Korean J. Chem. Eng., 38(10), 1971-1982 (2021)
DOI: 10.1007/s11814-021-0869-2

pISSN: 0256-1115
eISSN: 1975-7220

INVITED REVIEW PAPER

Machine learning-based discovery of molecules, crystals, and composites: A perspective review

Sangwon Lee[‡], Haeun Byun[‡], Mujin Cheon, Jihan Kim[†], and Jay Hyung Lee[†]

Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea

(*Received 13 April 2021 • Revised 14 June 2021 • Accepted 15 June 2021*)

Abstract−Machine learning based approaches to material discovery are reviewed with the aim of providing a perspective on the current state of the art and its potential. Various models used to represent molecules and crystals are introduced and such representations can be used within the neural networks to generate materials that satisfy specified physical features and properties. For problems where large database for structure-property map cannot be created, the active learning approaches based on Bayesian optimization to maximize the efficiency of a search are reviewed. Successful applications of these machine learning based material discovery approaches are beginning to appear and some of the notable ones are reviewed.

Keywords: Material Discovery, Machine Learning, Molecule Design, Crystal Design, Adaptive Experimental Design, Bayesian Optimization

INTRODUCTION

In the past few years, there has been increasing interest in research pertaining to applying machine learning to material discovery, spurred by the recent advancements in deep learning [1-3] as well as the large amount of available databases of molecules and materials [4]. Given the importance of artificial intelligence and big data, data-driven discovery is already thought as the "fourth paradigm of science" [5]. The aim of this perspective article is to summarize the recent advancements in the field of machine learning and applications to materials discovery. We first present the different contexts where machine learning tools are applied to materials discovery, where implementations can vary depending on the nature of the inputs and the amount of data available. Also, relevant higher-level machine learning concepts are presented with an overview of some widely used representations of molecules and materials, and some of the relevant applications are summarized. For further illustration, designs of crystalline materials and composites are specifically discussed.

1. Machine Learning and Material Discovery

Machine learning is a field of computer science that extracts knowledge and patterns from a set of data. In supervised learning, the best-known branch of machine learning, one tries to map the input representation to the desired outputs in myriad of different applications (e.g., classification of the number images). In general, classic machine learning, without deep learning, uses high-level, manually constructed features, while deep learning uses relatively low-level features. To illustrate this difference, an example in materials science can be drawn where higher-level abstract concepts such as void fraction, density, surface area of materials can be used as inputs in classical machine learning, whereas modern machine learning techniques like the deep neural networks use the actual positions of the atoms that comprise the materials. In general, high-level features are often material properties (e.g., density, void fraction) that can be calculated from low-level features (e.g., atomic positions, bond connectivity). This difference stems from the fact that deep learning uses multiple layers to build a hierarchy of features, starting from low-level features like atom positions to higher ones. This can be an important advantage, as the extraction of low-level features from the data is easier than that of high-level features. The power of deep learning comes from the fact that high-level features can be automatically learned (i.e., representation learning) during the training process. It should be pointed out that the features of learned representations from the deep learning model may not be human-interpretable like a property of the materials, but nonetheless be still useful because it contains the compressed data information. The schematic that illustrates the high-level relationship between machine learning and material science is shown in Fig. 1.

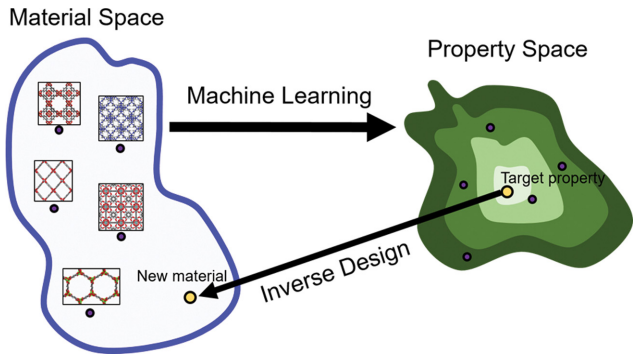

Fig. 1. Abstraction of relationship between machine learning and materials.

[†]To whom correspondence should be addressed.
E-mail: jihankim@kaist.ac.kr, jayhlee@kaist.ac.kr
[‡]Equally contributing authors.

As seen from Fig. 1, machine learning can map the material space to the property space (e.g., drug similarity, surface area, formation energy or abstract vectors). By using the mapping functionality, one can predict desired properties directly or find/generate materials of desired property via inverse mapping from property space to material space (inverse design).

2. Contexts of Material Discovery Problems

Material discovery problems can vary widely in context and complexity. In simple cases like composite design, the input space may simply be the composition of various known components of the composite, e.g., resins, hardener, accelerant, etc. The input may also include some operating parameters (e.g., temperature) relevant for the manufacturing process. More generally, one may try to design a new material with a specific set of targeted properties. In the case of crystalline materials, the problem is more complicated, compared to simple molecules, as such materials have more parameters (e.g., angles) and the generation algorithm needs to abide by periodic boundary conditions. A typical input space for such a problem may be huge, given the large choices of atoms as well as branching structures. Efficient handling of the large input space is where machine learning can be primarily helpful, as we saw in the case of AlphaGo [3].

Yet, another important challenge is to relate the molecular structure to targeted properties, i.e., to build the structure-property relationship. For this, experimental data and molecular simulations can be used. However, depending on the property being targeted, such data may not be readily available, at least not in the quantities needed, or easy to produce, thus impeding the discovery process. For example, in composite design, it may be very difficult to gather a large set of data to build a recipe-property relationship that is valid over a wide range. If sufficient structure-property data cannot be accessed, the structure-property map built from the limited data would carry significant uncertainties, which may be quantified and used in experimental design or running time-consuming molecular simulations. In this case, the role of machine learning may be to suggest a set of candidates to try in experiment or molecular simulation, in an iterative manner, so that the whole discovery process can be sped up. As such, the key would be to control the *exploration vs exploitation* balance, which requires a model that carries the uncertainty information.

MOLECULE AND CRYSTAL DESIGN

1. Representation of Molecules and Crystals

In most machine learning problems, the proper representation of the input to the machine learning model is essential in improving the performance. It is not intuitively obvious on how to best represent molecules and crystals, and as such, this is still the majority of on-going research. In general, molecules and crystals are different in that the former can be accurately described using a finite number of atoms, whereas the latter is, in principle, an infinitely extended material with periodic boundary conditions. As such, different representations are being investigated for what is essentially two different types of materials. Although human extracted high-level features, such as pore diameter and surface area of porous materials, can be used to train machine learning models, we focus more on the end-to-end atomic level representations in this section.

In 2012, Rupp et al. created a simple molecular representation named Coulomb matrix (CM) [6], where the CM is constructed using the pairwise distance of atoms and the nuclear charges of the atoms in the molecule. CM is invariant with respect to translation and rotation because the distances between atoms do not change for these operations. Using the CM representation, in their work, molecular atomization energy was successfully predicted with kernel rigid regression (KRR) [7]. Hansen et al. suggested an extension of the CM named "bag of bonds" (BoB) [8], which was inspired by the concept of bag-of-words used in natural language processing (NLP). In this representation, the elements of CM are stored in a specific bag according to particular bond types (e.g., C-C, C-N) and the stored values are concatenated with zero padding with equal size for all the bags. The simplified molecular-input line-entry system (SMILES) [9,10], which has already been widely used in biochemical society, is another popular representation of molecules. Useful concepts and machine learning model of NLP (e.g., recurrent neural network (RNN)) can be applied to the SMILES representation because it can be treated as similar to language with its own unique syntax. In addition, the SMILES representation is actively used for generative models (e.g., generation of small molecules with user-desired properties) because facile conversion from SMILES can be to molecular structures and vice versa. More recently, molecular graph [11,12] representation has been gaining considerable traction, where the nodes and edges in a molecular graph are represented by vectors that contain information of the chemical (e.g., atom types, bond types, bond lengths). Fundamentally, the graph representation of molecules can be determined uniquely. However, the permutation variant of the adjacency matrix of graphs can be a problem from the perspective of machine learning, because this problem leads to the non-uniqueness of the materials. Recent advances in graph neural networks [13,14] have solved the problems. In graph neural networks, the invariances of graphs are imposed by the neural network architecture. Other than these, there are other types of input representations such as the smooth overlap of atomic positions (SOAP) [15] and Fourier series of atomic radial distribution (FR) [16], and more likely, this will be a field of interest for the foreseeable future.

