

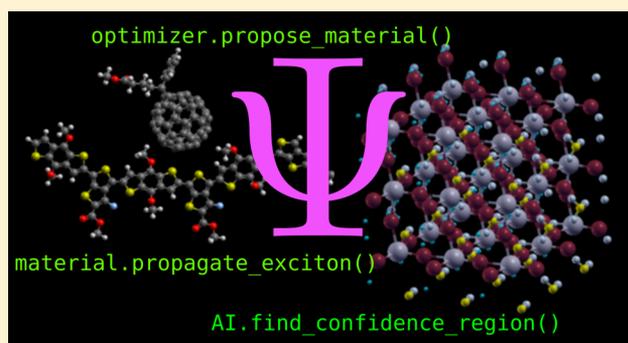
# Wave Functions, Density Functionals, and Artificial Intelligence for Materials and Energy Research: Future Prospects and Challenges

Martín A. Mosquera,<sup>†</sup> Bo Fu,<sup>‡</sup> Kevin L. Kohlstedt,<sup>†</sup> George C. Schatz,<sup>†</sup> and Mark A. Ratner<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry and the Materials Research Center, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United States

<sup>‡</sup>Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United States

**ABSTRACT:** Semiconducting materials, crystalline or amorphous, feature a diverse family of emergent transient properties (excitons, free carriers, plasmons, polarons, etc.) of interest to energy science, which are observed (indirectly or directly) in carefully designed experiments. Theoretical methods, which provide detailed and accurate information about the excitations of small molecules, have trouble with large systems because of computational limitations, such that a thorough selection of algorithms plays a crucial role. With a wide range of research opportunities in mind, in this Perspective we consider, from a *first-principles* perspective, the techniques available to calculate optical and electronic properties of materials and discuss (i) challenges in density-functional and wave function methods for materials and energy science, (ii) a method developed by us for describing excited-state phenomena (which consists of the linear response analysis of perturbed initial states), and (iii) opportunities for using machine learning in computational and theoretical chemistry studies.



Elucidating the principles and detailed molecular processes underlying the operation of optoelectronic (semiconducting) devices, such as photovoltaics (PV), photodiodes, and phototransistors, is among the primary objectives pursued by energy scientists and engineers. The rational design of these devices depends on a deep understanding of molecular structures and excited states at large length scales where charge transport processes operate, and it is also essential to forge a close link between theory and experiment.

Numerous efficient experimental techniques are able to monitor *both* full-gap particles such as electrons/holes, excitons, charge carriers, and also subgap (or intraband gap) quasiparticles such as (bi)polarons, plasmons, and plexcitons. To understand experimental observations, in *first-principles* theoretical/computational chemistry settings, scientists utilize methods based on ground-state and linear response theories that offer crucial information about the systems being studied. These techniques often provide specific data (such as minimum-energy geometries, equilibrium molecular trajectories, energy levels, among others) that complement understanding of the phenomena underlying the operation of semiconducting materials.

With the goal of simulating (in a parameter-free fashion and with relatively low computational demand) all the physicochemical processes taking place in the realistic operation of semiconducting materials, a significant challenge in theory development

for energy applications is the integration of advanced theoretical and computational techniques. In this spirit, despite requiring user-provided parameters, phenomenological/semiempirical/coarse-grained approaches are computationally inexpensive alternatives that can capture the essence of the physics and chemistry of relevance. Hence, we believe these approaches can serve as guides for progress in ab initio theory beyond equilibrium. In addition, the emergence of fields like machine learning is opening diverse research opportunities that in combination with ab initio quantum chemical methods will lead to exciting discoveries in energy and materials science.

In this Perspective we present our reflections on future directions of quantum chemical modeling and computation of semiconducting materials for energy research. We discuss (i) challenges in methods based on wave function and density functional theories; (ii) our recent experience on estimating absorption spectra of excited states, and related challenges; and (iii) the role and future of artificial intelligence/machine learning for applied quantum chemistry.

*Density Functional Methods.* The selection of the best quantum chemistry method for a particular problem often can be

Received: October 27, 2017

Accepted: December 7, 2017

Published: December 7, 2017

The emergence of fields like machine learning is opening diverse research opportunities that in combination with ab initio quantum chemical methods will lead to exciting discoveries in energy and materials science.

thought of as a refined art. The majority of electronic structure calculations can be categorized into three methods: correlated-wave function theory (CWFT), semiempirical methods, and density-functional theory (DFT). It is of no surprise to the reader that the latter is currently the most widely employed. Alternatives such as CWFT methodologies can be quite accurate, yet computationally expensive. Nonetheless, embedding techniques,<sup>1,2</sup> i.e., inserting a CWFT region into a material that is otherwise described by inexpensive DFT models, are extending the applicability of CWFT methods to solids, and it is possible that simulations of thousands of atoms will be accessible by CWFT methods, or even millions when quantum computers become available.<sup>3</sup> Currently, however, density-functional approaches permit mid-to-large scale (almost parameter-free) computer simulations. Linear-scaling methods<sup>4–6</sup> also hold promise in further extending the applicability of both DFT and CWFT methodologies. Lastly, for large chemical database/library studies (such as material/drug discovery), semiempirical methods, although requiring careful parametrization, are quite convenient.

Density functionals enjoy widespread success for the calculation of features that correlate well with device performance, such as band structure, optical properties in the linear regime, among others. However, conventional density functionals are often developed for ground-state properties. In general, the challenges and future prospects we discuss in this work concern approaching, with ab initio methods, regimes far from the ground state (excited-state dynamics, for instance) and equilibrium (as in the case of quantum ratchets<sup>7,8</sup>), which are computationally expensive and often unexplored but may offer valuable insight into the operation of semiconducting devices.

