Title

Big Data in a Nano World: A Review on Computational, Data-Driven Design of Nanomaterials Structures, Properties, and Synthesis

Journal

ACS Nano, 16(12)

ISSN

1936-0851

Authors

Yang, Ruo Xi

McCandler, Caitlin A

Andriuc, Oxana

et al.

Publication Date

2022-12-27

DOI

10.1021/acsnano.2c08411

ACSNANO

www.acsnano.org

REVIEW

Big Data in a Nano World: A Review on Computational, Data-Driven Design of Nanomaterials Structures, Properties, and Synthesis

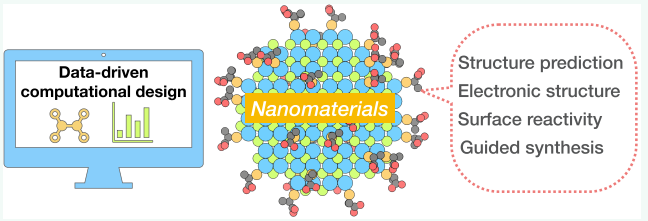
Ruo Xi Yang,* Caitlin A. McCandler, Oxana Andriuc, Martin Siron, Rachel Woods-Robinson, Matthew K. Horton, and Kristin A. Persson*

Cite This: *ACS Nano* 2022, 16, 19873−19891

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The recent rise of computational, data-driven research has significant potential to accelerate materials discovery. Automated workflows and materials databases are being rapidly developed, contributing to high-throughput data of bulk materials that are growing in quantity and complexity, allowing for correlation between structural−chemical features and functional properties. In contrast, computational data-driven approaches are still relatively rare for nanomaterials discovery due to the rapid scaling of computational cost for finite systems. However, the distinct behaviors at the nanoscale as compared to the parent bulk materials and the vast tunability space with respect to dimensionality and morphology motivate the development of data sets for nanometric materials. In this review, we discuss the recent progress in data-driven research in two aspects: functional materials design and guided synthesis, including commonly used metrics and approaches for designing materials properties and predicting synthesis routes. More importantly, we discuss the distinct behaviors of materials as a result of nanosizing and the implications for data-driven research. Finally, we share our perspectives on future directions for extending the current data-driven research into the nano realm.



KEYWORDS: *data, databases, nanomaterials, materials design, computation, electronic structure, surfaces, synthesis*

INTRODUCTION

The predictive power of quantum mechanics combined with the efficiency of density functional theory has revolutionized our ability to describe microscopic phenomena in materials. During past decades, first-principles computations have become an indispensable part of materials design, with applications ranging from energy harvesting, conversion, and storage, to quantum information and drug design.[1−5] With the advancement of modern supercomputers, computational capacity today enables high-throughput screening and data-driven in silico exploration of novel materials. At the same time, there has been a corresponding increase in scientific publications and a remarkable amount of data which fuels the application of artificial intelligence to interpret patterns, predict properties, and steer the directions of materials design. Materials science has entered the fourth paradigm: data-intensive scientific discovery.

Today, automated workflows empower the growth of materials databases, e.g., Materials Project, NOMAD, and the Open Quantum Materials Database (OQMD), and others (Table 1).[6−10] These databases have expanded their capacity from basic properties like structure, total energy, formation energy, and band structures, to evaluating phase behavior, vibrational, dielectric, elastic, and spectral properties. In turn, such data are used in data-driven approaches, where machine learning models are trained to map the composition and structure to the targeted properties. While much of the data are obtained within the mean-field approximation of density

Received: August 22, 2022
Accepted: November 8, 2022
Published: November 15, 2022

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

Table 1. General Open-Source Databases for Solids[a]

General databases	Electronic structure	Thermodynamics	Mechanical	Spectra	Synthesis/surface
Materials Project;[6] materialsproject.org;API available	Band structure, DOS, magnetic moment, charge densities, bonding and chemical environment, Fermi surfaces dielectric tensor	Phonon dispersion, Hull stability, formation energy, phase diagram, Pourbaix diagram (aqueous stability)	Elastic tensor, piezoelectric tensors, equation of states	XANES, XAFS, EXAFS, absorption spectra XRD	Surface energies, Wulff shape, suggested substrates, synthesis recipe
MPContribs(community-contributed data including experimental data); mpcontribs.org;API available	Band structure (including GW), effective masses, electronic transport, band alignment, DOS, Hubbard U/J	Formation energy, disordered structures, experimental phase and formation energies	Elastic constant		Surfaces, adsorbates
NOMAD;[7] nomad-lab.eu;API available	Band structure, DOS, eigenvalues, angular spin momentum	Phonon dispersion, heat capacity, Helmholtz free energy	Stress tensor	Oscillator strength, transition dipole moment	
AFLOW;[18] aflowlib.org;API available	Band structure, DOS, magnetic properties	Formation energy, Debye temperature, vibrational free energy, heat capacity	Bulk modulus, Poisson ratio, elastic anisotropy		
OQMD;[8] oqmd.org;API available	Band gap	Phase diagrams, formation energy			
Materials Cloud;[19] www.materialscloud.org;API available	Band structure, DOS, magnetization, Wannier functions, band inversion strength,[20] (anti)ferromagnetism	Phonon dispersion		Molecular IR/Raman spectroscopy	Exfoliation energy (2D mat.)[21,22]
NREL Materials Database;[23−25] materials.nrel.gov	Band gap (including GW), VBM/CBM, effective masses, magnetic moment	Formation energy, Hull stability		XRD	
Computational Materials Repository;[26] cmr.fysik.dtu.dk	Band structure (including GW), effective masses, polarizabilities, band alignment	Formation energy, Hull stability	Elastic constant	Absorption spectra	Adsorption energy, surface energy
JARVIS;[27] jarvis.nist.gov;API available	Band gap, effective masses, magnetic moments, dielectric constants	Formation energy	Elastic tensor, piezoelectric tensors, equation of states	Frequency-dependent dielectric function, infrared intensities, spectroscopy limited maximum efficiency, XRD	Exfoliation energies
MatCloud;[28] http://matcloud.cnic.cn;API available	Band structure, DOS, dielectric constant	Formation energy, vibrational entropy,Gibbs free energy, heat capacity	Elastic tensor, shear modulus, Young's modulus		Adsorption
MPDS[29] (including experimental data);mpds.io; API available	Band structure, DOS, effective mass, electrical conductivity, charge transfer, superconductivity, magnetic properties	Entropy thermal conductivity, Debye temperature, heat capacity, thermal expansion, Seebeck coefficient, phase transition temperature	Elastic moduli, sound velocity, compressibility	Optical absorption, work functions, refractive index	Decomposition
Materiae;[30,31] materiae.iphy.ac.cn;API available	Band structure (SOC), DOS (SOC), SOC band gap, topological class				
phonondb;[31,32] phonondb.mtl.kyoto-u.ac.jp		Phonon dispersion, phonon DOS, entropy, free energy, heat capacity, Gruneisen parameters			
MatNavi [33] (literature and experimental data);mits.nims.go.jp	Band structure, DOS, charge density	Phase diagrams (T), thermal capacity, diffusion coefficient	Tensile strength (metal), proof stress	XRD, XAS	Surface tension

[a]The list of properties intends to be as exhaustive as possible, but due to the high diversity of data, we only listed properties with more than 500 data points. Databases are marked if they include experimental results or an API (Application Programming Interface) for programmatic access.

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

functional theory (DFT), such as generalized gradient approximation (GGA) and extensions thereof, more accurate methodologies such as GW and time-dependent DFT (TDDFT) can complement the large data sets which together provide useful trends and descriptors. Robust, predictive models require a large, systematic, and diverse set of high-quality input data covering the target properties, emphasizing the urgency and importance of curating databases containing such information. However, the majority of data-driven explorations have focused on crystalline bulk materials or molecular crystals, where periodic DFT codes run efficiently.

Owing to the rapid development in materials synthesis and imaging techniques, materials at the nanometer scale—including quantum dots, atomically thin nanoplates, nanowires, and nanocrystalline thin films (Figure 1)—have generated

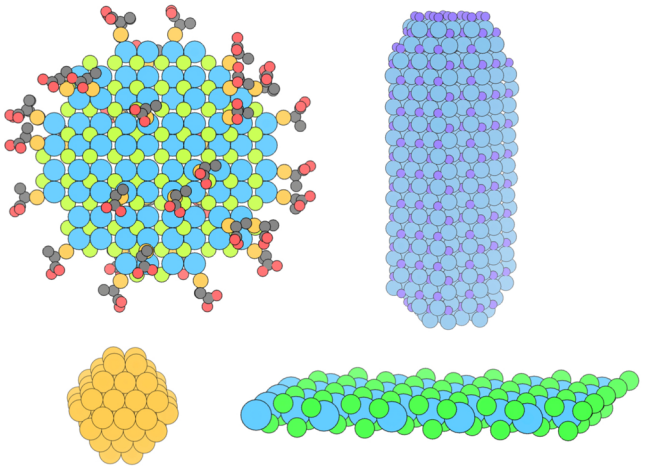


Figure 1. Common model nanomaterials. Top left, nanocrystals with ligands; top right, nanorods; bottom left, nanoclusters; bottom right, 2D nanoplatelets.

tremendous excitement and are impacting a range of technological applications including optoelectronics, catalysis, and energy storage. While such nanomaterials, spanning 1–100 nm, can offer precise control[11] over composition, size, and shape, they also exhibit large surface areas, a range of morphologies, structural reconstruction, defects, and associated quantum confinement effects which result in behaviors distinct from their bulk counterparts.[12−17] The increased complexity, as well as the lack of long-range order, has posed challenges for traditional first-principles methods. However, recent data-driven approaches have proven effective in navigating the vast parameter space, aiding structure determination and property predictions. In this review, we discuss the progress on functional nanomaterials design, including structure, properties, and synthesis, specifically focusing on data-intense research. We summarize the commonly used computational methods, high-throughput studies, descriptors, and machine learning approaches, discuss the challenges brought by the complexities of nanomaterials, and comment on the opportunities for future direction.

DATA-DRIVEN DESIGN OF NANOMATERIALS

Nanostructured materials, including nanocrystals, nanorods, nanoplates, nanoclusters, and nanocrystalline thin films, have gained prominence in recent years due to rapid advancements in materials synthesis techniques. This class of materials offers great physical and chemical tunability with enhanced performance across systems ranging from inorganic semiconductors to metals and molecular crystals.[34−36] As a result of the large surface area-to-bulk ratio as well as strong interactions with the external environment,[37,38] nanomaterials, as compared to the bulk, are more prone to defects and surface reconstruction.[39,40] To passivate the material, surface surfactants and ligands as well as core−shell architectures are employed; however, these techniques in turn introduce additional complexity as well as tunability. For traditional first-principles modeling and associated property screens, where the cost of computation scales rapidly with the number of atoms necessary to describe the material, such chemical and structural complexity presents challenges.

While periodic DFT codes perform efficiently on reasonably well-ordered bulk materials with up to hundreds of atoms/unit cell, many nanomaterials require a much larger number of atoms to account for, e.g., hybrid, defective, and disordered structures, as well as complex surface chemistry, depending on the details of the environment (pH, chemical potential, salts, and solvent).

In the following, we discuss the current status of data-driven approaches to predicting materials structure and functionality, specifically focusing on bridging the gap between bulk to nanomaterials (Figure 2).

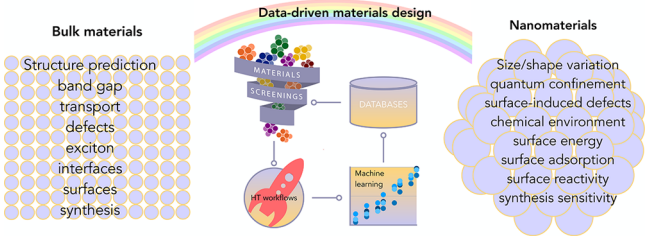


Figure 2. Properties relevant for general bulk materials and additional properties that are unique to nanomaterials as a result of nanosizing. We note that the properties in bulk are also native to the nanomaterials. The middle panel is the data-driven methods including database query, materials screening, HT workflows, and machine learning.

Nanomaterials Structure Prediction. In nanomaterials, the size and morphology provide extra degrees of freedom that add tunability of properties as well as complexity of the available phase space. To explore and direct synthesis efforts in this phase space, computational structure prediction can provide a map of stable, low-energy synthesizable configurations.

Nanoparticles are one class of nanomaterials that have a wide range of shapes and sizes. Nanoclusters are a subclass of nanoparticles, delineated as such due to their ultrasmall sizes. Nanoclusters comprise up to 100−150 atoms (<1 nm), while nanoparticles generally refer to larger structures between 1 and 100 nm. This distinction originates from the irregularity of nanoclusters, which have stronger quantum confinement effects, off-lattice atomic arrangements, variable bond lengths, and more diversity in their resulting properties. On the other hand, nanoparticles typically exhibit atomic configurations and symmetries that resemble the bulk crystal structure, which in turn correlates with more predictable properties.

Identifying stable nanocluster geometries computationally is important for comparing to experimental measurements and

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

directing synthesis efforts; however, the time and length scales will generally determine the level of theory that is computationally tractable. Nanoclusters are small enough to be calculated with DFT, including dispersion corrections, hybrid exchange−correlation functionals, spin−orbit coupling, implicit/explicit solvation, and stabilizing ligands for increased accuracy. Typically, however, DFT is too expensive to consider for calculating a nanoparticle, an array of nanoclusters, or a simulation over many time steps. For this reason, less computationally expensive methods are typically used to estimate the structure and energies, including semi-empirical tight binding methods or interatomic potentials.[41]

The potential energy surface (PES) of a nanostructure dictates the global minima, i.e., the most stable atomic configurations given the composition. Sampling the entire PES via first-principles methods in the complex geometry is computationally infeasible; hence, a number of structure generation methods have been developed. The major techniques to generate plausible structures are genetic algorithms,[42] basin hopping,[43] random structure searching,[?] and simulated annealing.[44] Less standard methods include particle swarm algorithms[44] and other variants of these broad methods including tree growth algorithms,[45] Tabu ("taboo") algorithms,[46] and ab initio random structure searching (AIRSS).[47] Several reviews have previously highlighted these techniques and their advantages for certain systems.[48−50] Furthermore, some of these algorithms have been adapted to screen for properties beyond thermodynamic stability, such as vertical electron affinity.[51,52]

An efficient method to rapidly generate structures is to simply reuse and recalculate structures that were previously found for similar elements. As an example of this approach, Sokol et al. conducted a data mining study with the idea that binary heteropolar materials have similar energy orderings to one another and showed that the low-energy structures of $(ZnO)_n$ clusters exhibit similar energy orderings when applied to other binary heteropolar materials such as $(AlN)_n$.[53] The tendency of certain elements to behave similarly at the nanoscale can be quantified by comparing the relative ordering of stabilities for structures with each element. The similarity of different elements at the nanoscale was quantified by comparing the energy rankings of clusters with similar atomic arrangements at the nanoscale for all elements in the Quantum Cluster database.[54] A study conducted by Chaves et al. is another example of a systematic study of transition metal nanoclusters, where they used structures from literature as a starting point and then conducted a Revised Basin Hopping Monte Carlo structure search.[55−57]

An interatomic potential (IAP) is an empirical parametrization of a potential energy surface that can be fitted to nanoparticles. Recently, machine-learned potentials (MLPs) specifically applied to nanoparticle systems have become increasingly sophisticated and can rival traditional atomistic methods. Machine learning potential regression methods bring benefits over more traditional physics-based methods in that the models can be built agnostically to the material, applied to a diverse set of input systems, and have been shown to improve the topology of the PES.[41,58,59] It is important to note that, like any parametrized model, MLPs are fundamentally limited by the quality of their training data. Many, on the order of 50000, first-principles calculations of nanoclusters, including their structures, energies, and forces, are required to fit IAP models with very good accuracy,[59,60] with the possibility of reducing that number to ∼500 calculations by using active learning that iteratively queries the energy landscape in regions of failure and updates the training pools.[61] IAPs excel at small structures for which we can sample many configurations, and large structures in which bulk-like behavior and surface facets dominate. On the other hand, medium-sized structures, on the order of a few hundred atoms large, are challenging because they fall between the two regimes—they are typically too large for first-principles approaches and too disordered to model as a bulk crystal with surface facets. Bridging the data gap between these two length scales will be the key to future developments in nanoparticle structure prediction.