Compared to small molecules, the representation of crystalline materials can be more complicated because of periodicity in crystals and the non-uniqueness of the unit cell selections. These problems are resolved generally by considering the chemically local environment with the treatment of periodic boundary condition and obtaining normalized quantities within that environment. Valle et al. suggested a crystal fingerprint that relates pair correlation function and diffraction patterns [17]. The fingerprints of the crystals are obtained by calculating distances between different atom types and concatenating over the dimension for all possible atom type pairs. The similarity between different crystal structures can be measured using the crystal fingerprint with the cosine similarity metric. Schütt et al. suggested using partial radial distribution function (PRDF) [18] as a suitable representation for crystalline systems. The PRDF is obtained by calculating the radial distribution function (RDF) of all possible element pairs and concatenating the RDFs.

Some researchers took representations that worked on small molecules and expanded their usage to crystal systems. Faber et al. [19] extended CM to Ewald sum matrix, extending Coulomb-like matrix and sine matrix to account for periodic systems. The Ewald sum matrix and extended Coulomb-like matrix take into account long-range interactions in the periodic system via the Ewald summation. The sine matrix replaces the long-range electrostatic interaction to arbitrary function that is periodic with respect to the lattice vector. Xie et al. extended the molecular graph to the periodic system using undirected multigraph and named it "crystal graph" [20]. The crystal graph considers the periodicity when the edges are built. To utilize the crystal graph as an input, they developed an ANN architecture named crystal graph convolutional neural network (CGCNN). The invariance of unit cell choice is attained by the normalization of the pooling operation in the CGCNN. The CGCNN can take different sized inputs, such as different number of atom types and number of atoms, and encode them to the same size vector, which allows comparisons of diverse type of input crystals.

It is conceivable that the representation can vary based not only in their periodicity but their properties. One way to utilize machine learning in chemistry is to predict the potential energy surface using artificial neural networks and derive the neural network potential (NNP). In the area of NNP, the symmetry function [21-23] is widely used, and it is designed to predict the potential energy and atomic forces of the molecule. As such, the symmetry function is differentiable with respect to atomic position in order to calculate the forces from the potential energy. The symmetry function is calculated from the local environment of the atoms, and the representation also can be used in crystal system if the building of neighbors list of the atoms considers the periodic boundary condition. Even though the symmetry function aims to predict the potential energy surface, this representation can also predict other properties. The variants of the symmetry function have been also suggested [24,25].

It is possible that the neural network can predict potential energy surface without symmetry functions from the basic information regarding the elements and atomic configurations, where the materials representation also can be incorporated into the ANN architecture. Schütt et al. developed a specialized neural network architecture named SchNet [26,27] that uses embeddings of atoms for representation, which are trained during the learning process. In the SchNet, the local environment of atoms within the cutoff radius is considered in special network architecture named interaction blocks and filter-generating network. Due to the flexibility of SchNet and the representation, it can be used for variable number of species and number of atoms and periodic system. In addition, SchNet has better performance than other models for larger size of training set due to the characteristics of the representation learning.

2. Structure-property Maps

In practice, one can list all the important properties inherent to a given molecule or a material and repeat this exercise for a large database of molecules/materials to accumulate a large dataset that can potentially let us formulate a coherent relationship between the "structure" and its "property". With the facility in which one can obtain data, both computationally via molecular simulations and experimentally through accumulation of data, many researchers have used this information to construct structure-property maps for many different molecules/materials. For example, in the case for porous materials and zeolites in particular, Lin et al. used data from large-scale screening of over 100,000 materials to unearth structure-property relationship between the zeolite structures and various properties relevant to carbon dioxide capture [28]. There have been many studies conducted on metal-organic frameworks (MOFs) to elucidate the relationship between the large chemical space found within MOFs and some of the properties that relate to various applications such as gas storage and separations [29-31].

Due to the linkage that exists between the structures and the properties, one can exploit this relationship to inversely design molecules/materials that contain the user-desired properties. In more conventional direct analysis, one starts from the molecules/materials and then derives the associated desired properties. Given that the direction of the inverse design is opposite to the direct analysis, one can conceivably save significant time and effort in finding and focusing on just the user-desired molecules/materials without wasting resources in analyzing sub-optimal chemical structures. With the advent of machine learning/deep learning models, there has been much more effort to generate molecules/materials via VAE, GAN, RL and other neural networks to design the materials inversely. This type of workflow will continue to be popularized in many applications, including drug discovery, photovoltaics, and other applications.

3. Applications

Scientific and engineering applications for nanomaterials are quite diverse and can differ based on the type of material (e.g., inorganic crystals, organic crystals, nanoporous materials, proteins, drugs). For small molecules, the application that interests AI researchers is drug discovery. Due to the enormously large chemical space for drug molecules [32], inverse design with machine learning is an enticing tool to facilitate the discovery of useful drugs. Using the SMILES representation of drug molecules, research work for both the forward model (property prediction) and inverse model (drug discovery) has been widely demonstrated [33]. More recently, inverse molecule design for drugs is being actively studied with the introduction and success of generative models in deep learning. Some of these generative models include generative adversarial network (GAN) [34], variational autoencoder (VAE) [35], and generative recurrent neural network (RNN). Gómez-Bombarelli et al. developed a VAE that consists of RNN encoder and decoder with the SMILES representation [36]. Using this ANN, they successfully generated molecules with user-desired property by optimizing the encoded molecule vectors in the latent space (Fig. 2(a)). Many different generative models are being continuously developed to produce molecules with user-desired properties, making this an exciting field for research [33,37-47].

For solid and inorganic materials, many works have been conducted on structure, composition, property prediction and inverse design. Zhou et al. [48] developed an unsupervised machine learning model named Atom2Vec, in which they extracted the important features from the atomic environment of the crystal from the materials database. The learned atomic vectors captured the conventional trends of atoms (e.g., family and valence trend), and using these vectors, the formation energy of elpasolites ABC_2D_6 was well

