ROYAL SOCIETY
OF CHEMISTRY

PAPER

Check for updates

Predicting performance limits of methane gas storage in zeolites with an artificial neural network†

Sangwon Lee, [ID] Baekjun Kim and Jihan Kim [ID] *

Performance limit is an important metric that can determine the commercial feasibility of a given class of materials. Here, an artificial neural network is used to generate crystalline nanoporous materials (*i.e.* shapes) in the energy space, and molecular simulations are conducted directly on a large number of these generated shapes to obtain the performance limits. A test application of methane storage in zeolites reveals excellent agreement in the methane working capacity values of top performing structures between the real zeolites and the shapes generated from our neural network. Moreover, even when the real zeolites with the highest working capacity values are excluded from the input training sets, our neural network generates shapes that yield large working capacity values that are close to the true maximum value. With further development, this method can be extended to other applications and materials, thereby providing rational guidance to construct reasonable performance targets for future development of materials.

Introduction

Governments and businesses around the world have started many research initiatives to fund projects that aim to develop nanoporous materials for a wide range of applications including carbon capture,[1] gas separations/storage,[2,3] sensors,[4] catalysis,[5] and batteries.[6] Unfortunately, an enormous amount of money is spent on projects that ultimately fail to deliver the material that can reach the intended performance targets. Overall, it would be extremely beneficial if there were an easy way to accurately estimate the performance limits (*i.e.* maximum possible performance regarding a property of interest) for a given class of materials prior to funding the research. This would better inform the funding agencies as well as the entire scientific community on whether it is feasible to discover the "magic" material that can reach the desired target.

In practice, performance limits for nanoporous materials can be estimated using two approaches: (1) experimentally synthesize as many materials as possible and infer the performance limits from the available data and/or (2) computationally generate as many hypothetical materials as possible, screen through them, and deduce the materials' performance limits from the best performing structures.[7–9] Unfortunately, both of these approaches have drawbacks that make them inadequate when it comes to accurately assessing the performance limits. First, the amount of data available from the experimental approach is

often too small to conclude whether new materials can outperform the current best ones. Moreover, there is a catch-22 situation where a large number of experimental data is needed to reliably evaluate the performance limits that cannot be obtained without the research projects being funded in the first place. Second, although there have been successful computational studies focused on screening thousands of materials and consequently identifying the optimal material and its performance,[10–13] most of the materials that are currently being generated from these computational studies closely resemble (both chemically and structurally) the set of experimental materials that already exist. Accordingly, it is difficult to generate unique materials, leading to the dearth of novel computational materials that can perhaps outperform the current best materials.

In this work, a different approach based on artificial neural networks (ANNs) is pursued to tackle this important problem of evaluating materials' performance limits. With the advent of deep learning and enhanced computing power, ANNs have found their way into many different fields and applications.[14–16] Here, a test application of methane storage is selected to illustrate our method. Recently, the U.S. Department of Energy (D.O.E.) set a target of 315 volumes of methane at STP (standard temperature and pressure) of working capacity (WC, difference between adsorption at 65 bar and desorption at 5.8 bar) for methane storage in a technology known as adsorbed natural gas vehicle (or ANGV).[17] Unfortunately, after many years of research, none of the materials has come close to reaching the target set by the D.O.E.[7,18,19] As such, it is not clear whether there exist materials that can potentially exceed the D.O.E. target or if this is an improbable task, making it difficult to determine whether ANGV is a viable technology worth pursuing.

Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of South Korea. E-mail: jihankim@kaist.ac.kr

Although there have been past studies that generated small molecules *via* ANN,[20] as far as we know no one has successfully generated crystalline nanoporous materials using ANN presumably due to its complex nature. Given this difficulty, we use abstract representations of nanoporous materials, not in the conventional materials space of atoms/molecules, but in the potential energy space of the gas molecule of interest (in the case for ANGV, methane). Moreover, conventional classical molecular simulations can be directly applied to these ANN generated potential energy shapes to evaluate the methane adsorption properties. Subsequently, in practice, the methane storage performance limits can be deduced by screening through a vast number of potential energy shapes generated by the ANN. In this work, we have generated over one million energy shapes to obtain reliable statistics regarding the performance limits and as such, molecules such as methane that lack quadrupole moments were deemed ideal for this purpose given that their classical molecular simulation wall times are relatively quick.

To test our method, pure silica crystalline zeolites were chosen as the test case materials to evaluate the limits of methane storage. Pure silica zeolites are comprised of tetrahedral SiO_4 units and with only two atom types, they possess relatively simple chemistry and topology compared to other porous materials such as metal–organic frameworks (MOFs) and polymers. Currently, there are only 228 experimentally synthesized zeolite frameworks in the IZA database.[21] However, Deem and coworkers devised an algorithm to computationally generate hundreds of thousands of hypothetical pure silica zeolites, and as a result have reported most of the feasible zeolite topologies.[22,23] Subsequently, their predicted crystallography open database (PCOD) of pure silica zeolites and the IZA database serve as excellent candidate sets to test our ANN method given that the performance limit for methane storage will likely be located from structures within these two sets and can be obtained from molecular simulations. This is not the case with other (more complex) materials such as MOFs and polymers where an exhaustive list of all possible structures does not exist, making it difficult to verify the accuracy of the performance limit values obtained from our ANN.

