three functionals[17] ($T_{\mathrm{s}}[\rho]$, $E_{\mathrm{H}}[\rho]$ and $E_{\mathrm{xc}}[\rho]$) where the first two can be explicitly defined and calculated, and the unknown leftovers are pushed into the last one. To do so, they introduced an auxiliary system of electrons not interacting with themselves (the KS particles) that, when subjected to a local potential $v_{\mathrm{KS}}(\mathbf{r})$, yields the same ground-state charge density of the real system of interacting electrons in the external potential $v_{\mathrm{ext}}(\mathbf{r})$. Such construction provides a definition of what v_{KS} is, and allows one to define T_{s} as the kinetic energy of the non-interacting KS particles, that is, a trivial second derivative of their single-particle orbitals (thus, T_{s} is an implicit functional of the density but an explicit functional of the KS orbitals). $E_{\mathrm{H}}[\rho]$ is then set to be the classical, electrostatic ('Hartree') energy of the charge density ρ. This leaves to the 'exchange correlation' (xc) functional $E_{\mathrm{xc}}[\rho]$ the challenge to recover the exact energy by adding the missing parts of the kinetic energy of the interacting electrons not captured by the KS particles, and of the electron–electron interactions not captured by Hartree electrostatics. Crucially, T_{s} and E_{H} contribute to a large fraction of $F[\rho]$, albeit at the price of reintroducing orbitals to calculate T_{s}; the remaining xc functional E_{xc} has to be approximated and will determine the accuracy of the calculations.

The ground-state energy can then be obtained by direct minimization of the total-energy functional $E_{v_{\mathrm{ext}}}[\rho]$, or equivalently by the Euler–Lagrange equations associated with the variational principle. These are known as Kohn–Sham equations and, assuming discrete occupied states, are:

$$\left[-\frac{1}{2}\nabla^2 + v_{\mathrm{ext}}(\mathbf{r}) + v_{\mathrm{H}}(\mathbf{r}) + v_{\mathrm{xc}}(\mathbf{r})\right]\varphi_n(\mathbf{r}) = \varepsilon_n\varphi_n(\mathbf{r}), \qquad (1)$$

Box 2 | Green's function methods

Green's function (GF) methods[65,66] provide a systematic approach to address electronic excitations (for example, band structures or optical spectra) and, in principle, total energies. Here the fundamental quantity is $G(\mathbf{r}t,\mathbf{r}'t')$, the one-particle GF describing the time propagation of an interacting system when one electron is added at $\mathbf{r}'t'$ and then removed at $\mathbf{r}t$ (vice versa for the hole; either case addresses a charged excitation). The GF obeys the Dyson equation $G=G^0+G^0\Sigma_{xc}G$, where $\Sigma_{xc}[G]$ is the electronic self-energy capturing the many-body complexity, and G^0 is the GF of the Hartree Hamiltonian $-\frac{1}{2}\nabla^2+v_{ext}+v_H$. For discrete states it leads to

$$[-\frac{1}{2}\nabla^2+v_{ext}+v_H]\,f_n(\mathbf{r})+\int d\mathbf{r}'\;\Sigma_{xc}(\mathbf{r},\mathbf{r}',\varepsilon_n)f_n(\mathbf{r}')=\varepsilon_n f_n(\mathbf{r}), \quad (3)$$

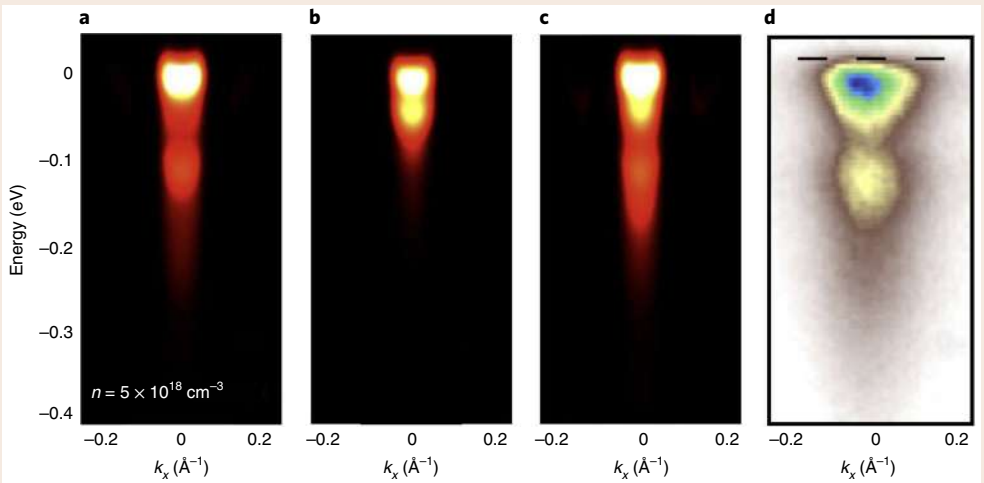
$$G(\mathbf{r},\mathbf{r}',\omega)=\sum_n\frac{f_n(\mathbf{r})f_n^*(\mathbf{r}')}{\omega-\varepsilon_n\pm i0^+}, \quad (4)$$

where equation (3) is an eigenvalue problem for the Dyson orbitals $f_n(\mathbf{r})$ that, through equation (4), provide the Fourier transform of the one-particle GF. It is remarkable that these coupled equations actually mirror the KS equations (1) and (2) for orbitals and density. The key difference lies in having a non-local and dynamical self-energy $\Sigma_{xc}(\mathbf{r},\mathbf{r}',\omega)$ rather than a local exchange-correlation potential.

The eigenvalues ϵ_n—the poles of the GF—are total energy differences for the addition or removal of an electron from the interacting system. The Dyson orbitals $f_n(\mathbf{r})$ are neither orthogonal nor normalized; for the occupied ones (that can be many more than N), the sum of their weights integrates to N. In the thermodynamic limit, the poles of the GF merge into a branch-cut continuum, giving rise to more complex spectral features, such as the broadening of the non-interacting peaks (hence, 'quasiparticles') and the emergence of satellites and side bands (see Fig. 7.2 in ref. [66]). These can be experimentally detected by means of photoemission spectroscopy[61].

Many-body perturbation theory constructs approximations to the self-energy by means of Feynman diagrams[66]; only certain classes are included, giving rise to many and ever-more-complex approaches. The most used one is the GW approximation[114,183,196], $\Sigma_{xc}=iGW$, where W is the screened Coulomb interaction in the random-phase approximation[66]. Most of the time, GW is applied as a one-shot correction (G_0W_0) on top of KS states and energies. Self-consistency can improve the accuracy of excitation energies[197] and, recovering the missing particle conservation[198], also of the total energies. Extensions beyond GW add additional diagrams to Σ_{xc}, from vertex corrections[199] to cumulant expansions[200]. Diagrams describing the interaction with other excitations can also be added, and can be very relevant to predict or interpret experiments, as shown in the figure below.

The G_2 Green's function describes the dynamics of two particles (electrons and holes) added to the system and contains, among other, information about neutral excitations and response functions required to address optical properties and other spectroscopies. Taking the variation of G with respect to an arbitrary external perturbation leads[65,66] to the Bethe–Salpeter equation (BSE), a Dyson-like equation for the two-particle correlation function $L=-G_2+GG$. Besides the bare exchange coming from the Hartree potential, the kernel of the BSE contains $K_{xc}=\frac{\delta\Sigma_{xc}}{\delta G}$, paralleling the exchange-correlation kernel $f_{xc}=\frac{\delta v_{xc}}{\delta\rho}$ of TDDFT[193]. In the GW approximation, and neglecting the dependence of W on G, the BSE kernel becomes $K_{xc}=iW$, accounting for electron–hole interactions. This vertex equation can then be cast in the form of an eigenvalue problem for a two-particle Hamiltonian that can address excitonic effects[66]. In practical calculations[172,173], it requires as a preliminary step the GW orbital energies and the screened Coulomb interaction W.



Interacting excitations. a–d, The momentum-resolved spectral function of anatase[195] calculated including the interactions of electrons with other excitations, that is, phonons (**a**), plasmons (**b**) and both (**c**), and compared with experiments (**d**). The dashed line in **d** is the Fermi energy for the doped sample. Credit: figure adapted with permission from ref. [195], APS

$$\rho(\mathbf{r})=\sum_n^{occ}|\varphi_n(\mathbf{r})|^2, \quad (2)$$

where v_H and v_{xc} are the functional derivatives with respect to $\rho(\mathbf{r})$ of E_H and E_{xc}, φ_n and ε_n are the nth KS eigenstate and eigenvalue, and the sum runs over the occupied states. The solution of the KS equations needs to be self-consistent, since v_H and v_{xc} depend on the charge density $\rho(\mathbf{r})$.

Exact properties. If the exact xc functional were known, ground-state total energies and charge densities would be exact. Quantities directly related to these—first-order and higher-order derivatives of

the energy, or expectation values of local single-particle operators—would also be exact. So, at the Born–Oppenheimer level, structural, elastic, dielectric, vibrational and thermodynamical properties of materials[13,18] could all be described correctly, provided zero-point motion[19] and finite-temperature anharmonicity[20] were accounted for. In contrast, only a single spectral property would be explicitly reproduced by the KS particles: in a molecule (and in a solid, provided the vacuum level could be accessed[21]), exact KS DFT would recover the true highest occupied molecular orbital (HOMO) as its highest occupied KS eigenvalue, since this latter determines the asymptotic decay of the charge density, which needs to be exact. The exact KS HOMO is also the opposite of the exact ionization potential, owing to Janak's theorem and the piecewise linearity of the total energy with respect to the total number of electrons[21]. No other KS level has a direct physical interpretation: even the position of the exact KS lowest unoccupied molecular orbital (LUMO) is incorrect, making the KS gap different from the physical one[22] by the so-called derivative discontinuity (see below). Nevertheless, the optimized-effective-potential approach[23] shows that the KS potential is the variational-best local and static approximation to the electronic self-energy (Box 2); this lends support to using the KS eigenvalues as orbital energies and band structures (since we do not have the exact KS functional, this is, except for the HOMO, a double approximation).

In a finite system, one can take the alternative Δ-SCF approach, where the ionization potential, electron affinity and HOMO–LUMO gap are obtained (exactly in exact DFT) as total-energy differences from separate self-consistent field (SCF) calculations performed with $N-1$, N and $N+1$ electrons, rather than as eigenvalues of a single calculation. For commonly used functionals, Δ-SCF breaks down in the thermodynamic limit of infinite systems and the results asymptotically approach the KS gap[21,24,25]. The exact functional leads instead to a potential that has a discontinuous (uniform in space) jump when an infinitesimal fraction of an electron is added to the system[21]. Intuitively, this derivative discontinuity is present because in exact KS DFT, the incorrect LUMO needs to become the correct HOMO as soon as it is infinitesimally filled, something that is very difficult to capture in approximate functionals and is the missing quantity to be added to the exact but unphysical KS gap to retrieve the correct fundamental one[21,22]. Such requirement makes exact KS DFT highly non-trivial[21] and challenging to approximate or improve. Generalized KS schemes where the KS potential itself becomes non-local—such as meta-generalized-gradient approximations (meta-GGAs)[26], DFT+U[27] or hybrid functionals[28]—mitigate some of these failures[14,24]. Last, we note that, owing to the first Hohenberg–Kohn theorem, there exist in principle functionals of the density for any physical observable of the ground-state wavefunction, and not only those mentioned earlier. Nevertheless, these functionals are in general not known, and an independent variational principle for these might not exist.