Cluster expansions, methods typically applied to bulk crystals with well-defined lattices, have emerged as a possible approach to rapidly estimate the energies of nanoparticles. Cluster expansions have effectively modeled the surface and adsorbates of nanoparticles.[62,63] An off-lattice cluster expansion formalism, the Atomic Cluster Expansion (ACE), has recently been developed by Drautz.[64] PACE, the performant open-source implementation of ACE, has exhibited comparable accuracy and performance for small Cu and Si clusters as other MLPs.[60] Other equivariant MLPs that have recently emerged and show excellent accuracy include the Neural Equivariant Interatomic Potential (NeQuip),[65] Allegro,[66] and MACE[67] (a message-passing variant of ACE).

We emphasize the usefulness of diligent data reporting standards, as combining data from several individual studies can provide a superior starting point for further development by furnishing informed guesses of stable structures when considering a new system. Curating quality data on the energies and geometries of nanoparticles will act as innovation multipliers across the broad range of nanomaterial design efforts. Furthermore, the record of prior tuned/trained IAP speeds up the development of potentials, thus improving the efficiency of structure prediction. We note here the example of OpenKIM and NIST interatomic potential repositories, which have archives of previously fitted potentials for many systems.[68−70] Databases such as these provide excellent avenues to build on existing work and close the gap from bulk to nanostructures.

DESIGN METRICS OF MATERIALS: BULK TO NANOMATERIALS

In searching for materials with favorable properties for a specific application, generally, multiple metrics are of importance. A consideration of computational cost and robustness/accuracy is optimized typically through a funnel-type investigation where multiple filtering steps are needed, applying increasingly strict criteria and/or computationally intense properties toward the end. In the following, we discuss several commonly used criteria used in data-driven searches for materials in applications where nanosizing is prevalent (Figure 4). We note the use of nanocrystals (NCs) as model nanomaterials throughout the discussion.

Stability. Stability commonly provides the first screening metric in high-throughput (HT) materials design and can be employed for both bulk and nanomaterials. Relevant quantities for stability include formation energy ΔH_f and E_{hull} (energy above convex hull) where the formation energies ΔH_f are calculated using elemental chemical potentials to determine if the compound is favored to form from the constituent elements. In contrast, the energy above hull (E_{hull}) determines the distance between the energy of the compound and the

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

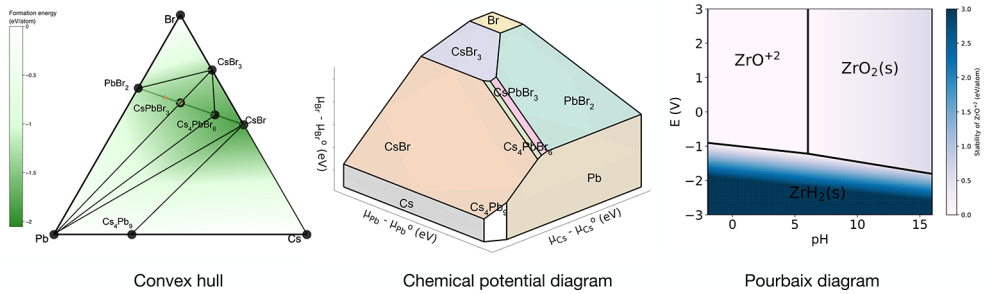


Figure 3. Diagrams for finding stable compounds. The convex hull diagram denotes the relative energy (shown as a color map) of a composition with respect to the lowest energy boundary of other phases or linear combinations of those phases. The chemical potential diagram looks at the same phase stability by tuning the chemical potential of the constituent elements, indicating the relative stability of a phase at a certain chemical environment. Both the convex hull and chemical potential diagram are generated for $CsPbBr_3$ using the Materials Project phase diagram tool. The Pourbaix diagram is a way to identify stable nanostructures in an aqueous environment as the pH and voltage vary (example generated using the Pourbaix diagram app on Materials Project for Zr).

geometric, bounded shape consisting of all phases within the relevant chemical space that exhibit energy lower than any other phase or linear combination of phases at the respective compositions (Figure 3). E_{hull} is a more demanding metric than the formation energy and hence is often used as a descriptor for stability against decomposition to other phases.[71−73] A similar metric to E_{hull} is the decomposition enthalpy ΔH_d to measure the stability against other phases.[74] Phase transitions into alternative structures at fixed composition can be elucidated via the occurrence of phonon imaginary frequencies,[75] and solid−solid polymorphic phase transition may occur when the external conditions enable a low activation energy barrier.[76,77]

While these quantities provide a basic understanding of the energetics, they do not always translate to synthesizability. DFT calculations are carried out at 0 K, and at elevated T, vibrational and configurational entropy contribute toward the free energy. Recently, Aykol et al.[78] showed that synthesis of any material, nanometric or bulk, is subject to the enthalpy limit of the analogous amorphous state, a useful result as many nanomaterials exhibit high enthalpies of formation. Indeed, the prolific existence of diverse carbon nanostructures is enabled by the abnormally high enthalpy of amorphous carbon, allowing for a large window of metastability.

Since nanocrystals possess large surface-to-volume ratio, specific surface facets relevant to the morphology, size, and polymorph will impact stability as well as functionality,[17,79−81] and large surface areas lend themselves to the stabilization by a chemical environment (Figure 3), e.g., chemical potential, surface reconstruction, and/or ligation. For instance, Singh et al. used Pourbaix diagrams as a filtering tier for photocathodes, removing over 95% of the materials under consideration, thus highlighting the importance of including aqueous stability metrics (Figure 3).[73,82,83] Finally, while entropic differences between bulk solids are often neglected, increased defect populations and various atomic configurations in nanostructures can contribute to significant entropy effects, leading to improved thermodynamic stability.[84,85] Overall, the stability of nanomaterials is closely related to the size, shape, and surface chemistry and must be understood for individual materials. To the best of our knowledge, HT studies for the stability of nanostructures are rare.[86] As a step in the right direction, the construction of phase diagrams, including a chemical environment and the corresponding surface energies of the underlying bulk materials,[17] will be highly beneficial for the initial screening.

Band Gap. In optoelectronics and photocatalysis applications, the ideal band gap is one of the most determining descriptors for functionality. For solar cells, in particular, the thermodynamic limit states that the theoretical efficiency of a light absorber reaches its maximum at a band gap of 1.3−1.4 eV.[87] A direct band gap is commonly favored over an indirect band gap for more efficient absorption or emission as it does not require phonons to compensate for the momentum transfer. In other optoelectronic applications, band gap requirements can differ widely; e.g., a band gap across the whole visible light range (1.65−3 eV) is considered desirable for LEDs, while, for transparent conductors, band gaps larger than the energy of visible light threshold (3 eV) are preferred.[88,89]

HT computational studies generally use band gaps calculated by the efficient and low-cost functional GGA-PBE as the first step, with subsequent corrections to account for the underestimation of E_g,[90,91] or followed by additional screening of more expensive hybrid functional calculations to narrow down the candidates.[71,72,92,93] While hybrid functionals or GW-theory-based approaches offer improved accuracy, the computational cost is orders of magnitude higher and requires expert knowledge to converge co-dependent parameters. While some work has used GW band gaps either as the middle or the last step in the screening process, there are—to the best of our knowledge—no robust automated workflows.[94,95] Recently, van Setten et al. identified a correlation between the number of empty bands and energy cutoffs in the self-energy, and a better description of the experimental band gap by the G_{00}^{GN}@PBE (with plasmon-pole approximation) than the linear model.[96] Rasmussen et al. further investigated numerical methods to correct the error in the linearized quasi-particle self-energy.[97] Both these efforts toward an HT GW workflow stress the importance of an accurate starting point at the Kohn−Sham energy.

With large amounts of theoretically calculated band gaps, machine learning help elucidate the structure−property relation. Emphasizing the importance of physically meaningful descriptors, various studies[98−100] have shown excellent mapping of bulk band gaps onto compositional representations for the materials. Single-value scalars such as band gap, ionization energy, and electron affinity are commonly used descriptors,[101] and even the densities of states have been used

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

as a complex input feature for mapping the relationship between materials and their electronic structure.[102−104]

In nanostructures, increased band gaps are expected as a result of quantum confinement,[105] which follows the relation of $\Delta E_g \sim 1/R^2$ for an infinite well (Figure 4).[106−108] Different

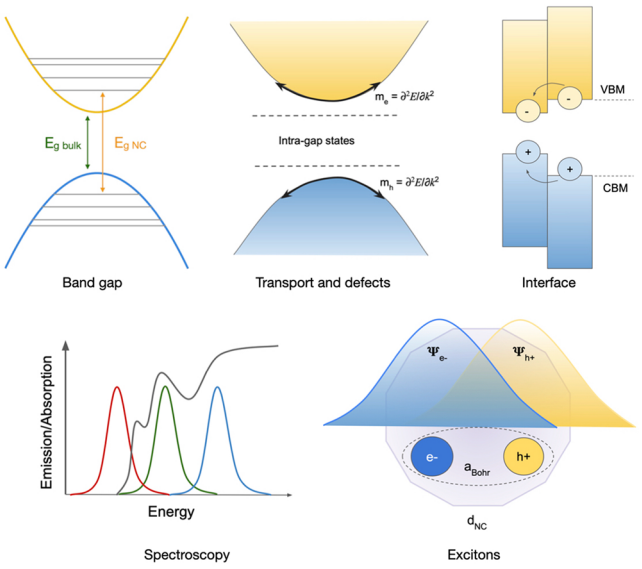


Figure 4. Schematics of relevant design metrics in electronic structure screening of solid materials. E_g stands for band gap, and NC stands for nanocrystals. m_e and m_h are the effective mass of electrons and holes, respectively. VBM and CBM refer to the valence band minimum and conduction band maximum. Ψ_e and Ψ_h refer to the electron and hole wave functions, respectively; and a_{Bohr} is the Bohr radius of the exciton. d_{NC} stands for the nanocrystal size.

schemes are applied depending on whether the regime of confinement is weak, intermediate, or strong, as the size of the NC approaches the size of the exciton Bohr radius, resulting in different relations of ΔE_g with respect to the size.[16,17,109−111] The dimensionality of the nanostructure also affects how the band gap behaves with regard to size. For example, in 2D materials, the band gap is confined in the out-of-plane direction but remains bulk-like in-plane, offering an extended design space.[112] These works show the value of developing physics-based descriptors that connect bulk properties to the specifics of the nanocrystal symmetry and dimensionality.

Transport. The flow of charged carriers are crucial in determining the efficiency of an operating semiconductor (Figure 4). The quantum mechanics of carrier mobility are described by the transport equations with different levels of approximations from semiempirical to fully first-principles.[113−119] In the simplest Drude model, the mobility is inversely proportional to the effective mass: $\mu = \frac{q\bar{\tau}}{m^*}$, where q is the charge, $\bar{\tau}$ is the mean free time, and m^* is the effective mass.[120] While multiple scattering mechanisms, including phonon, impurity, defect, and surface scattering, are responsible for slowing down the mean free time, effective masses are a key element in determining the intrinsic carrier mobilitiy, μ. The effective masses of electrons and holes in semiconductors are derived from the dispersion of the electronic structure, precisely, the inverse of the second derivative of the energy with respect to wavenumber, e.g., the effective mass of the carriers, m_e^* and m_h^*, for electrons and holes, respectively: $\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}$, in the parabolic approximation. In a charge-based device, m^* should ideally be small as light carrier mass facilitates carrier mobility, and thus more efficient carrier extraction in the case of a light absorber, or fast recombination in the case of a light emitter. In thermoelectrics, small m^* and band degeneracy are favored to deliver a high power factor.[88,121,122] Many studies have considered effective mass in the screening steps as an additional filter to band gap, with $m^* \leq 0.1m_0$, in which m_0 refers to the free electron mass.[88,92,93]

In nanostructures where quantum confinement effects are present, the effective masses derived from the bulk band structures can be inaccurate (or undefined) due to the modified band structure.[123] When the crystal size is limited, the electron (hole) wave functions are confined within the crystal boundaries and the effective mass becomes energy-dependent[124,125] As a general trend, the effective mass is expected to increase as crystal size decreases due to reduced dispersion at band edges; however, as shown in a literature comparison of ZnS NCs, this is not always the case experimentally due to crystal orientations, impurities, and other factors.[123] Carrier mobilities have been shown to decrease with NC size,[126] most likely due to enhanced surface scattering; in these cases, surface passivation is useful to reduce carrier trapping.[127] It is also noted that many thin films grown with low-cost methods introduce nanocrystalline domains or grains, both within the film and at surfaces or interfaces, and the nanostructure of these films can dominate properties such as electronic transport and optical absorption. For example, in many nanocrystalline thin films (e.g., Ag,[128] CdSe,[129] and ZnO,[130] among others) conductivity is decreased from that expected in a bulk material due to variable range hopping, in which thermally activated charge carriers "hop" between localized states in different crystalline domains.[131,132] Hence, multiple factors contribute to carrier scattering in nanomaterials; therefore, the effective masses derived from the parent bulk can only be used as an estimation rather than a prediction for the final transport.

Defects. While the electronic structure of the parent crystal serves as the foundation for understanding the electronic behavior of materials, defects should not be overlooked, as they are ubiquitously present in any material—both bulk and nanomaterials (Figure 4). Point defects, including vacancies, substitutional defects, and impurities can introduce charge transition levels either within the band or in the gap. The latter is considered a "deep" defect that acts as a recombination center that traps electrons or holes, hindering the energy conversion process, and hence deteriorating the device performance. To determine the configuration, concentration, and charge transition levels of these defects present by no means trivial tasks, requiring large supercell calculations to simulate the dilute limit.[133] In the case of charged defects, various charge states need to be considered for which additional corrections are needed to account for the periodic image effect.[134,135] Due to the increased complexity of such calculations, defect levels are usually present at the last step of the high-throughput screening projects.[93,94,136] Computationally less demanding methods such as the atomic-orbital-based method (LCAO) was used by Kuhar et al. to investigate the electronic structure, and specifically, the density of states (DOS), in the presence of vacancies. Compounds with mid-gap energy states were ruled out, aiming at filtering out those with detrimental intrinsic defects.[93] Another study by Hinuma

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

et al. examined more specifically the formation energy diagram in crystalline Ca_2ZnN_2, $CaZn_2N_2$, and $CaMg_2N$ at the last step of the screening process and provided guidance for synthesis conditions to achieve favorable defect concentrations. Kumagai et al. developed codes for HT point-defect calculations to characterize oxygen vacancies in oxides and were able to identify electronic structure characteristics that affect the defect formation energy.[136] We note that high-throughput descriptors such as the hydrogen interstitial defect energy (i.e., the "hydrogen descriptor")[137] and the branch point energy[138−140] have been proposed to assess doping type in semiconductors.