Materials and methods

For each unit cell that describes a given zeolite material, a methane potential energy grid can be obtained using force field-based classical molecular simulations. Specifically, a three-dimensional spatial energy grid is superimposed on top of the unit cell and at each grid point, the methane molecule is placed and its Lennard-Jones potential energy with all of the zeolite atoms are computed[24] (see ESI†). Since molecular simulations can be conducted directly using these energy grids and lattice constants, (the two together will be called "energy shapes" from this point forth) if our ANN can successfully generate *realistic*, *hypothetical* zeolite energy shapes, then the materials themselves are not needed to compute the methane adsorption properties to identify the performance limits for zeolites. These energy shapes possess similar characteristics to

pore geometry/shapes discussed in other studies.[25,26] However, these shapes have an added dimension of energy, which allows one to compute the adsorption properties of a specific gas molecule, whereas this is not possible with just the pore geometry information. Fig. 1 depicts the relationship between the materials and the energy shapes for four prominent zeolites.

Next, we provide an overview of our ANN. Between the two main types of the generative ANN (variational auto-encoder (VAE),[27] generative adversarial network (GAN)[28]), GAN is chosen as our neural network due to its capability of generating more realistic samples. GAN consists of a discriminator and a generator, where the goal of the discriminator is to differentiate between the real data and the fake data generated from the generator. On the other hand, the generator acts to deceive the discriminator by progressively creating more and more realistic samples. Overall, this scheme leads to adversarial learning and realistic samples are generated as the two systems become better trained to beat one another. To generate the energy shapes, we developed a new type of GAN, which we named Energy Shape GAN (ESGAN).

The overall schematics of ESGAN are shown in Fig. 2 (a more detailed architecture of the discriminator and generator can be found in Fig. S1†).

The task of the discriminator (D) in our ESGAN is to differentiate between the real (IZA + hypothetical) and the ANN-generated zeolites from the three-dimensional methane energy grids. On the other hand, the generator (*i.e.* **G**, see Fig. 2) attempts to deceive the discriminator by generating more and more realistic methane energy grids. The generator is comprised of two sub-generators: (1) energy grid generator (**g**) and (2) lattice constant generator (**h**). For all of the zeolite
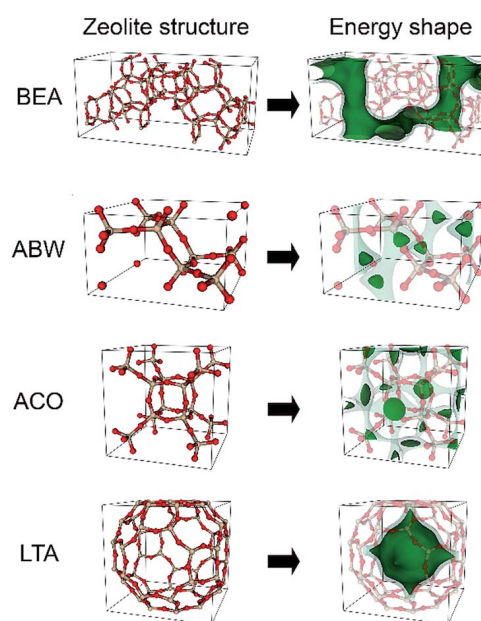

Fig. 1 Left column illustrates the unit cells of four prominent zeolites (BEA, ABW, ACO, and LTA). Right column shows the corresponding methane potential energy profiles (*i.e.* shapes) computed from molecular simulations (dark green: low energy, light green: higher but accessible energy, white: inaccessible regions).
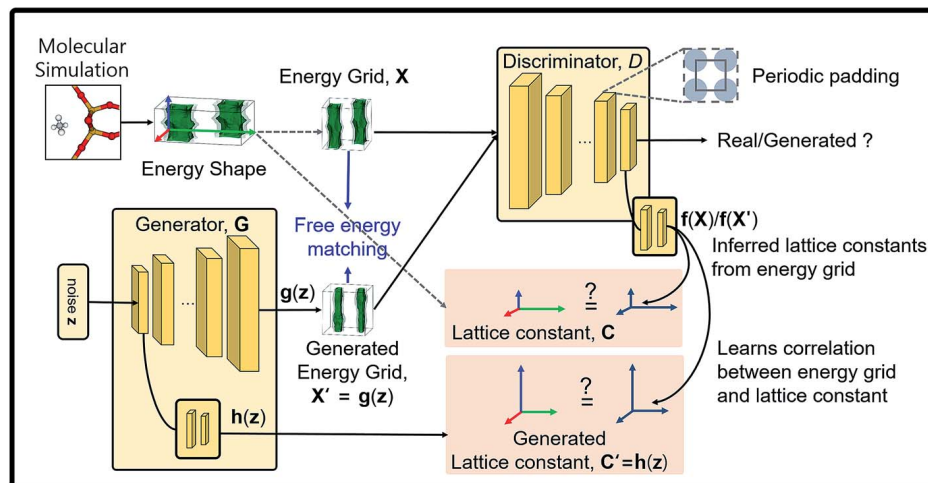
ESGAN