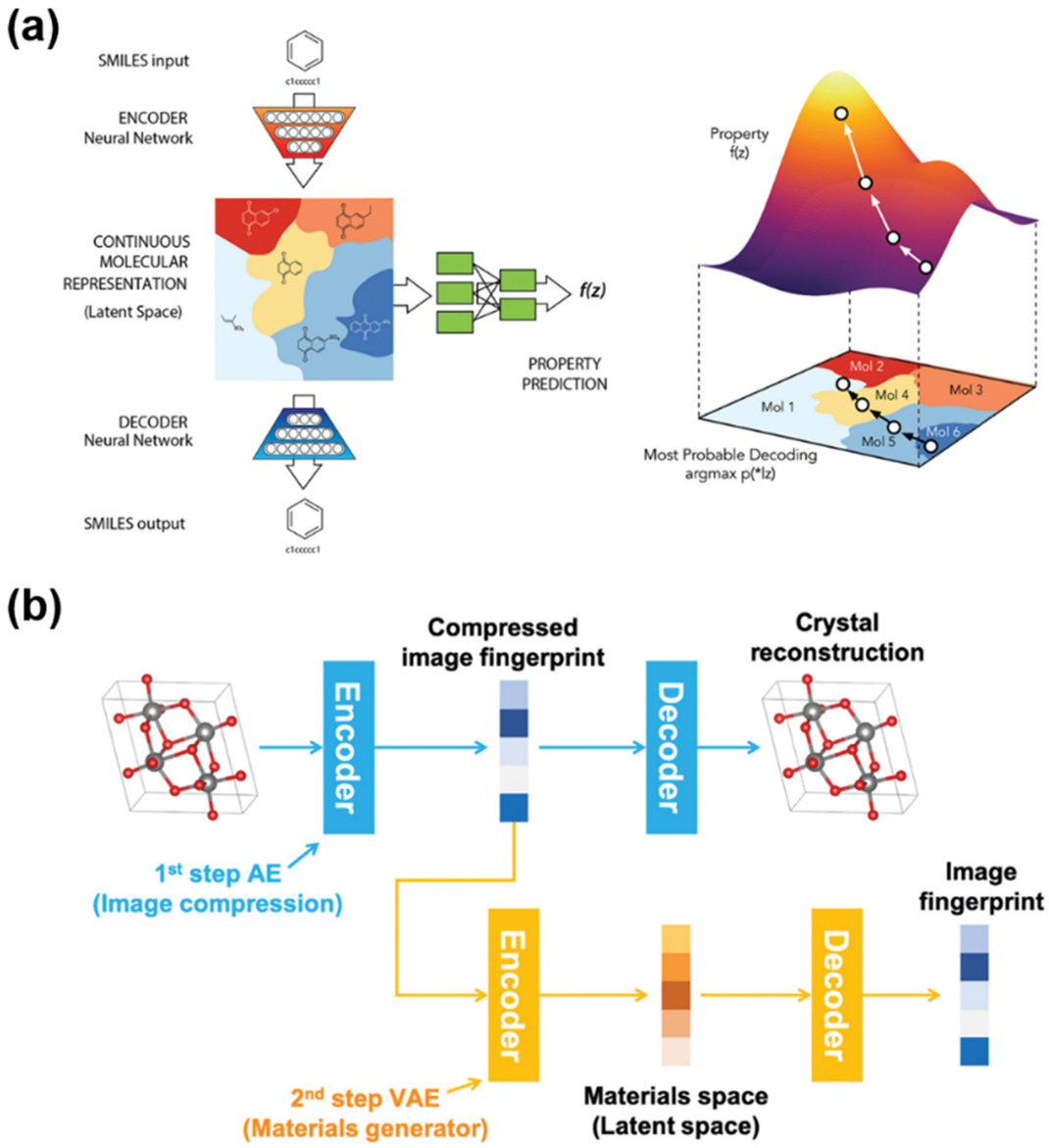

Fig. 2. (a) Inverse design of small molecules. SMILES strings are encoded to the latent vector and optimized for optimal property over latent space. Then the optimal latent vector is decoded to SMILES. Figure adapted from Ref. [36] Copyright 2018 with permission from American Chemical Society. (b) ANN architecture for the inverse design of inorganic crystals. The hierarchical VAE encodes 3D image obtained from inorganic crystal to latent vector. The latent vector is optimized for optimal property and the optimal latent vector is decoded to 3D image. Figure adapted from Ref. [50] Copyright 2019 with permission from Elsevier.

predicted with concatenated atom vector as representation (concatenation of atom vectors of A, B, C and D). Ziletti et al. [49] demonstrated that their deep learning model can predict the crystal structure for both perfect crystal and defective crystals, while conventional classification method was shown to be highly sensitive to the imperfection of the crystals (e.g., random displacements and vacancies).

Recently, inverse design studies using generative models for inorganic materials have been conducted. Noh et al. [50] developed for the first time a VAE based model for inverse design of inorganic solid. They used 3D images as the representation and these images are calculated by applying Gaussian function to crystal structures. In order to utilize the 3D images for inverse design, they developed the Image-Based Materials Generator (iMatGen), which is a hierarchical two-step VAE (Fig. 2(b)). Inverse design using iMatGen was applied to vanadium oxide materials for proof of concept and forty completely new V_xO_y structures were found. Kim et al. [51] developed a GAN-based model that is memory-efficient and uses inversion-free representation. They directly used fractional atomic coordinates concatenated to cell parameters that is similar to the point cloud. Their model was applied to predict promising Mg-Mn-O ternary materials and twenty-three new structures with reasonable stability and band gap were discovered. In addition to inverse design, a study has also been conducted to predict the synthesizability of inorganic solids. Jang et al. [52] developed for the first time a deep learning model that can predict the

synthesizability of inorganic materials from the structure. They used CGCNN as a model and applied positive and unlabeled machine learning techniques for the training of the model, because only positive labels exist due to the nature of experimental data. The CGCNN was trained to predict crystal-likeness scores (CLscore), a metric for synthesizability, and showed 86.2% true positive rate for experimental papers published between 2015 and 2019. It also confirmed that 71 of the 100 high-scoring virtual substances were already synthesized and published in the literature.

Nanoporous materials such as zeolites, metal-organic frameworks (MOFs), and covalent organic frameworks (COFs) contain many pores that are on the order of few Angstroms to nanometers. Due to the large surface area and pore volume, these materials are seen as promising materials for wide range of energy and environmental related applications. Fernandez et al. introduced large-scale quantitative structure-property relationship (QSPR) to predict the methane uptake in the MOFs [53] and predict high performing carbon capture MOFs [54]. Simon et al. predicted Xenon/Krypton selectivity using random forest algorithm with structural descriptors such as void fraction and crystal density [55]. In addition, they introduced a new descriptor called Voronoi energy that provides the average energy of Xenon at the accessible Voronoi nodes. They conducted hybrid screening by combining molecular simulation and the developed prediction model. From the screening result, they found two most selective materials that are already synthesized but not tested for the Xe/Kr separations. Chung et al. developed an efficient searching algorithm on an MOF database to find high performing pre-combustion CO_2 capture MOFs [56]. In this work, MOFs are represented by building block components such as the type of inorganic node, organic linkers, functional groups and information of interpenetration. These descriptors are optimized using genetic algorithm to find the high performing MOFs. The algorithms are shown to be more efficient than brute force search, and one of discovered MOFs showed better CO_2 working capacity than any previously published MOF under the process conditions used in this study.

Machine learning can also be used to predict the atomic property of atoms in MOFs. Raza et al. developed a deep learning model that can predict partial charge of atoms in MOFs [57]. The crystal graphs were used as representations, and message passing natural network (MPNN) was used as a machine learning model to consider local bonding environments in MOFs. The MPNN were trained using DFT-derived DDEC charges as training data and showed reasonable prediction performance and fast evaluation speed (mean absolute deviation on test set: 0.025, runtime: ~3 s).

Recently, there have been attempts to conduct inverse design of MOF using generative models. Yao et al. presented for the first time an MOF inverse design methodology using generative models [58]. Due to the large size and complexity of the MOF structures, they used a different approach other than using the atomic information directly. They decomposed the MOFs into topology, metal cluster vertex, organic vertex and organic edge. The first three components are represented as the categorical variable from already known data (e.g., topology database), and for organic edges, the graph representation is used as normal small molecules. This representation limits diversity because it uses only existing topologies and vertices, but can be considered reasonable for complex materials such as MOFs. They developed a supramolecular variational autoencoder (SmVAE) that can map these representations to latent space. The MOFs were optimized in the latent space to find MOFs to be used in CO_2 separation applications. As a result, promising MOFs were found that are competitive against best-performing MOFs and zeolites ever reported.