The versatility in calculations over many systems of interest, with good trade-off between weighing computational cost and accuracy, are the main reasons for DFT techniques being preferred. However, systematic errors do exist, especially with conjugated polymer systems because of self-interaction errors overestimating torsional barriers. For instance, for polyenes and polydiacetylenes, torsional barriers calculated with different functionals can differ by over 5 kcal/mol.<sup>9</sup> Finely tuned range-separated functionals have provided robust calculations of electronic couplings and torsional angles of conjugated organic molecules.<sup>10</sup> Recently, these functionals have shown promise in predicting charge-transfer characteristics at metal–organic interfaces, which bodes well for DFT methods in predicting properties for the next generation of hybrid organic/inorganic devices.<sup>11,12</sup>

**Motivation.** To describe the properties of a photovoltaic system, in principle and ideally, we should find the solution (given an initial state) to a statistical collection of time-dependent (TD) Schrödinger/Dirac equations, in which electrons and nuclei are treated on the same footing (as quantum mechanical particles). This task is achievable only for very small systems that are barely related to real devices. Beyond this point, in

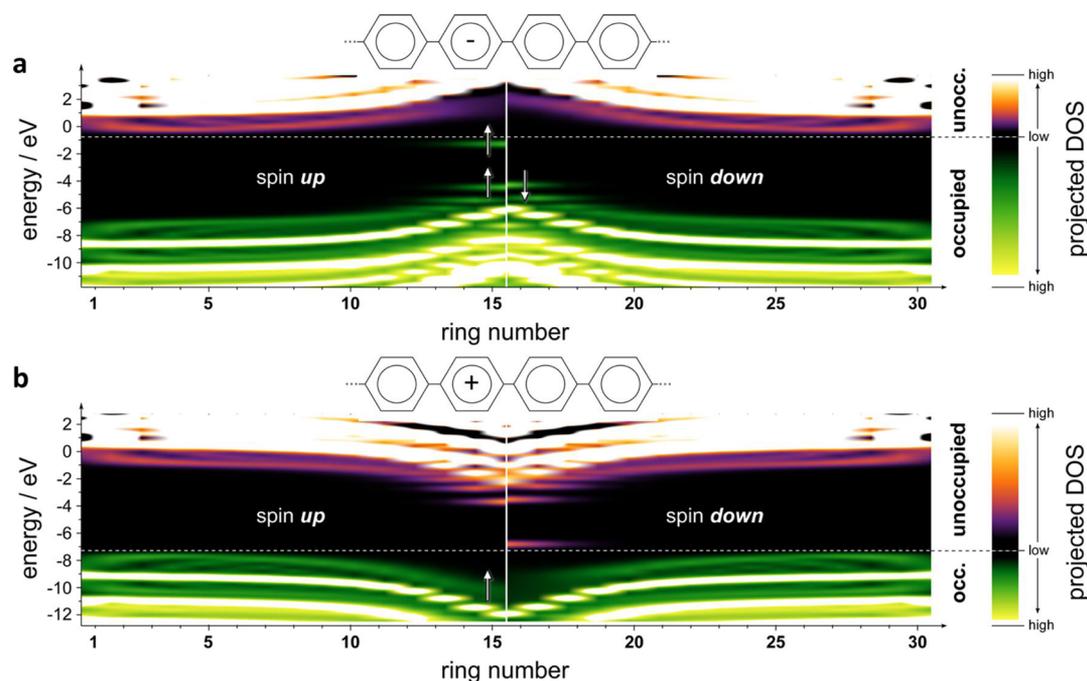
terms of molecular size, we need useful models that reduce complexity of the physical picture via assumptions, principles, and/or transformations. A quite common assumption is the Born–Oppenheimer (BO) approximation, in which electronic degrees of freedom are treated separately from the nuclei and the exchange of energy between electrons and nuclei is simplified. This approximation can be refined in practice (in a post-BO fashion) by the careful implementation of electron–nuclei couplings (e.g., nuclear gradients of electronic orbitals) and nuclear dynamics algorithm, such as surface hopping.<sup>13</sup>

Even within the BO picture, solving the *electronic* Schrödinger equation remains an unsolvable problem: The number of floating-point operations dramatically (or exponentially) increases as we slightly increase demand for wave functions/orbitals of higher quality (see ref 14 and references therein). To avoid this problem (TD) DFT methods offer numerous ways to control the computational scaling (usually at the expense of systematic improvability). Kohn–Sham theory is the most commonly used flavor of DFT model. Provided a proper approximation to its missing piece, the exchange–correlation (XC) energy, the Kohn–Sham framework allows the user to employ an improved form of Hartree–Fock equations, where the system is described with a single determinant wave function, to compute ground-state properties of interest.

To improve the applicability of XC approximations, ground-state functional development is perhaps one of the most active areas in DFT research. Functionals are found by satisfying exact constraints (the more the better) or by fitting trial functionals to ab initio and experimental data. In either case, we miss a priori predictors that tell us how well a newly developed functional will perform; broadly speaking, high-level reference calculations and experimental validation are required to attain confidence. In contrast to ground-state DFT research, functional development in TD density functional theory (TDDFT) is rare and more difficult. In general, algorithms use the ground-state XC kernel to compute absorption spectra and related quantities.

**Challenges.** Many PV devices utilize either conjugated organic molecules or hybrid metal–organic crystals, e.g. perovskites. While periodic (plane-wave) DFT studies with GGA (PBE) functionals dominate the latter,<sup>15–18</sup> isolated-molecule DFT allows for a wider variety of long-range corrected and hybrid functionals that are significantly expensive for plane-wave periodic calculations. Yet, in the case of metal–organic DFT studies, careful selections must be made regarding treatment of core electrons in the metal (pseudopotentials), spin–orbit coupling, and post-DFT corrections, etc. For example, in studies of lead and tin perovskites, the GW post-DFT correction improves estimation of the band structure and light absorption properties (by properly approximating the electron–hole propagator, and by extracting the relevant information from it).<sup>19</sup>

DFT XC functionals have been more broadly tested for organic conjugated molecules for both neutral and charged states.<sup>20</sup> For the most part, band gap and frontier orbital energy levels have been well-predicted by hybrid functionals like B3LYP. Molecular couplings and hence mobilities have been more difficult for DFT approximations, often underestimating the overlaps between the molecules. Semiempirical methods (see extended Hückel methods) have had more success in predicting the orbital overlaps between the disordered molecules in organic photovoltaic (OPV) and organic light-emitting diode (OLED) devices. In order to predict the power conversion efficiency (PCE) for new materials needed in PV devices, DFT approximations must



**Figure 1.** Ring-projected density-of-states (DOS) plots for an ionized 30-unit oligomer of PPP. (a) Top: Subplot for the anion species. Spin-up channel DOS is shown on the left for rings 1 through 15, whereas spin-down channel DOS is shown for the remaining rings on the right; the ring-projected DOS for each channel is mirror-symmetric. The separation between occupied and unoccupied levels is indicated by the dashed line, and the white vertical line represents the mirror plane of the molecule. The intensity of each state is represented using the color gradient on the right: white for high and black for low contributions. The white arrows represent occupation of the HOMO and HOMO–1 levels. (b) Same subplot as in panel a for the cation species. Reprinted from ref 21. Copyright 2017 American Chemical Society.

be able to adaptively account for the variety of organic materials. Polymerization length dependence, donor–acceptor energy offsets, and excited (polaronic) states are all standing in the way of current DFT methods being an accurate predictor of material properties.

