https://doi.org/10.1038/s43588-023-00471-w

# Structure prediction and materials design with generative neural networks

Da Yan, Adam D. Smith & Cheng-Chien Chen



The prediction of stable crystal structures is an important part of designing solid-state crystalline materials with desired properties. Recent advances in structural feature representations and generative neural networks promise the ability to efficiently create new stable structures to use for inverse design and to search for materials with tailored functionalities.

Solid-state crystalline materials can be accurately modeled using first-principles quantum mechanical calculations, as long as their underlying crystal structures are specified. For materials of unknown structure, however, predicting their properties requires an additional step of crystal structure prediction (CSP). CSP aims to discover stable and metastable structures given only the chemical formula (and the number of atoms in the unit cell), by locating the Gibbs free energy minima under certain thermodynamic pressure and temperature conditions. Achieving this task requires accurate computation of the potential energy surface and a powerful optimization technique. Efficient optimization methods such as evolutionary algorithms and particle swarm optimization have led to the discovery of various new materials[1,2].

A typical CSP task scans through thousands of structures and performs the corresponding energy calculations. It is computationally expensive, especially for materials with ternary or quaternary (or higher-order) compositions. For this reason, large-scale materials discovery remains challenging. More recently, generative models have provided promising new opportunities to tackle these challenges, because once they are trained, they can produce new structures much faster than traditional CSP techniques can[3,4]. However, developing generative models for CSP is highly non-trivial, given that it requires an invertible representation to map a three-dimensional (3D) crystal to the feature space, and a corresponding reverse mapping. It also needs a target database that is statistically representative of the system of interest. Despite these non-trivial problems, several recent studies[5–10] have demonstrated the feasibility of using generative neural networks for the efficient and accurate prediction of new stable crystal structures.

## Generative models

Machine learning models can generally be categorized into discriminative models and generative models. Discriminative models focus on predicting the data labels and drawing boundaries in the feature space. On the other hand, generative models focus on explaining how the data was generated and trying to model how data is placed throughout the space. Therefore, although discriminative models can achieve a

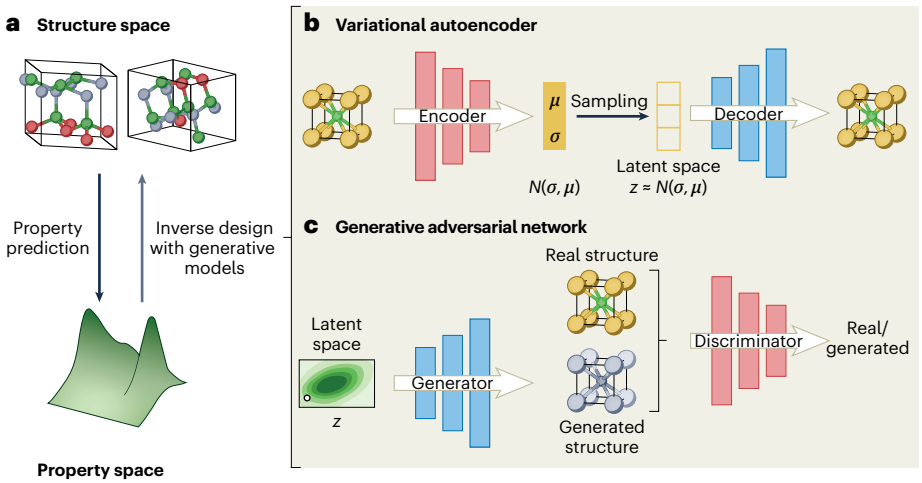


**Fig. 1 | Material property prediction and inverse design. a**, Schematic showing traditional material property prediction from the structure space to the property space (downward arrow), and inverse material design from the property space back to the structure space (upward arrow). **b, c**, Two popular generative models for structure prediction: the variational autoencoder (VAE) (**b**) and the generative adversarial network (GAN) (**c**). The VAE consists of an encoder that transforms the input sample feature vector to a latent distribution space, and a decoder that reconstructs the sample given the hidden distribution. The VAE also models the latent space vector $z$ from a normal distribution $N(\mu, \sigma)$ with a mean $\mu$ and a standard deviation $\sigma$. GAN uses a generator to transform a random noise variable into the generated sample, and a discriminator to distinguish whether a sample is real or generated.