Density functional practice. To perform an actual calculation, one needs to address the numerical solution of the KS equations and to approximate the xc functional. The numerical solution[29] involves algorithmic choices (for example, direct minimization or iterative diagonalization); representation of the orbitals, with a wide variety of basis sets possible (atomic or numerical orbitals, wavelets, plane waves, linearized augmented plane waves, real-space grids), either with an all-electron or a pseudopotential treatment of the electron-ion interactions[30]; and numerical methods (for example, Monkhorst–Pack sampling and finite-temperature smearing). This variety translates to the many and widely used electronic-structure codes available nowadays (see, for example, Table 1 of ref. [31]). For materials simulations, plane- and augmented-wave approaches—naturally embodying periodicity—are some of the most commonly found. Localized basis sets become advantageous when dealing with large-scale applications, for which linear-scaling methods have been developed and applied[32] (standard DFT calculations are cubic scaling, due to the orthonormality constraint of the KS orbitals in equation (1)). Linear scaling is also achieved if the non-interacting kinetic energy $T_s[\rho]$ is approximated directly[33]. Finally, stochastic approaches[34], even with sublinear scaling, have been proposed and extended to excited-state properties.

The quest for approximations to the xc functional is a more fundamental one. The local-density approximation (LDA) was already made in the original KS formulation[17], applying Fermi's suggestion to the xc functional rather than to the kinetic-energy functional. Albeit unbeknown at the time, LDA satisfies an exact sum rule for the xc hole[35]; it is this exact constraint, and the intuition of building the non-interacting KS system to approximate the kinetic energy, that has allowed DFT to describe real materials with predictive accuracy. The treatment of spin polarization and magnetism is very challenging, and extensions that introduce spin-resolved densities, functionals and spinors for non-collinear magnetism are then used[16]. Improving on LDA and its spin-resolved extension, which get so many things correct in the first instance, is a challenge. One approach is to introduce the gradients of the charge density, and then the kinetic-energy density, while satisfying as many exact and asymptotic constraints as possible. These GGAs[36] and meta-GGAs[26] represent currently—especially with PBEsol and SCAN—some of the most used all-purpose functionals for materials simulations[13]; still, challenges remain[37]. Alternatively, one can train on higher-accuracy calculations, as is done for B3LYP[38], the Minnesota functionals[39] or those based on machine learning (for example, Bayesian error estimation or neural networks). Direct approximations of the KS potential, rather than the functional, have also been attempted: these include mBJ[40] (a meta-GGA) and GLLB-SC[41] (estimating the derivative discontinuity); both are used to improve bandgap predictions. Also, the quality of xc functionals can be improved, with substantial computational costs, through the adiabatic-connection fluctuation-dissipation theorem[23,42].

Remarkably, most functionals are incorrect in one of the simplest cases possible—one-electron systems—where the Hartree and xc terms should cancel exactly, as it happens in Hartree–Fock, to deliver the same solution as the Schrödinger equation[43]. The orbital-density-dependent Perdew–Zunger correction[44] addresses this self-interaction error (SIE) in many-electrons systems, but may nevertheless require downscaling[45]. A number of authors have connected SIEs to the lack of piecewise linearity of the approximate total energy as a function of the number of electrons[21], often heuristically extended to fractional occupations of deeper orbitals or manifolds[43,46,47]. Broadly speaking, short-ranged hybrids or Hubbard functionals[27,28,46] improve total energies, while other range-separated or Koopmans functionals[25,48,49] improve spectral properties. In fact, the success of hybrid functionals[28,50,51], obtained by adding an appropriate fraction of full or range-separated Fock exchange, can also be rationalized as correcting SIEs and imposing some piecewise linearity.

The local nature of many approximations (the exact xc potential is local, but depends non-locally on the density everywhere) affects other key properties, from image potentials at metal surfaces to the $1/r$ decay of $v_{KS}(\mathbf{r})$ and the Rydberg series in atoms and molecules to long-range van der Waals interactions. These latter interactions are particularly important for materials—from two-dimensional heterostructures to molecular and organic solids—but notable progress has been made, and empirical or first-principles formulations have emerged[52] that have greatly improved predictive accuracy.

The challenges. Some of the most important materials for scientific advance and technological innovation remain challenging. Examples include mixed-valence systems[53], magnetic materials[54–56], lanthanides and actinides, defects and dopants; state-of-the-art

Table 1 | An overview of selected materials properties that can be obtained from DFT ground-state calculations

Materials properties	Models and theories	Electronic-structure toolbox
Atomic and cell geometries at fixed volume or pressure[13,166]	Hellmann–Feynman theorem	Total energy, forces and stresses
Zero-temperature stability[141–143]: formation energies, elastic constants, defects concentrations[167]	Equations of state (Murnaghan, Birch, …), convex hulls, self-consistent chemical potentials	Total energies, forces and stresses for all different phases, defects, periodic-boundary corrections for charged defects
Chemistry and reactivity[84,85], surface science[88]	Potential-energy surfaces, transition-state theory, volcano plots, kinetic Monte Carlo, rate equations, conical intersections, Marcus theory, Franck–Condon principle	Total energies and forces, van der Waals functionals, nudged-elastic-band method, constrained DFT, non-adiabatic dynamics (surface hopping, Ehrenfest)
Phonon dispersions and thermomechanical properties[18], thermal and electrical transport[19], superconductivity[168]	Linear-response theory, quasi-harmonic approximation, Grüneisen parameters, Boltzmann transport equation, equilibrium/non-equilibrium Green's functions, Allen–Dynes formula, Migdal–Eliashberg equations, superconducting DFT	Density functional perturbation theory and $2n+1$ theorem for el–ph, ph–ph interactions, Born effective charges, dielectric tensor
Dielectric[18], magnetic[123] and topological properties[125], ferroelectrics[92] and multiferroics[169]	Linear-response theory, modern theory of polarization and of magnetization, electric enthalpy, model Hamiltonians, topological invariants	Berry phases, maximally localized Wannier functions, Wilson loops, spin–orbit coupling
Magnetic phases[54], magnetic anisotropy, spin waves[55], skyrmions[56]	Spin Hamiltonians (Ising, Heisenberg, Dzyaloshinskii–Moriya), paramagnetism as ensemble average	Spin–orbit coupling, non-collinear magnetism, force theorem, spin-density functionals
Thermodynamic ensembles[71,77,78]: finite-temperature properties and Helmholtz or Gibbs free energies[81], transport coefficients[108]	Molecular dynamics, Monte Carlo, thermodynamic integration, metadynamics, path-integral molecular dynamics, Green–Kubo relations	Born–Oppenheimer and Car–Parrinello molecular dynamics, density functional perturbation theory, thermostats and barostats, ring-polymer mapping
Thermodynamic ensembles[88,129,130,170]: composition, chemical potential, partial pressure	Lattice Hamiltonians, Monte Carlo, mean-field approximation, cluster variation method, model entropies, special quasirandom structures	Cluster expansions, computational alchemy
Electrochemistry[86,89], pH[90], operando studies	Grand-canonical simulations for electrons and ions, embedding, double-layer and diffuse-layer models	Computational hydrogen electrode, model solvents and electrolytes, Poisson–Boltzmann solvers, ensemble simulations, coupling with reservoirs
Macroscopic mechanical properties[109,110] (strength, fracture, and plasticity), soft matter, biomolecules	Multiscale simulations, QM-MM, dislocation dynamics, effective volumes	Force fitting, learn-on-the-fly, neural networks and kernel-regression potentials
Microscopies[171]: STM, AFM, TEM	Tersoff–Hamann model, electron scattering	Local density of states, first-principles molecular dynamics, all-electron charge density and potential, PAW reconstructions

If the exact functional were known, these properties would be exactly reproduced, provided all the microscopic phenomenology were included. The second column points to the broader concepts, models and phenomenology from physics, chemistry and materials science; the third column highlights the electronic-structure quantities, algorithms and theories that are typically coded. Band-structure properties require more advanced methods, but are often approximated using KS states. STM, scanning tunnelling microscopy; AFM, atomic force microscopy; TEM, transmission electron microscopy; QM-MM, quantum mechanics - molecular mechanics; el, electron; ph, phonon; PAW, projector augmented wave.

implementations often require individual attention for each material studied. For magnetic systems, model Hamiltonians fitted on DFT calculations can be solved exactly, allowing for progress when the degrees of freedom of the model can be decoupled from the environment. Even more challenging are materials displaying strong correlations and the breakdown of the Fermi-liquid picture of quasiparticles (for example, with fluctuating localized/delocalized f electrons, heavy-fermion behaviour, Hubbard side bands); dynamical mean-field theory (DMFT) and its extensions[57–60] are often the tool of choice. And there are properties that escape completely current capabilities—high-temperature superconductivity in copper oxides being one of the most notable cases.

In addition, new and accurate instruments and facilities (from highly resolved photoemission[61] to time-resolved pump–probe experiments[62,63] to free-electron lasers) call for additional theoretical developments and implementations[64–67]. To provide some guidance in this complex landscape we summarize in Box 1 many of the electronic-structure methods in use today, and in Box 2 Green's function methods, of central relevance for the calculation of spectroscopic properties.

Predictions of materials properties and spectroscopies

The power of first-principles computational materials science has now become apparent owing to the wide range of properties and spectroscopies that can be studied for systems of increasing complexity[66,68]. As guidance, we sketch in Tables 1 and 2 the very rich phenomenology that can be addressed nowadays, providing pointers to the models, theories and electronic-structure toolbox that can be used to make these predictions. In the following, we present some selected subjects that are most relevant to materials, and highlight in Figs. 1 and 2 some milestone approaches that are now widely used.

Thermodynamic ensembles and realistic conditions. Volume and pressure are not only the simplest thermodynamic variables to control in periodic-boundary conditions but also some of the most powerful, owing to the relevance and phenomenological richness of high-pressure physics. In fact, first-principles simulations have had some of their early and notable impact in the field of materials under high pressure[69,70], effortlessly imposing extreme conditions that might not even be possible in earthly experiments.