Fig. 2 Overall schematics of ESGAN.

materials used in this work, the unit cell dimensions are transformed from Cartesian to fractional coordinates to standardize the size of the energy grids for all the zeolites (both real and GAN generated ones). By implementing additional layers to the discriminator, an additional function ($\mathbf{f}$) is created that can predict the lattice constants from the fractional energy grid due to their correlated nature. The distributions of the generated lattice constants were compared with those of the real zeolite materials, yielding reasonable agreement (see ESI Fig. S2†).

In machine learning, the loss function (or cost function) of a model is designed to improve the target functionality by minimizing this loss function. Due to the adversarial nature of learning of standard GANs, the discriminator and the generator have their own distinct loss functions (denoted as $L_{\mathrm{GAN,D}}$ and $L_{\mathrm{GAN,G}}$, respectively in this work). In ESGAN, additional loss functions are introduced to create more realistic energy shapes. The first is the lattice inference loss function $L_{\mathrm{Lattice,D}}$ for the discriminator and $L_{\mathrm{Lattice,G}}$ for the generator. These loss functions help the generator create accurate lattice constants. And the second is the loss function $L_{\mathrm{FE-matching}}$ used for free energy matching that helps the generator create proper energy values.

The specific form of the discriminator loss function is given by

$$L_D = L_{\mathrm{GAN,D}} + L_{\mathrm{Lattice,D}} \tag{1}$$

$$L_{\mathrm{GAN,D}} = \mathrm{E}_{\mathbf{X} \sim P_X}[H(\delta, D(\mathbf{X}))] + \mathrm{E}_{\mathbf{z} \sim P_z}[H(0, D(\mathbf{X}'))] \tag{2}$$

$$L_{\mathrm{Lattice,D}} = \mathrm{E}_{\mathbf{X} \sim P_X}[H(\mathbf{C}, \mathbf{f}(\mathbf{X}))] \tag{3}$$

where L_D is the total loss function of the discriminator, δ is the smoothed label, $H(p, q)$ is the binary cross entropy given as $H(p,q) = -p \log q - (1-p) \log(1-q)$, $\mathbf{X}$ is the energy grid of the real zeolite, $\mathbf{C}$ is the lattice constant of the real zeolite, P_X is the distribution of the energy grid of the real zeolite, $\mathbf{z}$ is the noise source for the generator, $P_{\mathbf{z}}$ is the noise distribution, $\mathbf{X}' = \mathbf{g}(\mathbf{z})$ is the generated energy grid, $\mathbf{C}' = \mathbf{h}(\mathbf{z})$ is the generated lattice

constants, $D(\mathbf{X})$ is the probability that $\mathbf{X}$ is from the real data, and $\mathbf{f}(\mathbf{X})$ is the inferred lattice constants from the energy grid. And in our notation, if the inputs H are tensors, the value of H is calculated element-wise and averaged over the dimensions. With proper training, both $\mathrm{E}_{\mathbf{x} \sim P_X}[H(\delta, D(\mathbf{X}))]$ and $\mathrm{E}_{\mathbf{z} \sim P_z}[H(0, D(\mathbf{X}'))]$ become minimized, which in turn minimizes $L_{\mathrm{GAN,D}}$. And this happens ideally when $D(\mathbf{X}) = \delta$ and $D(\mathbf{X}') = 0$ because given the form of $H(p, q)$, these terms become minimized when $q = p$ for given p values. One-side label smoothing[29] with $\delta = 0.8$ is applied to $L_{\mathrm{GAN,D}}$ to improve the convergence of ESGAN. The minimization of $L_{\mathrm{GAN,D}}$ leads to higher values for $D(\mathbf{X})$ and lower values for $D(\mathbf{X}')$. It should be pointed out that these losses are used only to update the discriminator function D and $\mathbf{f}$ with the fixed generator. The second term ($L_{\mathrm{Lattice,D}}$) is the binary cross entropy loss between the lattice constant ($\mathbf{C}$) and inferred lattice constant ($\mathbf{f}(\mathbf{X})$) from the corresponding energy grid $\mathbf{X}$. It should be pointed out that $L_{\mathrm{Lattice,D}}$ is only updated with $\mathbf{X}$ and $\mathbf{C}$ from the real data. By minimizing $L_{\mathrm{Lattice,D}}$, the function $\mathbf{f}$ learns the correlation between the energy grid and lattice constants.