Even though linear response techniques are essential for excited-state phenomena, the information they provide could also be determined in the same way as ground-state densities are found by standard computational packages. Such extended self-consistent DFT techniques, even though missing for applications, could offer quantities like exciton densities and couplings between them. Given these quantities, we could use them to determine how transient species evolve under the action of driving external and internal fields. Second, even though relatively affordable electron–phonon coupling algorithms are available<sup>22</sup> to estimate lifetimes of free carriers in perfect periodic crystals, or with a few defects, for amorphous systems we miss *first-principles* computational methodologies that extend beyond semiclassical and Marcus theories and solve more demanding models (based on Liouville quantum dynamics) that account for the generation and recombination of carriers. Amorphous systems of current interest include glassy polymers<sup>23</sup> (for which it is difficult to sample the conformational space) and perovskite-like systems with significant concentration of defects.<sup>24</sup>

Functionals commonly used in plane-wave packages underestimate the transport gap of the system; the method “DFT+U” (which allows the user to constrain the occupation numbers of problematic orbitals) fixes gap underestimation at the expense of an extra system-dependent parameter. A single electron or hole that is added to a macroscopic solid is not completely described by the lowest unoccupied molecular orbital (LUMO) or highest occupied molecular orbital (HOMO) of the neutral

system but by frontier orbitals that account for the nonlocal interaction that the added electron/hole is subject to. Traditional transport models assume the energy levels are frozen (at the neutral system) and use the HOMO/LUMO for stochastic transport simulations. In these models it is also assumed that the KS wave function is the true wave function of the system. To improve upon this assumption, new research could tell us how to extract information rigorously from the evolution of the TD density, or current-density, directly. To achieve this we could employ TDDFT and TD XC functionals.

First-principles solution to the band gap underestimation problem could enhance the description of transient (subgap) polaron species, which in OPV systems do not benefit from a frozen-level description used in semiconducting crystals. Figure 1 shows, for example, the ring-dependent, B3LYP, projected density of states of an ionized (+ or – charged, both in the ground state) poly(para-phenylene) (PPP) oligomer, with 30 rings.<sup>21</sup> For the anion, a localized level (HOMO) appears in the middle of the chain, while for the cation, a localized hole emerges in the same region. In the top subplot, the white arrows represent the occupation of the HOMO and HOMO–1 levels of the chain, and the single arrow at the bottom subplot indicates that two, HOMO and HOMO–1, electrons are removed from the anion. This example illustrates that the frontier orbitals of charged OPV polymers significantly relax (we use the term “sub-gap”, or “intragap”, to refer to them) and so do their energy spacings, affecting the polaron optical gap. In comparison to pure density-functionals such as GGAs, hybrid functionals, to a lesser degree, alleviate description of these subgap states. However, they are much more expensive for periodic systems; thus, reducing costs is an outstanding challenge. Additionally, for polaronic species and their excitations,

we also require development of algorithms to compute electron–phonon coupling in disordered systems.

*Wave Function Theory: Coupled Clusters and Static/Strong Correlation.* More fundamental wave function-based options are available to directly address the problem of finding solutions to the nonrelativistic Schrödinger equation. Because of high computational costs, however, further progress is needed to enable their use in large-scale electronic-structure calculations for energy applications.

Coupled cluster (CC) theory and methods<sup>25</sup> are powerful examples of wave function-based alternatives. Usually, these methods start from a single reference, create determinants by exciting occupied orbitals, and generate a combination of these excited determinants that depends on a collection of coefficients. When a set of nonlinear equations is solved, the coefficients of this combination are determined, and properties of interest can be calculated.

Size-extensiveness is a quite positive, and important, feature of CC methodologies (the CC energy grows linearly with respect to the number of particles in the thermodynamic limit<sup>26</sup>) and is systematically improvable (augmenting the size of the trial space usually leads to more accuracy). One might argue that KS-DFT leads to systematically improvable expansions of the energy, but at a much higher cost than in coupled cluster methods.<sup>27</sup> Because of the rich algebra<sup>28</sup> of CC methods, they require, in general, more computer coding efforts than DFT algorithms. Thus, a significant challenge in the CC field is also reducing the computational costs to compete against DFT methods, or bridge both DFT and CC to achieve systematically improvable XC functionals. To achieve this, CC methods developed within a divide-and-conquer perspective<sup>29</sup> could be very promising tools to achieve linear-scaling cost.

*Challenges.* Understanding reactions at the gas–surface interface via computation can assist the design and optimization of photocatalytic processes. CC and DFT methods are both commonly used for molecular systems that can be described by single references. There are problems, nonetheless, in which single determinantal techniques are insufficient.<sup>30</sup> Reactive dynamics of molecules on surfaces falls in this category, where molecular collisions and transition states play decisive roles. It is particularly difficult to describe, in real-time, radical species that are scattered off from surfaces because this phenomenon needs a large-scale multireference description, which, to date, is unavailable. Nonetheless, the relevance of simulating photocatalytic reactions should serve as continuous motivation to pursue speeding-up development in this area of multireference methods.

Complete active space methods,<sup>31</sup> or their derivatives, could achieve accurate description of bond-breaking reactions on surfaces. These techniques are capable of switching from the standard single-reference regime (accessible with traditional DFT/CWFT algorithms) into the difficult case of multireference wave functions. Although their costs of computation are still high, there is promise in these methods, especially if combined with DFT/CC/artificial intelligence ideas, to offer universal first-principles methods for treating reactive molecular interactions.