There has been excellent recent progress in the automation of defect calculations. Several packages provide point-defect calculation workflows, for instance, PyCDT, PyLada and PyDEF, which enable the automatic calculation of defect formation energy with image charge corrections.[141−143] With the help of these computational frameworks, broader and systematic defect calculations are made possible, which in turn enables future databases of charged defect properties across chemical systems and structures. Collection of defect information will lead to improved understanding and trends of how to rationally design functional materials, e.g., in optoelectronics and quantum materials applications. The Quantum Point Defect (QPOD) database is an example that contains more than 1900 defect systems comprising various charge states of intrinsic point defects in a set of 2D semiconductors and insulators.[144]

For NCs, the large surface-to-bulk ratio and often non-stoichiometry composition further induce more defects in addition to the ones existing in bulk, where trap states are often formed within the band gap.[37,38,145] It is nontrivial to directly compare the defect formation energies calculated for the bulk to nanostructures, but we note a few key differences. First, unlike bulk materials, the defect formation energies in NCs are location-dependent, where the defect concentrations are likely higher on the surface than in the core.[146] This means that the defect levels may be tuned by merely changing the size of the NCs.[147] Second, dangling bonds on the surface introduce localized surface states that are often deep traps, which are not captured by the bulk defect calculation.[37,148,149] Third, the formation of NCs typically happens in colloidal solutions, meaning the chemical potentials may be considered as ion pairs instead of solid precipitates.[38] These different behaviors present challenges for directly using bulk defect energetics, and additional considerations for surface defects must be given as materials approach the nanoscale.

Exciton Binding Energy. Another design-relevant electronic structure property is the exciton binding energy, E_b, which is the minimum energy required to separate an electron and hole pair. Depending on the application, the ideal E_b differs. While a small E_b is beneficial for the separation of electron−hole pairs in a solar cell for efficient carrier extraction,[150] a large E_b may be preferred for a larger quantum yield in a light-emitting diode.[151] Formally computing the exciton energy involves calculating the bound states of the electron−hole pair via Bethe−Salpeter equation (BSE), a computationally demanding method that is challenging to automate. In the hydrogen-like Wannier−Mott approximation assuming band edge excitation, the exciton binding energy is simplified as $E_b = \frac{\mu R_y}{m_0 \epsilon^2}$, where R_y is the Rydberg constant, μ is the reduced mass of the electron and hole $(m_e^* m_h^*)/(m_e^* + m_h^*)$, and ϵ is the dielectric constant (often ϵ_∞, the high-frequency dielectric constant, is used). The exciton radius equals $m_0 \epsilon a_H/\mu$, where a_H is the hydrogen Bohr radius. The quantity ϵ, including ϵ_0 (static dielectric constant) and ϵ_∞, can be obtained from density functional perturbation theory (DFPT) calculation. Currently, ϵ calculated for >7000 materials are tabulated on Materials Project, allowing for analyses of bulk trends over structure and composition.[152] Furthermore, optical spectral information, e.g., photoluminescence emission and UV/vis absorption spectra, are often used for inferring the exciton binding energy. Computational spectra that include many-body interaction (GW plus BSE) generally compare well with experimentally measured spectra for semiconductors, where the energy difference between the absorption onset and electronic band gap determines the exciton binding energy: $E_b = E_g - E_{opt}$.[153−158] Study by Yang et al. also showed that the band-to-band vertical optical transitions obtained at the independent-particle approximation with an HSE band gap correction compare well with experiments, providing a good starting point for accurate high-throughput optical spectra.[159]

Transferring trends in exciton binding energy between bulk and nanosized materials should be exercised with caution. In structures that are smaller than the Bohr radius of the exciton, the wave functions of the electrons and holes become spatially confined, and thus the overlap is strengthened, leading to increased exciton binding energy (Figure 4).[16,160−162] Furthermore, the exciton fine structures caused by exchange interaction depends closely on the size and shape of the crystals,[160,163,164] and can reverse the energy order of the dark/bright states.[165] Notably, the dielectric constant also changes due to reduced screening in the finite-size system, resulting in enhanced exciton binding energy.[166]

Interfaces. Interfacial alignment in nanomaterials has a fundamental impact on materials performance and device functionality; however, the electronic structure of an interface is intimately connected to the stability of a given nanomaterial with respect to its external environment. One simple yet useful metric of thermodynamic and chemical stability in bulk interfaces is the Materials Project's Interfacial Reaction Calculator,[6,167] in which "reactants" are the materials in contact with one another. The reaction energy descriptor E_{rxn} indicates whether a stable solid−solid interface is favorable to form under canonical thermodynamic conditions, as well as the degree of instability of any unintentional product. This descriptor has been predictive in polycrystalline and nanocrytalline thin film interfaces such as $CdTe-Cu_xZn_{1-x}S$,[168] determining an unstable interface and accurately predicting the products observed experimentally, and has promise for assessing the stability of other nanocrystalline interfaces.

In electronic devices, sufficient alignment of band extrema (CBM and VBM) between neighboring materials is crucial to ensure efficient charge transfer (Figure 4). In semiconductor heterojunctions and metal−semiconductor junctions, the relative energies of the CBM, VBM, and work functions dictate whether electrons and holes can flow efficiently across the interface. For instance, for a water splitting reaction to occur, the CBM should be higher than the proton reduction potential and VBM lower than the oxidation potential of water.[169,170]

DFT slab calculations are often used to compute band offsets; however, the approach comes with caveats that are exacerbated in nanomaterials.[171−173] Since detailed knowledge of chemical as well as structural features of interfaces is rarely

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

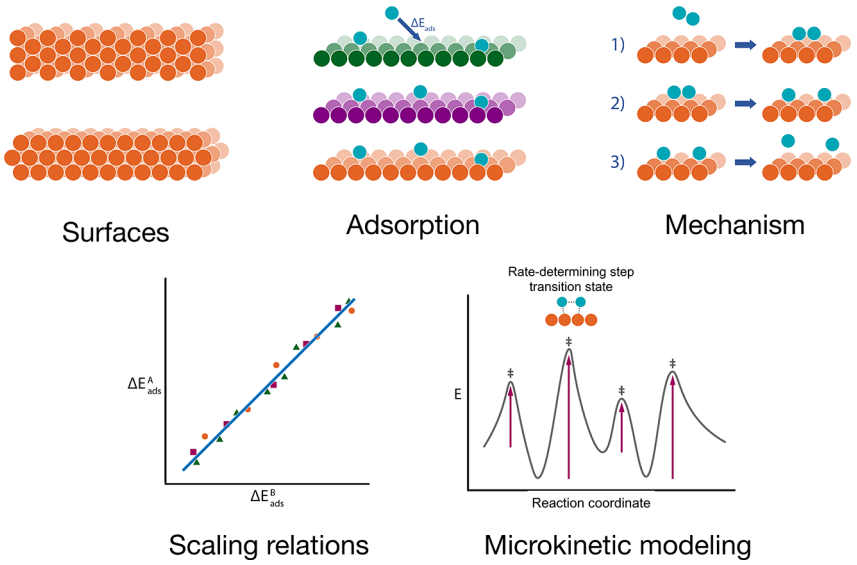


Figure 5. Schematics of relevant phenomena and modeling approaches for surface reactivity: identifying relevant surfaces, adsorption sites, and energies; elucidating reaction mechanisms; scaling relations in adsorption energy; and microkinetic modeling by investigating the transition states.

available, results based on explicit first-principles calculations should be interpreted with caution, employing reliable reference states. Within DFT, the absolute values of the eigenstates should not be directly compared as they depend on the pseudopotentials, basis sets, and functionals. The vacuum should be the common reference level to compare across different codes and databases.

Band alignment in real materials is surface-specific and hence highly dependent on the facet of the exposed surface, the surface morphology, and interfacial defects; however, there is not yet a consensus among the DFT community about how to treat nanocrystalline thin film interfaces.[174] Additionally, in solid−liquid interfaces the direct comparison of reference levels ignores the fact that the surface dipoles between the semiconductor and liquid can impact the band edges and qualitatively alter the comparison with experimental results.[175] State-of-the-art computational frameworks have recently been developed to take into account the aqueous−solid interface to obtain the correct electrostatic alignment.[176,177] However, these approaches rely on a case-by-case examination of a specific system, hence difficult to apply in HT workflows for systematic data production.

Surface Adsorption. The adsorption of a molecule to a solid surface is relevant for many applications including catalysis and sensors, thus providing one of the most widely used descriptors of surface reconstruction, reactivity, and catalytic activity. The adsorption energy is essential in determining the mechanism of formation and breaking of the bonds, and it depends on the geometries as well as on the electronic properties of the surface and adsorbates (Figure 5).[178] Therefore, mapping out the adsorption energy for different sites and adsorbates using computational methods provides rich information on the geometrical and chemical determinants for active adsorption sites. Indeed, we have seen a steady growth of HT workflow development and data of adsorption energy for transition metal surfaces for catalyst design.[179−181]

Standardized and reproducible workflows can expedite the screening. An HT workflow for computing adsorption energies on solid surfaces was developed by Montoya et al.[182] using the codes Atomate and Pymatgen.[6,183−185] The workflow allows for (i) generation of slabs corresponding to all of the unique terminations in a bulk, (ii) identification of adsorption sites using a triangulation network of surface sites, and (iii) DFT geometry optimizations with automated error corrections[6,185] for all of the empty slabs and adsorption structures. Adsorption energies computed using the workflow with the revised PBE (RPBE) functional[186] for a series of transition metals showed close quantitative agreement with experimental values.

Inorganic semiconductors are widely used as photocatalysts, and their surface chemistries are more complex than metals due to less homogeneous electronic density at the surface, dipole formation, and strong site dependence. An automated DFT adsorption workflow was developed by Andriuc et al.[187] This workflow, while building on the previous Atomate version,[182] features computational improvements in efficiency and robustness using relaxed structures, frozen core slabs, and a slab−adsorbate distance optimization step. The computed metrics include semiconductor-specific properties for photocatalysis (e.g., the p-band center, or the elemental makeup of the conduction band minimum and the valence band maximum).

Compared to inorganic systems, hybrid materials often exhibit more complex structures, with nonhomogeneous electron densities at surfaces that require more careful handling. An HT DFT framework for screening metal−organic frameworks (MOFs) was developed by Rosen et al. for heterogeneous catalysis applications, where two different methods are used to identify adsorbate sites: a geometry-based one meant to identify adsorption sites on under-coordinated atoms and a potential energy grid method designed to find low-energy adsorption sites.[188]

Despite these advances, challenges remain pertaining to (i) rotational degrees of freedom for larger, more complex adsorbate molecules, (ii) concentration-dependent behavior associated with the adsorbate coverage, and (iii) the high computational cost of DFT for larger systems.

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

To accelerate learning at a reduced cost, machine learning methods are being rapidly developed.[189−194] For instance, Tran et al. developed GASpy (Generalized Adsorption Simulator for Python), a dynamic workflow for DFT adsorption data.[195] Their "smart database" approach implemented surrogate-based optimization and active machine learning concepts which allow for the exploration of a broader chemical space at a reduced cost. Each adsorption site was represented by a vector of the atomic number, Pauling electronegativity, coordination number, and median adsorption energy of the adsorbate on the pure element.[195,196] Among other noteworthy data resources, The Open Catalyst Project[197] was designed as a community effort toward developing machine learning models. The database (OC20) contains 1.2 million DFT relaxations of adsorption structures built from different adsorbates and surfaces of metallic materials, including relaxation trajectories, charge analyses, and orbital information. Using these data, including randomly perturbed structures, the Project aims to use machine learning to replace or augment DFT simulations. Notably, the performed baseline graph neural network (GNN) models would require 10 orders of magnitude larger data set to achieve the desired accuracy in predicting the total energy against DFT, proving the necessity of both increasing the data quantity and improving the physics-informed representations.[197] The recently released OC22 data set, including more than 60*K* oxides calculated by DFT, improved in accuracy in the prediction of total energies and forces by joint training of the OC20 data set, suggesting the importance of including a variety of data sets of complex systems.[198]

These challenges call for further development in building comprehensive data sets: the surface energy and adsorption energy are the most essential information in determining the exposed facets, the shape of the nanocrystal, and the possible active sites, providing valuable descriptors for compositions in training ML models. Standardized workflows and data sets for these properties will be instrumental.

Surface Reactivity. Surface reactivity can be described by microkinetic modeling (Figure 5), which requires the thermodynamic properties of absorbed species and the rate constants.[199−201] The rate of surface-mediated bond disassociation rely on computing adsorption and transition-state energies, a computationally demanding parameter if using traditional quantum chemical methods. These energies then propagate in physical models to predict the rates of reactive events. The large parameter space involved in these processes can be reduced to low-dimensional "descriptor space" via scaling relations of the adsorption energy, thereby enabling HT screening based on a few metrics such as adsorption energies and electronic structure properties (Figure 5).[202−204] Automated methods based on machine learning have been shown to be efficient in navigating reaction pathways by combining DFT adsorption energies and scaling relations.[205]

Despite recent advances, there are challenges in modeling nanomaterials surface reactivity. First, surface reactivity is highly affected by particle size. A study by Li et al. showed there exists an optimal chemical ratio and size of a FePt nanoparticle for CO oxidation using DFT.[206] In another computational study, Yudanov et al. elucidated that the CO adsorption energy is correlated with the Pd nanocluster size.[207] It was found that CO interaction weakens as particle becomes smaller up to a critical size and that CO adsorption energy approaches the bulk adsorption energy as the size increases. Second, the specificity of active sites and the chemical environment influences surface reactions. Understanding of site activity requires a priori knowledge of the atomic structure and reaction site calculated by DFT, which can deviate substantially from real materials interfaces, where surface passivation, reconstruction, and defects are prevalent. A wide span of parameters, from pressure, temperatures, and pH, to reactants/products species and concentration, and to solvents and electrolytes can influence the surface chemistry and topology, making it extremely difficult to derive one prediction which holds for all conditions.[169,203] Hence, surface reactivity trends with surface and interfacial conditions may prove more useful, as exact information about the surface states may not be available. For data-driven design of materials with target reactivity, there is a pressing need for well-structured and transferable databases with reliable metadata that reports surface- and site-related properties, as well as chemical environments and kinetics.[208−210]

In summary, designing functional nanomaterials present challenges as well as opportunities for data-driven research. Current HT studies mostly consider bulk material properties due to computational constraints, and hence one must bear in mind their limitations when transferring the knowledge into the nano domain. Indeed, one set of materials/properties may respond differently to nanosizing than others. To accurately describe the nano phenomena can be computationally demanding even for one system, much less in high-throughput fashion. In order to overcome this challenge, systematic reporting and curation of experimentally measured nanostructure data, including but not limited to shape-, size-, surface-, and defect-dependent properties, are needed. A few databases of experimental characterization at the nanoscale are available, such as electron energy loss spectrum[211,212] and refractive index[213] etc., but these resources are often for small sets of materials. Indeed, the development of robust, efficient, and accurate computational workflows is hampered by the lack of systematic, well-curated public experimental data to benchmark against for determination of a reasonable level of theory and numerical parameters. Therefore, progress in systematic reports of experiment synthesis and characterization will facilitate data-driven materials research, both from an experimental and a computational perspective.