The loss function of the generator is given by

$$L_G = L_{\mathrm{GAN,G}} + \lambda L_{\mathrm{Lattice,G}} + L_{\mathrm{FE-matching}} \tag{4}$$

$$L_{\mathrm{GAN,G}} = \mathrm{E}_{\mathbf{z} \sim P_z}[H(1, D(\mathbf{X}'))] = \mathrm{E}_{\mathbf{z} \sim P_z}[H(1, D(\mathbf{g}(\mathbf{z})))] \tag{5}$$

$$L_{\mathrm{Lattice,G}} = \mathrm{E}_{\mathbf{z} \sim P_z}[H(\mathbf{f}(\mathbf{X}'), C')] = \mathrm{E}_{\mathbf{z} \sim P_z}[H(\mathbf{f}(\mathbf{g}(\mathbf{z})), \mathbf{h}(\mathbf{z}))] \tag{6}$$

where L_G is the total loss for the generator, $\mathbf{f}(\mathbf{X}')$ is the inferred lattice constant from the generated energy grid $\mathbf{X}'$, λ is the scale factor. Similar to the case of the discriminator, these losses are used only to update the energy grid generator $\mathbf{g}$ and lattice constant of the generator $\mathbf{h}$ at fixed D and $\mathbf{f}$. With proper training, $\mathrm{E}_{\mathbf{z} \sim P_z}[H(1, D(\mathbf{g}(\mathbf{z})))]$ becomes minimized, which in turn minimizes $L_{\mathrm{GAN,G}}$. And this happens ideally when $D(\mathbf{g}(\mathbf{z})) = 1$. The minimization of $L_{\mathrm{GAN,G}}$ leads to higher values for $D(\mathbf{g}(\mathbf{z}))$, indicating that the energy grid generator is updated to make a more realistic energy grid because D cannot be changed in the

This journal is © The Royal Society of Chemistry 2019

J. Mater. Chem. A, 2019, **7**, 2709–2716 | **2711**

generator learning step. Similarly, $L_{\text{Lattice,G}}$ is minimized by modifying **h** and **g** because the inference function **f** that is learned from real **X** and **C** cannot be modified during the generator updating steps. Due to this fact, **g** and **h** are updated to generate **X**′ and **C**′ that have a similar correlation between **X** and **C**. The scale factor λ is selected as 0.1 to prevent overfitting to current inference function **f**. The loss functions and volume distributions are shown in Fig. S2.†

Initially, energy shapes generated from ESGAN revealed that the methane potential energy distributions of the generated zeolites possessed far different profiles compared to real zeolites. To remedy this, the loss function for free energy matching was added to the loss of ESGAN, which significantly improved the energy distributions. The feature matching was originally introduced by Salimans et al.[29] where the original paper matched the activation values of the intermediate layer between the real and the fake data. In this work, we used Helmholtz free energy as the feature given that it is a property that can be calculated directly from the energy grid. If we treat the interactions between methane of the united atom model and framework as a function of methane position, we can represent the canonical partition function of a methane gas molecule in a framework as

$$Q = \frac{1}{\Lambda^3} \int_{\mathbb{R}^3 \in V} \exp(-\beta\phi(\mathbf{r}))\mathrm{d}\mathbf{r} \tag{7}$$

where Q is the canonical partition function, Λ is the de Broglie thermal wavelength, k_{B} is the Boltzmann constant, β is $1/k_{\text{B}}T$, T is the temperature of the system, V is the unit cell volume, $\mathbf{r}$ is the position of methane and $\phi(\mathbf{r})$ is the interaction energy between methane and the framework at the given methane position $\mathbf{r}$.

We can separate the partition function into two terms, ideal term and the excess term, as shown below.

$$Q = Q_{\text{id}} \, Q_{\text{ex}} \tag{8}$$

$$Q_{\text{id}} = \frac{V}{\Lambda^3} \tag{9}$$

$$Q_{\text{ex}} = \frac{1}{V} \int_{\mathbb{R}^3 \in V} \exp(-\beta\phi(\mathbf{r}))\mathrm{d}\mathbf{r} \tag{10}$$

Subscripts id and ex stand for ideal and excess terms respectively. Because the expression of the excess partition function is the same as the spatial average value of the Boltzmann factor of the interaction energy over the unit cell volume, we can approximate the value of the excess partition function from the energy grid values as

$$Q_{\text{ex}} \approx \frac{1}{N} \sum_i^N \exp(-\beta E_i) \tag{11}$$

where N is the number of grid points and E_i is the ith value of the energy grid. From statistical mechanics, we can express the Helmholtz free energy A as a negative logarithm of the canonical partition function

$$\beta A = -\log Q \tag{12}$$

and if we define the reduced Helmholtz free energy as $\beta A = A^*$, the excess part of reduced free energy can be written as below.