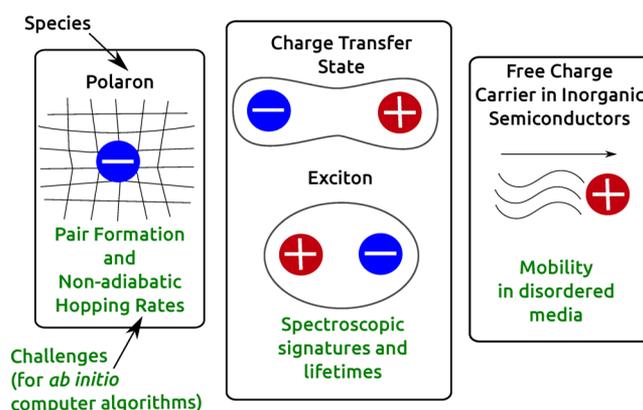
Another challenge in the computation of strongly correlated systems is finding the correct *ab initio* description of *d* and *f* bands in transition-metal oxides; standard XC functionals for plane-wave calculations overdelocalize key orbitals. To date, the method DFT+U offers numerous options to control orbital localization.<sup>32</sup> This requires judicious choice of the DFT+U model and a few parameters, which can be estimated by imposing electron occupation constraints.<sup>33</sup> We believe further

research in this area could result in assimilation of these powerful techniques by new generations of pure-density functional methods.

*Search for Methods for Excited-State Phenomena.* Excited-state dynamics models that significantly depart from the ground-state

Excited-state dynamics models that significantly depart from the ground-state regime are necessary to gain insight into transport at a junction.

regime are necessary to gain insight into transport at a junction. For example, following successful hot exciton splitting in polymer solar cells, transient species such as (hot) polarons may lie on excited state surfaces, with the ground-state playing a minor role. Although significant efforts and resources are invested in developing ground-state functionals, there are plentiful opportunities in extending DFT/CWFT methods and algorithms to accurately estimate dynamical properties of crucial relevance such as (Figure 2) charge-carrier lifetimes in



**Figure 2.** Schematic representation of relevant species in solar electricity generation and a few challenges for first-principles computational methods. Left, polaron species and relevant challenges: description of polaron-pair formation/dissociation rates and nonadiabatic hopping probabilities. Center, excitons: prediction of their absorption spectra and conversion into charge-transfer states and vice versa. Right, charge carrier: mobility in amorphous inorganic semiconductors.

inorganic, disordered, semiconducting environments; polaron (pair) formation and relaxation in organic molecular materials; exciton dynamics; charge transfer in electronically excited coherent systems, etc.

Regarding the interaction between electronic degrees of freedom in molecules and light, commonly available linear response TDDFT methods are frequently used to produce simplified, singly excited wave functions, which can be applied as a starting point in the study of excited-state dynamics. But, to date, improving this type of wave function is computationally expensive, and new methods are required to address this issue [excited states are more properly described by combinations of determinants, featuring multiple (double, triple, etc.) orbital excitations]. For exciton dynamics, even though it would be preferred to know the pure density matrix of excitons (or their collection of excited-state wave functions), the conventional linear response formalism only produces transition elements (e.g., dipole matrix elements) from ground to excited state. Pure

excited-state information (multipoles, polarizability tensors, etc.) of large systems by computational means remains elusive.

Producing the absorption spectrum of an excited state in an affordable manner is also challenging within TDDFT. Motivated by this, we noticed that a simple modification of the standard linear response TDDFT algorithm<sup>35</sup> enables prediction of the absorption spectrum of excitons. We proposed that by starting electronic propagation from a perturbed initial state one can extract information about transitions (by absorption/emission of a photon) between excited states. This idea is founded on exact quantum mechanics. In the context of TDDFT calculations, we perturb the ground-state orbitals using a weighted sum of virtual orbitals, where these weights are given by transition elements describing the transition from the ground state to a base excited state (that we choose). By reanalyzing the response of the system to a weak perturbing field, we calculate the strengths of electronic transitions from the base excited state. In practice, the algorithm needs only perturbation of orbitals and reapplication of the conventional linear-response eigenvalue problem. We tested this methodology against experimental data in oligomers<sup>34</sup> and small organic molecules.<sup>35</sup> For example, for poly(3-hexylthiophene) (P3HT) oligomers we noticed (Figure 3) the algorithm captures convergence of the

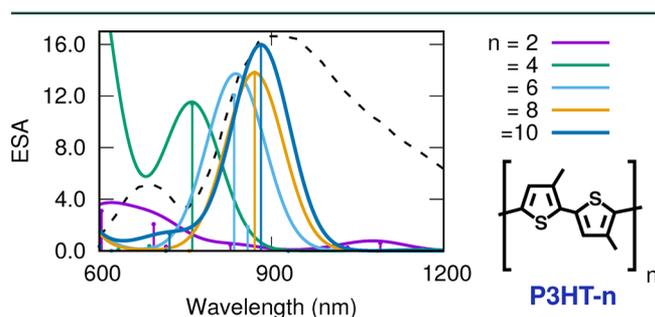


Figure 3.  $S_1$  absorption spectra of P3HT oligomers. The dashed and solid lines represent experimental data and broadened calculated absorption strengths, respectively. The vertical lines indicate positioning of the discrete oscillator strengths; the solid lines were broadened with Gaussian functions ( $\sigma = 50$  nm). Adapted from ref 34. Copyright 2017 American Chemical Society.

exciton absorption as the number of units is increased. Similarly, we observed close agreement with experiment for oligomers of poly(benzodithiophene-thieno[3,4-*b*]thiophene) (PTB7) and poly(2-methoxy-5-(2-ethylhexyloxy)-1,4-phenylenevinylene) (MEH-PPV). We are currently working on extending the mentioned quantum mechanical principle to estimate other properties such as nonadiabatic couplings, excited-state densities, and spectroscopy of coherent electronic superpositions.

**Challenges.** Obtaining spectroscopic information via full linear response TDDFT (with standard XC functionals) is slightly more expensive than a regular ground-state self-consistent cycle, with nearly cubic scaling. For systems with thousands of atoms this scaling could be reduced in a parameter-free fashion. It is challenging, however, to maintain or improve accuracy while simplifying the theory for cost-saving purposes. To date, density-functional tight binding offers some of the best (parameter-dependent) alternative protocols to achieve this goal.<sup>36</sup> With respect to fundamental challenges, excited-state dynamics needs new theory for implementing (excited-state) couplings of vibronic/phononic and magnetic (spin-orbit, for example) characters and couplings involving electronic wave function gradients (nonadiabatic couplings).