Table 2 | A selected overview of spectroscopic properties that can be obtained from DFT ground-state calculations or from excited-state spectral formulations

Spectroscopic properties	Models and theories	Electronic-structure toolbox
Vibrational spectroscopies[18,19]: infrared, Raman and hyper-Raman, SERS, SEIRAS, and SFG, IETS, HREELS, INS	Vibrations from harmonic and adiabatic approximations, anhamonicity, Placzek and Albrecht approximation, Fermi golden rule	Phonon frequencies, dispersions and lifetimes, electron–phonon matrix elements. Born effective charges, frequency-dependent electric tensor, polarizability tensor, electric enthalpy, electric-field gradients
'Charged excitation' spectroscopies[66,114]: direct and inverse photoemission, ARPES[61], UPS, STS[171]	Quantum field theory of electron–electron and electron–boson interactions, diagrammatic perturbation theory, Hedin's equations, one-step and three-step models, Tersoff–Hamann model	Green's functions and self-energies, GW approximation, vertex corrections, cumulants, electron–phonon self-energies, DMFT, Koopmans spectral functionals, hybrids/range-separated/dielectric-dependent functionals
'Neutral excitation' spectroscopies[66]: optical absorption and luminescence[172,173], Stokes shifts, colour and reflectivity[174], nonlinear optics and second harmonic generation[175], (non-resonant) IXS, EELS, resonant Raman, Auger recombination[176]	Light–matter interaction, electron scattering, linear and nonlinear response, first- and second-order Fermi golden rules, excited-state relaxations, trichromatic theory, Kramers–Heisenberg	TDDFT, Casida equations, Bethe–Salpeter equation, optical ($\mathbf{q}=0$) and finite-$\mathbf{q}$ limits of the dielectric response, real-time propagation
Magnetic spectroscopies[113]: NMR, EPR, quadrupolar coupling, g-tensor, hyperfine couplings, Knight shifts, MCD, muon spin spectroscopy, magnons[55]	Effective nuclear Hamiltonians, linear-response theory, modern theory of magnetization	Gauge-including projector augmented waves, Berry phases and converse NMR, all-electron reconstruction, electric-field gradients, TDDFT
X-ray spectroscopies: XPS[114], XAS[177,178] (XANES, EXAFS), RIXS, Compton scattering, XMCD	Light–matter interaction, Fermi golden rule, crystal-field splitting, multiplets and model Hamiltonians, dipole and impulse approximations	Full and half core-hole pseudopotentials, total energies, electronic states from DFT, GW, Bethe–Salpeter and wavefunction methods
Real-time[64,65,179], ultrafast, pump–probe[62,63,106] and two-dimensional spectroscopies	Out-of-equilibrium dynamics, time-dependent Boltzmann transport, Keldysh non-equilibrium Green's function, Kadanoff–Baym equations	DFT, real-time propagation TDDFT, time-dependent current DFT, Ehrenfest dynamics, non-equilibrium Green's functions
Quantum optics[67,180,181]	Field quantization	QED DFT, QED many-body perturbation theory

See also Table 1 footnote. SERS, surface-enhanced Raman spectroscopy; SEIRAS, surface-enhanced infrared absorption spectroscopy; SFG, sum-frequency generation; IETS, inelastic electron tunnelling spectroscopy; HREELS, high-resolution electron energy loss spectroscopy; INS, inelastic neutron scattering; ARPES, angle-resolved photoemission spectroscopy; UPS, ultraviolet photoemission spectroscopy; STS, scanning tunnelling spectroscopy; IXS, inelastic X-ray scattering; EELS, electron energy loss spectroscopy; NMR, nuclear magnetic resonance; EPR, electron paramagnetic resonance; MCD, magnetic circular dichroism; XPS, X-ray photoelectron spectroscopy; XAS, X-ray absorption spectroscopy; XANES, X-ray absorption near-edge structure; EXAFS, extended X-ray absorption fine structure; RIXS, resonant inelastic X-ray scattering; XMCD, X-ray magnetic circular dichroism; QED, quantum electrodynamics; **q**, quasi-momentum transfer.

Temperature represents the first major computational challenge, since it requires probing atomic vibrations and motion, and a breakthrough came with Car–Parrinello molecular dynamics[71], evolving simultaneously the nuclear and electronic degrees of freedom through an effective Lagrangian[72]. Other approaches followed, based on robust direct-minimization algorithms[73] or efficient iterative ones[74]; these could be applied equally well to insulators or metals[74,75]. Thermostats and barostats can also couple simulations[76] to reservoirs to maintain the desired temperature or pressure. First-principles molecular dynamics considers every nucleus as a distinguishable entity, and so the resulting statistics is the Maxwell–Boltzmann one of classical particles; imposing the correct Bose–Einstein statistics requires path-integral sampling[77] or coloured-noise thermostats[78]. In the harmonic approximation, the dynamics of crystalline solids is that of an ensemble of independent oscillators, for which the Bose–Einstein partition function can be calculated analytically. The capability to calculate phonon dispersions with density functional perturbation theory[18,79,80] or with finite differences allows one to derive the Helmholtz free energy and, in the quasi-harmonic approximation, the temperature dependence of the lattice parameter and the elastic constants[18]. Stronger anharmonicity can be tackled with molecular dynamics, where free energies can be obtained from thermodynamic integration[81], with higher-order interatomic force constants or by sampling the phonon Green's function[20]. Collective variables and metadynamics[82] allow one to capture the most relevant degrees of freedom and determine their free energy landscape.

In alloys, the dependence on composition can be studied by calculating free energies as a function of chemical potential to determine phase coexistence and composition–temperature phase diagrams (see 'Building model Hamiltonians' below); if one of the phases is liquid, full molecular dynamics simulations are required[69]. Ensemble simulations can also be applied to paramagnetic phases[83]. Surface science has been another early driver towards the exploration of materials under realistic conditions: first rationalizing the descriptors driving reactivity[84], exploring the complex potential energy surfaces of molecular chemisorption[85,86] and surface reconstructions[87], and extending the dependence on chemical potentials to the partial pressure of a gas in equilibrium with a bulk or with a surface[88] or to electrochemical potentials[86,89] and pH[90]. Fittingly, the first high-throughput search for novel materials was driven by the quest for heterogeneous catalysts[91].

Multiscale and multiphysics. The scope of materials properties that can be predicted from first-principles can be broadened further by multiscale or multiphysics simulations. The potential-energy surface of crystals can be fitted to first-principles calculations, to study phase transitions, for example, in bulk ferroelectrics[92] or in complex geometries[93]. Such an approach has now been greatly expanded owing to progress in machine learning, where the potential-energy surface can be learnt on-the-fly[94,95], fit to neural networks[96] or featurized in kernel-regression methods[97] building on vast amounts of consistent calculations. Wannier Hamiltonians[98] can capture the building blocks of electronic structure, providing tight-binding

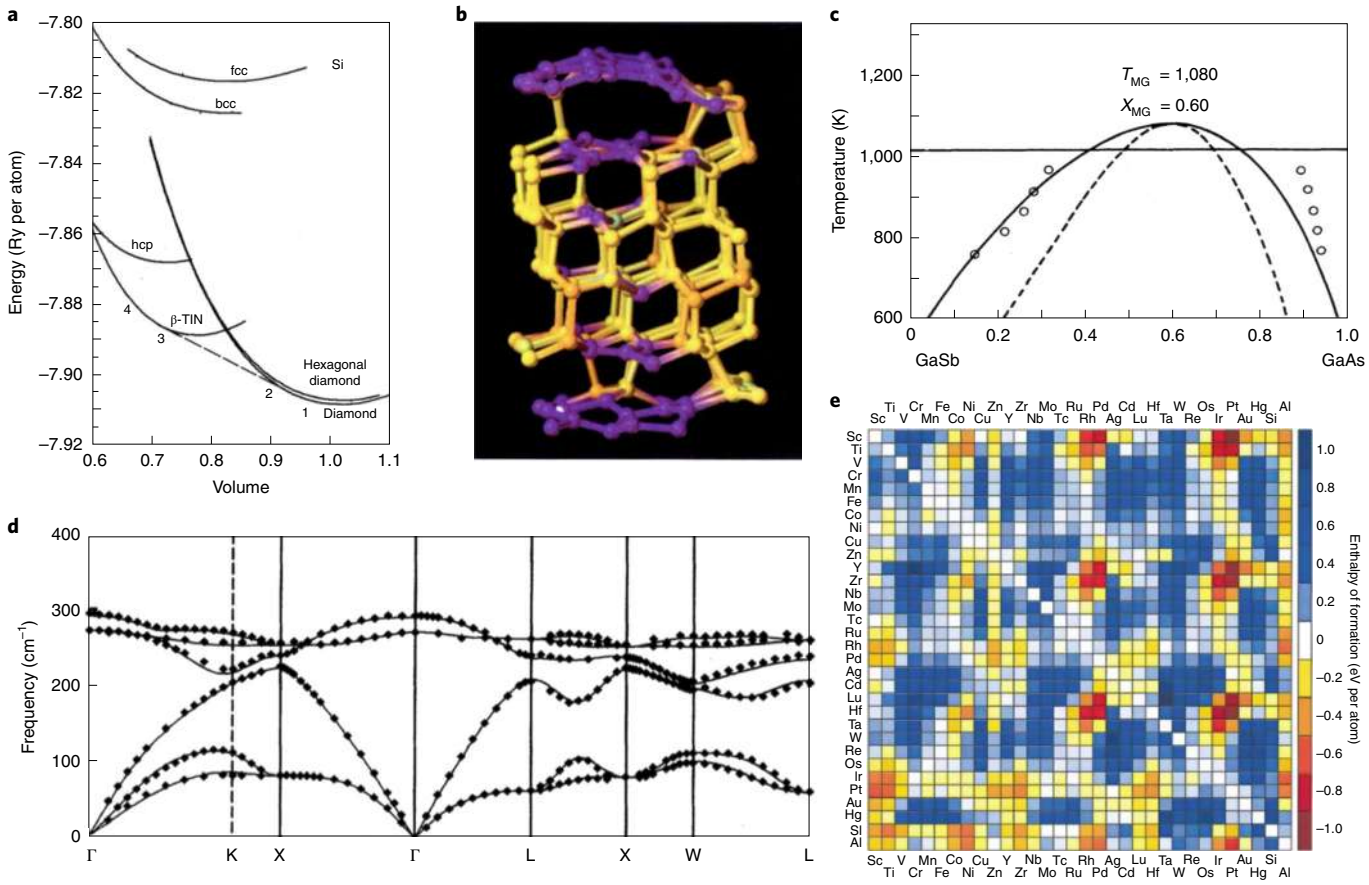


Fig. 1 | An overview of some key first-principles approaches to predict materials properties. a, The total-energy planewave pseudopotential method: equations of state for silicon in six different phases, highlighting a compression path (1–2–3–4) and the common-tangent transition (2–3) from the diamond to the β-tin structure[166]. **b**, Car–Parrinello molecular dynamics: graphitization of diamond (diamond and graphitic bonding in yellow and purple, respectively)[182]. **c**, First-principles thermodynamics: phase diagram of $Ga_{1-x}As_xSb$ (coexistence and spinodal curves, and composition X_{MG} and temperature T_{MG} of the miscibility gap) (ref. [129]). **d**, Density functional perturbation theory: phonon dispersions in GaAs (ref. [79]). **e**, High-throughput materials discovery: formation enthalpies of multinary alloys[91]. Credit: panels reproduced with permission from: **a**, ref. [166], APS; **b**, ref. [182], Springer Nature Ltd; **c**, ref. [129], APS; **e**, ref. [91], APS. Panel **d** adapted with permission from ref. [79], APS

parametrizations from first principles that can be used, for example, to study device operations; state-of-the-art methods can also directly access the properties of realistic nanostructures[99–101] and interfaces[62,102].