direct forward mapping of crystal structure to material property, they are unable to realize inverse design in the way that generative models can (Fig. 1a). In essence, a generative model learns the distribution of the data itself, and then samples new data instances from the learned distribution, which enables the exploration of a more diverse output of crystal structures. In the domain of inverse design, two types of generative model are popularly used for crystalline materials[3,4]: a variational autoencoder (VAE) and a generative adversarial network (GAN) (Fig. 1b,c). A key characteristic of generative models is their ability to map any data point in the learned latent feature space back to its corresponding crystal structure in the materials space. Generative models could provide better compositional and structural diversity than substitution-based enumeration in high-throughput calculations, and better structural generation efficiency than conventional CSP techniques.

A VAE, which is composed of an encoder and a decoder, is trained to minimize the reconstruction error between the decoded data and the input data. However, instead of encoding an input as a single point in the latent space, the encoder treats the input as (the parameters of) a distribution over the latent space. A new data point $z$ from the latent space can then be sampled and decoded. The encoded distributions are typically chosen to be normal $N(\mu, \sigma)$ with a mean $\mu$ and a standard deviation $\sigma$. Representative VAE structure predictors include iMatGen[11], the Fourier-transformed Crystal Properties (FTCP) framework[12], and the Constrained Crystals Deep Convolutional Generative Adversarial Network (CCDCGAN)[13]. Specifically, FTCP adds a target-learning branch to map latent points to target properties (with an additional property-mapping loss). CCDCGAN employs a VAE to learn a reverse map from a latent 2D crystal representation back to crystal structures, and then uses the representation to train a GAN to generate new crystal structures. VAEs are relatively easy to train, and they provide more diversified structures that cover the distribution better than other generative models. However, VAEs have the potential drawback of a low output validity rate (that is, some results may not be valid). In computer vision tasks, VAE models may produce unrealistic, blurry samples, partly owing to the assumption that the embedding feature space follows a Gaussian distribution.

A GAN, which uses the idea of a minimax game in game theory, uses two networks: the generator, $G$, which transforms a random variable $z$ in the latent space into the generated sample $G(z)$, and the discriminator, $D$, which distinguishes whether a sample is real or generated. A GAN is trained so that $G$ maximizes the probability that $D$ misclassifies a generated sample as a real one (that is, $G$ wants to 'fool' $D$ as much as possible), and that $D$ makes the best possible distinction between real and generated samples. This allows $G$ to learn the distribution of the real data. In addition, conditional GANs allow the extension of the latent variable $z$ with desired conditions, such as user-desired composition, as in the Composition-Conditioned Crystal GAN[5]. Another network branch is often added to predict the property of $G(z)$ and to include this prediction in the loss to produce crystal structures with the desired property, as in CCDCGAN[13]. This idea is similar to the target-learning branch of the FTCP framework[12]. CrystalGAN[6] further utilizes a cross-domain GAN to generate complex ternary palladium–hydrogen–nickel structures starting from simpler binary Pd–H and Ni–H structures. In computer vision tasks, GANs generally produce better photo-realistic images. However, compared to a VAE, a GAN is more difficult to train, as it can exhibit issues such as non-convergence (where the model parameters oscillate and fail to converge), mode collapse (where the generator produces limited sample varieties), and diminished gradient (where the discriminator is so successful that the generator gradient vanishes and learns nothing). Thus, overall, balancing the generator and discriminator in a GAN is essential to prevent overfitting.

## Feature representation

Typically, the structures in materials databases are stored in the Crystallographic Information File (CIF) format, which often serves as the input for feature representation. Representing a discrete crystal structure in the continuous latent feature space is the first important step towards structure prediction using generative models. There are currently two main approaches: continuous 3D voxel representation, in which encoders and decoders respectively prepare 2D crystal graphs and reconstruct 3D voxel images, and matrix representation, wherein crystal structure features such as lattice parameters, atomic occupation coordinates, and elemental properties are separated into different matrix rows and columns.