TOWARD PREDICTIVE SYNTHESIS

While the modeling and data-driven predictions of materials properties and their correlations with the underlying structure have made significant advances in the past decade, predictive synthesis is still in its nascency. The phase space of possible structures is high-dimensional, and the reaction pathways between competing phases are driven by both thermodynamic and kinetic factors. Minor changes in the intricate interplay of experimental conditions, including reactants, concentrations, temperature, pressure, humidity, etc., can translate to distinct phases and unwanted byproducts.

As noted in Section Stability, the thermodynamic energy surface of all phases and multiphase equilibrium can be inferred from a computed convex-hull phase diagram (Figure 6). Phases that are calculated to be close to or on the convex hull are predicted to be thermodynamically metastable/stable at low temperatures but may not be easily synthesized, for a variety of reasons. In this context, a survey of the ICSD entries showed that the experimentally found compounds can exhibit energies greater than 100 meV/atom above the convex hull,

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

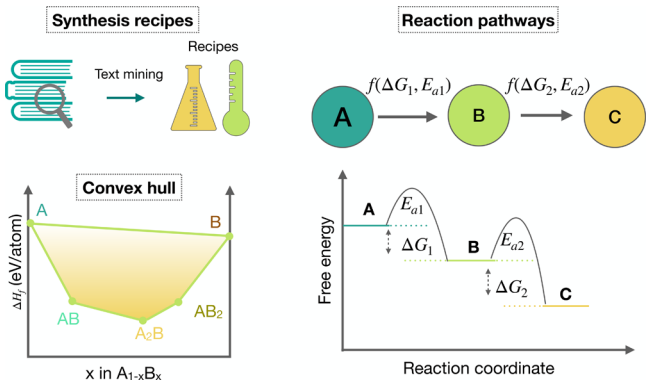


Figure 6. Schematic for data-driven methods for guided synthesis in inorganic solids, including natural language processing, convex hull stability, and reaction pathway prediction. ΔG represents free energy difference, and E_a represents activation barriers. Right panel adapted with permission from ref 214. Copyright 2021 The Authors under Creative Commons Attribution 4.0 International License, published by Springer Nature.

depending on the anion chemistry. Some of these materials are synthesized under conditions where the environmental parameters competitively lower their free energy and then remain kinetically trapped at lower temperatures.[215] Thus, the proximity to the convex hull provides a guide to synthesizability, but environmental factors as well as kinetics influence the final results.

Combining the timeline of experimentally discovered compounds with their stability determined by the convex-hull diagram, the likelihood of synthesis of a target material can be approximated using graph networks and machine learning models.[216] Via the application of pathfinding algorithms, McDermott et al. predicted the most probable reaction pathways to a target material using the DFT-calculated ground-state bulk energy including a machine-learned vibrational entropy[217] of all stable phases for the particular chemical systems (Figure 6).[214] Selectivity rules based on chemical potential maps[218] have been shown to inform the selection of precursors to maximize the yield of the target materials. Information about competing phases, such as the amorphous limit[78] or entropically favored materials,[219] can help to exclude materials from consideration that are unlikely to be successfully synthesized.

While a reaction pathway determines whether the synthesis route is possible, a synthesis recipe must also include environmental conditions. While the formation of crystal is controlled by nucleation and growth, the process depends on many factors, such as nucleation barrier, surface free energy, and the free energy of the reaction pathway.[220] Information about the synthesis environment, such as solution salts and impurity ions, has been shown to impact the kinetics of structure selection.[221] Often times, the description of the synthesis conditions is embedded in the literature, making it difficult to directly mine the data. To unravel the vast information, efforts using natural language processing are underway to harness the treasure trove of published synthesis recipes.[222−225] These machine-learning-based approaches have the potential to deconvolute kinetic and thermodynamic factors, especially when combined with thermodynamic predictions of phase competition. However, notably, published work is highly biased toward successful outcomes[226], and greater efforts to capture the entire experience of experimental synthesis are needed.

Successful predictions resulting in synthesized materials are still limited. The reasons are 2-fold. First, it remains unclear in experiments which factors precisely determine successful synthesis outcomes. A wide range of experimental synthesis methods exist, and the determining parameters may differ in each method. Second, compared to extensive research on understanding their properties, materials modeling has paid less attention to how materials are made in the real world, resulting in a knowledge gap between the theory and experiments.

Rationalizing synthesis is an area of materials science and chemistry that is in dire need of harnessing the knowledge that is today largely embedded as human experience. Automated robotics laboratories and associated data infrastructure will become a crucial component toward this goal, delivering standardized, consistent, and successful as well as unsuccessful "dark reactions" synthesis data and results.[226−234] Availability of such data enables the correlation between conditions and outcomes, which can be analyzed via, e.g., classification methods.[224] Ontology, as a branch of computer science, has also been applied to semantically represent materials informatics for machine interpretability.[235,236] The HT capabilities of robot-driven laboratories will contribute large amounts of quality synthesis data to the aforementioned models, enabling the predictive simulation of possible scenarios.

DATA-DRIVEN NANOMATERIALS SYNTHESIS

In addition to bulk synthesis, nanomaterials pose another set of synthesis parameters including but not limited to size, shape, and ligand chemistry.[237,238] Experimental observations show that the phase stabilities of NCs are dependent on sizes, where the metastable bulk phases can be stabilized by a more favorable surface energy as compared to the bulk ground state.[17,239−242] Questions associated with bulk materials synthesis still exist for nanomaterials, adding complexity with regard to controlling these parameters. Minor changes in synthesis conditions (e.g., temperature, mole ratio, and the length of alkyl chains) of nanomaterials can lead to significant variations in composition, dimensionality, morphology, short-range order, etc., all of which impact the electronic and optical properties.[243−246]

How to easily access this information becomes the accelerator for nanomaterials synthesis. Nanoparticle databases have emerged in an effort to assemble a vast collection of metadata that includes both experimental and theoretical data. Among multiple databases (Table 2), we highlight a few as

Table 2. Nanostructure Focused Databases

Nanomaterial databases	Data sets
Cambridge Structure Database[248]	Small-molecule organic and metal−organic crystal structures
Quantum Cluster Database[54]	50*k* total DFT calculated structures for nanoclusters with 3−55 atoms
Catalysis-Hub[263]	100*k* DFT adsorption and reaction energies
Open Catalyst Project[197]	1.2 M DFT adsorption structures
Materials Cloud[22,264,265]	2D materials and porous materials
Quantum Point Defect[144]	Point defects in 2D materials

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

excellent sources of data. The NOMAD (Novel Materials Discovery) Laboratory is a curated database of shared data from both experimental and computational studies and includes a search engine, a Python API, and over 2 M clusters in the database.[7] The Open Catalysis Project is a computational database for catalyst structures, with 1.3 million molecular relaxations that are open-source for training purposes.[247] The Cambridge structure database (CSD) contains experimentally observed and computationally verified small-molecule organic and metal−organic structures that have been growing steadily since 1965 and now have over 1 M structures, with over 50% of those involving at least one metal atom.[248] The Quantum Cluster Database is an online repository of DFT calculated pure element nanoclusters with over 50*K* total clusters containing between 3 and 55 atoms.[54] The clusters comprise a collection of clusters found in literature in addition to clusters generated with a genetic algorithm. An extensive list can be found in a review by Panneerselvam and Choi.[249]

On the other hand, while there are experimental databases (e.g., caNanoLab,[250] NR,[251] and eNanoMapper[252]) specifically for biological/medical applications, including physicochemical properties such as size, shape, composition, purity, surface area, toxicity, and environmental effects (reviews can be found in refs 253 and 254), a similar resource is not available for the purpose of materials science.

Nanomaterials databases face the challenges of variability in information and how to represent each structure with the appropriate metadata. A theoretically calculated database could include the following: atomic species, atomic coordinates, effective charge, implicit solvent, calculation software, functional, basis set, energy, forces, band gap, electronic structure, etc. An experimentally observed nanoparticle entry in a database could include atomic species, atomic coordinates, counterions present, stabilizing ligands, synthesis recipe, concentrations of reagents, temperature, and percent yield and likely contains information on the average particle size and distribution of sizes. The large available space of features and descriptors adds challenges to the already daunting task of materials synthesis data curation.

Efforts have been made to create universal descriptors. Yan et al. have made efforts to normalize the annotations of nanoparticle databases by creating a procedure to generate 2142 nanodescriptors for each nanomaterial in their open database, Pubvinas.[255] These descriptors include physicochemical properties and bioactivities like cytotoxicity and protein adsorption. While Pubvinas has created fixed descriptors for their database, NOMAD has created a posteriori data sorting technique, meaning that their database can continue adding descriptors if necessary depending on the added data.[7] The challenges associated with general materials data, volume, variety, velocity, and veracity, call for efficient approaches in its exchange and usage, and we encourage the reader to explore references by Draxl and Scheffler[7,256] for more details. We note that the data challenge becomes even more pronounced in nanomaterials, where data-driven research is nascent and the property−synthesis−characterization landscape increases in complexity. This field should also benefit from careful handling of expanding data sets by using the "FAIR" data principles (Findable, Accessible, Interoperable, and Repurposable), a guideline in major databases to enhance data reusability as it becomes crucial infrastructure for physics, chemistry, and materials science.[257]

The ultimate goal, for experiments, is to provide a highly interpretable set of guidelines that indicate how different tunable factors will influence the products. For synthesis, the most important descriptors are the reaction conditions that led to the experimentally synthesized product and the relative thermodynamic energies of candidate structures. Phase diagrams and intelligent data visualization are great tools for this purpose. In these phase diagrams, partial pressure, chemical potentials, pH, temperature, etc., provide tuning knobs, enabling the identification of promising regions in high-dimensional parameter space for fabricating a target compound.[258−261]

There is an opportunity for more development of easily interpretable guidelines on a larger scale. As an example, bulk materials databases like the Materials Project have made significant strides in providing synthesis-relevant data using Phase Diagram and Pourbaix Diagram analysis applications.[6,83,262] Similar analysis applications to facilitate searches through reaction space for nanomaterials would be highly desirable; however, the underlying data are today unavailable. With the rise of data-driven research in nanomaterial synthesis, we expect a rise in the quantity, quality, accessibility, and interpretability of data for nanomaterials synthesis and properties.

CONCLUSION

The development in data-driven materials research has surged in the past decade, creating an interdisciplinary domain of physics, chemistry, software, and data science. Materials databases are collecting more complex data with far-reaching potentials for functional materials design and guided synthesis.

Nanomaterials, on the other hand, exhibit behaviors distinct from underlying bulk structures, which originates from the confined nature and large surface area. The electronic structure and chemical activity of nanomaterials are controlled by the size and morphology of the crystals, the active surface sites, and the chemical environment. Synthesis is controlled by both thermodynamics and kinetic processes, both of which are highly sensitive to the external environment. There is a nontrivial gap to bridge between bulk and nanomaterials predictive modeling, which data-driven approaches are well-suited for, given the multidimensional, complex landscape. For future development of materials data, the information particularly helpful will be size-, surface-, and environment-specific properties, including surface energies, adsorption sites, various phase diagrams, and reaction networks.

As large-scale computational methods and high-throughput automated laboratories with data curation investments become more mainstream, we believe progress in the accessibility, quantity, accuracy, and complexity of nanomaterials data will profoundly advance this field. We note the opportunities in curating and disseminating experimental results with annotated metadata, as complementary to existing computational resources. The large number of raw data generated by automated laboratories contains highly reproducible parameters and includes a variety of information such as synthesis conditions, which can then be used by machine learning approaches and feeds back into the design process. Given the paradigm shift toward data-centric computational materials research, we are optimistic about the outlook for nanomaterials research in utilizing big data to its fullest potential.

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

AUTHOR INFORMATION

Corresponding Authors

Ruo Xi Yang − *Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0000-0001-8225-5856; Email: ruoxiyang@lbl.gov

Kristin A. Persson − *Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States; Molecular Foundry, Energy Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0000-0003-2495-5509; Email: kapersson@lbl.gov

Authors

Caitlin A. McCandler − *Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States; Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0003-2616-310X

Oxana Andriuc − *Department of Chemistry, University of California, Berkeley, California 94720, United States; Liquid Sunlight Alliance and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0000-0002-4011-3339

Martin Siron − *Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States; Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0002-4562-7814

Rachel Woods-Robinson − *Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0000-0001-5009-9510

Matthew K. Horton − *Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States; Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0001-7777-8871

Complete contact information is available at: https://pubs.acs.org/10.1021/acsnano.2c08411

Author Contributions

The team put together this review collaboratively. In particular: R.X.Y. wrote the Introduction, the electronic structure part of section Design Metrics of Materials: Bulk to Nanomaterials, part of section Data-Driven Nanomaterials Synthesis, and Conclusion, and was responsible for editing the final manuscript. C.A.M. wrote the section Nanomaterials Structure Prediction and part of the section Data-Driven Nanomaterials Synthesis; O.A. and M.S. wrote Surface Adsorption and Surface Reactivity in section Design Metrics of Materials: Bulk to Nanomaterials. R.W.-R. edited the subsections Interfaces and Transport, and M.K.H. edited the general text, particularly with regard to the general databases. K.A.P. conceptualized the scope of the review and edited the manuscript extensively.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was intellectually led by the Materials Project, which is funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract No. DE-AC02-05-CH11231: Materials Project Program KC23MP. Significant support was also provided (especially pertaining to the review of predictive synthesis) by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, for their support under Contract No. DE-AC02-05-CH11231 within the Data Science for Data-Driven Synthesis Science grant (KCD2S2). Supporting funding was received from the Liquid Sunlight Alliance, which is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Fuels from Sunlight Hub under Award No. DE-SC0021266. We thank Jason Munro, Matthew McDermott, and Mingjian Wen for insightful discussions and comments for this work.

VOCABULARY

convex hull, energy diagram that connects all ground state polymorphs of a specific chemical system, where points on the connected lines indicate thermodynamically stable phases, while points above the hull indicate thermodynamic instability; nanomaterial, material consisting of particles or constituents of nanoscale dimensions, in which typically at least one dimension is between 1 and 100 nm; quantum confinement, phenomena observed when the size of the particle is too small to be comparable to the wavelength of the electron, described by the quantum well model, and results in discrete energy levels in the valence band and conduction band and an increase of the band gap; reaction network, network of interlinking phases with the nodes representing a combination of phases and the edges representing the reaction pathways, being a complex thermodynamic phase space, where the reaction pathways can be represented by cost functions that combine thermodynamic and kinetic features; scaling relation, correlation between the adsorption energies of different adsorbates on various catalytic surfaces and often used to understand the reactivity of catalysts and to identify promising new catalysts

REFERENCES

(1) Jain, A.; Shin, Y.; Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **2016**, *1*, 15004.

(2) Luo, S.; Li, T.; Wang, X.; Faizan, M.; Zhang, L. High-throughput computational materials screening and discovery of optoelectronic semiconductors. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1489.