A notable example of multiphysics simulations is that of transport theories, where the microscopic dynamics of excitations is abstracted into the dynamics of their populations through master equations[103], such as the Boltzmann[104] or Wigner[105] transport equations, and extended also to non-equilibrium dynamics[106]. The finite excitation lifetimes due to the scattering mechanisms that limit transport—either extrinsic (such as defects) or intrinsic (most commonly, electron–phonon and phonon–phonon scattering)—can be obtained from Fermi's golden rule, for example, through the evaluation of the probability amplitudes for three-body events (these can be obtained from the third derivatives of the energy functional[19,107]). Green–Kubo relations can also provide transport coefficients directly from molecular dynamics[108].

Structural materials have seen the development of multiscale methods to understand fracture, adhesion, yield strength and other quantities associated with the materials microstructure. First-principles calculations can then provide microscopic parameters such as misfit volumes[109] for phenomenological models at different scales, atomic forces to learn on-the-fly[95], and stacking faults for dislocation dynamics[110].

Materials spectroscopies. Spectroscopies are central to the characterization of materials: matching accurate experimental measurements with first-principles predictions provides both validation and understanding, and a causal relation between electronic and atomic phenomena and their macroscopic effects[111,112]. We summarize most of the techniques accessible to first-principles simulations in Table 2, using a classification that tries to follow the electronic-structure methods employed. Broadly speaking, there are spectroscopies that can be addressed by DFT (for example, some vibrational[18,80] and magnetic[113] spectroscopies, often addressed with density functional perturbation theory), and those that require the calculations of electronic excitations (charged excitations[66,114], where the number of electrons in system varies, as in a photoemission experiment, or neutral excitations, where the number remains constant and electron–hole pairs are created, as in an absorption experiment[66]). This latter distinction recalls also the different gaps in materials, where the fundamental or transport gap is determined by the valence and conduction band edges, whereas the optical gap is lowered due to the presence of excitons. The correct description of excitations is thus also very relevant to transport[115,116], in addition to its formal connection at zero bias to linear response[65]. Owing to their importance and the requirement to go beyond DFT, we have dedicated part of Box 1 and the whole of Box 2 to spectral and non-equilibrium (time dependent) electronic-structure approaches.

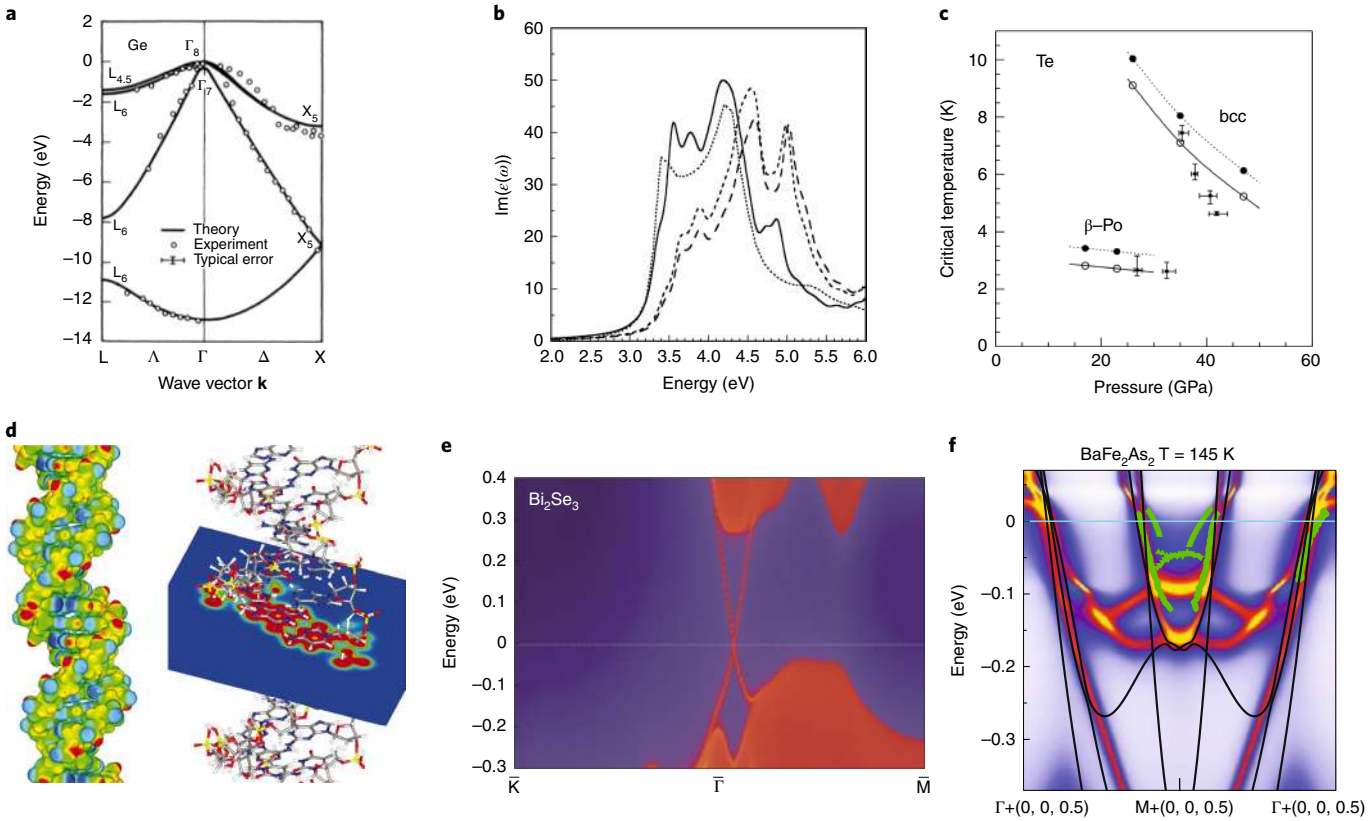


Fig. 2 | An overview of some key first-principles approaches to predict materials properties and spectroscopies. a, GW many-body perturbation theory: band structure of germanium[183]. **b**, Bethe–Salpeter equation: optical absorption of silicon[172] (in particular: experiments, dotted line; Bethe–Salpeter prediction, thick solid line). **c**, Electron–phonon coupling: pressure-dependent critical temperature for two different superconducting phases of tellurium in the bcc and β-polonium structures[168]. **d**, Linear-scaling DFT: 20-base-pair fragment of DNA (left panel, colour map of the electrostatic potential on a charge-density isosurface; right panel, colour map of the charge density in the plane of the hydrogen bonds of a base pair)[32]. **e**, Recursive Green's function method in a Wannier basis: topologically protected surface states in Bi_2Se_3 (ref. [117]). **f**, (Extended) DMFT: doping and temperature-dependent band structure of hole-doped $BaFe_2As_2$ (ref. [60]). Credit: panels reproduced with permission from: **a**, ref. [183], APS; **c**, ref. [168], APS; **e**, ref. [117], Springer Nature Ltd; **f**, ref. [60], Springer Nature Ltd. Panels adapted with permission from: **b**, ref. [172], APS; **d**, ref. [32], AIP

Electronic structure and topology. The topological properties of electrons in solids[117,118] have rapidly taken a prominent role in current condensed-matter physics and materials science. While these first manifested themselves with the integer and fractional quantum Hall effects, in electronic-structure methods geometric phases emerged with the modern theory of polarization[119], driven by the need to calculate the electronic polarization, a non-trivial task due to the fact that the position operator in a periodic solid is ill defined. It was shown that the polarization, when recast as an integral of the linear response to an electric field[120], measures an adiabatic flow of current. This can be computed as the Berry phase acquired by the wavefunctions under parallel transport[121], or, equivalently, from the displacements of the Wannier function centres[98], which represent an exact mapping of the dielectric response into classical electrostatics. A complementary effort has also given rise to the modern theory of magnetization[122,123]; notably, an implication of these theories is that the xc functional may also depend on macroscopic quantities such as the polarization[124]. The topological properties of materials—from Chern and quantum spin Hall insulators to Weyl semimetals—are nowadays very actively investigated with electronic-structure methods[125], even charting the entire landscape of known materials[126].

Materials' design and discovery

Owing to the accuracy and efficiency of first-principles simulations, and the exponential availability of computational power, not only has the range of calculated materials properties expanded dramatically but also the use of these to design and discover novel materials has become an extremely active and growing research area. We identify four main approaches to first-principles design and discovery, and briefly outline each one here: model Hamiltonians, structure prediction, high throughput and data driven.

Building model Hamiltonians. In this approach, calculated materials properties are mapped onto an explicit parametric model, which can then be explored more extensively than is possible with direct, first-principles calculations. One of the most famous examples comes from alloy theory in the form of the cluster expansion[127] (CE). The CE can address many problems in materials science that require the total energies of a large number of substitutional structures. Examples include the aforementioned composition–temperature phase diagrams of alloys, but also the search for the lowest-energy structure among substitutional configurations of lattice sites by different species, the calculation of the energetics of random alloys, and the stability of differently oriented superlattices, substitutional impurities or antiphase boundaries. In the CE, the energy of any atomic configuration on a lattice is written in a generalized Ising-like form in terms of products of single-site spin variables over clusters of atoms, where the coefficients of each term represent the effective interaction for that cluster. First-principles energies for selected sets of substitutional arrangements can then

be fit to the CE, yielding the interactions. The resulting parametric CE can then be used in statistical mechanics simulations (such as Monte Carlo) or enumeration algorithms to solve the problems listed above, or inspire the choice of special 'quasirandom' periodic structures that target the correlation functions of more complex disordered materials[128].

The use of first-principles calculations in the cluster expansion was pioneered in the 1980s and 1990s[129,130] and the approach has proven to be applicable to a wide range of systems, including metal alloys[131], semiconductor alloys[129] and oxides[132], but also vacancy ordering in superconductors[133] and batteries[134], surface ordering and precipitate shapes[135]. The CE is only an example of a much broader class of model-building approaches, but represents well the key idea of leveraging DFT calculations to parametrize a simpler Hamiltonian that can then be more extensively explored.

Structure prediction. The fundamental paradigm of materials science involves the structure–property relation: the structure of a material (at all length scales) ultimately controls its properties. First-principles calculations require structures as input, and hence calculations can be aided by collections of experimental crystal structures. However, in the attempt to predict new, as-yet-undiscovered compounds, one is immediately faced with the challenge of predicting crystal structures a priori. The editor of *Nature* famously wrote in 1988: "One of the continuing scandals in the physical sciences is that it remains impossible to predict the structure of even the simplest crystalline solids from a knowledge of their composition"[136]. In these subsequent years, dramatic progress has been made, and there now exists a collection of computational methods capable of predicting a wide array of crystal-structure types, often using only first-principles calculations and no other input. The structure-prediction problem is made difficult because of the very large dimensionality for the space of possible crystal-structure symmetries, cell shapes and basis vectors. Most of the structure-prediction methods are based on the idea of attempting to minimize the energy over this high-dimensional space, where the minimization is performed using an efficient optimization algorithm.