$$A_{\text{ex}}^* = -log \, Q_{\text{ex}} \tag{13}$$

Because the ideal part of the Helmholtz free energy is independent of the energy grid, we can simply neglect the ideal part of the free energy. To match the free energy distribution between the real and the generated energy shapes, the loss function form is chosen as below

$$\begin{aligned} L_{\text{FE−matching}} = \; & |\text{E}_{\mathbf{X} \sim P_\mathbf{X}}[A_{\text{ex,real}}^*] - \text{E}_{\mathbf{z} \sim P_z}[A_{\text{ex,gen}}^*]| \\ & + |\text{Std}_{\mathbf{X} \sim P_\mathbf{X}}[A_{\text{ex,real}}^*] - \text{Std}_{\mathbf{z} \sim P_z}[A_{\text{ex,gen}}^*]| \end{aligned} \tag{14}$$

where Std is the standard deviation function. The subscript real means that value calculated from real energy grid **X** and subscript gen means that value calculated from the generated energy grid **X**′. The value of the Helmholtz free energy is very sensitive to the value of the low energy region of the energy grid. Therefore, if the generator tries to match the Helmholtz free energy distribution of the real energy grid, the values of the low energy region become more reasonable. The effect of free energy matching is described in the ESI Fig. S6.†

Also due to the crystalline nature of the zeolites, a new type of padding was implemented. The default padding of the convolution layers is zero padding, but with zero padding, the periodicity of the energy grid is only captured indirectly. As such, it is difficult to discriminate between the real energy grid and generated energy grid directly from the point of view of periodicity even though the generated grids should have periodicity strictly enforced. In order to remedy this, a new type of padding was applied, which we named periodic padding (illustration shown in Fig. S5†), to all the convolution layers in the discriminator. Subsequently, the discriminator directly captures the difference between periodic and aperiodic energy grids, resulting in an increased likelihood of generating periodic samples. It was observed that in the absence of periodic padding, the generated energy grids were for the most part non-periodic with slow convergence when it comes to learning.

It is well established that the standard GANs do not possess a universal metric to determine convergence (whereas some GANs have their own specific convergence criteria.[30,31] To determine the convergence of our model, Jensen–Shannon divergence (JSD) was estimated from the properties of the generated energy shape. JSD was selected for a couple of reasons. First, the training of GANs minimizes the JSD between the distribution of real data and the generated data[32] and second, the JSD is always defined to be numerically stable. As mentioned above, the JSD cannot be obtained directly from the energy shapes, the probability function of energy shapes was approximated as a probability function of properties. To estimate the probability of properties, in every 10 000 learning steps, 10 000 energy shapes were generated and the probability was estimated as a function of the Henry coefficient, void fraction and heat of adsorption using a multidimensional

2712 | *J. Mater. Chem. A*, 2019, **7**, 2709–2716

This journal is © The Royal Society of Chemistry 2019

histogram. It turns out that the approximated JSD is insensitive to sampling. And although the absolute value of JSD is a function of bins used in the histogram calculation, the order of JSD rank was almost identical. Because JSDs showed oscillatory behavior even after convergence, a model checkpoint that has minimum JSD was selected.

Results and discussion

The PCOD zeolite set contains 3 31 163 zeolites, from which 99 204 are deemed to have sufficiently high methane uptakes (*i.e.* methane $K_H > 10^{-6}$ mol kg^{-1} Pa^{-1}) and considered to have large working capacity. From this reduced set, 63422 orthorhombic zeolite materials (63324 PCOD and 98 IZA) were selected and their energy shapes were obtained using molecular simulations to create an initial training set for the ESGAN. By choosing just the orthorhombic cells, the lattice constants' dimensions reduced from six (three unit cell lengths and three unit cell angles) to three (lengths only), making it easier for the ESGAN to reproduce the lattice constants accurately. Moreover, given that the zeolite material with the highest methane working capacity is orthorhombic (*i.e.* zeolite ACO, WC = 156.74 cm^3 cm^{-3}), the performance target/limit remains the same in this reduced set.
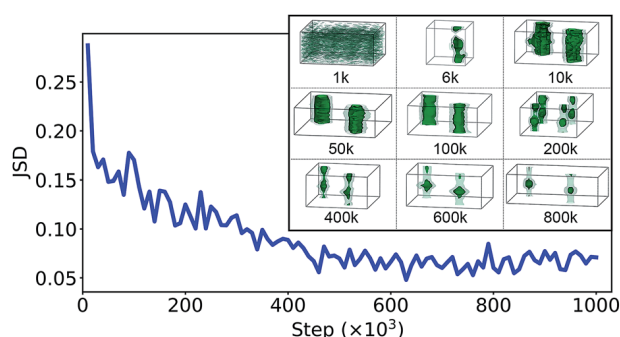

Fig. 3 Jensen–Shannon divergence as a function of learning steps (inset: snapshot evolution of a single shape from 1 K to 800 K steps).