A different theory being currently explored that goes beyond the extent of linear response methods is real-time TDDFT.<sup>37</sup> Propagations in real time hold promise on providing more detailed information about the evolution of transient species of purely excited-state character. Standard XC potentials (which are based on ground-state DFT), however, miss important memory effects and lead to spurious shifts<sup>38</sup> in predicted transient spectra. In addition to reducing computational costs, eliminating over-relaxation effects in real-time TDDFT algorithms is an open challenge. Development in real-time TDDFT, we believe, might be complemented by TD CWFT algorithms, which are uncommon.

**New Theories, Algorithms, and Artificial Intelligence.** Computational packages depend on tens to hundreds of internal settings (tolerances, energy shifts, basis set size, etc.) that determine the cost and performance of algorithms. For materials discovery, the number of possible molecules is extremely large, as is the number of physicochemical conditions under which molecules interact. Guided by empirical rules and chemical principles, researchers select and control a careful combination of all these factors to study a finite set of systems of interest. By means of computer-assisted frameworks,<sup>39</sup> such as machine learning (ML), we can study large sets of potential materials and accelerate discovery. Provided a training set of input–output (descriptor–property) pairs, ML offers a wide selection of inexpensive functions that can identify patterns within the training set and determine internal rules to predict new patterns. This flexibility of ML tools makes them quite useful for materials discovery (Figure 4).

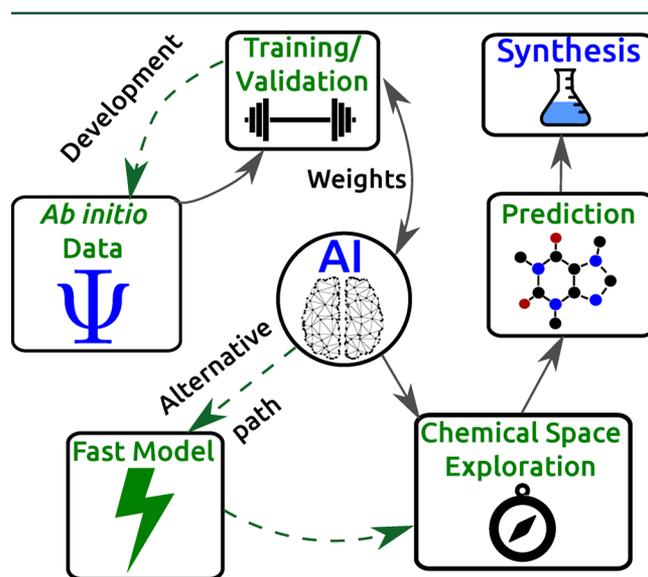


Figure 4. Basic steps in materials discovery fueled by quantum chemistry and AI. Starting with data determined by *first-principles* methods, an AI system is judiciously trained. Then, an exploration of the chemical space is carried out to propose a new material for synthesis. Traditionally, functionals and/or semiempirical models are developed by using known *ab initio* data in a closed loop between data and training, as in density functional development, for example. An alternative is indicated, which consists of using AI algorithms to generate fast phenomenological/semiempirical methods to explore the chemical space.

Materials discovery has been traditionally pursued by means of quantitative structure–property relationship models,<sup>40,41</sup> which directly associate single molecule structural descriptors

with properties of experimental interest, without requiring data from theoretical or computational sources. The success of these methods is inspiring development and application of AI-based tools that contribute with reliable quantum chemical methods,<sup>42</sup> and inexpensive kinetic models,<sup>43</sup> to improve the screening process and increase the amount of data that can be processed, i.e., high throughput. The most successful of these methodologies have been validated by experimental measurements.<sup>44</sup> As materials discovery becomes robust enough to process more complex phenomena and constantly increasing amounts of data, whether through theoretical and computational models and/or experimental measurements (which could be provided by automated laboratories), the screening moves from just a data-driven approach toward a closed-loop, high-throughput methodology. Although input–output relations can be studied by numerous AI techniques such as artificial neural networks, supervised ML, and clustering, single-molecule descriptors can be insufficient for many systems, especially those found in biology, where intermolecular interactions are dominant, so information about the molecular environment may be needed.<sup>45</sup>

A major focus of high-throughput computational discovery lies on artificial intelligence (AI) applications to find candidate materials or molecules for solar cells, electronic circuitry, and drug synthesis, among others. Depending on each case, this is achieved by optimizing properties such as optical gap, lifetimes, singlet–triplet splittings, etc.<sup>43,46–50</sup> In the case of OLEDs, for example, metal-free molecules can lead to significantly less expensive display screens. For example, ref 43 reports a successful integration of computational tools such as TDDFT and artificial neural networks and collaboration with synthetic chemists to produce promising blue OLED candidates. This methodology can of course explore larger chemical spaces and/or implement other forms of ab initio methods to achieve discovery of new candidates for synthesis.

To search for new materials, ML algorithms are trained with reliable physicochemical data (e.g., from DFT/CWFT calculations and/or experimental measurements) containing specific properties of different molecules. The new best material is searched by proposing and assessing candidates that share chemical trends, i.e., a set of descriptors, with the systems considered for training (for example, candidate molecules should have the same types of ligands used in the training library). AI algorithms for these applications combine chemoinformatics<sup>51–53</sup> and advanced data-processing techniques to achieve discovery. ML methods sharply accelerate the quest for candidates, whereas searching only with DFT/CWFT calculations would take much longer. We thus expect the use of ML will likely continue to grow.

At more fundamental levels, other applications of ML are aimed at speeding-up theoretical calculations.<sup>54,55</sup> For example, AI techniques may adapt simplified phenomenological models (tight-binding, Hubbard, etc.) for the computation of molecular properties.<sup>56–58</sup> Under this intriguing and interesting possibility, we believe phenomenological models could gain momentum by AI-enhanced capabilities to predict features of interest in semiconductors. For subgap problems, one can picture cases in which model parameters are estimated by a ML method according to the transient species of interest, so properties such as energy levels and coupling constants are properly tuned according to molecular environment. Furthermore, AI algorithms can be designed to speed up simulations of OPV blends by computing molecular transport networks and identifying efficient pathways for charge carriers to reach the electrodes.<sup>59</sup>

In the near future, an AI could also create a new density functional, ab initio technique, and/or proper phenomenological model for a given application.