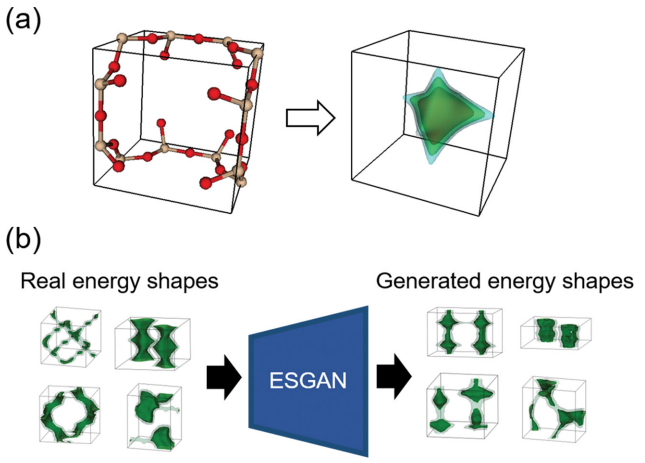

Fig. 3. (a) The energy shape calculation of the zeolite. The green surface represents the potential energy contour between methane gas and the zeolite. The interaction energy is calculated using Lennard-Jones potential, (b) Schematics of energy shape generation of ESGAN.

Recently, Lee et al. predicted the performance limit of methane gas storage in zeolites by generating a hypothetical potential energy surface using developed generative adversarial network architecture (ESGAN) [59]. In the work, the interaction energy between methane and the zeolite is stored in a three-dimensional grid (energy shape) (see Fig. 3(a)) and the ESGAN is trained to generate an energy shape similar to the real one (see Fig. 3(b)). For the generated energy shape, grand canonical Monte Carlo simulations were conducted to obtain the working capacity. Even without high performing zeolites in training set, ESGAN successfully generated high performing energy shape with almost top performance. Kim et al. developed unique ANN called ZeoGAN that can generate nanoporous zeolite structures [60]. The ZeoGAN is based on the ESGAN architecture, but unlike its predecessor, ZeoGAN takes the material shape as an input additionally (see Fig. 4). The material shape is a three-dimensional grid that contains atomic position smeared by Gaussian function. They also successfully generated zeolites with user-desired property (in this work, heat of adsorption) by adding a penalty function to the generator of ZeoGAN.

In general, given the high computational cost associated as well as millions of materials available, machine learning can provide an alternative way to computationally identify the optimal materials for the given application. It remains to be seen whether the various approaches that do not utilize AI/machine learning can be compatible or adversarial to the current machine learning works.

MATERIALS BY ADAPTIVE DESIGN

In the previous section, we summarized the machine learning

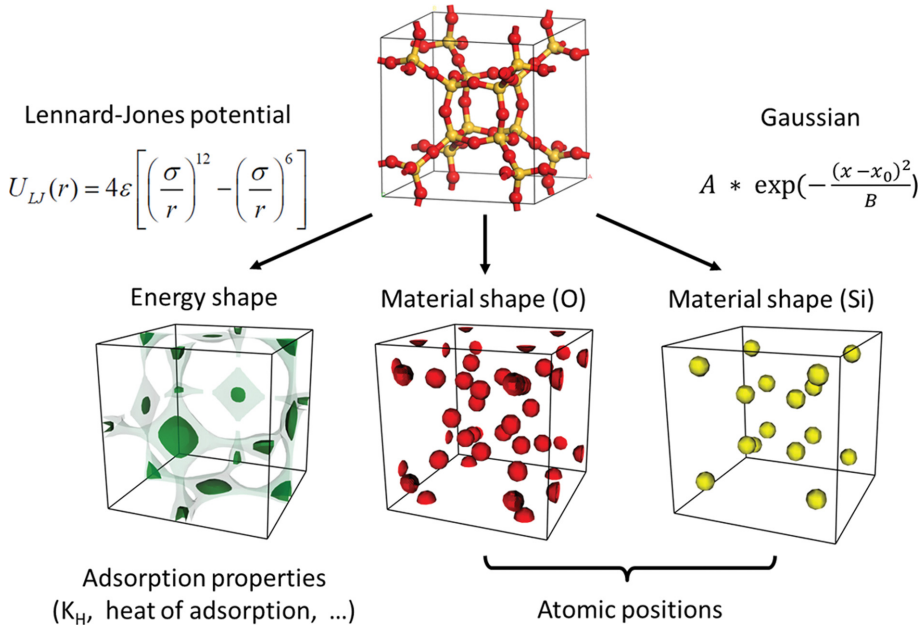

Fig. 4. The calculation of energy shape and material shape in ZeoGAN. The energy shape is calculated like ESGAN and the material shapes are calculated using the Gaussian function (highest value at atomic positions).

approaches for predicting various material properties or even designing new molecular structures given large databases, for example, those generated by high throughput DFT calculations. However, in practice, many material problems involve multicomponent compounds, solid solutions, and defects, which the current state-of-the-art machine learning tools are not best suited to handle [61]. In such a case, it may be very difficult to build a large database from high throughput calculations, as such simulations may not be possible or demand high computational resources. Alternatively, one can generate data by experiments but experimental datasets are typically limited since conducting experiments can be time-consuming and expensive. Therefore, a method to efficiently search materials with target properties given a small dataset is needed.

One of the emerging methodologies to address this issue is *Bayesian optimization* (BO), also studied under the name of *surrogate-based optimization* [62] and *efficient global optimization* [61]. Bayesian optimization is an experimental design strategy which adaptively guides the next experiments, or calculations, by balancing the merits of exploring high-uncertainty regions that have not been previously sampled and exploiting previously explored regions where materials with good properties can be found with a high level of certainty. It provides a general framework for iteratively finding the desired materials with the minimum number of experiments. Bayesian optimization has been successfully applied to design problems of various materials including organic-inorganic molecules, alloys, functional oxides, and polymers [61,63,64]. In the following, the basic concept of the BO is introduced in section 3.1, and its extensions for further effective experimental design are surveyed in section 3.2. Finally, several BO applications in materials design are presented in section 3.3.

1. Bayesian Optimization

Bayesian optimization, which is a strategy for finding a global optimum of an objective function as efficiently as possible, is usually used when it is expensive and time-consuming to evaluate the objective function. The high efficiency of Bayesian optimization in terms of data requirements has been shown in several articles [65-67]. This efficiency is achieved by optimizing the balance between exploration and exploitation of the search space based on Bayes' theorem. Bayes' theorem states that the posterior probability of model M from observed data E is related to the likelihood of the data and the prior belief of the model. Mathematically, the theorem says.

$$P(M|E) \propto P(E|M)P(M).$$

This provides a way to quantify the posterior distribution of the model starting from the prior belief by combining it with the likelihood of the observations. Bayesian optimization has two main components to in using the theorem: a surrogate model and an acquisition function.

A surrogate model tries to mimic an objective function from data/observations to represent the unknown objective function over the design space X. A more commonly used form of the surrogate model is a Gaussian process (GP). Gaussian process models provide the prediction mean m(x) as well as its variance σ(x) over the design space based on given data. A Gaussian process is defined by its mean function m(x) and covariance function k(x, x′):

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad x, x' \in X.$$

The mean function m(x) represents the expected value of function f at input x, whereas the covariance function k(x, x′), often referred to as the kernel function, expresses the smoothness of the function, i.e., how much it varies from the mean and how its deviations are correlated. The most popular choice for the kernel functions is the radial basis function (RBF). The RBF kernel has the

following mathematical form:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\theta^2}|\mathbf{x}-\mathbf{x}|^2\right)$$

The above implicitly assumes that the correlation decreases in a negative exponential manner as the distance between two points $\mathbf{x}$ and $\mathbf{x}'$ increases. The length-scale parameter θ is a hyperparameter which has to be tuned for each application. Since the RBF is infinitely differentiable, the implicit assumption is that the underlying objective function is smooth [68].