Fig. 3 shows the evolution of sample energy shape snapshots as a function of the learning step along with its JSD curve. Initially, the generated energy shape (Fig. 3 1k image from the inset) resembles a typical noise distribution but with enhanced learning steps, it slowly evolves into a shape that looks more like a zeolite.

Finally, similar to the GANs that generate human faces, explicit visualization of the output images is important as a sanity check. As such, Fig. 4 includes side-by-side comparisons between randomly chosen energy shapes generated from our ESGAN and real zeolite shapes. The ESGAN does a good job of capturing the essential qualities of real zeolites and diverse selections of 1D, 2D, and 3D pore channels are generated.

To provide additional evidence that the energy shapes generated by the ESGAN yield are realistic, Henry's constants of methane (K_H), void fractions (VF), heat of adsorption (HoA), uptakes at 5.8 bar ($U_{5.8bar}$) and 65 bar (U_{65bar}) were computed for 1 00 000 energy shapes generated from the ESGAN. The distributions between real and generated zeolites (Fig. 5A–F) indicate good overall agreement across all property dimensions. To discern the similarity between multiple correlated properties, structure–property maps[33,34] were obtained for both real and generated zeolites (Fig. 5G, H).

The similar trends observed between the two distributions indicate that the ESGAN manages to capture the essential properties of real zeolites and does not generate anomaly data points. Finally, the maximum property values of the real and the generated energy shapes are summarized in Table 1. It is noteworthy that the difference in the methane working capacity between real and the ESGAN generated zeolites is only 1.7%. Overall, these data demonstrate the predictive power of the ESGAN in evaluating the accurate performance limits. All relevant data are included in the ESI† S2.

Previously, we established that the top performing energy shapes from the ESGAN possessed working capacity values that agree well with the top values from the real zeolites. On the other hand, this result might not be too surprising given that the initial training set included zeolites with the highest
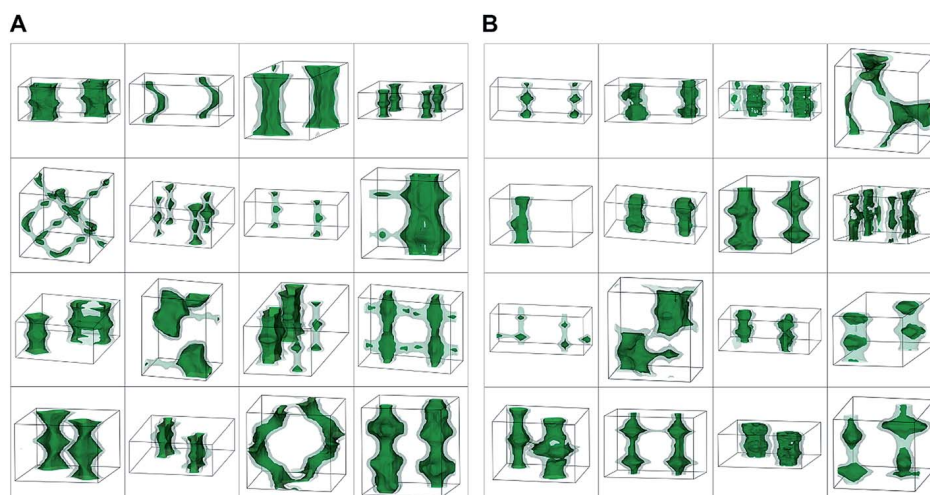