Successful methods have used a variety of approaches, such as genetic algorithms, particle swarm methods, molecular dynamics minima hopping, random structure searches and Monte Carlo techniques. A detailed description of these, along with substantial examples of their successful application to materials discovery, has recently been presented[137]. These approaches can also be applied to the structure–solution problem (that is, the solution of crystal structures given experimental diffraction data) by minimizing a cost function that is a combination of the first-principles total energy and the match to the experimental diffraction pattern[138].

High-throughput first-principles calculations. Another successful and widely used approach to materials design is that of high-throughput (HT) first-principles calculations[1,91]. Here the focus shifts from calculating the property of interest of one compound at a time to computing it at scale for many compounds. In the extreme, one can even attempt to compute properties for all known or predicted compounds. HT calculations are made possible due to the automation of workflows in performing first-principles calculations, robust rates of automation for these calculations with little or no manual intervention, and the applicability of first-principles methods to most or all of the periodic table, owing to the availability of verified libraries of pseudopotentials[10,74].

Some of the first HT calculations already combined ideas from machine learning and data-driven approaches: from the search, in combination with an evolutionary algorithm, for the most stable four-component alloys out of the 192,016 face-centred cubic (fcc) and body-centred cubic (bcc) structures that can be constructed out of 32 different metals[91], to the Pareto-optimal set in compressibility, stability and cost for alloys out of a HT dataset of thousands of metallic compounds[139], to crystal-structure prediction through heuristic-rule extraction on a large library of calculated properties[140]. HT first-principles calculations have been used to build large and powerful databases of materials structures and properties, the earliest examples being the Materials Project[141], AFLOWlib[142] and the Open Quantum Materials Database (OQMD)[143]. Each of these is based on DFT-calculated properties of experimentally observed and of novel, predicted compounds. These databases include not only properties obtained from standard total-energy calculations (such as energy, density of states, electronic structure, magnetic moments) but also have been increasingly used to curate additional datasets for more complex properties (for example, elastic and piezoelectric tensors, thermoelectric properties, surface energies, phonon dispersions and X-ray absorption near-edge spectroscopy spectra). Their utility and versatility has been demonstrated by their application to design and discover novel materials in a vast array of applications: strengthening precipitates in structural alloys, lithium (and beyond) battery materials, efficient thermoelectric materials and solar thermochemical water-splitting materials, to name a few. Also, these databases are by no means the only examples and many new efforts are emerging—both delivering[144–146] or consolidating curated data[147].

In discovery efforts involving large datasets, a screening strategy is often employed: one starts with a large pool of candidate compounds, and applies more and more stringent criteria to downselect the most promising compounds. Selection criteria prioritize quantities that are relatively straightforward and inexpensive to calculate, and only later introduce more computationally complex, and hence expensive, quantities for further refinement. A very common criterion is the zero-temperature ground-state phase stability, which is identified through a convex hull approach (Box 3). Such a screening strategy is illustrated in Fig. 3, for the case of robust and synthesizeable photocatalysts for CO_2 reduction[148], but the literature offers many examples that have been summarized in several reviews[1,149,150].

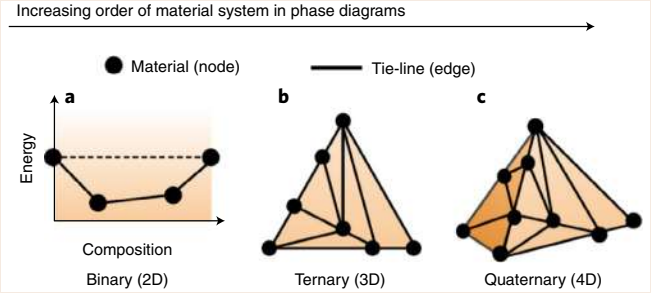
Data-driven approaches. Machine learning and data-driven approaches can greatly aid the search for new materials. Materials informatics approaches are generally based on three distinct components: a resource of materials data, a representation to quantitatively describe each material and machine learning algorithms to discover patterns within the data or to predict the properties of new materials. The use of HT computations for data collection has the great benefit of improving systematically and exponentially the size of the data pool, and the discovery or prediction portion of the workflow is greatly facilitated by open-source, state-of-the-art algorithms and software packages both in electronic-structure and machine learning.

The key challenge, therefore, lies in the construction of a material representation. This representation is a set of quantitative attributes that describe the relevant characteristics for the materials property of interest. Work in this field has generally fallen into two categories of representations: those that are functions of composition only, and those that additionally contain structural attributes of the crystallographic unit cell. Many of the composition-only attribute sets are primarily based on statistics of the properties of constituent elements. For instance, a material could be described just from the fraction of each element present and various intuitive properties, such as the maximum difference in electronegativity, to model the formation energies, bandgaps or other properties[151–154]. Incorporating structure-dependent attributes is an obvious route to improve the material representation. Examples of these

Box 3 | Convex hull construction

The convex hull is a powerful graphical construct for determining the phase stability of compounds (most commonly, at zero temperature). For a compound to be thermodynamically stable, it must not only be lower in energy than all other phases at the same composition but also be lower in energy than all linear combinations of phases. If one identifies the lowest-energy linear combinations of phases for each composition of a system, then the set of all such energies forms a convex hull. Phases that are on the convex hull are thermodynamically stable, and those that have an energy that lies above the convex hull are unstable or metastable. For a binary A–B system, there is only one composition variable, and so the convex hull can be represented as a two-dimensional (2D) object on a composition-energy plot (panel **a**). For an N-dimensional system, there are $N-1$ independent compositions, and so the convex hull becomes an N-dimensional object. In ternary systems, the convex hull is 3D, but it is common to take a 2D projection of the convex hull on a Gibbs triangle, eliminating the energy axis (panel **b**). Similarly, in quaternary systems, the 4D convex hull is often represented as a 3D projection on a Gibbs tetrahedron (panel **c**). If one is given a set of DFT-calculated energies for a given system, determination of the convex hull is straightforward using a variety of standard algorithms. The construction of convex hulls is automated in many of the high-throughput DFT databases (for example, the Materials Project[141], OQMD[143] and AFLOWlib[142]).

The convex hull also provides a simple, straightforward way to assess the stability of new, predicted phases: their energies may be compared with those on the existing convex hull, and a phase is stable if its energy lies below; the convex hull and all its derivatives then must be updated to include the new, stable phase.



Convex hulls. a, Convex hull for a binary system, connecting the lowest-energy phases at each composition with tie-lines. **b**,**c**, Projections of the 3D (**b**) and 4D (**c**) convex hulls for a ternary or quaternary system on a Gibbs triangle or tetrahedron[161]. Credit: figure adapted with permission from ref. [161], under a Creative Commons License CC BY 4.0

representations include the smooth overlap of atomic orbitals[155], the pair radial distribution function and the Coulomb matrix[156], with machine learning algorithms then built to predict scalar or tensorial quantities or fields. Representations based on Voronoi tessellation of crystal structures have been suggested, using the resulting attributes of the Voronoi polyhedral shape and connectivities[157]. A recent improvement comes from representations based on crystal graphs in a convolutional neural network[158]. These (and recently introduced variants[159,160]) are highly versatile approaches that can lead to accurate predictions, and involve learning material properties directly from graph-like representations of crystal structures ('crystal graphs').

Data-driven machine learning models of materials properties have seen a large number of applications targeted at the discovery of novel materials with promising properties; many of these models have also been used to direct HT searches, by steering these efficiently towards more promising compounds. In fact, the discovery rate for new and stable compounds has been greatly accelerated by the use of machine learning[156,160]. Recent studies[161,162] have shown how ideas from network theory can be applied to HT datasets, by mapping these data onto graph representations and analysing the topological features of this graph. This approach allows for a top-down study of the organizational structure of networks of materials, based on the interaction between materials themselves. A large-scale network has been defined by the convex hull of essentially all compounds in the OQMD dataset; analysis of this 'complete phase-stability network of all inorganic materials' shows that it is a densely connected complex network of 21,000 thermodynamically stable compounds (nodes) interlinked by 41 million tie-lines (edges) defining their two-phase equilibria. Since the edges of the graph represent two-phase equilibria, the number of edges coming out of a node can be thought of as a measure of the stability of the compound. Hence, the connectivity of nodes in this phase-stability network allows for the derivation of a rational, data-driven metric for material reactivity, the 'nobility index'.[162] Another illustration of the power of this network theory approach involves assigning a time stamp to each node/compound associated with the year of its

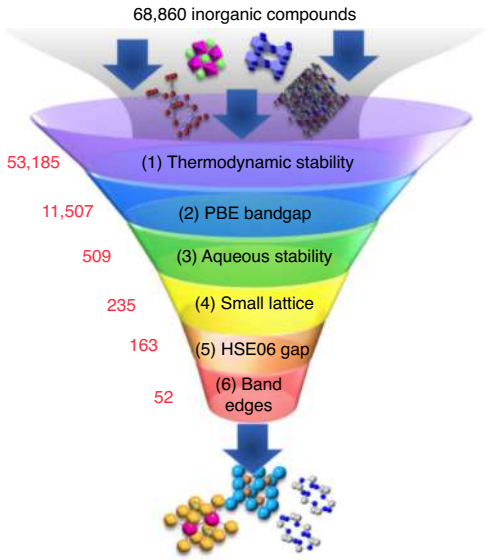


Fig. 3 | An illustration of the high-throughput screening approach. A large number of candidate materials are chosen from experimental or computational databases, and a sequence of screening calculations reduces their number down to a small set of candidates with the most promising properties[148]. Credit: figure reproduced with permission from ref. [148], under a Creative Commons License CC BY 4.0

experimental discovery. This allows one to determine the graph at any previous point in time, and to determine whether the topology of the graph shows any indication of the imminent discovery of a compound in the years before its synthesis. Training a machine learning model of this time evolution in terms of the topological features of the graph allows one to predict the likelihood that hypothetical, computer-generated materials will be amenable to successful experimental synthesis[161].

Outlook

As simulations can suggest novel materials, in addition to predicting novel properties of experimentally known materials, key challenges become the assessment of thermodynamic stability, synthesis conditions, manufacturability and tolerance of the predicted properties to intrinsic and extrinsic defects. DFT estimates might ultimately need to be augmented by more advanced electronic-structure methods or machine learning algorithms to improve accuracy, and by seamless deployment of computational materials science methods to address realistic conditions—from vibrational entropies to defects' concentrations to applied electrochemical potentials. Materials synthesis will be further driven by the freedom afforded by many combinatorial templates—from perovskites to half-Heuslers to metal–organic frameworks—with automated experiments and automated simulations starting to handshake, and human or artificial intelligence deciding which experiment is required by a simulation, and which simulation is needed by an experiment.