In iMatGen[11], for example, an encoder is first trained to compress images of 3D voxels (for lattice parameters and atomic positions) into a single image, from which a decoder then reconstructs the crystal structure. CCDCGAN[13] considers a 3D voxel representation using a lattice autoencoder, which first converts the atomic positions onto a voxel grid with smearing; the voxel grids are further transformed into 1D vectors, which are encoded into a 2D crystal graph. In the Composition-Conditioned Crystal GAN[5], a point cloud representation is utilized to decrease the memory requirements substantially, by constructing a 2D matrix representation of unit cell parameters and atomic fractional coordinates. The FTCP framework[12] revisits the point cloud representation by considering both real-space and reciprocal-space features, with a Fourier transformed elemental property matrix and Miller indices, all in 2D matrix form. In the above representations, a reversible mapping from the latent space back to the materials space is necessary.

In addition to reversibility between the latent space and the materials space, structure feature representation in principle needs the property of invariance. Because of the underlying crystal symmetry groups, a representation undergoing translation, rotation, or permutation of the crystal axes (for example) should be invariant, meaning that it still represents the same latent space data point. In practice, however, a fully reversible and invariant scheme is still absent in the literature[12]. Other crystal feature presentation ideas should be explored, such as E(3)-equivariant graph neural networks[7], which can preserve known system properties under equivariant transformations. Another potential issue with feature representation concerns the loss of fidelity during a reverse mapping. In other words, when a latent space point is reversely mapped back to the materials space, the resulting crystal structure is not identical to the original one. To what extent a given input structure can be fully reconstructed and how significant is the error caused by the fidelity loss in the reconstruction needs to be more carefully determined, for instance, via uncertainty quantification.

## Training data

Current generative models employ mainly experimental databases such as the Inorganic Crystal Structures Database[8], and computational databases such as the Materials Project[9]. To ensure sufficient structural and elemental diversity, additional high-throughput calculations with element substitutions in known crystal structures or data augmentation techniques are often performed to achieve a statistically representative data distribution.

For example, the training data in iMatGen[11] was created by taking 25 unique vanadium–oxygen compositions from the Materials Project and

substituting these compositions into 10,981 binary metal structures, also in the Materials Project. This led to the rediscovery of 26 out of 31 existing $V_xO_y$ structures in the Materials Project, and 40 previously unidentified new structures with energies of around 80 meV per atom above the convex hull. CCDCGAN[13] studied Bi–Se systems, whereas the Materials Project database contains only 17 known Bi–Se materials. The training data were generated similarly, where 10,981 prototype structures were substituted with Bi–Se, by limiting the maximum number of atoms per unit cell to 20 and the maximum lattice constant to 10 Å; 9,810 structures were converged in subsequent first-principles calculations. The Composition-Conditioned Crystal GAN[5] studied Mg–Mn–O systems. The training data were also generated by element substitution of ternary compounds in the Materials Project, with an initial dataset of 1,240 structures with 112 compositions. The training data were further augmented by adding rotated and translated unit cells, as well as supercell structures. This data augmentation yielded 1,000 structures for each composition, leading to 112,000 Mg–Mn–O structures. After the training, the GAN model was utilized to create 9,300 unique structures for high-throughput calculations, where 23 new Mg–Mn–O crystals were discovered. These studies indicate that current generative models are capable of generating the structures of real materials and also of producing new stable ones. However, for a given material system, each generative model needs to be trained individually with data augmentation, since existing databases remain too small to develop a comprehensive and generic generative model that works for all materials.

Naturally, the quality of the training data has a crucial role in determining the performance of the resulting generative neural networks. Sufficient data (meaning, $10^5$ to $10^6$) with high structural diversity (meaning, $10^3$ to $10^4$) are needed, otherwise the model may be biased during the training. In addition to element substitution of known structures in open materials databases, data augmentation (for example, utilizing invariance of the unit cell[12]) and active transfer learning[10] could be helpful. Conventional CSP optimization techniques could also help to alleviate the problem by searching both stable and metastable structures for the training of generative models. The resulting structures produced from generative models can in turn be supplied as seed structures in CSP searches. Finally, achieving the inverse design of properties beyond the formation energy requires a corresponding database (for instance, for mechanical, electronic, and thermal transport properties). For example, the FTCP[12] framework has attempted to target properties such as the bandgap and thermoelectric power factor. The lack of a relevant property database can potentially be mitigated by high-throughput density functional theory or force-field molecular dynamics calculations, or by machine learning simulations such as the Crystal Graph Convolutional Neural Network (CGCNN)[14], which can provide rapid property prediction once the underlying crystal structures are known.