Fig. 4 (A) Real energy shapes of zeolites from Deem's database. (B) Generated energy shapes from the ESGAN.
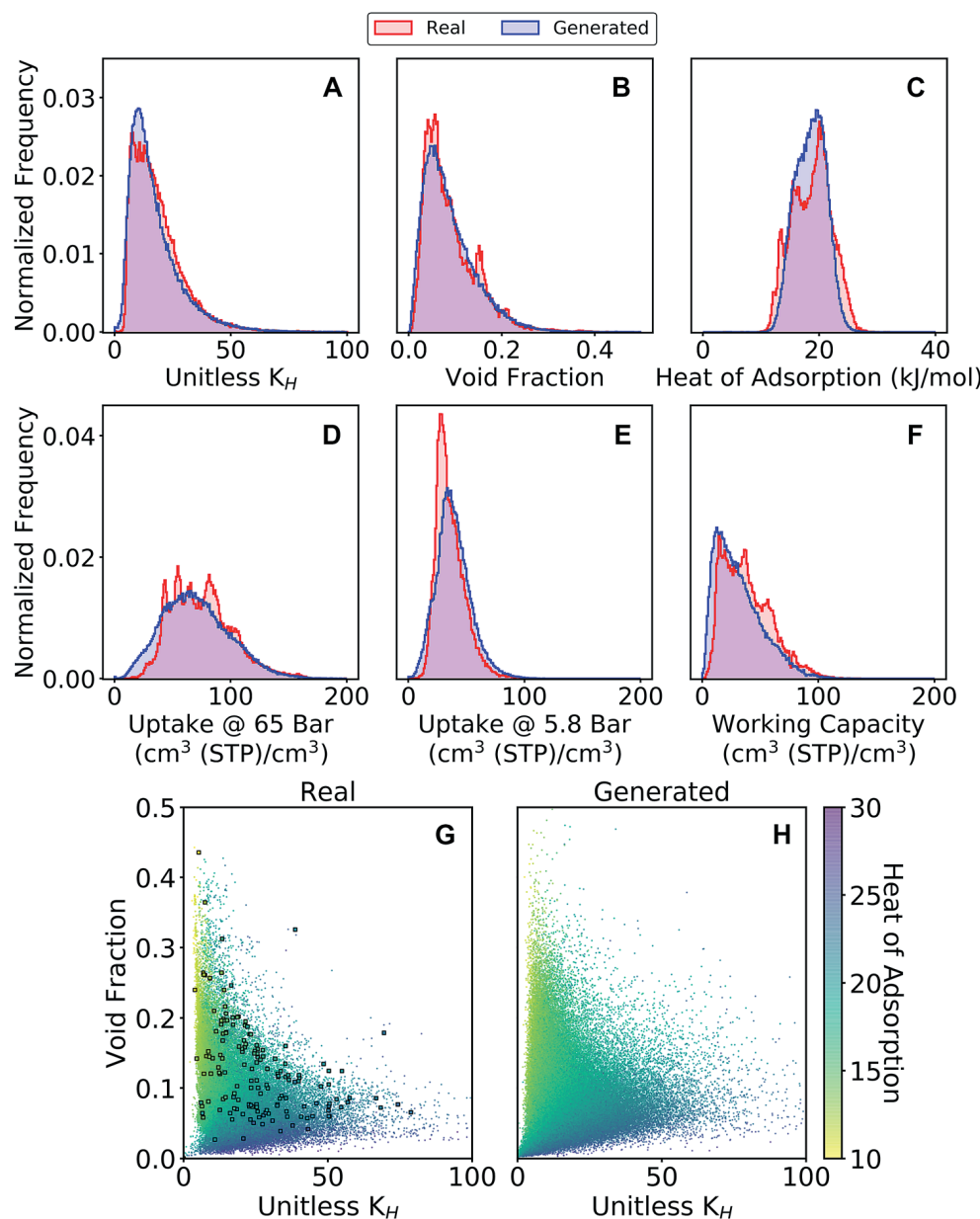
Fig. 5 Distributions between 62886 real and 94157 ESGAN generated zeolites (both excluding non-porous zeolites) for (A) K_H, (B) void fraction, (C) heat of adsorption, (D) 65 bar methane uptake, (E) 5.8 bar methane uptake, (F) working capacity. (G and H) Structure–property map comparisons between real and generated zeolites. Squares in G indicate IZA structures that have methane $K_H > 10^{-6}$ mol kg^{-1} Pa^{-1} (total 158 structures). All were taken at $T = 298$ K.

Table 1 The maximum property values of the real and the generated energy shapes

Sets			
Property	Real	Generated	Difference (%)
K_H	398.0	400.0	0.5
VF	0.55	0.53	3.6
HoA (kJ mol^{-1})	32.2	29.3	9.0
WC (cm^3 cm^{-3})	156.74	159.41	1.7
$U_{5.8bar}$ (cm^3 cm^{-3})	143.9	133.56	7.2
U_{65bar} (cm^3 cm^{-3})	232.76	224.82	3.4

working capacity values. To evaluate the effect of excluding the top performing zeolites from the training sets, nine additional subsets of the initial input training set were created where each set was capped by a zeolite with the maximum methane working capacity from set #1 = 50 cm^3 cm^{-3} to set #9 = 130 cm^3 cm^{-3} in increments of +10 cm^3 cm^{-3}, for the respective nine sets (Fig. 6). For each set, the ESGAN was trained independently and used to generate 1 00 000 energy shapes. In all the cases, the generated energy shapes with the highest working capacity outperformed the best zeolites from the input training sets. The differences are prominent in all the cases for example, set #6 with a maximum input zeolite of WC = 99.99 cm^3 cm^{-3} yields ESGAN generated shapes with the highest working capacity value of