On a more fundamental level, achieving predictive accuracy in the simulations and capturing the realistic complexity of materials remain grand challenges. Not only magnetism or correlations push our ingenuity, but also ever-more-advanced experimental techniques test our capability to address time-resolved or ultrafast processes, to describe correctly the nano- and microstructure of materials, or to sample configurations and processes across length scales or timescales. Some of these latter capabilities are essential to understand synthesis, processing, manufacturing and ultimately failure. Last, simulations can make meaningful predictions only when the underlying theory is understood and coded, although there is also the possibility of observing unexpected behaviour emerging from interactions that are correctly captured, and for data mining pointing to unexpected correlations.

The sustained performance scaling in computing hardware and the emergence of disruptive accelerators for exponentially expensive tasks—such as neuromorphic and quantum computing—together with those driven by artificial intelligence that are already becoming established[94,97] make for one easy prediction. Electronic-structure simulations will keep increasing their relevance and impact in accelerating, streamlining and focusing our efforts in materials design and discovery, delivering ever-more-complex capabilities for predicting and characterizing properties and performance, and addressing those, and those of the devices that are built out of them, in ever-more-realistic conditions and environments. This acceleration mirrors the one of information-and-communication technologies, rather than that of physical infrastructures; is compounded by the novel ideas and algorithms that enter the field; and, at variance with physical infrastructure, can be instantaneously shared worldwide. The impact of the field is already apparent in its acceptance and uptake by the experimental community, and in the remarkable predictions made—from ultrafast thermal transport[163] to electron–phonon mediated superconductivity in hydrides[164] to the emergence of flat bands in twisted-bilayer graphene[165]—that have inspired even more remarkable experiments.

Against this background it is then very fitting to note that while physical infrastructures and major installations for research—from synchrotrons to radiotelescopes to supercomputers—are well established, and rightly so, in our scientific society, the support and planning for the computational infrastructures, from the widely used scientific software[31] that powers the research presented here, to the verification of codes and the validation of theories, to the dissemination and curation of computational data, tools and workflows, and the associated career models these entail and require are only just, belatedly, emerging. And users and beneficiaries of these capabilities now go well beyond core practitioners, underlying the importance of robustness and reliability in electronic-structure simulations, and education in their scope and limits—something this Review aims to contribute to.

We conclude by recalling that it is often said that human ages take their name from the materials that characterize them, and the computational design and discovery of novel materials, as enablers of the technologies that power our economy and sustain our society, will be be firmly at the centre stage for the coming decades. What to do with them will remain, for better or for worse, a human decision.

Received: 5 October 2020; Accepted: 19 April 2021;
Published online: 27 May 2021

References

1. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
2. Bartók, A. P. et al. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **3**, e1701816 (2017).
3. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
4. Di Ventra, M. & Traversa, F. L. Perspective: memcomputing: leveraging memory and physics to compute efficiently. *J. Appl. Phys.* **123**, 180901 (2018).
5. Barkoutsos, P. K. et al. Quantum algorithms for electronic structure calculations: particle–hole Hamiltonian and optimized wave-function expansions. *Phys. Rev. A* **98**, 022322 (2018).
6. Pribram-Jones, A., Gross, D. A. & Burke, K. DFT: a theory full of holes? *Rev. Phys. Chem.* **66**, 283–304 (2015).
7. Hohenberg, P. & Kohn, W. Inhomogeneus electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
8. van Noorden, R., Maher, B. & Nuzzo, R. The top 100 papers. *Nature* **514**, 550–553 (2014).
9. van Setten, M. J. et al. GW100: benchmarking G_0W_0 for molecular systems. *J. Chem. Theory Comput.* **11**, 5665–5687 (2015).
10. Lejaeghere, K. et al. Reproducibility in density functional theory calculations of solids. *Science* **351**, aad3000 (2016).
11. Rangel, T. et al. Reproducibility in G_0W_0 calculations for solids. *Comput. Phys. Commun.* **255**, 107242 (2020).
12. de Jong, M. et al. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, 150009 (2015).
13. Isaacs, E. B. & Wolverton, C. Performance of the strongly constrained and appropriately normed density functional for solid-state materials. *Phys. Rev. Mater.* **2**, 063801 (2018).
14. Borlido, P. et al. Large-scale benchmark of exchange-correlation functionals for the determination of electronic band gaps of solids. *J. Chem. Theory Comput.* **15**, 5069–5079 (2019).
15. Marzari, N. Materials modelling: the frontiers and the challenges. *Nat. Mater.* **15**, 381–382 (2016).
16. Parr, R. G. & Yang, W. *Density-Functional Theory of Atoms and Molecules* (Oxford Science Publications, 1989).
17. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
18. Baroni, S., de Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515–562 (2001).
19. Giustino, F. Electron–phonon interactions from first principles. *Rev. Mod. Phys.* **89**, 015003 (2017).
20. Errea, I. et al. Quantum crystal structure in the 250-kelvin superconducting lanthanum hydride. *Nature* **578**, 66–69 (2020).
21. Perdew, J. P. & Levy, M. Physical content of the exact Kohn–Sham orbital energies: band gap and derivative discontinuities. *Phys. Rev. Lett.* **51**, 1884–1887 (1983).
22. Godby, R. W., Schlüter, M. & Sham, L. J. Trends in self-energy operators and their corresponding exchange-correlation potentials. *Phys. Rev. B* **36**, 6497–6500 (1987).
23. Casida, M. E. Generalization of the optimized-effective-potential model to include electron correlation: a variational derivation of the Sham–Schlüter equation for the exact exchange-correlation potential. *Phys. Rev. A* **51**, 2005–2013 (1995).
24. Perdew, J. P. et al. Understanding band gaps of solids in generalized Kohn–Sham theory. *Proc. Natl Acad. Sci. USA* **114**, 2801–2806 (2017).
25. Nguyen, N. L., Colonna, N., Ferretti, A. & Marzari, N. Koopmans-compliant spectral functionals for extended systems. *Phys. Rev. X* **8**, 021051 (2018).
26. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).
27. Anisimov, V. I., Zaanen, J. & Andersen, O. K. Band theory and Mott insulators: Hubbard *U* instead of Stoner *I*. *Phys. Rev. B* **44**, 943–954 (1991).

28. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).
29. Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64**, 1045–1097 (1992).
30. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
31. Ghiringhelli, L. M. et al. Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. *npj Comput. Mater.* **3**, 46 (2017).
32. Skylaris, C.-K., Haynes, P. D., Mostofi, A. A. & Payne, M. C. Introducing ONETEP: linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **122**, 084119 (2005).
33. Wang, Y. A. & Carter, E. A. in *Theoretical Methods in Condensed Phase Chemistry* (ed. Schwartz, S. D.) 117–184 (Springer, 2002).
34. Baer, R., Neuhauser, D. & Rabani, E. Self-averaging stochastic Kohn–Sham density-functional theory. *Phys. Rev. Lett.* **111**, 106402 (2013).
35. Gunnarsson, O. & Lundqvist, B. I. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B* **13**, 4274–4298 (1976).
36. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
37. Fu, Y. & Singh, D. J. Applicability of the strongly constrained and appropriately normed density functional to transition-metal magnetism. *Phys. Rev. Lett.* **121**, 207201 (2018).
38. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
39. Peverati, R. & Truhlar, D. G. Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Phil. Trans. R. Soc. A* **372**, 20120476 (2014).
40. Tran, F. & Blaha, P. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.* **102**, 226401 (2009).
41. Gritsenko, O., van Leeuwen, R., van Lenthe, E. & Baerends, E. J. Self-consistent approximation to the Kohn–Sham exchange potential. *Phys. Rev. A* **51**, 1944–1954 (1995).
42. Görling, A. Hierarchies of methods towards the exact Kohn–Sham correlation energy based on the adiabatic-connection fluctuation-dissipation theorem. *Phys. Rev. B* **99**, 235120 (2019).
43. Cohen, A. J., Mori-Sanchez, P. & Yang, W. Insights into current limitations of density functional theory. *Science* **321**, 792–794 (2008).
44. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
45. Pederson, M. R., Ruzsinszky, A. & Perdew, J. P. Communication: self-interaction correction with unitary invariance in density functional theory. *J. Chem. Phys.* **140**, 121103 (2014).
46. Cococcioni, M. & Gironcoli, S. D. Linear response approach to the calculation of the effective interaction parameters in the LDA+U method. *Phys. Rev. B* **71**, 035105 (2005).
47. Kronik, L. & Kümmel, S. Piecewise linearity, freedom from self-interaction, and a Coulomb asymptotic potential: three related yet inequivalent properties of the exact density functional. *Phys. Chem. Chem. Phys.* **22**, 16467–16481 (2020).
48. Dabo, I. et al. Koopmans' condition for density-functional theory. *Phys. Rev. B* **82**, 115121 (2010).
49. Refaely-Abramson, S., Jain, M., Sharifzadeh, S., Neaton, J. B. & Kronik, L. Solid-state optical absorption from optimally tuned time-dependent range-separated hybrid density functional theory. *Phys. Rev. B* **92**, 081204 (2015).
50. Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982–9985 (1996).
51. Brawand, N. P., Vörös, M., Govoni, M. & Galli, G. Generalization of dielectric-dependent hybrid functionals to finite systems. *Phys. Rev. X* **6**, 041002 (2016).
52. Stöhr, M., Van Voorhis, T. & Tkatchenko, A. Theory and practice of modeling van der Waals interactions in electronic-structure calculations. *Chem. Soc. Rev.* **48**, 4118–4154 (2019).
53. Cococcioni, M. & Marzari, N. Energetics and cathode voltages of $LiMPO_4$ olivines (M=Fe,Mn) from extended Hubbard functionals. *Phys. Rev. Mater.* **3**, 033801 (2019).
54. Blügel, S., Weinert, M. & Dederichs, P. H. Ferromagnetism and antiferro-magnetism of 3*d*-metal overlayers on metals. *Phys. Rev. Lett.* **60**, 1077–1080 (1988).
55. Niu, Q. & Kleinman, L. Spin-wave dynamics in real crystals. *Phys. Rev. Lett.* **80**, 2205–2208 (1998).
56. Heinze, S. et al. Spontaneous atomic-scale magnetic skyrmion lattice in two dimensions. *Nat. Phys.* **7**, 713–718 (2011).
57. Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
58. Kotliar, G. et al. Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.* **78**, 865–951 (2006).
59. Biermann, S., Aryasetiawan, F. & Georges, A. First-principles approach to the electronic structure of strongly correlated systems: combining the GW approximation and dynamical mean-field theory. *Phys. Rev. Lett.* **90**, 086402 (2003).
60. Werner, P. et al. Satellites and large doping and temperature dependence of electronic properties in hole-doped $BaFe_2As_2$. *Nat. Phys.* **8**, 331–337 (2012).
61. Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).
62. Falke, S. M. et al. Coherent ultrafast charge transfer in an organic photovoltaic blend. *Science* **344**, 1001–1005 (2014).
63. Sangalli, D., Dal Conte, S., Manzoni, C., Cerullo, G. & Marini, A. Nonequilibrium optical properties in semiconductors from first principles: a combined theoretical and experimental study of bulk silicon. *Phys. Rev. B* **93**, 195205 (2016).
64. Ullrich, C. A. *Time-Dependent Density-Functional Theory: Concepts and Applications* (Oxford Univ. Press, 2012).
65. Stefanucci, G. & van Leeuwen, R. *Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction* (Cambridge Univ. Press, 2013).
66. Martin, R. M., Reining, L. & Ceperley, D. *Interacting Electrons Theory and Computational Approaches* (Cambridge Univ. Press, 2016).
67. de Melo, P. M. M. C. & Marini, A. Unified theory of quantized electrons, phonons, and photons out of equilibrium: a simplified ab initio approach based on the generalized Baym–Kadanoff ansatz. *Phys. Rev. B* **93**, 155102 (2016).
68. Giustino, F. *Materials Modelling Using Density Functional Theory* (Oxford Univ. Press, 2014).
69. Alfè, D., Gillan, M. J. & Price, G. D. The melting curve of iron at the pressures of the Earth's core from ab initio calculations. *Nature* **401**, 462–464 (1999).
70. Cavazzoni, C. et al. Superionic and metallic states of water and ammonia at giant planet conditions. *Science* **283**, 44–46 (1999).
71. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
72. Parrinello, M. & Rahman, A. Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.* **45**, 1196–1199 (1980).
73. Teter, M. P., Payne, M. C. & Allan, D. C. Solution of Schrödinger's equation for large systems. *Phys. Rev. B* **40**, 12255–12263 (1989).
74. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
75. Marzari, N., Vanderbilt, D. & Payne, M. C. Ensemble density-functional theory for ab initio molecular dynamics of metals and finite-temperature insulators. *Phys. Rev. Lett.* **79**, 1337–1340 (1997).
76. Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications* (Academic-Press, 2001).
77. Marx, D. & Parrinello, M. Structural quantum effects and three-centre two-electron bonding in CH_5^+. *Nature* **375**, 216–218 (1995).
78. Ceriotti, M., Bussi, G. & Parrinello, M. Langevin equation with colored noise for constant-temperature molecular dynamics simulations. *Phys. Rev. Lett.* **102**, 020601 (2009).
79. Giannozzi, P., de Gironcoli, S., Pavone, P. & Baroni, S. Ab initio calculation of phonon dispersions in semiconductors. *Phys. Rev. B* **43**, 7231–7242 (1991).
80. Gonze, X. & Lee, C. Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B* **55**, 10355–10368 (1997).
81. Milman, V. et al. Free energy and entropy of diffusion by ab initio molecular dynamics: alkali ions in silicon. *Phys. Rev. Lett.* **70**, 2928–2931 (1993).
82. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
83. Körmann, F., Dick, A., Grabowski, B., Hickel, T. & Neugebauer, J. Atomic forces at finite magnetic temperatures: phonons in paramagnetic iron. *Phys. Rev. B* **85**, 125104 (2012).
84. Hammer, B. & Norskov, J. K. Why gold is the noblest of all the metals. *Nature* **376**, 238–240 (1995).
85. De Vita, A., Štich, I., Gillan, M. J., Payne, M. C. & Clarke, L. J. Dynamics of dissociative chemisorption: Cl_2/Si(111)-(2×1). *Phys. Rev. Lett.* **71**, 1276–1279 (1993).
86. Nørskov, J. K. et al. Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J. Phys. Chem. B* **108**, 17886–17892 (2004).
87. Lazzeri, M., Vittadini, A. & Selloni, A. Structure and energetics of stoichiometric TiO_2 anatase surfaces. *Phys. Rev. B* **63**, 155409 (2001).
88. Reuter, K., Frenkel, D. & Scheffler, M. The steady state of heterogeneous catalysis, studied by first-principles statistical mechanics. *Phys. Rev. Lett.* **93**, 116105 (2004).