(3) Pollice, R.; dos Passos Gomes, G.; Aldeghi, M.; Hickman, R. J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C. T.; Yao, Z.; Aspuru-Guzik, A. Data-Driven Strategies for Accelerated Materials Design. *Acc. Chem. Res.* **2021**, *54*, 849−860.

(4) Mbaye, M. T.; Pradhan, S. K.; Bahoura, M. Data-driven thermoelectric modeling: Current challenges and prospects. *J. Appl. Phys.* **2021**, *130*, 190902.

(5) Sendek, A. D.; Ransom, B.; Cubuk, E. D.; Pellouchoud, L. A.; Nanda, J.; Reed, E. J. Machine Learning Modeling for Accelerated Battery Materials Design in the Small Data Regime. *Adv. Energy Mater.* **2022**, *12*, 2200553.

(6) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002.

(7) Draxl, C.; Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* **2019**, *2*, 036001.

(8) 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* **2013**, *65*, 1501−1509.

(9) Allen, F. H. The Development, Status and Scientific Impact of Crystallographic Databases. *Acta Cryst. A* **1998**, *54*, 758−771.
(10) Zagorac, D.; Müller, H.; Ruehl, S.; Zagorac, J.; Rehme, S. Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *J. Appl. Crystallogr.* **2019**, *52*, 918−925.
(11) Volk, S.; Yazdani, N.; Wood, V. Manipulating Electronic Structure from the Bottom-Up: Colloidal Nanocrystal-Based Semiconductors. *J. Phys. Chem. Lett.* **2020**, *11*, 9255−9264.
(12) Takagahara, T.; Takeda, K. Theory of the quantum confinement effect on excitons in quantum dots of indirect-gap materials. *Phys. Rev. B* **1992**, *46*, 15578−15581.
(13) Xia, J.-B.; Cheah, K. W. Quantum confinement effect in thin quantum wires. *Phys. Rev. B* **1997**, *55*, 15688−15693.
(14) Allan, G.; Delerue, C.; Lannoo, M. Quantum confinement in amorphous silicon layers. *Appl. Phys. Lett.* **1997**, *71*, 1189−1191.
(15) Jiang, J.; Sun, L.; Gao, B.; Wu, Z.; Lu, W.; Yang, J.; Luo, Y. Structure dependent quantum confinement effect in hydrogen-terminated nanodiamond clusters. *J. Appl. Phys.* **2010**, *108*, 094303.
(16) Zhao, Q.; Hazarika, A.; Schelhas, L. T.; Liu, J.; Gaulding, E. A.; Li, G.; Zhang, M.; Toney, M. F.; Sercel, P. C.; Luther, J. M. Size-Dependent Lattice Structure and Confinement Properties in $CsPbI_3$ Perovskite Nanocrystals: Negative Surface Energy for Stabilization. *ACS Energy Lett.* **2020**, *5*, 238−247.
(17) Yang, R. X.; Tan, L. Z. Understanding size dependence of phase stability and band gap in $CsPbI_3$ perovskite nanocrystals. *J. Chem. Phys.* **2020**, *152*, 034702.
(18) Curtarolo, S.; Setyawan, W.; Hart, G. L.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O.; Mehl, M. J.; Stokes, H. T.; Demchenko, D. O.; Morgan, D. AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **2012**, *58*, 218−226.
(19) Talirz, L.; et al. Materials Cloud, a platform for open computational science. *Sci. Data* **2020**, *7*, 299.
(20) Vitale, V.; Pizzi, G.; Marrazzo, A.; Yates, J. R.; Marzari, N.; Mostofi, A. A. Automated high-throughput Wannierisation. *npj Comput. Mater.* **2020**, *6*, 66.
(21) Marrazzo, A.; Gibertini, M.; Campi, D.; Mounet, N.; Marzari, N. Relative Abundance of Z2 Topological Order in Exfoliable Two-Dimensional Insulators. *Nano Lett.* **2019**, *19*, 8431−8440.
(22) Mounet, N.; Gibertini, M.; Schwaller, P.; Campi, D.; Merkys, A.; Marrazzo, A.; Sohier, T.; Castelli, I. E.; Cepellotti, A.; Pizzi, G.; Marzari, N. Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* **2018**, *13*, 246−252.
(23) Stevanović, V.; Lany, S.; Zhang, X.; Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Phys. Rev. B* **2012**, *85*, 115104.
(24) Lany, S. Band-structure calculations for the 3 d transition metal oxides in G W. *Phys. Rev. B* **2013**, *87*, 085112.
(25) Lany, S. Semiconducting transition metal oxides. *J. Phys.: Condens. Matter* **2015**, *27*, 283203.
(26) Landis, D. D.; Hummelshoj, J. S.; Nestorov, S.; Greeley, J.; Dulak, M.; Bligaard, T.; Norskov, J. K.; Jacobsen, K. W. The Computational Materials Repository. *Comput. Sci. Eng.* **2012**, *14*, 51−57.
(27) Choudhary, K.; et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Comput. Mater.* **2020**, *6*, 173.
(28) Yang, X.; Wang, Z.; Zhao, X.; Song, J.; Zhang, M.; Liu, H. MatCloud: A high-throughput computational infrastructure for integrated management of materials simulation, data and resources. *Comput. Mater. Sci.* **2018**, *146*, 319−333.
(29) Villars, P.; Cenzual, K.; Gladyshevskii, R.; Iwata, S.Pauling File: Toward a Holistic View. In *Materials Informatics*, 1st ed.; Isayev, O., Tropsha, A., Curtarolo, S., Eds.; Wiley, 2019; pp 55−106.
(30) He, Y.; Jiang, Y.; Zhang, T.; Huang, H.; Fang, C.; Jin, Z. SymTopo: An automatic tool for calculating topological properties of nonmagnetic crystalline materials. *Chinese Phys. B* **2019**, *28*, 087102.
(31) Zhang, T.; Jiang, Y.; Song, Z.; Huang, H.; He, Y.; Fang, Z.; Weng, H.; Fang, C. Catalogue of topological electronic materials. *Nature* **2019**, *566*, 475−479.
(32) Hinuma, Y.; Pizzi, G.; Kumagai, Y.; Oba, F.; Tanaka, I. Band structure diagram paths based on crystallography. *Comput. Mater. Sci.* **2017**, *128*, 140−184.
(33) Tanifuji, M.; Matsuda, A.; Yoshikawa, H.Materials Data Platform - a FAIR System for Data-Driven Materials Science. In *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, Toyama, Japan, Jul. 7−11, 2019; IEEE, 2019; pp 1021−1022. DOI: 10.1109/IIAI-AAI.2019.00206
(34) Talapin, D. V.; Lee, J.-S.; Kovalenko, M. V.; Shevchenko, E. V. Prospects of Colloidal Nanocrystals for Electronic and Optoelectronic Applications. *Chem. Rev.* **2010**, *110*, 389−458.
(35) Kovalenko, M. V.; et al. Prospects of Nanoscience with Nanocrystals. *ACS Nano* **2015**, *9*, 1012−1057.
(36) Kagan, C. R.; Lifshitz, E.; Sargent, E. H.; Talapin, D. V. Building devices from colloidal quantum dots. *Science* **2016**, *353*, eaac5523.
(37) Goldzak, T.; McIsaac, A. R.; Van Voorhis, T. Colloidal CdSe nanocrystals are inherently defective. *Nat. Commun.* **2021**, *12*, 890.
(38) ten Brinck, S.; Zaccaria, F.; Infante, I. Defects in Lead Halide Perovskite Nanocrystals: Analogies and (Many) Differences with the Bulk. *ACS Energy Lett.* **2019**, *4*, 2739−2747.
(39) Kirakosyan, A.; Chinh, N. D.; Sihn, M. R.; Jeon, M.-G.; Jeong, J.-R.; Kim, D.; Jang, J. H.; Choi, J. Mechanistic Insight into Surface Defect Control in Perovskite Nanocrystals: Ligands Terminate the Valence Transition from Pb^{2+} to Metallic Pb^0. *J. Phys. Chem. Lett.* **2019**, *10*, 4222−4228.
(40) Smart, T. J.; Takenaka, H.; Pham, T. A.; Tan, L. Z.; Zhang, J. Z.; Ogitsu, T.; Ping, Y. Enhancing Defect Tolerance with Ligands at the Surface of Lead Halide Perovskites. *J. Phys. Chem. Lett.* **2021**, *12*, 6299−6304.
(41) Ghasemi, S. A.; Amsler, M.; Hennig, R.; Roy, S.; Goedecker, S.; Umrigar, C.; Genovese, L.; Lenosky, T.; Morishita, T.; Nishio, K. The energy landscape of silicon systems and its description by force fields, tight binding schemes, density functional methods and Quantum Monte Carlo methods. *Phys. Rev. B* **2010**, *81*, 214107.
(42) Johnston, R. L. Evolving better nanoparticles: Genetic algorithms for optimizing cluster geometries. *Dalton Trans* **2003**, 4193.
(43) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.
(44) Pannetier, J.; Bassas-Alsina, J.; Rodriguez-Carvajal, J.; Caignaert, V. Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature* **1990**, *346*, 343−345.
(45) Chen, M.; Dixon, D. A. Tree Growth—Hybrid Genetic Algorithm for Predicting the Structure of Small (TiO2)n, n = 2−13, Nanoclusters. *J. Chem. Theory Comput.* **2013**, *9*, 3189−3200.
(46) Glover, F. Tabu Search—Part I. *ORSA Journal on Computing* **1989**, *1*, 190−206.
(47) Tan, L.; Pickard, C. J.; Yu, K.; Sapelkin, A.; Misquitta, A. J.; Dove, M. T. Structures of CdSe and CdS Nanoclusters from Ab Initio Random Structure Searching. *J. Phys. Chem. C* **2019**, *123*, 29370−29378.
(48) Ciobanu, C. V.; Wang, C.-Z.; Ho, K.-M.*Atomic Structure Prediction of Nanostructures, Clusters and Surfaces*; Wiley-VCH, 2012.
(49) Rossi, G.; Ferrando, R. Searching for low-energy structures of nanoparticles: a comparison of different methods and algorithms. *J. Phys.: Condens. Matter* **2009**, *21*, 084208.
(50) Woodley, S. M.; Day, G. M.; Catlow, R. Structure prediction of crystals, surfaces and nanoparticles. *Philos. Trans. R. Soc. A* **2020**, *378*, 20190600.

(51) Marom, N.; Kim, M.; Chelikowsky, J. R. Structure Selection Based on High Vertical Electron Affinity for TiO 2 Clusters. *Phys. Rev. Lett.* **2012**, *108*, 106801.
(52) Bhattacharya, S.; Sonin, B. H.; Jumonville, C. J.; Ghiringhelli, L. M.; Marom, N. Computational design of nanoclusters by property-based genetic algorithms: Tuning the electronic properties of (TiO 2) n clusters. *Phys. Rev. B* **2015**, *91*, 241115.
(53) Sokol, A. A.; Catlow, C. R. A.; Miskufova, M.; Shevlin, S. A.; Al-Sunaidi, A. A.; Walsh, A.; Woodley, S. M. On the problem of cluster structure diversity and the value of data mining. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8438.
(54) Manna, S.; Hernandez, A.; Wang, Y.; Lile, P.; Liu, S.; Mueller, T.A Database of Low-Energy Atomically Precise Nanoclusters. *ChemRxiv Preprint (Materials Chemistry*, 2021. 2021-0fq3q. https://chemrxiv.org/engage/chemrxiv/article-details/61a6c9ad836f652a90836fce (accessed 2022-10-28).
(55) Chaves, A. S.; Piotrowski, M. J.; Da Silva, J. L. F. Evolution of the structural, energetic, and electronic properties of the 3d, 4d, and 5d transition-metal clusters (30 TM $_n$ systems for n = 2−15): a density functional theory investigation. *Phys. Chem. Chem. Phys.* **2017**, *19*, 15484−15502.
(56) Chaves, A. S.; Rondina, G. G.; Piotrowski, M. J.; Tereshchuk, P.; Da Silva, J. L. F. The Role of Charge States in the Atomic Structure of Cun and Ptn (n = 2−14 atoms) Clusters: A DFT Investigation. *J. Phys. Chem. A* **2014**, *118*, 10813−10821.
(57) Rondina, G. G.; Da Silva, J. L. F. Revised Basin-Hopping Monte Carlo Algorithm for Structure Optimization of Clusters and Nanoparticles. *J. Chem. Inf. Model.* **2013**, *53*, 2282−2298.
(58) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. Performance and Cost Assessment of Machine Learning Interatomic Potentials. *J. Phys. Chem. A* **2020**, *124*, 731−745.
(59) Weinreich, J.; Römer, A.; Paleico, M. L.; Behler, J. Properties of α -Brass Nanoparticles. 1. Neural Network Potential Energy Surface. *J. Phys. Chem. C* **2020**, *124*, 12682−12695.
(60) Lysogorskiy, Y.; Oord, C. v. d.; Bochkarev, A.; Menon, S.; Rinaldi, M.; Hammerschmidt, T.; Mrovec, M.; Thompson, A.; Csányi, G.; Ortner, C.; Drautz, R. Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon. *npj Comput. Mater.* **2021**, *7*, 97.
(61) Loeffler, T. D.; Manna, S.; Patra, T. K.; Chan, H.; Narayanan, B.; Sankaranarayanan, S. Active Learning A Neural Network Model For Gold Clusters & Bulk From Sparse First Principles Training Data. *ChemCatChem.* **2020**, *12*, 4796−4806.
(62) Cao, L.; Li, C.; Mueller, T. The Use of Cluster Expansions To Predict the Structures and Properties of Surfaces and Nanostructured Materials. *J. Chem. Inf. Model.* **2018**, *58*, 2401−2413.
(63) Li, C.; Raciti, D.; Pu, T.; Cao, L.; He, C.; Wang, C.; Mueller, T. Improved Prediction of Nanoalloy Structures by the Explicit Inclusion of Adsorbates in Cluster Expansions. *J. Phys. Chem. C* **2018**, *122*, 18040−18047.
(64) Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **2019**, *99*, 014104.
(65) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat. Commun.* **2022**, *13*, 2453.
(66) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B.Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. *arXiv Preprint (Computational Physics)*, 2022. arXiv:2204.05249. https://arxiv.org/abs/2204.05249 (accessed 2022-10-28).
(67) Batatia, I.; Kovács, D. P.; Simm, G. N. C.; Ortner, C.; Csányi, G.MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *arXiv Preprint (Statistics: Machine Learning)*, 2022. arXiv:2206.07697. https://arxiv.org/abs/2206.07697 (accessed 2022-10-28).
(68) Tadmor, E. B.; Elliott, R. S.; Sethna, J. P.; Miller, R. E.; Becker, C. A. The potential of atomistic simulations and the knowledgebase of interatomic models. *JOM* **2011**, *63*, 17−17.
(69) Becker, C. A.; Tavazza, F.; Trautt, Z. T.; Buarque de Macedo, R. A. Considerations for choosing and using force fields and interatomic potentials in materials science and engineering. *Curr. Opin. Solid State Mater. Sci.* **2013**, *17*, 277−283.
(70) Hale, L. M.; Trautt, Z. T.; Becker, C. A. Evaluating variability with atomistic simulations: the effect of potential and calculation methodology on the modeling of lattice and elastic constants. *Modelling Simul. Mater. Sci. Eng.* **2018**, *26*, 055003.
(71) Yan, Q.; Yu, J.; Suram, S. K.; Zhou, L.; Shinde, A.; Newhouse, P. F.; Chen, W.; Li, G.; Persson, K. A.; Gregoire, J. M.; Neaton, J. B. Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 3040−3043.
(72) Jin, H.; Zhang, H.; Li, J.; Wang, T.; Wan, L.; Guo, H.; Wei, Y. Data-Driven Systematic Search of Promising Photocatalysts for Water Splitting under Visible Light. *J. Phys. Chem. Lett.* **2019**, *10*, 5211−5218.
(73) Singh, A. K.; Montoya, J. H.; Gregoire, J. M.; Persson, K. A. Robust and synthesizable photocatalysts for CO2 reduction: a data-driven materials discovery. *Nat. Commun.* **2019**, *10*, 443.
(74) Bartel, C. J.; Weimer, A. W.; Lany, S.; Musgrave, C. B.; Holder, A. M. The role of decomposition reactions in assessing first-principles predictions of solid stability. *npj Comput. Mater.* **2019**, *5*, 4.
(75) Bartel, C. J. Review of computational approaches to predict the thermodynamic stability of inorganic solids. *J. Mater. Sci.* **2022**, *57*, 10475−10498.
(76) Kissinger, H. E. Reaction kinetics in differential thermal analysis. *Analytical chemistry* **1957**, *29*, 1702−1706.
(77) Ozawa, T. Estimation of activation energy by isoconversion methods. *Thermochimica acta* **1992**, *203*, 159−165.
(78) Aykol, M.; Dwaraknath, S. S.; Sun, W.; Persson, K. A. Thermodynamic limit for synthesis of metastable inorganic materials. *Sci. Adv.* **2018**, *4*, No. eaaq0148.
(79) Luo, W.; Hu, W. Gibbs free energy, surface stress and melting point of nanoparticle. *Physica B: Condensed Matter* **2013**, *425*, 90−94.
(80) Yang, R. X.; Tan, L. Z. First-Principles Characterization of Surface Phonons of Halide Perovskite CsPbI $_3$ and Their Role in Stabilization. *J. Phys. Chem. Lett.* **2021**, *12*, 9253−9261.
(81) Seif, M. N.; Beck, M. J. Surface excess free energies and equilibrium Wulff shapes in variable chemical environments at finite temperatures. *Appl. Surf. Sci.* **2021**, *540*, 148383.
(82) Singh, A. K.; Zhou, L.; Shinde, A.; Suram, S. K.; Montoya, J. H.; Winston, D.; Gregoire, J. M.; Persson, K. A. Electrochemical Stability of Metastable Materials. *Chem. Mater.* **2017**, *29*, 10159−10167.
(83) 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* **2012**, *85*, 235438.
(84) Guisbiers, G. Schottky Defects in Nanoparticles. *J. Phys. Chem. C* **2011**, *115*, 2616−2621.
(85) Yao, Y.; Dong, Q.; Brozena, A.; Luo, J.; Miao, J.; Chi, M.; Wang, C.; Kevrekidis, I. G.; Ren, Z. J.; Greeley, J.; Wang, G.; Anapolsky, A.; Hu, L. High-entropy nanoparticles: Synthesis-structure-property relationships and data-driven discovery. *Science* **2022**, *376*, No. eabn3103.
(86) Shi, H.; Barnard, A. S.; Snook, I. K. High throughput theory and simulation of nanomaterials: exploring the stability and electronic properties of nanographene. *J. Mater. Chem.* **2012**, *22*, 18119.
(87) Shockley, W.; Queisser, H. J. Detailed Balance Limit of Efficiency of *p-n* Junction Solar Cells. *J. Appl. Phys.* **1961**, *32*, 510−519.
(88) Li, Y.; Yang, K. High-throughput computational design of organic−inorganic hybrid halide semiconductors beyond perovskites for optoelectronics. *Energy Environ. Sci.* **2019**, *12*, 2233−2243.