The acquisition function suggests next points for experiment, or simulation, based on the information provided from the surrogate model constructed using the prior collected data. The recommendation is made to balance exploration (trying points in a new region, which usually comes with huge uncertainty) and exploitation (optimizing within the region of prior data, which provides predictions of high confidence), which enables active learning of the objective function. For this, acquisition functions are designed to be large near the points potentially having high values of the objective function. There are three most commonly used forms of the acquisition function (Fig. 5): possibility of improvement (PI), expected improvement (EI), and upper confidence bound (UCB). Among three candidates, the most popular choice is the EI, which is the expected size of improvement upon the currently found maximum or minimum:

$$\mathrm{EI}(\mathrm{x}) = \mathbb{E}(\max\{\mathrm{f}(\mathbf{x}) - \mathrm{f}(\mathbf{x}^+), 0\})$$

where $\mathrm{f}(\mathbf{x}^+)$ is the value of the best sample so far. Jones et al. [69] derived the following closed form expression of the EI:

$$\mathrm{EI}(\mathbf{x}) = (\mu(\mathbf{x}) - \mathrm{f}(\mathbf{x}^+))\Phi(\mathrm{Z}) + \sigma(\mathbf{x})\phi(\mathrm{Z})$$

$$\text{where } \mathrm{Z} = \frac{\mu(\mathbf{x}) - \mathrm{f}(\mathbf{x}^+)}{\sigma(\mathbf{x})}$$

where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the mean and the standard deviation of the GP posterior at $\mathbf{x}$, respectively. $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard cumulative distribution function and the standard normal probability density function, respectively. The EI acquisition function consists of two terms: the first term can be increased by increasing the mean and the second term can be increased by increasing the variance. This explicitly encodes a balance between exploitation and exploration, respectively. A popular variant of the EI function is the knowledge gradient (KG) function, which was introduced in [70]. In the KG function, $\mathrm{f}(\mathbf{x}^+)$ is replaced by the minimum over all the data including the training and search space. The exact computation of the KG is more costly than that of the EI. Choosing a good acquisition function is crucial for the performance of Bayesian optimization.

The Bayesian optimization algorithm can be outlined as follows:

1) Construct a surrogate model from the data in-hand

2) Select a next experiment point which maximizes the acquisition function

3) Conduct the recommended experiment and acquire a new datapoint

4) Go back to 1

These steps are repeated until a chosen convergence criterion is met or computational resources are exhausted. Fig. 6 shows a visual representation of Bayesian optimization procedure with an example of 1-D function.

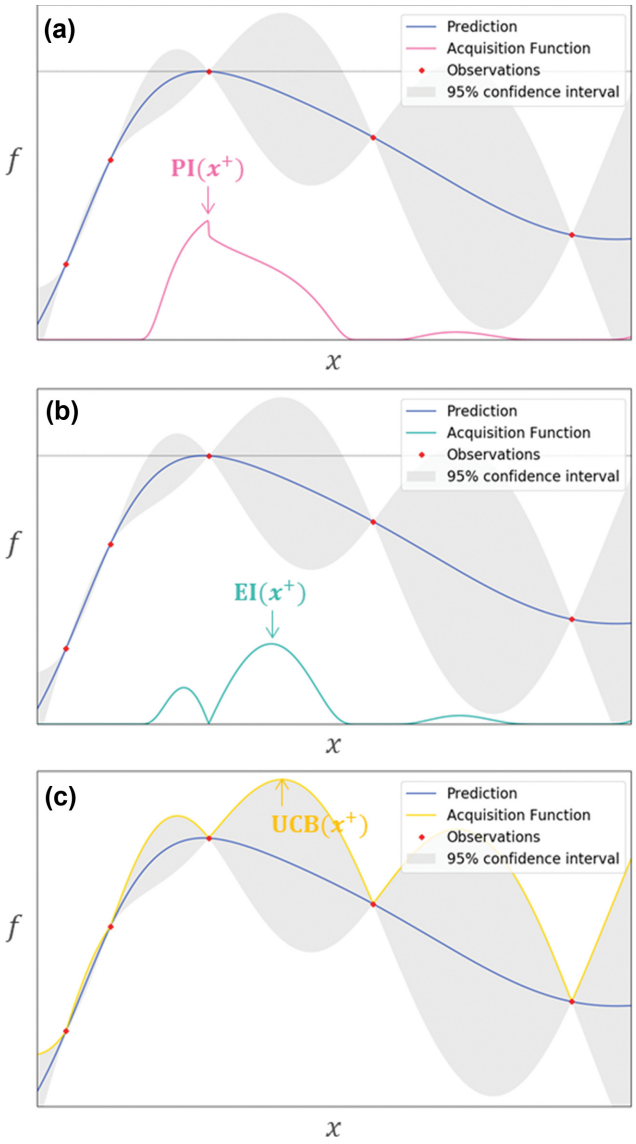

Fig. 5. Three different acquisition functions: (a) The probability of improvement is the probability that a query point will exceed the current maximum. (b) The expected improvement is the expected size of improvement upon the current maximum. (c) The upper confidence bound is a fixed number of standard deviations from the function mean.

2. Adaptive Experimental Design with Bayesian Optimization

To effectively apply the BO to experimental design for accelerated material search, several additional factors need to be considered. First, many real-world material design problems involve multiple objectives. For example, in polymer material design, one may wish to find a polymer that has high strain and stress but low viscosity. In such a case, one needs to consider multiple objectives that potentially conflict with each other. One possible formulation is to maximize the polymer strain, subject to the constraint that the viscosity of the polymer is below a certain threshold, leading to

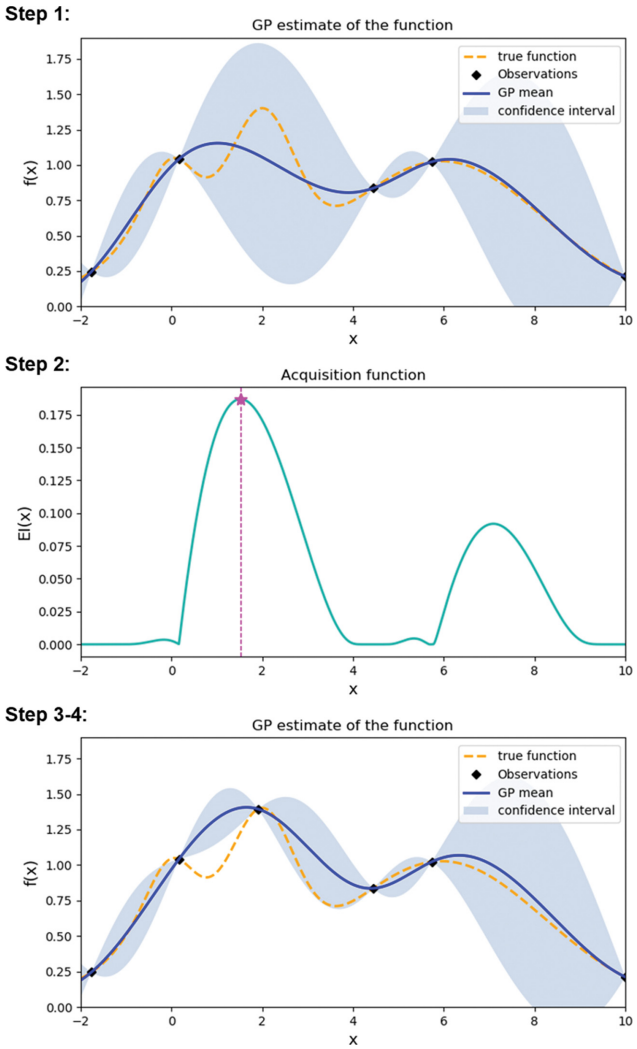

Fig. 6. Bayesian optimization procedure.

a constrained optimization problem. Additionally, it may be desired to test several experimental conditions simultaneously in the lab. Development of Bayesian optimization techniques for considering multiple objectives, constraints, and parallel evaluations is an active research area, which is briefly surveyed in the following sections.