The ideas above also suggest a scenario in which an algorithm decides or suggests internal parameters or settings<sup>60,61</sup> and type of density functional. In the near future, an AI could also create a new density functional, ab initio technique, and/or proper phenomenological model for a given application. However, a significant demanding aspect for future AI methods is to successfully scale up the system size (from atomic to cellular dimensions) while preserving accuracy and the description of emergent properties.<sup>62</sup>

Successful cases of AI optimization often involve computationally affordable (single-point energy) molecular libraries. As implied in previous sections, it is challenging to mine time-dependent ab initio data (especially nuclear dynamics<sup>63</sup>) of large systems because of high computational cost. In these cases we believe computer-generated theory could be useful, yet this remains unexplored to our knowledge. This path, however, needs strong synergy between rigorous theory and physical laws and conditions.

**Summary and Future Outlook.** We considered the advantages and challenges in wave function theory and density functional theories for applications in energy research. Besides the potential roles of artificial intelligence tools, there are numerous opportunities to develop first-principles computer algorithms to simulate optical, excited-state, and nuclear dynamics in disordered environments: molecular materials, imperfect inorganic crystals, liquid electrolytes, etc. For these cases, quantifying and classifying the behavior of pure excited-state species could be beneficial for the optimization of semiconductor performance.

On the basis of our discussion, we believe density-focused theory could evolve into new forms of computational chemistry with extended applicability, where future developments might be significantly assisted by artificial intelligence algorithms and new generations of electronic circuitry. Furthermore, computer-generated theoretical models might have a promising future because a large number of them can be created, tested, and screened according to their accuracy. However, generating models may require careful integration of wave function, density functional, and phenomenological theories and development of confidence estimation (which can be developed on the basis of high-level ab initio calculations and experimental validation).

Finally, future applications for new theoretical advancements may also include frontier topics such as quantum entanglement<sup>64</sup> and the engineering of quantum coherence,<sup>65</sup> which require methods for excited-state dynamics, to maximize energy conversion and storage. Progress in this context could benefit considerably from strong integration into large screening projects such as the materials genome initiative,<sup>66</sup> machine-learning-assisted materials discovery based on failed reactions,<sup>67</sup> Harvard clean energy project,<sup>50</sup> among others.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [ratner@northwestern.edu](mailto:ratner@northwestern.edu). Phone: +1-847-491-5371.

ORCID 

Martín A. Mosquera: 0000-0003-2170-5651

Kevin L. Kohlstedt: 0000-0001-8045-0930

George C. Schatz: 0000-0001-5837-4740

## Notes

The authors declare no competing financial interest.

## Biographies

**Martín Mosquera** is a Postdoctoral Associate in the Department of Chemistry, Northwestern University. He received his BS in Chemical Engineering at the Universidad del Valle, Colombia, and Ph.D. degree in Chemistry from Purdue University. His current research interests include excited-state dynamics, perovskite-like materials, and artificial intelligence (<http://martinmt.com>).

**Bo Fu** is a Ph.D. student in the Department of Physics and Astronomy at Northwestern University. His research interest is focused on using atomic-scale modeling and density functional theory to understand physical and chemical processes at the molecule–metal interface, including electrochemical process, sequential transmission, and Raman spectroscopy.

**Kevin Kohlstedt** is a Research Professor of Chemistry at Northwestern University. He received his BS from University of Kansas and Ph.D. from Northwestern University. His interests include using statistical models to understand mesoscale phenomena in disordered materials. He is on a long-term quest to extract the perfect cup of coffee (<http://sites.northwestern.edu/kevkohls/kevin-kohlstedt/>).

**George Schatz** is a Professor of Chemistry at Northwestern University. He received his BS degree in chemistry at Clarkson University and Ph.D. at Caltech. He specializes in the optical, structural, and thermal properties of nanomaterials, including plasmonic nanoparticles, catalysts, DNA and peptide self-assembled nanostructures, and carbon-based materials (<http://sites.northwestern.edu/schatz/>).

**Mark Ratner** obtained his BA and Ph.D. degrees at Harvard University and Northwestern University, respectively, and is an Emeritus Professor of Chemistry and of Materials Science and Engineering at Northwestern University. He is interested in structure and function at the nanoscale and the theory of fundamental chemical processes (<http://chemgroups.northwestern.edu/ratner/>).

## ACKNOWLEDGMENTS

Electronic structure methods development (M.A.M., M.A.R., and G.C.S.) was funded by the Department of Energy, Office of Basic Energy Sciences, under Grant DE-SC0004752. Machine learning applications (K.L.K., M.A.R., and G.C.S.) were supported as part of the ANSER center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001059. Electronic structure applications (B.F., G.C.S., and M.A.R.) were supported by the Air Force Office of Scientific Research (MURI Grant FA9550-14-1-0003).

## REFERENCES

- (1) Cheng, J.; Yu, K.; Libisch, F.; Dieterich, J. M.; Carter, E. A. Potential functional embedding theory at the correlated wave function level. 2. Error sources and performance tests. *J. Chem. Theory Comput.* **2017**, *13*, 1081–1093.
- (2) Snegov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. The polarizable embedding coupled cluster method. *J. Chem. Phys.* **2011**, *134*, 104108.
- (3) Kais, S. Quantum information for quantum chemistry. *Adv. Chem. Phys.* **2014**, *1*.

(4) Flocke, N.; Bartlett, R. J. A natural linear scaling coupled-cluster method. *J. Chem. Phys.* **2004**, *121*, 10935–10944.

(5) Chwee, T. S.; Szilva, A. B.; Lindh, R.; Carter, E. A. Linear scaling multireference singles and doubles configuration interaction. *J. Chem. Phys.* **2008**, *128*, 224106.

(6) Taube, A. G.; Bartlett, R. J. Rethinking linearized coupled-cluster theory. *J. Chem. Phys.* **2009**, *130*, 144112.