(89) Brunin, G.; Ricci, F.; Ha, V.-A.; Rignanese, G.-M.; Hautier, G. Transparent conducting materials discovery using high-throughput computing. *npj Comp. Mater.* **2019**, *5*, 63.
(90) Tran, F.; Blaha, P. Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential. *Phys. Rev. Lett.* **2009**, *102*, 226401.
(91) Morales-García; Valero, R.; Illas, F. An Empirical, yet Practical Way To Predict the Band Gap in Solids by Using Density Functional Band Structure Calculations. *J. Phys. Chem. C* **2017**, *121*, 18862−18866.
(92) Zhao, X.-G.; Yang, J.-H.; Fu, Y.; Yang, D.; Xu, Q.; Yu, L.; Wei, S.-H.; Zhang, L. Design of Lead-Free Inorganic Halide Perovskites for Solar Cells via Cation-Transmutation. *J. Am. Chem. Soc.* **2017**, *139*, 2630−2638.
(93) Kuhar, K.; Pandey, M.; Thygesen, K. S.; Jacobsen, K. W. High-Throughput Computational Assessment of Previously Synthesized Semiconductors for Photovoltaic and Photoelectrochemical Devices. *ACS Energy Lett.* **2018**, *3*, 436−446.
(94) Hinuma, Y.; Hatakeyama, T.; Kumagai, Y.; Burton, L. A.; Sato, H.; Muraba, Y.; Iimura, S.; Hiramatsu, H.; Tanaka, I.; Hosono, H.; Oba, F. Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis. *Nat. Commun.* **2016**, *7*, 11962.
(95) Ma, X.-Y.; Lewis, J. P.; Yan, Q.-B.; Su, G. Accelerated Discovery of Two-Dimensional Optoelectronic Octahedral Oxyhalides via High-Throughput *Ab Initio* Calculations and Machine Learning. *J. Phys. Chem. Lett.* **2019**, *10*, 6734−6740.
(96) van Setten, M. J.; Giantomassi, M.; Gonze, X.; Rignanese, G.-M.; Hautier, G. Automation methodologies and large-scale validation for GW: Towards high-throughput GW calculations. *Phys. Rev. B* **2017**, *96*, 155207.
(97) Rasmussen, A.; Deilmann, T.; Thygesen, K. S. Towards fully automated GW band structure calculations: What we can learn from 60.000 self-energy evaluations. *npj Comput. Mater.* **2021**, *7*, 22.
(98) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2016**, *2*, 16028.
(99) Ward, L.; et al. Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **2018**, *152*, 60−69.
(100) 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.* **2015**, *114*, 105503.
(101) Wang, V.; Liang, Y.-Y.; Kawazoe, Y.; Geng, W.-T.High-Throughput Computational Screening of Two-Dimensional Semiconductors. *arXiv Preprint (Condensed Matter: Mesoscale and Nanoscale Physics)*, 2018. arXiv:1806.04285. https://arxiv.org/abs/1806.04285 (accessed 2022-10-28).
(102) Broderick, S. R.; Aourag, H.; Rajan, K. Data mining density of states spectra for crystal structure classification: An inverse problem approach: Data Mining Density of States Spectra. *Statistical Analy Data Mining* **2009**, *1*, 353−360.
(103) Ben Mahmoud, C.; Anelli, A.; Csányi, G.; Ceriotti, M. Learning the electronic density of states in condensed matter. *Phys. Rev. B* **2020**, *102*, 235130.
(104) Bang, K.; Yeo, B. C.; Kim, D.; Han, S. S.; Lee, H. M. Accelerated mapping of electronic density of states patterns of metallic nanoparticles via machine-learning. *Sci. Rep.* **2021**, *11*, 11604.
(105) Brus, L. Electronic wave functions in semiconductor clusters: experiment and theory. *J. Phys. Chem.* **1986**, *90*, 2555−2560.
(106) Efros, A. L.; Efros, A. L. Interband absorption of light in a semiconductor sphere. *SPIE Milestone Ser.* **1982**, *16*, 772−775.
(107) Brus, L. E. A simple model for the ionization potential, electron affinity, and aqueous redox potentials of small semiconductor crystallites. *J. Chem. Phys.* **1983**, *79*, 5566−5571.
(108) Brus, L. E. Electron−electron and electron-hole interactions in small semiconductor crystallites: The size dependence of the lowest excited electronic state. *J. Phys. Chem.* **1984**, *80*, 4403−4409.
(109) Talapin, D. V.; Nelson, J. H.; Shevchenko, E. V.; Aloni, S.; Sadtler, B.; Alivisatos, A. P. Seeded Growth of Highly Luminescent CdSe/CdS Nanoheterostructures with Rod and Tetrapod Morphologies. *Nano Lett.* **2007**, *7*, 2951−2959.
(110) Tilchin, J.; Rabouw, F. T.; Isarov, M.; Vaxenburg, R.; Van Dijk-Moes, R. J. A.; Lifshitz, E.; Vanmaekelbergh, D. Quantum Confinement Regimes in CdTe Nanocrystals Probed by Single Dot Spectroscopy: From Strong Confinement to the Bulk Limit. *ACS Nano* **2015**, *9*, 7840−7845.
(111) Even, J.; Pedesseau, L.; Katan, C. Understanding Quantum Confinement of Charge Carriers in Layered 2D Hybrid Perovskites. *ChemPhysChem* **2014**, *15*, 3733−3741.
(112) Smith, A. M.; Nie, S. Semiconductor Nanocrystals: Structure, Properties, and Band Gap Engineering. *Acc. Chem. Res.* **2010**, *43*, 190−200.
(113) Bardeen, J.; Shockley, W. Deformation Potentials and Mobilities in Non-Polar Crystals. *Phys. Rev.* **1950**, *80*, 72−80.
(114) Madsen, G. K.; Singh, D. J. BoltzTraP. A code for calculating band-structure dependent quantities. *Comput. Phys. Commun.* **2006**, *175*, 67−71.
(115) Faghaninia, A.; Ager, J. W.; Lo, C. S. *Ab initio* electronic transport model with explicit solution to the linearized Boltzmann transport equation. *Phys. Rev. B* **2015**, *91*, 235123.
(116) Pizzi, G.; Volja, D.; Kozinsky, B.; Fornari, M.; Marzari, N. BoltzWann: A code for the evaluation of thermoelectric and electronic transport properties with a maximally-localized Wannier functions basis. *Comput. Phys. Commun.* **2014**, *185*, 422−429.
(117) Madsen, G. K.; Carrete, J.; Verstraete, M. J. BoltzTraP2, a program for interpolating band structures and calculating semi-classical transport coefficients. *Comput. Phys. Commun.* **2018**, *231*, 140−145.
(118) Poncé, S.; Margine, E.; Verdi, C.; Giustino, F. EPW: Electron−phonon coupling, transport and superconducting properties using maximally localized Wannier functions. *Comput. Phys. Commun.* **2016**, *209*, 116−133.
(119) Ganose, A. M.; Park, J.; Faghaninia, A.; Woods-Robinson, R.; Persson, K. A.; Jain, A. Efficient calculation of carrier scattering rates from first principles. *Nat. Commun.* **2021**, *12*, 2222.
(120) Drude, P. Zur Elektronentheorie der Metalle. *Ann. Phys.* **1900**, *306*, 566−613.
(121) Chen, W.; Pöhls, J.-H.; Hautier, G.; Broberg, D.; Bajaj, S.; Aydemir, U.; Gibbs, Z. M.; Zhu, H.; Asta, M.; Snyder, G. J.; Meredig, B.; White, M. A.; Persson, K.; Jain, A. Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment. *J. Mater. Chem. C* **2016**, *4*, 4414−4426.
(122) Park, J.; Xia, Y.; Ganose, A. M.; Jain, A.; Ozoliņš, V. High Thermoelectric Performance and Defect Energetics of Multipocketed Full Heusler Compounds. *Phys. Rev. Applied* **2020**, *14*, 024064.
(123) Kolahi, S.; Farjami-Shayesteh, S.; Azizian-Kalandaragh, Y. Comparative studies on energy-dependence of reduced effective mass in quantum confined ZnS semiconductor nanocrystals prepared in polymer matrix. *Mater. Sci. Semicond. Process.* **2011**, *14*, 294−301.
(124) Zhou, A. P.; Sheng, W. D. Electron and hole effective masses in self-assembled quantum dots. *Eur. Phys. J. B* **2009**, *68*, 233−236.
(125) Rodríguez-Mas, F.; Ferrer, J.; Alonso, J.; Valiente, D.; Fernández de Avila, S. A Comparative Study of Theoretical Methods to Estimate Semiconductor Nanoparticles' Size. *Crystals* **2020**, *10*, 226.
(126) Liu, Y.; Gibbs, M.; Puthussery, J.; Gaik, S.; Ihly, R.; Hillhouse, H. W.; Law, M. Dependence of Carrier Mobility on Nanocrystal Size and Ligand Length in PbSe Nanocrystal Solids. *Nano Lett.* **2010**, *10*, 1960−1969.
(127) Biswas, S.; Dutta, B.; Bhattacharya, S. Dependence of the carrier mobility and trapped charge limited conduction on silver nanoparticles embedment in doped polypyrrole nanostructures. *J. Appl. Phys.* **2013**, *114*, 143701.
(128) Majeed Khan, M. A.; Kumar, S.; Ahamed, M.; Alrokayan, S. A.; AlSalhi, M. S. Structural and thermal studies of silver nanoparticles and electrical transport study of their thin films. *Nanoscale Res. Lett.* **2011**, *6*, 434.