2-1. Multiple Objectives

In many practical optimization problems involving multiple objectives, the objectives tend to be conflicting, such that all the objectives cannot be optimized individually. Therefore, the goal in the traditional multi-objective optimization literature is to identify the set of optimal solutions, called a Pareto set or Pareto front. To address the multiple objectives in the framework of BO, several variants of the standard BO have been proposed.

One class of approaches uses the transformations of the multi-objective problem into a single-objective problem via scalarization, which allows using the standard BO. ParEGO [71] uses random scalarization to recover the whole Pareto front. At each iteration of the algorithm, a weight vector is drawn at random from a uniform distribution, which is used to compute a scalarized single-objective function. The next evaluation point is then chosen by maximizing an acquisition function using the surrogate model fitted on the single-objective. ParEGO is simple and fast, but the scalarization approach has a limitation in that it works only when the Pareto curve is convex [72].

SMSego [73] is an alternative approach that utilizes the hypervolume contribution to decide which point is evaluated next. The hypervolume measure is one of the most popular quality indicators for the assessment of Pareto front approximations, indicating the size of the dominated space with respect to a given reference point. In SMSego, to compute the hypervolume contribution, an optimistic estimate of the objectives is used in a UCB fashion. Pareto active learning (PAL) [74] is similar to SMSego but provides a theoretical guarantee on the algorithm's sampling cost to achieve a desired accuracy.

Expected hypervolume improvement (EHI) [75] and sequential uncertainty reduction (SUR) [76] are two similar methods, based on another approach for multi-objective Bayesian optimization. They use the EHI as an acquisition function, which is a natural extension of the EI to a multi-objective setting. However, computing the expected increment of the hypervolume is very expensive, and thus EHI and SUR are computationally intractable in practical cases with more than three objectives. On the other hand, Hernandez et al. [77] proposed a predictive entropy search (PESMO), which is based on reducing the entropy of the posterior distribution over the Pareto front. The PESMO acquisition function is defined as a sum of objective-specific acquisition functions, which allows for decoupled evaluation scenarios, to identify the most difficult objectives that require more evaluations. The computational cost of PESMO scales linearly with the number of objectives, while the other methods incur exponential costs.

2-2. Constraints

Several BO formulations have been proposed for handling constraints. Schonlau et al. [78] suggested multiplying the EI by the probability of feasibility, i.e., the probability that the point is feasible. The resulting acquisition function will become zero where there is a very low likelihood of feasibility. For considering multiple constraints, the product of the probabilities of feasibility of the individual constraints can be used. One concern for this method is that one of the product terms may dominate, to prevent the algorithm from exploring points on the constraint boundary where the true optimum may lie [79]. To overcome this limitation, Sasena et al. [80] proposed a penalty method, where a large constant (i.e., a penalty) is added to the acquisition function to prevent the search from choosing samples in the infeasible region. It has been shown that the penalty method can find samples closer to the constraint boundary than the probability method [79].

Another method for constraint handling uses the concept of the 'expected violation (EV)' [81]. The EV is calculated in the same way as the EI function and it is vector-valued for problems with multiple constraints. This method first computes the infinity norm of the EV vectors at all the candidate points and then considers the candidates with norm values less than a user-specified threshold as 'acceptable candidates'. The expected improvement is then evaluated only for the set of acceptable candidate points. In addition, augmented Lagrangian methods have been proposed [82,83] and several approaches to deal with discontinuous or binary con-

straints have been studied [84,85].

2-3. Batch Experiments

In many applications, it is possible and also desirable to perform multiple evaluations (e.g., in high throughput experiments) in parallel. For that, a 'batch BO' formulation, which guides the next batch of experiments to evaluate at once, is needed. Several studies on batch BO have been conducted, most of which involve the explicit use of the predictive distributions from the surrogate model. The simulation matching algorithm of Azimi et al. [86] uses the posterior distribution of the surrogate function to generate Monte Carlo simulations of a sequential policy, so as to design a set of experiments that matches the expected behavior of a sequential policy as closely as possible. Here, the sequential policy refers to the standard BO where one experiment is selected at each iteration sequentially. Similarly, in [87] and [88], the predictive distribution is used to generate fake observations.

On the other hand, the so called exploratory approaches [89,90] use the variance of the predictive distribution for guiding an exploratory search. Contal et al. [90] proposed the Gaussian process upper confidence bound and pure exploration (GP-UCB-PE), which combines the one UCB strategy and K-1 Pure Exploration to guide batches of size K. Pure exploration is introduced to reduce the uncertainty about f in order to improve the guesses of the UCB procedure. The exploratory approach is based on the property that the posterior variance does not depend on the values of the observations. In addition, other approaches for batch BO have been studied in the context of multi-armed bandits [91], Bayesian networks [92], and robust optimization [93].

3. Applications

Adaptive design based on Bayesian optimization has been successfully applied to a wide variety of materials search problems, including those for organic-inorganic molecules, alloys, photovoltaics, and polymers, through which it has proven its performance. In this section, we introduce several applications focusing on how the standard or extended BO techniques presented in the previous sections are employed in materials search and which features are used as the input in the respective applications. A more in-depth review of Bayesian optimization in materials design can be found in [94] and [95].

Xue et al. [61] applied an adaptive design strategy to find the composition of NiTi-based shape memory alloys with minimum thermal hysteresis. They considered $Ni_{50-x-y-z}Ti_{50}Cu_xFe_yPd_z$ alloys with x, y, and z compositions, which leads to a potential search space of ~800,000 compositions. Instead of using the x-y-z compositions as features, Xue et al. selected several features for the surrogate model, including Waber-Cromer pseudopotential radii, Pauling electronegativity, metallic radius, valence electron number, Clementi's atomic radii and Pettifor chemical scale, based on their prior knowledge of alloy materials. In that study, they compared different surrogate models, a GP model and a support vector regression (SVR) model in combination with bootstrap sampling to estimate model uncertainty, and interestingly, found that the SVR with bootstrapping outperforms the GP model. In addition, different design strategies were compared, and it turned out that SVR-KI and SVR-EGO gave better performance than other surrogate model-acquisition function pairs on the training set. Balachandran et al. showed similar results regarding the choice of surrogate models and acquisition functions for the composition optimization of M_2AX compounds [96]. Using the SVR-KI pair, nine design iterations (including four alloys at a time) were performed, starting with an original training set of 22 alloys. As a result, Xue et al. successfully synthesized 14 alloys with smaller thermal hysteresis than any of the original dataset.

Talapatra et al. [97] treated a multi-objective design problem of MAX phases with maximal bulk modulus and minimal shear modulus, explored via DFT calculations. Talapatra et al. employed EHI as an acquisition function for the multi-objective Bayesian optimization. Instead of using a single GP model as the surrogate model, they proposed the Bayesian model averaging (BMA) approach, which uses a marginalized aggregation of all the potential models (e.g., models based on different feature sets), weighted by their posterior probability of being the correct model. This work considered the conventional MAX phases with M_2AX and M_3AX_2, of which design space consists of 403 MAX phases (see Fig. 7). The 15 potential features were considered in this study, including the valence electron concentration, electron to atom ratio, atomic number, the order of the MAX phase, and so on. The results proved that the proposed BMA approach can successfully auto-select the best feature set. This approach might be useful especially when the number of available data is too small to select good features.