## Conclusion and perspectives

Since existing materials feature representations are based mainly on images or encoded latent feature vectors, ConvNets, or multi-layer perceptrons, dominate the neural network models under study. Even though architectures such as CGCNN[14] have emerged for forward mapping from materials to their latent features for property prediction, an explicit formulation converting them back to 3D crystal structures in inverse design remains an open problem. Although there are challenges in developing future generative models for large-scale structure prediction and materials design, once trained, these models could expedite materials discovery by orders of magnitude when compared to conventional optimization techniques. Ideally, generative models should be able to tackle the inverse design of small problems where data is limited in quantity, as well as large problems with many crystallographic degrees of freedom. Therefore, more demonstrations of generative models with less training data or broader scope in training compositions and structures are important areas of future study.

Meanwhile, many studies have applied graph convolutional networks to generate organic molecules[15], where physics-inspired models such as flow and diffusion models have been employed alongside VAEs and GANs. It would be interesting to test whether these models are also adaptable for inverse design of crystal structures. Besides generative models, reinforcement learning, such as in the Monte Carlo tree search, has been utilized to discover material structures with desired properties[16], by performing guided structural growth (one atom at a time, for example) along promising directions. Finally, it is important to address whether the machine learning models can provide a measure of the synthesizability of the discovered hypothetical materials for practical applications. For example, FTCP[12] has addressed the synthesizability issue by examining whether the generated structures exist in the experimental Inorganic Crystal Structures Database. More research along these lines is needed.

**Da Yan** [1] ✉, **Adam D. Smith** [2] ✉ & **Cheng-Chien Chen** [2] ✉
[1]Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA. [2]Department of Physics, University of Alabama at Birmingham, Birmingham, AL, USA.
✉e-mail: yanda@uab.edu; smitha20@uab.edu; chencc@uab.edu

Published online: 03 July 2023

## References

1. Chen, W.-C., Schmidt, J. N., Yan, D., Vohra, Y. K. & Chen, C.-C. *npj Comput. Mater.* **7**, 114 (2021).
2. Peng, F. et al. *Phys. Rev. Lett.* **119**, 107001 (2017).
3. Chen, L., Zhang, W., Nie, Z., Li, S. & Pan, F. *J. Mater. Inform.* **1**, 4 (2021).
4. Noh, J., Gu, G. H., Kim, S. & Jung, Y. *Chem. Sci.* **11**, 4871–4881 (2020).
5. Kim, S., Noh, J., Gu, G. H., Aspuru-Guzik, A. & Jung, Y. *ACS Cent. Sci.* **6**, 1412–1420 (2020).
6. Nouira, A., Sokolovska, N. & Crivello, J.-C. CrystalGAN: learning to discover crystallographic structures with generative adversarial networks. Preprint at https://arxiv.org/abs/1810.11203 (2018).
7. Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. Preprint at https://arxiv.org/abs/2207.09453 (2022).
8. Levin, I. *NIST Inorganic Crystal Structure Database (ICSD)* (National Institute of Standards and Technology, 2018).
9. Jain, A. et al. *APL Mater.* **1**, 011002 (2013).
10. Kim, Y. et al. *npj Comput. Mater.* **7**, 140 (2021).
11. Noh, J. et al. *Matter* **1**, 1370–1384 (2019).
12. Ren, Z. et al. *Matter* **5**, 314–335 (2022).
13. Long, T. et al. *npj Comput. Mater.* **7**, 66 (2021).
14. Xie, T. & Grossman, J. C. *Phys. Rev. Lett.* **120**, 145301 (2018).
15. Sanchez-Lengeling, B. & Aspuru-Guzik, A. *Science* **361**, 360–365 (2018).
16. Dieb, T. M. et al. *Sci. Technol. Adv. Mater.* **18**, 498–503 (2017).

## Acknowledgements

D.Y. acknowledges support from ARDEF 1ARDEF21 03 (ADECA) and from National Science Foundation (NSF) Award OAC-2106461. A.D.S. and C.-C.C. are supported by NSF Award RII Track-1 Future Technologies and Enabling Plasma Processes Project OIA-2148653. C.-C.C. also acknowledges support from NSF Award DMR-2142801.

## Author contributions

All authors developed the content outline, wrote the manuscript, and created the figure.

## Competing interests

The authors declare no competing interests.