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

(129) Yu, D.; Wang, C.; Wehrenberg, B. L.; Guyot-Sionnest, P. Variable Range Hopping Conduction in Semiconductor Nanocrystal Solids. *Phys. Rev. Lett.* **2004**, *92*, 216802.
(130) Benton, B. T.; Greenberg, B. L.; Aydil, E.; Kortshagen, U. R.; Campbell, S. A. Variable range hopping conduction in ZnO nanocrystal thin films. *Nanotechnology* **2018**, *29*, 415202.
(131) Hill, R. M. Variable-range hopping. *physica status solidi (a)* **1976**, *34*, 601−613.
(132) Tsigankov, D. N.; Efros, A. L. Variable Range Hopping in Two-Dimensional Systems of Interacting Electrons. *Phys. Rev. Lett.* **2002**, *88*, 176602.
(133) Zhang, S.; Northrup, J. Chemical potential dependence of defect formation energies in GaAs: Application to Ga self-diffusion. *Phys. Rev. Lett.* **1991**, *67*, 2339−2342.
(134) Freysoldt, C.; Grabowski, B.; Hickel, T.; Neugebauer, J.; Kresse, G.; Janotti, A.; Van de Walle, C. G. First-principles calculations for point defects in solids. *Rev. Mod. Phys.* **2014**, *86*, 253−305.
(135) Kumagai, Y.; Oba, F. Electrostatics-based finite-size corrections for first-principles point defect calculations. *Phys. Rev. B* **2014**, *89*, 195205.
(136) Kumagai, Y.; Tsunoda, N.; Takahashi, A.; Oba, F. Insights into oxygen vacancies from high-throughput first-principles calculations. *Phys. Rev. Materials* **2021**, *5*, 123803.
(137) Yim, K.; Youn, Y.; Lee, M.; Yoo, D.; Lee, J.; Cho, S. H.; Han, S. Computational discovery of p-type transparent oxide semiconductors using hydrogen descriptor. *npj Comput. Mater.* **2018**, *4*, 17.
(138) Van de Walle, C. G. Universal alignment of hydrogen levels in semiconductors and insulators. *Physica B: Condensed Matter* **2006**, *376*−*377*, 1−6.
(139) Schleife, A.; Fuchs, F.; Rödl, C.; Furthmüller, J.; Bechstedt, F. Branch-point energies and band discontinuities of III-nitrides and III-/II-oxides from quasiparticle band-structure calculations. *Appl. Phys. Lett.* **2009**, *94*, 012104.
(140) Woods-Robinson, R.; Broberg, D.; Faghaninia, A.; Jain, A.; Dwaraknath, S. S.; Persson, K. A. Assessing High-Throughput Descriptors for Prediction of Transparent Conductors. *Chem. Mater.* **2018**, *30*, 8375−8389.
(141) Goyal, A.; Gorai, P.; Peng, H.; Lany, S.; Stevanović, V. A computational framework for automation of point defect calculations. *Comput. Mater. Sci.* **2017**, *130*, 1−9.
(142) Broberg, D.; Medasani, B.; Zimmermann, N. E.; Yu, G.; Canning, A.; Haranczyk, M.; Asta, M.; Hautier, G. PyCDT: A Python toolkit for modeling point defects in semiconductors and insulators. *Comput. Phys. Commun.* **2018**, *226*, 165−179.
(143) Stoliaroff, A.; Jobic, S.; Latouche, C. PyDEF 2.0: An Easy to Use Post-treatment Software for Publishable Charts Featuring a Graphical User Interface. *J. Comput. Chem.* **2018**, *39*, 2251−2261.
(144) Bertoldo, F.; Ali, S.; Manti, S.; Thygesen, K. S. Quantum point defects in 2D materials - the QPOD database. *npj Comput. Mater.* **2022**, *8*, 56.
(145) Zheng, X.; Hou, Y.; Sun, H.-T.; Mohammed, O. F.; Sargent, E. H.; Bakr, O. M. Reducing Defects in Halide Perovskite Nanocrystals for Light-Emitting Applications. *J. Phys. Chem. Lett.* **2019**, *10*, 2629−2640.
(146) Luo, S.; Li, M.; Fung, V.; Sumpter, B. G.; Liu, J.; Wu, Z.; Page, K. New Insights into the Bulk and Surface Defect Structures of Ceria Nanocrystals from Neutron Scattering Study. *Chem. Mater.* **2021**, *33*, 3959−3970.
(147) Nadupalli, S.; Repp, S.; Weber, S.; Erdem, E. About defect phenomena in ZnO nanocrystals. *Nanoscale* **2021**, *13*, 9160−9171.
(148) Houtepen, A. J.; Hens, Z.; Owen, J. S.; Infante, I. On the Origin of Surface Traps in Colloidal II−VI Semiconductor Nanocrystals. *Chem. Mater.* **2017**, *29*, 752−761.
(149) du Fossé, I.; ten Brinck, S.; Infante, I.; Houtepen, A. J. Role of Surface Reduction in the Formation of Traps in *n* -Doped II−VI Semiconductor Nanocrystals: How to Charge without Reducing the Surface. *Chem. Mater.* **2019**, *31*, 4575−4583.
(150) Yamada, Y.; Nakamura, T.; Endo, M.; Wakamiya, A.; Kanemitsu, Y. Photocarrier Recombination Dynamics in Perovskite $CH_3NH_3PbI_3$ for Solar Cell Applications. *J. Am. Chem. Soc.* **2014**, *136*, 11610−11613.
(151) Li, F.; Yang, L.; Cai, Z.; Wei, K.; Lin, F.; You, J.; Jiang, T.; Wang, Y.; Chen, X. Enhancing exciton binding energy and photoluminescence of formamidinium lead bromide by reducing its dimensions to 2D nanoplates for producing efficient light emitting diodes. *Nanoscale* **2018**, *10*, 20611−20617.
(152) Petousis, I.; Mrdjenovich, D.; Ballouz, E.; Liu, M.; Winston, D.; Chen, W.; Graf, T.; Schladt, T. D.; Persson, K. A.; Prinz, F. B. High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. *Sci. Data* **2017**, *4*, 160134.
(153) Macfarlane, G.; McLean, T.; Quarrington, J.; Roberts, V. Exciton and phonon effects in the absorption spectra of germanium and silicon. *J. Phys. Chem. Solids* **1959**, *8*, 388−392.
(154) Rohlfing, M.; Louie, S. G. Excitonic Effects and the Optical Absorption Spectrum of Hydrogenated Si Clusters. *Phys. Rev. Lett.* **1998**, *80*, 3320−3323.
(155) Albrecht, S.; Reining, L.; Del Sole, R.; Onida, G. *Ab Initio* Calculation of Excitonic Effects in the Optical Spectra of Semiconductors. *Phys. Rev. Lett.* **1998**, *80*, 4510−4513.
(156) Marini, A.; Del Sole, R. Dynamical Excitonic Effects in Metals and Semiconductors. *Phys. Rev. Lett.* **2003**, *91*, 176402.
(157) Dvorak, M.; Wei, S.-H.; Wu, Z. Origin of the Variation of Exciton Binding Energy in Semiconductors. *Phys. Rev. Lett.* **2013**, *110*, 016402.
(158) Lee, J.-C.; Chai, J.-D.; Lin, S.-T. Assessment of density functional methods for exciton binding energies and related optoelectronic properties. *RSC Adv.* **2015**, *5*, 101370−101376.
(159) Yang, R. X.; Horton, M. K.; Munro, J.; Persson, K. A.High-throughput optical absorption spectra for inorganic semiconductors. *arXiv Preprint (Condensed Matter: Materials Science)*, 2022. arXiv:2209.02918, https://arxiv.org/abs/2209.02918 (accessed 2022-10-28).
(160) Efros, A. L.; Rosen, M.; Kuno, M.; Nirmal, M.; Norris, D. J.; Bawendi, M. Band-edge exciton in quantum dots of semiconductors with a degenerate valence band: Dark and bright exciton states. *Phys. Rev. B* **1996**, *54*, 4843−4856.
(161) Li, F.; Yang, L.; Cai, Z.; Wei, K.; Lin, F.; You, J.; Jiang, T.; Wang, Y.; Chen, X. Enhancing exciton binding energy and photoluminescence of formamidinium lead bromide by reducing its dimensions to 2D nanoplates for producing efficient light emitting diodes. *Nanoscale* **2018**, *10*, 20611−20617.
(162) Gélvez-Rueda, M. C.; Fridriksson, M. B.; Dubey, R. K.; Jager, W. F.; van der Stam, W.; Grozema, F. C. Overcoming the exciton binding energy in two-dimensional perovskite nanoplatelets by attachment of conjugated organic chromophores. *Nat. Commun.* **2020**, *11*, 1901.
(163) Becker, M. A.; et al. Bright triplet excitons in caesium lead halide perovskites. *Nature* **2018**, *553*, 189−193.
(164) Tang, Y.; Yin, C.; Jing, Q.; Zhang, C.; Yu, Z.-G.; Lu, Z.; Xiao, M.; Wang, X. Quantized Exciton Motion and Fine Energy-Level Structure of a Single Perovskite Nanowire. *Nano Lett.* **2022**, *22*, 2907−2914.
(165) Sercel, P. C.; Lyons, J. L.; Wickramaratne, D.; Vaxenburg, R.; Bernstein, N.; Efros, A. L. Exciton Fine Structure in Perovskite Nanocrystals. *Nano Lett.* **2019**, *19*, 4068−4077.
(166) Yang, Y.; Guo, W.; Wang, X.; Wang, Z.; Qi, J.; Zhang, Y. Size Dependence of Dielectric Constant in a Single Pencil-Like ZnO Nanowire. *Nano Lett.* **2012**, *12*, 1919−1922.
(167) Richards, W. D.; Miara, L. J.; Wang, Y.; Kim, J. C.; Ceder, G. Interface Stability in Solid-State Batteries. *Chem. Mater.* **2016**, *28*, 266−273.
(168) Woods-Robinson, R.; Ablekim, T.; Norman, A.; Johnston, S.; Persson, K. A.; Reese, M. O.; Metzger, W. K.; Zakutayev, A. Sputtered p-Type $Cu_xZn_{1-x}S$ Back Contact to CdTe Solar Cells. *ACS Appl. Energy Mater.* **2020**, *3*, 5427−5438.

(169) Pan, J.; Yan, Q. Data-driven material discovery for photocatalysis: a short review. *J. Semicond.* **2018**, *39*, 071001.
(170) Yin, W.-J.; Tang, H.; Wei, S.-H.; Al-Jassim, M. M.; Turner, J.; Yan, Y. Band structure engineering of semiconductors for enhanced photoelectrochemical water splitting: The case of TiO 2. *Phys. Rev. B* **2010**, *82*, 045106.
(171) Sun, W.; Ceder, G. Efficient creation and convergence of surface slabs. *Surf. Sci.* **2013**, *617*, 53−59.
(172) Park, J.-S.; Jung, Y.-K.; Butler, K. T.; Walsh, A. Quick-start guide for first-principles modelling of semiconductor interfaces. *J. Phys. Energy* **2019**, *1*, 016001.
(173) Tran, R.; Li, X.-G.; Montoya, J. H.; Winston, D.; Persson, K. A.; Ong, S. P. Anisotropic work function of elemental crystals. *Surf. Sci.* **2019**, *687*, 48−55.
(174) De Waele, S.; Lejaeghere, K.; Sluydts, M.; Cottenier, S. Error estimates for density-functional theory predictions of surface energy and work function. *Phys. Rev. B* **2016**, *94*, 235418.
(175) Ping, Y.; Sundararaman, R.; Goddard, W. A., III Solvation effects on the band edge positions of photocatalysts from first principles. *Phys. Chem. Chem. Phys.* **2015**, *17*, 30499−30509.
(176) Kharche, N.; Muckerman, J. T.; Hybertsen, M. S. First-Principles Approach to Calculating Energy Level Alignment at Aqueous Semiconductor Interfaces. *Phys. Rev. Lett.* **2014**, *113*, 176802.
(177) Pham, T. A.; Ping, Y.; Galli, G. Modelling heterogeneous interfaces for solar water splitting. *Nat. Mater.* **2017**, *16*, 401−408.
(178) Nørskov, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. Density functional theory in surface chemistry and catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 937−943.
(179) An, Q.; Shen, Y.; Fortunelli, A.; Goddard, W. A. QM-Mechanism-Based Hierarchical High-Throughput in Silico Screening Catalyst Design for Ammonia Synthesis. *J. Am. Chem. Soc.* **2018**, *140*, 17702−17710.
(180) Mamun, O.; Winther, K. T.; Boes, J. R.; Bligaard, T. High-throughput calculations of catalytic properties of bimetallic alloy surfaces. *Sci. Data* **2019**, *6*, 76.
(181) Takahashi, K.; Takahashi, L.; Le, S. D.; Kinoshita, T.; Nishimura, S.; Ohyama, J. Synthesis of Heterogeneous Catalysts in Catalyst Informatics to Bridge Experiment and High-Throughput Calculation. *J. Am. Chem. Soc.* **2022**, *144*, 15735−15744.
(182) Montoya, J. H.; Persson, K. A. A high-throughput framework for determining adsorption energies on solid surfaces. *npj Comp. Mater.* **2017**, *3*, 14.
(183) Mathew, K.; et al. Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **2017**, *139*, 140−152.
(184) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G.-M.; Hautier, G.; Gunter, D.; Persson, K. A. FireWorks: a dynamic workflow system designed for high-throughput applications: FireWorks: A Dynamic Workflow System Designed for High-Throughput Applications. *Concurrency Computat.: Pract. Exp.* **2015**, *27*, 5037−5059.
(185) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(186) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B* **1999**, *59*, 7413−7421.
(187) Andriuc, O.; Siron, M.; Montoya, J. H.; Horton, M.; Persson, K. A. Automated Adsorption Workflow for Semiconductor Surfaces and the Application to Zinc Telluride. *J. Chem. Inf. Model.* **2021**, *61*, 3908−3916.
(188) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine learning the quantum-chemical properties of metal−organic frameworks for accelerated materials discovery. *Matter* **2021**, *4*, 1578−1597.
(189) Ma, X.; Li, Z.; Achenie, L. E. K.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO 2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 3528−3533.
(190) Jinnouchi, R.; Asahi, R. Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. *J. Phys. Chem. Lett.* **2017**, *8*, 4279−4283.
(191) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-throughput screening of bimetallic catalysts enabled by machine learning. *J. Mater. Chem. A* **2017**, *5*, 24131−24138.
(192) Wang, M.; Zhu, H. Machine Learning for Transition-Metal-Based Hydrogen Generation Electrocatalysts. *ACS Catal.* **2021**, *11*, 3930−3937.
(193) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K.-i. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10*, 2260−2297.
(194) SchlexerLamoureux, P.; Winther, K. T.; GarridoTorres, J. A.; Streibel, V.; Zhao, M.; Bajdich, M.; Abild-Pedersen, F.; Bligaard, T. Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem.* **2019**, *11*, 3581−3601.
(195) Tran, K.; Palizhati, A.; Back, S.; Ulissi, Z. W. Dynamic Workflows for Routine Materials Discovery in Surface Science. *J. Chem. Inf. Model.* **2018**, *58*, 2392−2400.
(196) Tran, R.; Xu, Z.; Radhakrishnan, B.; Winston, D.; Sun, W.; Persson, K. A.; Ong, S. P. Surface energies of elemental crystals. *Sci. Data* **2016**, *3*, 160080.
(197) Chanussot, L.; et al. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catal.* **2021**, *11*, 6059−6072.
(198) Tran, R.; Lan, J.; Shuaibi, M.; Goyal, S.; Wood, B. M.; Das, A.; Heras-Domingo, J.; Kolluru, A.; Rizvi, A.; Shoghi, N.; Sriram, A.; Ulissi, Z.; Zitnick, C. L.The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysis. *arXiv Preprint (Condensed Matter: Materials Science)*, 2022. arXiv:2206.08917. https://arxiv.org/abs/2206.08917 (accessed 2022-10-28).
(199) Motagamwala, A. H.; Dumesic, J. A. Microkinetic Modeling: A Tool for Rational Catalyst Design. *Chem. Rev.* **2021**, *121*, 1049−1076.
(200) Ishikawa, A.; Tateyama, Y. A First-Principles Microkinetics for Homogeneous−Heterogeneous Reactions: Application to Oxidative Coupling of Methane Catalyzed by Magnesium Oxide. *ACS Catal.* **2021**, *11*, 2691−2700.
(201) Jørgensen, M.; Grönbeck, H. First-Principles Microkinetic Modeling of Methane Oxidation over Pd(100) and Pd(111). *ACS Catal.* **2016**, *6*, 6730−6738.
(202) Latimer, A. A.; Kulkarni, A. R.; Aljama, H.; Montoya, J. H.; Yoo, J. S.; Tsai, C.; Abild-Pedersen, F.; Studt, F.; Nørskov, J. K. Understanding trends in C−H bond activation in heterogeneous catalysis. *Nat. Mater.* **2017**, *16*, 225−229.
(203) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8*, 7403−7429.
(204) Rosen, A. S.; Notestein, J. M.; Snurr, R. Q. Identifying promising metal−organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory. *J. Comput. Chem.* **2019**, *40*, 1305−1318.
(205) Mazeau, E. J.; Satpute, P.; Blöndal, K.; Goldsmith, C. F.; West, R. H. Automated Mechanism Generation Using Linear Scaling Relationships and Sensitivity Analyses Applied to Catalytic Partial Oxidation of Methane. *ACS Catal.* **2021**, *11*, 7114−7125.
(206) Li, L.; Wang, Y.-Z.; Wang, X.-X.; Song, K.-K.; Jian, X.-D.; Qian, P.; Bai, Y.; Su, Y.-J. Size and Stoichiometry Effect of FePt Bimetal Nanoparticle Catalyst for CO Oxidation: A DFT Study. *J. Phys. Chem. C* **2020**, *124*, 8706−8715.
(207) Yudanov, I. V.; Genest, A.; Schauermann, S.; Freund, H.-J.; Rösch, N. Size Dependence of the Adsorption Energy of CO on Metal Nanoparticles: A DFT Search for the Minimum Value. *Nano Lett.* **2012**, *12*, 2134−2139.
(208) Bo, C.; Maseras, F.; López, N. The role of computational results databases in accelerating the discovery of catalysts. *Nat. Catal* **2018**, *1*, 809−810.