Griffiths et al. [98] employed a constrained Bayesian optimization to generate valid and synthesizable molecules. To avoid an

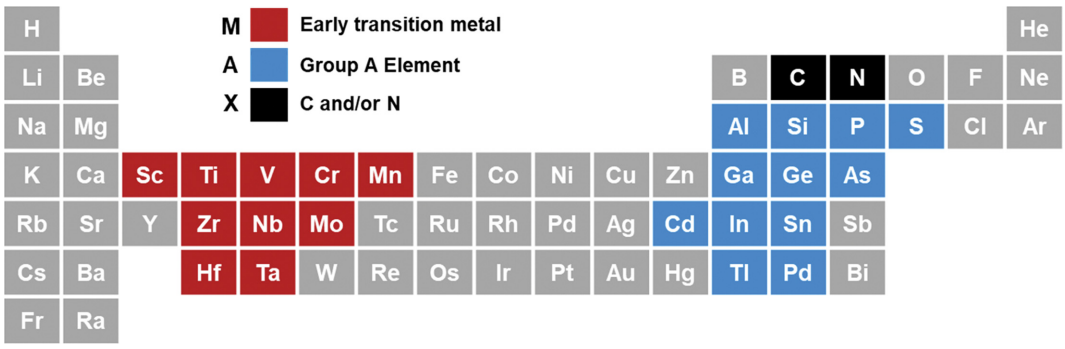

Fig. 7. Elements in the periodic table that react together to form the MAX phases, where M denotes an early transition metal, A is a group A element, and X is either C and/or N.

invalid molecular structure that might be present in the unconstrained approach, a penalty method for constrained BO was introduced. They compared the performance of constrained BO approach against the unconstrained BO using the examples of drug design and organic photovoltaics design. As a result, the constrained approach shows concrete improvement over the unconstrained one, in terms of both validity and quality of the generated molecules.

CONCLUSION

The enormous chemical space that exists is a blessing in the sense that there are endless possibilities when it comes to designing molecule or materials that can map to user-desired applications. Unfortunately, given such a large space, it is difficult to navigate through it in an efficient manner and, undoubtedly, there are regions within this chemical space that are completely unexplored at the moment. With the advent of machine learning and various neural networks, we are at the stage where efficient exploration of the chemical space is possible and, as such, one can obtain properties of molecules and materials much faster compared to experiments as well as conventional molecular simulations. Moreover, by utilizing the structure-property relationship that persists in many molecules/materials, it is possible to use inverse design to discover user-desired materials catered towards certain applications. The utility and type of methods deployed depends on the amount of data available as there are parallel efforts being made to account for different levels of available data. Eventually, these types of discovery efforts should be connected to ease in which one can synthesize these new materials, and this is an exciting field of research where machine learning is also being put to use. With many of these factors integrated together, it is conceivable that machine learning will play a dominant role in future materials design, discovery, and synthesis processes.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea funded by the Ministry of Science, ICT, & Future Planning under grant no. 2021R1A2C2003583 and 2021R1A2C2006083.

REFERENCES

1. Y. LeCun, Y. Bengio and G. Hinton, *Nature*, **521**, 436 (2015).
2. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg and D. Hassabis, *Nature*, **518**, 529 (2015).
3. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel and D. Hassabis, *Nature*, **529**, 484 (2016).
4. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, **559**, 547 (2018).
5. A. Agrawal and A. Choudhary, *APL Materials*, **4**, 053208 (2016).
6. M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, **108**, 058301 (2012).
7. T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning*, Springer, New York (2009).
8. K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *J. Phys. Chem. Lett.*, **6**, 2326 (2015).
9. D. Weininger, *J. Chem. Information Modeling*, **28**, 31 (1988).
10. D. Weininger, A. Weininger and J. L. Weininger, *J. Chem. Information Modeling*, **29**, 97 (1989).
11. S. Kearnes, K. McCloskey, M. Berndl, V. Pande and P. Riley, *J. Comput.-Aided Mol. Des.*, **30**, 595 (2016).
12. D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik and R. P. Adams, arXiv preprint arXiv:1509.09292 (2015).
13. A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez and J. Garcia-Rodriguez, arXiv preprint arXiv:1704.06857 (2017).
14. J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li and M. Sun, *AI Open*, **1**, 57 (2020)
15. A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B*, **87**, 184115 (2013).
16. O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp and A. Knoll, *Int. J. Quantum Chem.*, **115**, 1084 (2015).
17. M. Valle and A. R. Oganov, *Acta Crystallogr., Sect. A: Found. Crystallog.*, **66**, 507 (2010).
18. K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller and E. K. U. Gross, *Phys. Rev. B*, **89**, 205118 (2014).
19. F. Faber, A. Lindmaa, O. A. von Lilienfeld and R. Armiento, *Int. J. Quantum Chem.*, **115**, 1094 (2015).
20. T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, **120**, 145301 (2018).
21. J. Behler and M. Parrinello, *Phys. Rev. Lett.*, **98**, 146401 (2007).
22. J. Behler, *J. Chem. Phys.*, **134**, 074106 (2011).
23. J. Behler, *Int. J. Quantum Chem.*, **115**, 1032 (2015).
24. J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, **8**, 3192 (2017).
25. M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi and P. Marquetand, *J. Chem. Phys.*, **148**, 241709 (2018).
26. K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *J. Chem. Phys.*, **148**, 241722 (2018).
27. K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, **15**(1), 448 (2018).
28. L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk and B. Smit, *Nat. Mater.*, **11**, 633 (2012).
29. C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nat. Chem.*, **4**, 83 (2012).
30. D. A. Gómez-Gualdrón, C. E. Wilmer, O. K. Farha, J. T. Hupp and R. Q. Snurr, *J. Phys. Chem. C*, **118**, 6941 (2014).
31. C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr and B. Smit, *Energy Environ. Sci.*, **8**, 1190 (2015a).
32. A. Mullard, *Nature*, **549**, 445 (2017).
33. B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, **361**, 360 (2018).
34. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, *Commun. ACM*, **63**(11), 139 (2020).
35. D. P. Kingma and M. Welling, arXiv preprint arXiv:1312.6114 (2013).