89. Hörmann, N. G., Andreussi, O. & Marzari, N. Grand canonical simulations of electrochemical interfaces in implicit solvation models. *J. Chem. Phys.* **150**, 041730 (2019).
90. Persson, K. A., Waldwick, B., Lazic, P. & Ceder, G. Prediction of solid-aqueous equilibria: scheme to combine first-principles calculations of solids with experimental aqueous states. *Phys. Rev. B* **85**, 235438 (2012).
91. Jóhannesson, G. H. et al. Combined electronic structure and evolutionary search approach to materials design. *Phys. Rev. Lett.* **88**, 255506 (2002).
92. Zhong, W., Vanderbilt, D. & Rabe, K. M. First-principles theory of ferroelectric phase transitions for perovskites: the case of $BaTiO_3$. *Phys. Rev. B* **52**, 6301–6312 (1995).
93. Naumov, I. I., Bellaiche, L. & Fu, H. Unusual phase transitions in ferroelectric nanodisks and nanorods. *Nature* **432**, 737–740 (2004).
94. Csányi, G., Albaret, T., Payne, M. C. & De Vita, A. 'Learn on the fly': a hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys. Rev. Lett.* **93**, 175503 (2004).
95. Li, Z., Kermode, J. R. & De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
96. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
97. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
98. Marzari, N., Mostofi, A. A., Yates, J. R., Souza, I. & Vanderbilt, D. Maximally localized Wannier functions: theory and applications. *Rev. Mod. Phys.* **84**, 1419–1475 (2012).
99. Son, Y.-W., Cohen, M. L. & Louie, S. G. Half-metallic graphene nanoribbons. *Nature* **444**, 347–349 (2006).
100. Prezzi, D., Varsano, D., Ruini, A. & Molinari, E. Quantum dot states and optical excitations of edge-modulated graphene nanoribbons. *Phys. Rev. B* **84**, 041401 (2011).
101. Ruffieux, P. et al. On-surface synthesis of graphene nanoribbons with zigzag edge topology. *Nature* **531**, 489–492 (2016).
102. Pham, T. A., Ping, Y. & Galli, G. Modelling heterogeneous interfaces for solar water splitting. *Nat. Mater.* **16**, 401–408 (2017).
103. Poncé, S., Li, W., Reichardt, S. & Giustino, F. First-principles calculations of charge carrier mobility and conductivity in bulk semiconductors and two-dimensional materials. *Rep. Prog. Phys.* **83**, 036501 (2020).
104. Cepellotti, A. & Marzari, N. Thermal transport in crystals as a kinetic theory of relaxons. *Phys. Rev. X* **6**, 041013 (2016).
105. Simoncelli, M., Marzari, N. & Mauri, F. Unified theory of thermal transport in crystals and glasses. *Nat. Phys.* **15**, 809–813 (2019).
106. Bernardi, M., Vigil-Fowler, D., Lischner, J., Neaton, J. B. & Louie, S. G. Ab initio study of hot carriers in the first picosecond after sunlight absorption in silicon. *Phys. Rev. Lett.* **112**, 257402 (2014).
107. Gonze, X. & Vigneron, J.-P. Density-functional approach to nonlinear-response coefficients of solids. *Phys. Rev. B* **39**, 13120–13128 (1989).
108. Marcolongo, A., Umari, P. & Baroni, S. Microscopic theory and quantum simulation of atomic heat transport. *Nat. Phys.* **12**, 80–84 (2016).
109. Song, J. & Curtin, W. A. Atomic mechanism and prediction of hydrogen embrittlement in iron. *Nat. Mater.* **12**, 145–151 (2013).
110. Arsenlis, A. et al. Enabling strain hardening simulations with dislocation dynamics. *Model. Simul. Mater. Sci. Eng.* **15**, 553–595 (2007).
111. Pasquarello, A. & Car, R. Identification of Raman defect lines as signatures of ring structures in vitreous silica. *Phys. Rev. Lett.* **80**, 5145–5147 (1998).
112. Puzder, A., Williamson, A. J., Grossman, J. C. & Galli, G. Surface chemistry of silicon nanoclusters. *Phys. Rev. Lett.* **88**, 097401 (2002).
113. Pickard, C. J. & Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B* **63**, 245101 (2001).
114. Golze, D., Dvorak, M. & Rinke, P. The GW compendium: a practical guide to theoretical photoemission spectroscopy. *Front. Chem.* **7**, 377 (2019).
115. Neaton, J. B., Hybertsen, M. S. & Louie, S. G. Renormalization of molecular electronic levels at metal-molecule interfaces. *Phys. Rev. Lett.* **97**, 216405 (2006).
116. Thygesen, K. S. & Rubio, A. Renormalization of molecular quasiparticle levels at metal–molecule interfaces: trends across binding regimes. *Phys. Rev. Lett.* **102**, 046802 (2009).
117. Zhang, H. et al. Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 with a single Dirac cone on the surface. *Nat. Phys.* **5**, 438–442 (2009).
118. Vanderbilt, D. *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators* (Cambridge Univ. Press, 2018).
119. Resta, R. & Vanderbilt, D. *Theory of Polarization: A Modern Approach* 31–68 (Springer, 2007).
120. Resta, R. Theory of the electric polarization in crystals. *Ferroelectrics* **136**, 51–55 (1992).
121. King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Phys. Rev. B* **47**, 1651–1654 (1993).
122. Xiao, D., Shi, J. & Niu, Q. Berry phase correction to electron density of states in solids. *Phys. Rev. Lett.* **95**, 137204 (2005).
123. Thonhauser, T., Ceresoli, D., Vanderbilt, D. & Resta, R. Orbital magnetization in periodic insulators. *Phys. Rev. Lett.* **95**, 137205 (2005).
124. Gonze, X., Ghosez, P. & Godby, R. W. Density-polarization functional theory of the response of a periodic insulating solid to an electric field. *Phys. Rev. Lett.* **74**, 4035–4038 (1995).
125. Gresch, D. et al. Z2Pack: numerical implementation of hybrid Wannier centers for identifying topological materials. *Phys. Rev. B* **95**, 075146 (2017).
126. Bradlyn, B. et al. Topological quantum chemistry. *Nature* **547**, 298–305 (2017).
127. Sanchez, J., Ducastelle, F. & Gratias, D. Generalized cluster description of multicomponent systems. *Physica A* **128**, 334–350 (1984).
128. Zunger, A., Wei, S.-H., Ferreira, L. G. & Bernard, J. E. Special quasirandom structures. *Phys. Rev. Lett.* **65**, 353–356 (1990).
129. Ferreira, L. G., Wei, S.-H. & Zunger, A. First-principles calculation of alloy phase diagrams: the renormalized-interaction approach. *Phys. Rev. B* **40**, 3197–3231 (1989).
130. de Fontaine, D. Cluster approach to order-disorder transformations in alloys. *Solid State Phys.* **47**, 33–176 (1994).
131. Ozoliņš, V., Wolverton, C. & Zunger, A. Cu-Au, Ag-Au, Cu-Ag, and Ni-Au intermetallics: first-principles study of temperature-composition phase diagrams and structures. *Phys. Rev. B* **57**, 6427–6443 (1998).
132. Ceder, G., Kohan, A. F., Aydinol, M. K., Tepesch, P. D. & van der Ven, A. Thermodynamics of oxides with substitutional disorder: a microscopic model and evaluation of important energy contributions. *J. Am. Ceram. Soc.* **81**, 517–525 (1998).
133. de Fontaine, D., Ceder, G. & Asta, M. Low-temperature long-range oxygen order in $YBa_2Cu_3O_z$. *Nature* **343**, 544–546 (1990).
134. Wolverton, C. & Zunger, A. First-principles prediction of vacancy order-disorder and intercalation battery voltages in Li_xCoO_2. *Phys. Rev. Lett.* **81**, 606–609 (1998).
135. Wolverton, C. First-principles prediction of equilibrium precipitate shapes in Al-Cu alloys. *Phil. Mag. Lett.* **79**, 683–690 (1999).
136. Maddox, J. Crystals from first principles. *Nature* **335**, 201 (1988).
137. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
138. Meredig, B. & Wolverton, C. A hybrid computational–experimental approach for automated crystal structure solution. *Nat. Mater.* **12**, 123–127 (2013).
139. Bligaard, T. et al. Pareto-optimal alloys. *Appl. Phys. Lett.* **83**, 4527–4529 (2003).
140. 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).
141. Jain, A. et al. Commentary: The materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
142. Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mat. Sci.* **58**, 227–235 (2012).
143. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the Open Quantum Materials Database (OQMD). *JOM* **65**, 1501–1509 (2013).
144. Haastrup, S. et al. The computational 2D materials database: high-throughput modeling and discovery of atomically thin crystals. *2D Mater.* **5**, 042002 (2018).
145. Talirz, L. et al. Materials cloud, a platform for open computational science. *Sci. Data* **7**, 299 (2020).
146. Choudhary, K. et al. The Joint Automated Repository for Various Integrated Simulations (JARVIS) for data-driven materials design. *npj Comput. Mater.* **6**, 173 (2020).
147. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mat.* **2**, 036001 (2019).
148. Singh, A. K., Montoya, J. H., Gregoire, J. M. & Persson, K. A. Robust and synthesizable photocatalysts for CO_2 reduction: a data-driven materials discovery. *Nat. Commun.* **10**, 443 (2019).
149. Jain, A., Shin, Y. & Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **1**, 15004 (2016).
150. Hautier, G. Finding the needle in the haystack: materials discovery and design through computational ab initio high-throughput screening. *Comput. Mater. Sci.* **163**, 108–116 (2019).
151. Meredig, B. et al. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).
152. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 16028 (2016).
153. Ward, L. et al. Matminer: an open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60–69 (2018).

154. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.* **114**, 105503 (2015).
155. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
156. Faber, F. A., Lindmaa, A., von Lilienfeld, O. A. & Armiento, R. Machine learning energies of 2 million elpasolite (ABC_2D_6) crystals. *Phys. Rev. Lett.* **117**, 135502 (2016).
157. Ward, L. et al. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **96**, 024104 (2017).
158. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
159. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
160. Park, C. W. & Wolverton, C. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Phys. Rev. Mater.* **4**, 063801 (2020).
161. Aykol, M. et al. Network analysis of synthesizable materials discovery. *Nat. Commun.* **10**, 2018 (2019).
162. Hegde, V. I., Aykol, M., Kirklin, S. & Wolverton, C. The phase stability network of all inorganic materials. *Sci. Adv.* **6**, eaay5606 (2020).
163. Lindsay, L., Broido, D. A. & Reinecke, T. L. First-principles determination of ultrahigh thermal conductivity of boron arsenide: a competitor for diamond? *Phys. Rev. Lett.* **111**, 025901 (2013).
164. Li, Y., Hao, J., Liu, H., Li, Y. & Ma, Y. The metallization and superconductivity of dense hydrogen sulfide. *J. Chem. Phys.* **140**, 174712 (2014).
165. Bistritzer, R. & MacDonald, A. H. Moiré bands in twisted double-layer graphene. *Proc. Natl Acad. Sci. USA* **108**, 12233–12237 (2011).
166. Yin, M. T. & Cohen, M. L. Microscopic theory of the phase transformation and lattice dynamics of Si. *Phys. Rev. Lett.* **45**, 1004–1007 (1980).
167. Freysoldt, C. et al. First-principles calculations for point defects in solids. *Rev. Mod. Phys.* **86**, 253–305 (2014).
168. Mauri, F., Zakharov, O., de Gironcoli, S., Louie, S. G. & Cohen, M. L. Phonon softening and superconductivity in tellurium under pressure. *Phys. Rev. Lett.* **77**, 1151–1154 (1996).
169. Spaldin, N. A. & Ramesh, R. Advances in magnetoelectric multiferroics. *Nat. Mater.* **18**, 203–212 (2019).
170. de Gironcoli, S., Giannozzi, P. & Baroni, S. Structure and thermodynamics of Si_xGe_{1-x} alloys from ab initio Monte Carlo simulations. *Phys. Rev. Lett.* **66**, 2116–2119 (1991).
171. Tersoff, J. & Hamann, D. R. Theory of the scanning tunneling microscope. *Phys. Rev. B* **31**, 805–813 (1985).
172. Albrecht, S., Reining, L., Del Sole, R. & Onida, G. Ab initio calculation of excitonic effects in the optical spectra of semiconductors. *Phys. Rev. Lett.* **80**, 4510–4513 (1998).
173. Rohlfing, M. & Louie, S. G. Excitonic effects and the optical absorption spectrum of hydrogenated Si clusters. *Phys. Rev. Lett.* **80**, 3320–3323 (1998).
174. Prandini, G., Rignanese, G.-M. & Marzari, N. Photorealistic modelling of metals from first principles. *npj Comput. Mater.* **5**, 129 (2019).
175. Attaccalite, C., Grüning, M. & Marini, A. Real-time approach to the optical properties of solids and nanostructures: time-dependent Bethe–Salpeter equation. *Phys. Rev. B* **84**, 245110 (2011).
176. Kioupakis, E., Steiauf, D., Rinke, P., Delaney, K. T. & Van de Walle, C. G. First-principles calculations of indirect Auger recombination in nitride semiconductors. *Phys. Rev. B* **92**, 035207 (2015).
177. Gilmore, K. et al. Efficient implementation of core-excitation Bethe–Salpeter equation calculations. *Comput. Phys. Commun.* **197**, 109–117 (2015).
178. Calandra, M. et al. K-edge X-ray absorption spectra in transition-metal oxides beyond the single-particle approximation: shake-up many-body effects. *Phys. Rev. B* **86**, 165102 (2012).
179. Alonso, J. L. et al. Efficient formalism for large-scale ab initio molecular dynamics based on time-dependent density functional theory. *Phys. Rev. Lett.* **101**, 096403 (2008).
180. Tokatly, I. V. Time-dependent density functional theory for many-electron systems interacting with cavity photons. *Phys. Rev. Lett.* **110**, 233001 (2013).
181. Ruggenthaler, M. et al. Quantum-electrodynamical density-functional theory: bridging quantum optics and electronic-structure theory. *Phys. Rev. A* **90**, 012508 (2014).
182. De Vita, A., Galli, G., Canning, A. & Car, R. A microscopic model for surface-induced diamond-to-graphite transitions. *Nature* **379**, 523–526 (1996).
183. Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. *Phys. Rev. B* **34**, 5390–5413 (1986).
184. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (McGraw-Hill, 1989).
185. Booth, G. H., Grüneis, A., Kresse, G. & Alavi, A. Towards an exact description of electronic wavefunctions in real solids. *Nature* **493**, 365–370 (2013).
186. Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566–569 (1980).
187. Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **73**, 33–83 (2001).
188. Lathiotakis, N. et al. Density-matrix-power functional: performance for finite systems and the homogeneous electron gas. *Phys. Rev. A* **79**, 040501 (2009).
189. Ferretti, A., Dabo, I., Cococcioni, M. & Marzari, N. Bridging density-functional and many-body perturbation theory: orbital-density dependence in electronic-structure functionals. *Phys. Rev. B* **89**, 195134 (2014).
190. Gatti, M., Olevano, V., Reining, L. & Tokatly, I. Transforming nonlocality into a frequency dependence: a shortcut to spectroscopy. *Phys. Rev. Lett.* **99**, 057401 (2007).
191. Vanzini, M. et al. Design of auxiliary systems for spectroscopy. *Faraday Discuss.* **224**, 424–447 (2020).
192. Runge, E. & Gross, E. K. U. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **52**, 997–1000 (1984).
193. Petersilka, M., Gossmann, U. J. & Gross, E. K. U. Excitation energies from time-dependent density-functional theory. *Phys. Rev. Lett.* **76**, 1212–1215 (1996).
194. Lee, C.-C., Hsueh, H. C. & Ku, W. Dynamical linear response of TDDFT with LDA+U functional: strongly hybridized frenkel excitons in NiO. *Phys. Rev. B* **82**, 081106 (2010).
195. Caruso, F., Verdi, C., Poncé, S. & Giustino, F. Electron–plasmon and electron–phonon satellites in the angle-resolved photoelectron spectra of *n*-doped anatase TiO_2. *Phys. Rev. B* **97**, 165113 (2018).
196. Hedin, L. New method for calculating the one-particle Green's function with application to the electron gas problem. *Phys. Rev.* **139**, A796–A823 (1965).
197. van Schilfgaarde, M., Kotani, T. & Faleev, S. Quasiparticle self-consistent GW theory. *Phys. Rev. Lett.* **96**, 226402 (2006).
198. Schindlmayr, A., García-González, P. & Godby, R. W. Diagrammatic self-energy approximations and the total particle number. *Phys. Rev. B* **64**, 235106 (2001).
199. Del Sole, R., Reining, L. & Godby, R. W. GWΓ approximation for electron self-energies in semiconductors and insulators. *Phys. Rev. B* **49**, 8024–8028 (1994).
200. Guzzo, M. et al. Valence electron photoemission spectrum of semiconductors: ab initio description of multiple satellites. *Phys. Rev. Lett.* **107**, 166401 (2011).

Acknowledgements

We acknowledge all the students, researchers and colleagues that developed the theories, algorithms and codes underpinning the research sketched here and that could not always be explicitly cited, but are indirectly present through the references. We are grateful for support from the Swiss NSF for the National Centre for Competence in Research MARVEL on 'Computational Design and Discovery of Novel Materials' (N.M.), from the EU Commission for the MaX Centre of Excellence on 'Materials Design at the eXascale' under grant no. 824143 (A.F., N.M.), and from the US Department of Commerce and the National Institute of Standards and Technology as part of the Center for Hierarchical Materials Design (CHiMaD) under grant no. 70NANB14H012 (C.W.).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to N.M.

Peer review information *Nature Materials* thanks Graeme Watson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.