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

(209) Mendes, P. S. F.; Siradze, S.; Pirro, L.; Thybaut, J. W. Open Data in Catalysis: From Today's Big Picture to the Future of Small Data. *ChemCatChem.* **2021**, *13*, 836−850.
(210) Wulf, C.; Beller, M.; Boenisch, T.; Deutschmann, O.; Hanf, S.; Kockmann, N.; Kraehnert, R.; Oezaslan, M.; Palkovits, S.; Schimmler, S.; Schunk, S. A.; Wagemann, K.; Linke, D. A Unified Research Data Infrastructure for Catalysis Research − Challenges and Concepts. *ChemCatChem.* **2021**, *13*, 3223.
(211) Ewels, P.; Sikora, T.; Serin, V.; Ewels, C. P.; Lajaunie, L. A Complete Overhaul of the Electron Energy-Loss Spectroscopy and X-Ray Absorption Spectroscopy Database: eelsdb.eu. *Microsc Microanal* **2016**, *22*, 717−724.
(212) Chae, J. E.; Kim, J.-S.; Nam, S.-Y.; Kim, M. S.; Park, J. Introduction to the standard reference data of electron energy loss spectra and their database: eel.geri.re.kr. *Appl. Microsc.* **2020**, *50*, 2.
(213) Polyanskiy, M. N.*Refractive index database*. https://refractiveindex.info, (accessed 2022-10-28).
(214) McDermott, M. J.; Dwaraknath, S. S.; Persson, K. A. A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis. *Nat. Commun.* **2021**, *12*, 3097.
(215) Sun, W.; Dacek, S. T.; Ong, S. P.; Hautier, G.; Jain, A.; Richards, W. D.; Gamst, A. C.; Persson, K. A.; Ceder, G. The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2016**, *2*, No. e1600225.
(216) Aykol, M.; Hegde, V. I.; Hung, L.; Suram, S.; Herring, P.; Wolverton, C.; Hummelshøj, J. S. Network analysis of synthesizable materials discovery. *Nat. Commun.* **2019**, *10*, 2018.
(217) Bartel, C. J.; Millican, S. L.; Deml, A. M.; Rumptz, J. R.; Tumas, W.; Weimer, A. W.; Lany, S.; Stevanović, V.; Musgrave, C. B.; Holder, A. M. Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry. *Nat. Commun.* **2018**, *9*, 4168.
(218) Todd, P. K.; McDermott, M. J.; Rom, C. L.; Corrao, A. A.; Denney, J. J.; Dwaraknath, S. S.; Khalifah, P. G.; Persson, K. A.; Neilson, J. R. Selectivity in Yttrium Manganese Oxide Synthesis via Local Chemical Potentials in Hyperdimensional Phase Space. *J. Am. Chem. Soc.* **2021**, *143*, 15185−15194.
(219) Woods-Robinson, R.; Stevanović, V.; Lany, S.; Heinselman, K. N.; Horton, M. K.; Persson, K. A.; Zakutayev, A. Role of disorder in the synthesis of metastable zinc zirconium nitrides. *Phys. Rev. Materials* **2022**, *6*, 043804.
(220) Aykol, M.; Montoya, J. H.; Hummelshøj, J. S.Rational Solid-State Synthesis Routes for Inorganic Materials. *ChemRxiv Preprint (Theoretical and Computational Chemistry*, 2021. 13546484.v1. https://chemrxiv.org/engage/chemrxiv/article-details/60c7599d469df41f8ef458b4 (accessed 2022-10-28).
(221) Sun, W.; Jayaraman, S.; Chen, W.; Persson, K. A.; Ceder, G. Nucleation of metastable aragonite $CaCO_3$ in seawater. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 3199−3204.
(222) Kim, E.; Huang, K.; Saunders, A.; McCallum, A.; Ceder, G.; Olivetti, E. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem. Mater.* **2017**, *29*, 9436−9444.
(223) Kononova, O.; Huo, H.; He, T.; Rong, Z.; Botari, T.; Sun, W.; Tshitoyan, V.; Ceder, G. Text-mined dataset of inorganic materials synthesis recipes. *Sci. Data* **2019**, *6*, 203.
(224) Huo, H.; Rong, Z.; Kononova, O.; Sun, W.; Botari, T.; He, T.; Tshitoyan, V.; Ceder, G. Semi-supervised machine-learning classification of materials synthesis procedures. *npj Comput. Mater.* **2019**, *5*, 62.
(225) Oliynyk, A. O.; Antono, E.; Sparks, T. D.; Ghadbeigi, L.; Gaultois, M. W.; Meredig, B.; Mar, A. High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds. *Chem. Mater.* **2016**, *28*, 7324−7331.
(226) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533*, 73−76.
(227) Krishnadasan, S.; Brown, R. J. C.; deMello, A. J.; deMello, J. C. Intelligent routes to the controlled synthesis of nanoparticles. *Lab Chip* **2007**, *7*, 1434.
(228) Epps, R. W.; Felton, K. C.; Coley, C. W.; Abolhasani, M. Automated microfluidic platform for systematic studies of colloidal perovskite nanocrystals: towards continuous nano-manufacturing. *Lab Chip* **2017**, *17*, 4040−4047.
(229) Bai, Y.; Wilbraham, L.; Slater, B. J.; Zwijnenburg, M. A.; Sprick, R. S.; Cooper, A. I. Accelerated Discovery of Organic Polymer Photocatalysts for Hydrogen Evolution from Water through the Integration of Experiment and Theory. *J. Am. Chem. Soc.* **2019**, *141*, 9063−9071.
(230) Xie, Y.; Zhang, C.; Deng, H.; Zheng, B.; Su, J.-W.; Shutt, K.; Lin, J. Accelerate Synthesis of Metal−Organic Frameworks by a Robotic Platform and Bayesian Optimization. *ACS Appl. Mater. Interfaces* **2021**, *13*, 53485−53491.
(231) Li, Y.; Xia, L.; Fan, Y.; Wang, Q.; Hu, M. Recent advances in autonomous synthesis of materials. *Chem. Phys. Mater.* **2022**, *1*, 77−85.
(232) Li, J.; Li, J.; Liu, R.; Tu, Y.; Li, Y.; Cheng, J.; He, T.; Zhu, X. Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. *Nat. Commun.* **2020**, *11*, 2046.
(233) Epps, R. W.; Volk, A. A.; Reyes, K. G.; Abolhasani, M. Accelerated AI development for autonomous materials synthesis in flow. *Chem. Sci.* **2021**, *12*, 6025−6036.
(234) Aspuru-Guzik, A.; Persson, K.Materials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence. *Mission Innovation: Innovation Challenge 6*; Canadian Institute for Advanced Research, 2018.
(235) Zhao, S.; Qian, Q. Ontology based heterogeneous materials database integration and semantic query. *AIP Advances* **2017**, *7*, 105325.
(236) Takahashi, L.; Takahashi, K. Visualizing Scientists' Cognitive Representation of Materials Data through the Application of Ontology. *J. Phys. Chem. Lett.* **2019**, *10*, 7482−7491.
(237) Phan, H. T.; Haes, A. J. What Does Nanoparticle Stability Mean? *J. Phys. Chem. C* **2019**, *123*, 16495−16507.
(238) Wang, X.; Zhuang, J.; Peng, Q.; Li, Y. A general strategy for nanocrystal synthesis. *Nature* **2005**, *437*, 121−124.
(239) Levchenko, A. A.; Li, G.; Boerio-Goates, J.; Woodfield, B. F.; Navrotsky, A. TiO_2 Stability Landscape: Polymorphism, Surface Energy, and Bound Water Energetics. *Chem. Mater.* **2006**, *18*, 6324−6332.
(240) Satoh, N.; Nakashima, T.; Yamamoto, K. Metastability of anatase: size dependent and irreversible anatase-rutile phase transition in atomic-level precise titania. *Sci. Rep.* **2013**, *3*, 1959.
(241) Li, S.; Yang, G. W. Phase Transition of IIVI Semiconductor Nanocrystals. *J. Phys. Chem. C* **2010**, *114*, 15054−15060.
(242) Liu, L.; Zhao, R.; Xiao, C.; Zhang, F.; Pevere, F.; Shi, K.; Huang, H.; Zhong, H.; Sychugov, I. Size-Dependent Phase Transition in Perovskite Nanocrystals. *J. Phys. Chem. Lett.* **2019**, *10*, 5451−5457.
(243) Smith, D. K.; Korgel, B. A. The Importance of the CTAB Surfactant on the Colloidal Seed-Mediated Synthesis of Gold Nanorods. *Langmuir* **2008**, *24*, 644−649.
(244) Scarabelli, L.; Sánchez-Iglesias, A.; Pérez-Juste, J.; Liz-Marzán, L. M. A "Tips and Tricks" Practical Guide to the Synthesis of Gold Nanorods. *J. Phys. Chem. Lett.* **2015**, *6*, 4270−4279.
(245) Liz-Marzán, L. M.; Kagan, C. R.; Millstone, J. E. Reproducibility in Nanocrystal Synthesis? Watch Out for Impurities. *ACS Nano* **2020**, *14*, 6359−6361.
(246) Shamsi, J.; Urban, A. S.; Imran, M.; De Trizio, L.; Manna, L. Metal Halide Perovskite Nanocrystals: Synthesis, Post-Synthesis Modifications, and Their Optical Properties. *Chem. Rev.* **2019**, *119*, 3296−3348.
(247) Zitnick, C. L.et al.An Introduction to Electrocatalyst Design using Machine Learning for Renewable Energy Storage. *arXiv Preprint*

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891

(*Condensed Matter: Materials Science*), 2020, arXiv:2010.09435. https://arxiv.org/abs/2010.09435v1(accessed 2022-10-28).
(248) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. B Struct Sci. Cryst. Eng. Mater.* **2016**, *72*, 171−179.
(249) Panneerselvam, S.; Choi, S. Nanoinformatics: Emerging Databases and Available Tools. *IJMS* **2014**, *15*, 7158−7182.
(250) Gaheen, S.; Hinkal, G. W.; Morris, S. A.; Lijowski, M.; Heiskanen, M.; Klemm, J. D. caNanoLab: data sharing to expedite the use of nanotechnology in biomedicine. *Comput. Sci. Disc.* **2013**, *6*, 014010.
(251) Mills, K.; Ostraat, M. L.; Guzan, K.; Murry, D. The Nanomaterial Registry: facilitating the sharing and analysis of data in the diverse nanomaterial community. *Int. J. Nanomed.* **2013**, *8*, 7−13.
(252) Jeliazkova, N.; et al. The eNanoMapper database for nanomaterial safety information. *Beilstein J. Nanotechnol.* **2015**, *6*, 1609−1634.
(253) Pal, S.; Mondal, S.; Das, G.; Khatua, S.; Ghosh, Z. Big data in biology: The hope and present-day challenges in it. *Gene Reports* **2020**, *21*, 100869.
(254) Ji, Z.; Guo, W.; Sakkiah, S.; Liu, J.; Patterson, T.; Hong, H. Nanomaterial Databases: Data Sources for Promoting Design and Risk Assessment of Nanomaterials. *Nanomaterials* **2021**, *11*, 1599.
(255) Yan, X.; Sedykh, A.; Wang, W.; Yan, B.; Zhu, H. Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations. *Nat. Commun.* **2020**, *11*, 2519.
(256) Draxl, C.; Scheffler, M.Big Data-Driven Materials Science and Its FAIR Data Infrastructure. In *Handbook of Materials Modeling: Methods: Theory and Modeling*; Andreoni, W., Yip, S., Eds.; Springer: Cham, Switzerland, 2020; pp 49−73. DOI: 10.1007/978-3-319-42913-7_104-1.
(257) Wilkinson, M. D.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018.
(258) Beret, E. C.; Ghiringhelli, L. M.; Scheffler, M. Free gold clusters: beyond the static, monostructure description. *Faraday Discuss.* **2011**, *152*, 153.
(259) Bhattacharya, S.; Levchenko, S. V.; Ghiringhelli, L. M.; Scheffler, M. Stability and Metastability of Clusters in a Reactive Atmosphere: Theoretical Evidence for Unexpected Stoichiometries of Mg_MO_x. *Phys. Rev. Lett.* **2013**, *111*, 135501.
(260) Goldsmith, B. R.; Florian, J.; Liu, J.-X.; Gruene, P.; Lyon, J. T.; Rayner, D. M.; Fielicke, A.; Scheffler, M.; Ghiringhelli, L. M. Two-to-three dimensional transition in neutral gold clusters: The crucial role of van der Waals interactions and temperature. *Phys. Rev. Materials* **2019**, *3*, 016002.
(261) Dahl, J. C.; Wang, X.; Huang, X.; Chan, E. M.; Alivisatos, A. P. Elucidating the Weakly Reversible Cs−Pb−Br Perovskite Nanocrystal Reaction Network with High-Throughput Maps and Transformations. *J. Am. Chem. Soc.* **2020**, *142*, 11915−11926.
(262) Jain, A.; Hautier, G.; Ong, S. P.; Moore, C. J.; Fischer, C. C.; Persson, K. A.; Ceder, G. Formation enthalpies by mixing GGA and GGA + U calculations. *Phys. Rev. B* **2011**, *84*, 045115.
(263) Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T. Catalysis-Hub.org, an open electronic structure database for surface reactions. *Sci. Data* **2019**, *6*, 75.
(264) Davide, C.; Nicolas, M.; Marco, G.; Giovanni, P.; Nicola, M.The Materials Cloud 2D database (MC2D). Materials Cloud, https://archive.materialscloud.org/record/2022.84, (accessed 10−28−2022).
(265) Boyd, P. G.; et al.Data-driven design of metal-organic frameworks for wet flue gas CO2 capture. *Materials Cloud Archive*, 2019. materialscloud:2018.0016/v3.https://archive.materialscloud.org/record/2018.0016/v3 (accessed 2022-10-28).

https://doi.org/10.1021/acsnano.2c08411
ACS Nano 2022, 16, 19873−19891