(7) Kedem, O.; Lau, B.; Ratner, M. A.; Weiss, E. A. Light-responsive organic flashing electron ratchet. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8698.

(8) Reimann, P.; Grifoni, M.; Hänggi, P. Quantum ratchets. *Phys. Rev. Lett.* **1997**, *79*, 10.

(9) Sutton, C.; Körzdörfer, T.; Gray, M. T.; Brunfeld, M.; Parrish, R. M.; Sherrill, C. D.; Sears, J. S.; Brédas, J.-L. Accurate description of torsion potentials in conjugated polymers using density functionals with reduced self-interaction error. *J. Chem. Phys.* **2014**, *140*, 054310.

(10) Körzdörfer, T.; Brédas, J.-L. Organic electronic materials: Recent advances in the DFT description of the ground and excited states using tuned range-separated hybrid functionals. *Acc. Chem. Res.* **2014**, *47*, 3284–3291.

(11) Hofmann, O. T.; Rinke, P.; Scheffler, M.; Heimel, G. Integer versus fractional charge transfer at metal/(Insulator)/organic interfaces: Cu/(NaCl)/TCNE. *ACS Nano* **2015**, *9*, 5391–5404.

(12) Liu, Z.-F.; Egger, D. A.; Refaely-Abramson, S.; Kronik, L.; Neaton, J. B. Energy level alignment at molecule-metal interfaces from an optimally tuned range-separated hybrid functional. *J. Chem. Phys.* **2017**, *146*, 092326.

(13) Subotnik, J. E.; Jain, A.; Landry, B.; Petit, A.; Ouyang, W.; Bellonzi, N. Understanding the surface hopping view of electronic transitions and decoherence. *Annu. Rev. Phys. Chem.* **2016**, *67*, 387–417.

(14) Rassolov, V. A.; Garashchuk, S. Computational complexity in quantum chemistry. *Chem. Phys. Lett.* **2008**, *464*, 262–264.

(15) Motta, C.; El-Mellouhi, F.; Kais, S.; Tabet, N.; Alharbi, F.; Sanvito, S. Revealing the role of organic cations in hybrid halide perovskite  $\text{CH}_3\text{NH}_3\text{PbI}_3$ . *Nat. Commun.* **2015**, *6*, 7026.

(16) Giorgi, G.; Fujisawa, J.-I.; Segawa, H.; Yamashita, K. Cation role in structural and electronic properties of 3D organic-inorganic halide perovskites: A DFT analysis. *J. Phys. Chem. C* **2014**, *118*, 12176–12183.

(17) Leguy, A. M.; Goñi, A. R.; Frost, J. M.; Skelton, J.; Brivio, F.; Rodriguez-Martinez, X.; Weber, O. J.; Pallipurath, A.; Alonso, M. I.; Campoy-Quiles, M.; et al. Dynamic disorder, phonon lifetimes, and the assignment of modes to the vibrational spectra of methylammonium lead halide perovskites. *Phys. Chem. Chem. Phys.* **2016**, *18*, 27051–27066.

(18) Li, W.; Liu, J.; Bai, F.-Q.; Zhang, H.-X.; Prezhdoo, O. V. Hole trapping by iodine interstitial defects decreases free carrier losses in perovskite solar cells: A time-domain ab initio study. *ACS Energy Lett.* **2017**, *2*, 1270–1278.

(19) Umari, P.; Mosconi, E.; De Angelis, F. Relativistic GW calculations on  $\text{CH}_3\text{NH}_3\text{PbI}_3$  and  $\text{CH}_3\text{NH}_3\text{SnI}_3$  perovskites for solar cell applications. *Sci. Rep.* **2015**, *4*, 4467.

(20) Nayyar, I. H.; Batista, E. R.; Tretiak, S.; Saxena, A.; Smith, D. L.; Martin, R. L. Localization of electronic excitations in conjugated polymers studied by DFT. *J. Phys. Chem. Lett.* **2011**, *2*, 566–571.

(21) Heimel, G. The optical signature of charges in conjugated polymers. *ACS Cent. Sci.* **2016**, *2*, 309–315.

(22) Li, W.; Carrete, J.; Katcho, N. A.; Mingo, N. ShengBTE: A solver of the Boltzmann transport equation for phonons. *Comput. Phys. Commun.* **2014**, *185*, 1747–1758.

(23) Fryer, D. S.; Peters, R. D.; Kim, E. J.; Tomaszewski, J. E.; de Pablo, J. J.; Nealey, P. F.; White, C. C.; Wu, W.-I. Dependence of the glass transition temperature of polymer films on interfacial energy and thickness. *Macromolecules* **2001**, *34*, 5627–5634.

(24) Kim, J.; Lee, S.-H.; Lee, J. H.; Hong, K.-H. The role of intrinsic defects in methylammonium lead iodide perovskite. *J. Phys. Chem. Lett.* **2014**, *5*, 1312–1317.