36. R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams and A. Aspuru-Guzik, *ACS Cent. Sci.*, **4**, 268 (2018).
37. M. J. Kusner, B. Paige and J. M. Hernández-Lobato, ICML, PMLR (2017).
38. E. Putin, A. Asadulaev, Y. Ivanenkov, V. Aladinskiy, B. Sanchez-Lengeling, A. Aspuru-Gzik, and A. Zhavoronkov, *J. Chem. Information Modeling*, **58**, 1194 (2018).
39. M. H. S. Segler, T. Kogej, C. Tyrchan and M. P. Waller, *ACS Cent. Sci.*, **4**, 120 (2018).
40. G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias and A. Aspuru-Guzik, arXiv preprint arXiv:1705.10843 (2017).
41. A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper and A. Zhavoronkov, *Mol. Pharm.*, **14**, 3098 (2017).
42. M. Olivecrona, T. Blaschke, O. Engkvist and H. Chen, *J. Cheminformatics*, **9**, 48 (2017).
43. N. De Cao and T. Kipf, arXiv preprint arXiv:1805.11973 (2018).
44. N. W. A. Gebauer, M. Gastegger and K. T. Schütt, arXiv preprint arXiv:1810.11347 (2018).
45. D. Xue, Y. Gong, Z. Yang, G. Chuai, S. Qu, A. Shen, J. Yu and Q. Liu, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **9**, e1395 (2018).
46. Y. Li, L. Zhang and Z. Liu, *J. Cheminformatics*, **10**(1), 1 (2018).
47. M. Simonovsky and N. Komodakis, ICANN, Springer, Cham (2018).
48. Q. Zhou, P. Tang, S. Liu, J. Pan, Q. Yan and S.-C. Zhang, *Proc. Natl. Acad. Sci.*, **115**(28), E6411 (2018).
49. A. Ziletti, D. Kumar, M. Scheffler and L. M. Ghiringhelli, *Nat. Commun.*, **9**, 2775 (2018).
50. J. Noh, J. Kim, H. S. Stein, B. Sanchez-Lengeling, J. M. Gregoire, A. Aspuru-Guzik and Y. Jung, *Matter*, **1**(5), 1370 (2019).
51. S. Kim, J. Noh, G. H. Gu, A. Aspuru-Guzik and Y. Jung, *ACS Cent. Sci.*, **6**, 1412 (2020).
52. J. Jang, G. H. Gu, J. Noh, J. Kim and Y. Jung, *J. Am. Chem. Soc.*, **142**, 18836 (2020).
53. N. S. Bobbitt and R. Q. Snurr, *Mol. Simul.*, **45**(14-15), 1069 (2019).
54. M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji and T. K. Woo, *J. Phys. Chem. Lett.*, **5**, 3056 (2014).
55. C. M. Simon, R. Mercado, S. K. Schnell, B. Smit and M. Haranczyk, *Chem. Mater.*, **27**, 4459 (2015).
56. Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F. Stoddart, F. You, J. T. Hupp, O. K. Farha and R. Q. Snurr, *Sci. Adv.*, **2**(10), e1600909 (2016).
57. A. Raza, A. Sturluson, C. M. Simon and X. Fern, *J. Phys. Chem. C*, **124**, 19070 (2020).
58. Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr and A. Aspuru-Guzik, *Nat. Mach. Intell.*, **3**, 76 (2021).
59. S. Lee, B. Kim and J. Kim, *J. Mater. Chem. A*, **7**, 2709 (2019).
60. B. Kim, S. Lee and J. Kim, *Sci. Adv.*, **6**, eaax9324 (2020).
61. D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue and T. Lookman, *Nat. Commun.*, **7**(1), 1 (2016).
62. A. I. J. Forrester and A. J. Keane, *Prog. Aerosp. Sci.*, **45**, 50 (2009).
63. P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, **533**, 73 (2016).
64. S. Pruksawan, G. Lambard, S. Samitsu, K. Sodeyama and M. Naito, *Sci. Technol. Adv. Mater.*, **20**, 1010 (2019).
65. J. Mockus, *J. Glob. Optim.*, **4**, 347 (1994).
66. D. R. Jones, M. Schonlau and W. J. Welch, *J. Glob. Optim.*, **13**, 455 (1998).
67. S. Streltsov and P. Vakili, *J. Glob. Optim.*, **14**, 283 (1999).
68. C. E. Rasmussen and C. Williams, *Gaussian processes for machine learning*, MIT Press, Cambridge (2006).
69. D. R. Jones, M. Schonlau and W. J. Welch, *J. Glob. Optim.*, **13**, 455 (1998).
70. P. I. Frazier, W. B. Powell and S. Dayanik, *SICON*, **47**, 2410 (2008).
71. J. Knowles, *IEEE Trans. Evol. Comput.*, **10**, 50 (2006).
72. I. Das, *Nonlinear multicriteria optimization and robust optimality*, Rice University (1997).
73. W. Ponweiser, T. Wagner, D. Biermann and M. Vincze, *Multiobjective optimization on a limited budget of evaluations using model-assisted S-metric selection*, Springer, Berlin (2008).
74. M. Zuluaga, G. Sergent, A. Krause and M. Püschel, ICML, PMLR (2013).
75. M. Emmerich and J.-w. Klinkenberg, Rapport technique, Leiden University, 34, 7 (2008).
76. V. Picheny, *Stat. Comput.*, **25**, 1265 (2015).
77. D. Hernández-Lobato, J. Hernandez-Lobato, A. Shah and R. Adams, ICML, PMLR (2016).
78. M. Schonlau, *Computer experiments and global optimization*, University of Waterloo (1997).
79. M. J. Sasena, *Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations*, University of Michigan (2002).
80. M. Sasena, P. Papalambros and P. Goovaerts, *8th Multidiscip. Anal. Optim. Conf.*, 4921 (2000).
81. C. Audet, J. Denni, D. Moore, A. Booker and P. Frank, *8th Multidiscip. Anal. Optim. Conf.*, 4891 (2000).
82. B. Bichon, S. Mahadevan and M. Eldred, 50th AIAA/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf. (2009).
83. V. Picheny, R. B. Gramacy, S. Wild and S. L. Digabel, *ICONIP*, 1443 (2016).
84. H. Lee, R. Gramacy, C. Linkletter and G. Gray, *Pac. J. Optim.*, **7**, 467 (2011).
85. A. Basudhar, C. Dribusch, S. Lacaze and S. Missoum, *Struct. Multidiscip. Optim.*, **46**, 201 (2012).
86. J. Azimi, A. Fern and X. Z. Fern, NeurIPS (2010).
87. J. Bergstra, R. Bardenet, Y. Bengio and B. Kégl, NeurIPS, 24 (2011).
88. J. Azimi, A. Jalali and X. Fern, arXiv preprint arXiv:1202.5597 (2012).
89. M. Schonlau, W. J. Welch and D. R. Jones, *Lecture Notes-Monograph Series*, **34**, 11 (1998).
90. E. Contal, D. Buffoni, A. Robicquet and N. Vayatis, ECML PKDD, 225 (2013).
91. T. Desautels, A. Krause and J. W. Burdick, *J. Mach. Learn. Res.*, **15**, 3873 (2014).
92. J. Očenášek and J. Schwarz, *The state of the art in computational intelligence*, 61, Physica, Heidelberg (2000).
93. M. A. Taddy, H. K. H. Lee, G. A. Gray and J. D. Griffin, *Technometrics*, **51**, 389 (2009).
94. J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, *Npj Comput. Mater.*, **5**, 1 (2019).

95. T. Lookman, P. V. Balachandran, D. Xue, J. Hogden and J. Theiler, *Curr. Opin. Solid State Mater. Sci.*, **21**, 121 (2017).
96. P. V. Balachandran, D. Xue, J. Theiler, J. Hogden and T. Lookman, *Sci. Rep.*, **6**, 1 (2016).
97. A. Talapatra, S. Boluki, T. Duong, X. Qian, E. Dougherty and R. Arróyave, *Phys. Rev. Mater.*, **2**, 113803 (2018).
98. R.-R. Griffiths and J. M. Hernández-Lobato, arXiv preprint arXiv: 1709.05501 (2017).

Jihan Kim obtained his B.S. degree in Electrical Engineering and Computer Sciences from UC Berkeley in 2001. He received his M.S. and Ph.D. degree in Electrical and Computer Engineering from University of Illinois at Urbana-Champaign in 2004 and 2009 respectively. From 2009 to 2013, he was a postdoctoral researcher at Lawrence Berkeley National Laboratory. He joined KAIST in 2013 and is currently an associate professor in the Department of Chemical and Biomolecular Engineering. He has published more than 90 papers.

Jay Hyung Lee is currently a KEPCO Chair Professor at Korea Advanced Institute of Science and Technology (KAIST). He is also the director of Saudi Aramco-KAIST CO_2 Management Center. He received the AIChE CAST Computing in Chemical Engineering Award and was elected as an IEEE Fellow, an IFAC Fellow, and an AIChE Fellow. He was the 29th Roger Sargent Lecturer in 2016. He published over 200 manuscripts in SCI journals with more than 17000 Google Scholars citations. His research interests are in the areas of state estimation, model predictive control, planning/scheduling, and reinforcement learning with applications to energy systems and carbon management systems.