- (25) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291.
- (26) Bartlett, R. J. Coupled-cluster theory and its equation-of-motion extensions. *WIREs Comput. Mol. Sci.* **2012**, *2*, 126–138.
- (27) Bartlett, R. J.; Schweigert, I. V.; Lotrich, V. F. Ab initio DFT: getting the right answer for the right reason. *J. Mol. Struct.: THEOCHEM* **2006**, *771*, 1–8.
- (28) Hirata, S. Tensor contraction engine: Abstraction and automated parallel implementation of configuration-interaction, coupled-cluster, and many-body perturbation theories. *J. Phys. Chem. A* **2003**, *107*, 9887–9897.
- (29) Yang, W. Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **1991**, *66*, 1438.
- (30) Paldus, J.; Pittner, J.; Čársky, P. *Recent Progress in Coupled Cluster Methods*; Springer: New York, 2010; pp 455–489.
- (31) Roos, B. O. The complete active space self-consistent field method and its applications in electronic structure calculations. In *Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry Part 2*; Lawley, K. P., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2007; Vol. 69, pp 399–445.
- (32) Cococcioni, M. In *Correlated Electrons: From Models to Materials*; Pavarini, E., Koch, E., Anders, F., Jarrell, M., Eds.; Forschungszentrum Jülich: Jülich, Germany, 2012; Chapter 4.
- (33) Cococcioni, M.; De Gironcoli, S. Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 035105.
- (34) Mosquera, M. A.; Jackson, N. E.; Fauvell, T. J.; Kelley, M. S.; Chen, L. X.; Schatz, G. C.; Ratner, M. A. Exciton absorption spectra by linear response methods: application to conjugated polymers. *J. Am. Chem. Soc.* **2017**, *139*, 3728–3735.
- (35) Mosquera, M. A.; Chen, L. X.; Ratner, M. A.; Schatz, G. C. Sequential double excitations from linear-response time-dependent density functional theory. *J. Chem. Phys.* **2016**, *144*, 204105.
- (36) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a sparse matrix-based implementation of the DFTB method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.
- (37) Marques, M. A.; Castro, A.; Bertsch, G. F.; Rubio, A. Octopus: a first-principles tool for excited electron–ion dynamics. *Comput. Phys. Commun.* **2003**, *151*, 60–78.
- (38) Provorse, M. R.; Habenicht, B. F.; Isborn, C. M. Peak-shifting in real-time time-dependent density functional theory. *J. Chem. Theory Comput.* **2015**, *11*, 4791–4802.
- (39) Gil, Y.; Greaves, M.; Hendler, J.; Hirsh, H. Amplify scientific discovery with artificial intelligence. *Science* **2014**, *346*, 171–172.
- (40) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev.* **1995**, *24*, 279–287.
- (41) Bajorath, J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comp. Sci.* **2001**, *41*, 233–245.
- (42) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027–1044.
- (43) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **2016**, *15*, 1120–1127.
- (44) Hautier, G.; Jain, A.; Ong, S. P. From the computer to the laboratory: materials discovery and design using first-principles calculations. *J. Mater. Sci.* **2012**, *47*, 7317–7340.
- (45) Olivares-Amaya, R.; Amador-Bedolla, C.; Hachmann, J.; Atahan-Evrenk, S.; Sanchez-Carrera, R. S.; Vogt, L.; Aspuru-Guzik, A. Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics. *Energy Environ. Sci.* **2011**, *4*, 4849–4861.
- (46) Häse, F.; Valleau, S.; Pyzer-Knapp, E.; Aspuru-Guzik, A. Machine learning exciton dynamics. *Chem. Sci.* **2016**, *7*, 5139–5147.
- (47) Meredig, B.; Agrawal, A.; Kirklín, S.; Saal, J. E.; Doak, J.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 094104.
- (48) Meredig, B.; Wolverton, C. Dissolving the periodic table in cubic zirconia: data mining to discover chemical trends. *Chem. Mater.* **2014**, *26*, 1985–1991.
- (49) Hachmann, J.; Olivares-Amaya, R.; Jinich, A.; Appleton, A. L.; Blood-Forsythe, M. A.; Seress, L. R.; Roman-Salgado, C.; Treppe, K.; Atahan-Evrenk, S.; Er, S.; et al. Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry—the Harvard clean energy project. *Energy Environ. Sci.* **2014**, *7*, 698–704.
- (50) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2011**, *2*, 2241–2251.
- (51) Varnek, A.; Baskin, I. I. Chemoinformatics as a theoretical chemistry discipline. *Mol. Inf.* **2011**, *30*, 20–32.
- (52) Duch, W.; Swaminathan, K.; Meller, J. Artificial intelligence approaches for rational drug design and discovery. *Curr. Pharm. Des.* **2007**, *13*, 1497–1508.
- (53) Goldman, B. B.; Walters, W. P. Machine learning in computational chemistry. *Annu. Rev. Comput. Chem.* **2006**, *2*, 127–140.
- (54) Yao, K.; Herr, J. E.; Parkhill, J. The many-body expansion combined with neural networks. *J. Chem. Phys.* **2017**, *146*, 014106.
- (55) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **2013**, *15*, 095003.
- (56) Arsenault, L.-F.; Lopez-Bezanilla, A.; von Lilienfeld, O. A.; Millis, A. J. Machine learning for many-body physics: the case of the Anderson impurity model. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 155136.
- (57) Dral, P. O.; von Lilienfeld, O. A.; Thiel, W. Machine learning of parameters for accurate semiempirical quantum chemical calculations. *J. Chem. Theory Comput.* **2015**, *11*, 2120–2125.
- (58) Liu, J.; Qi, Y.; Meng, Z. Y.; Fu, L. Self-learning Monte Carlo method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, 041101.
- (59) Gagorik, A. G.; Savoie, B.; Jackson, N.; Agrawal, A.; Choudhary, A.; Ratner, M. A.; Schatz, G. C.; Kohlstedt, K. L. Improved scaling of molecular network calculations: the emergence of molecular domains. *J. Phys. Chem. Lett.* **2017**, *8*, 415–421.
- (60) Szaleniec, M.; Witko, M.; Tadeusiewicz, R.; Goclon, J. Application of artificial neural networks and DFT-based parameters for prediction of reaction kinetics of ethylbenzene dehydrogenase. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 145–157.
- (61) Balabin, R. M.; Lomakina, E. I. Neural network approach to quantum-chemistry data: accurate prediction of density functional theory energies. *J. Chem. Phys.* **2009**, *131*, 074104.
- (62) Luisi, P. L. Emergence in chemistry: chemistry as the embodiment of emergence. *Found. Chem.* **2002**, *4*, 183–200.
- (63) Cheung, D. L.; Troisi, A. Modelling charge transport in organic semiconductors: from quantum dynamics to soft matter. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5941–5952.
- (64) Monroe, C. Quantum information processing with atoms and photons. *Nature* **2002**, *416*, 238–246.
- (65) Scholes, G. D.; Fleming, G. R.; Chen, L. X.; Aspuru-Guzik, A.; Buchleitner, A.; Coker, D. F.; Engel, G. S.; van Grondelle, R.; Ishizaki, A.; Jonas, D. M.; et al. Using coherence to enhance function in chemical and biophysical systems. *Nature* **2017**, *543*, 647–656.
- (66) Jain, A.; Persson, K. A.; Ceder, G. Research update: the materials genome initiative: data sharing and the impact of collaborative ab initio databases. *APL Mater.* **2016**, *4*, 053102.
- (67) Raccuglia, P.; Elbert, K. C.; Adler, P. D.; 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.