2714 | *J. Mater. Chem. A*, 2019, **7**, 2709–2716

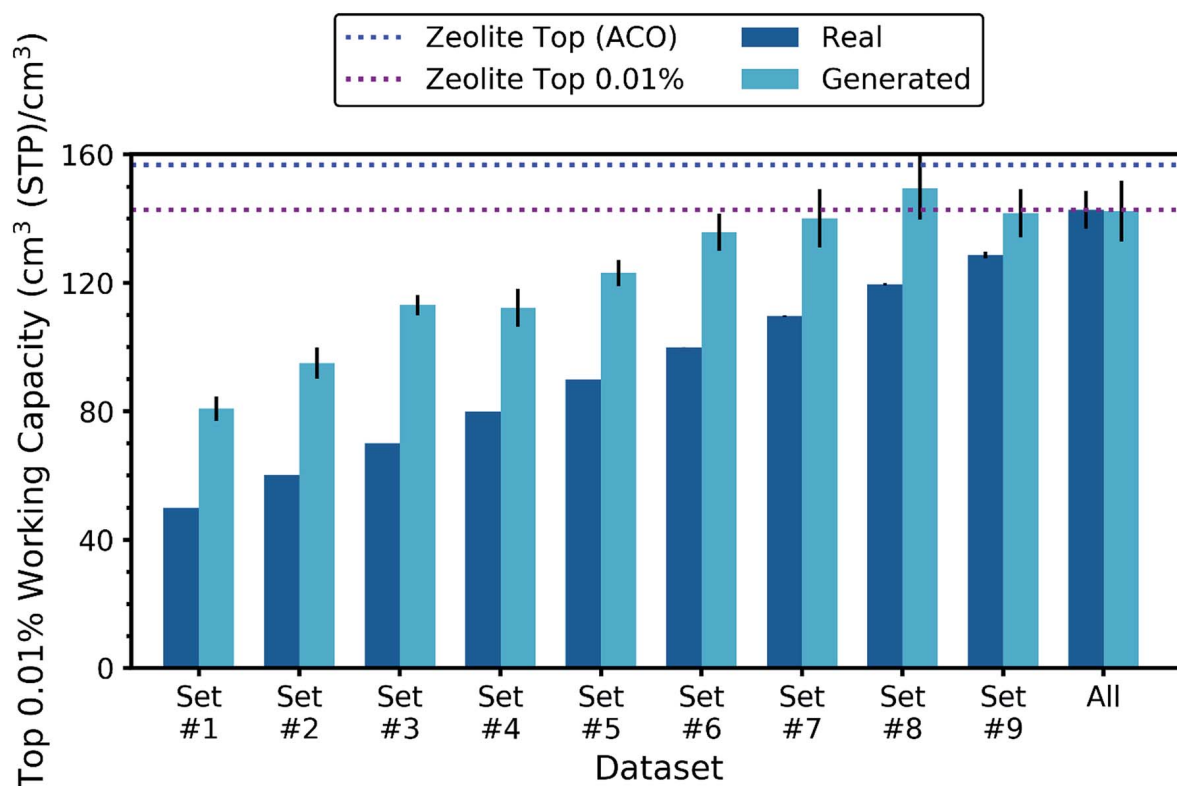This journal is © The Royal Society of Chemistry 2019

Fig. 6 Top 0.01% (*e.g.* top 10) working capacity for the input real zeolites and the output generated zeolites (1 00 000 for each set). The maximum working capacity for each input set ranges from 50 cm^3 cm^{-3} (set #1) to 130 cm^3 cm^{-3} (set #9) with the "All" set including all zeolites.

149.58 cm^3 cm^{-3}, which is very close to the true maximum value of 156.74 cm^3 cm^{-3} of zeolite ACO.

And what is interesting is that in all nine sets, the best energy shapes do not significantly exceed the true maximum value, indicating that the ESGAN can deduce the performance limits of methane storage in zeolites as long as the input set contains zeolites with reasonably high working capacity zeolites.

Conclusion

In this work, we have demonstrated for the first time that an ANN can be used to generate energy shapes of crystalline porous materials that yield accurate performance limits for methane storage. In future studies, the optimal energy shape generated from the ESGAN can perhaps be used to reverse-engineer the corresponding optimal material and serve as a blueprint for potential experimental synthesis. And although this work focused on methane adsorption properties in pure silica zeolite structures, it is conceivable that the conceptual framework can be applied to other gas molecules and other classes of nano-materials. Moreover, while we focused on the adsorption properties in this work, the energy shapes can be used to obtain diffusion properties and as such, this work can potentially be extended to investigate diffusion properties as well. Finally, these ideas can be extended to evaluate the materials' performance limits in other important applications such as catalysis, drug delivery, and sensors.

Data and materials availability

ESGAN code can be downloaded from https://github.com/Sangwon91/ESGAN.

Author contributions

S. L. devised the neural network and was in charge of the neural network part. B. K. conducted the molecular simulations. J. K. formulated the project. All authors contributed to exchanging ideas and writing the manuscript/SI.

Conflicts of interest

The authors declare no competing interests.

Acknowledgements

Notes and references

1 Y. S. Bae and R. Snurr, *Angew. Chem., Int. Ed.*, 2011, **50**, 11586–11596.

2 J. R. Li, R. J. Kuppler and H. C. Zhou, *Chem. Soc. Rev.*, 2009, **38**, 1477–1504.

3 J. A. Mason, M. Veenstra and J. R. Long, *Chem. Sci.*, 2014, **5**, 32–51.

4 M. G. Campbell, S. Liu, T. Swager and M. Dinca, *J. Am. Chem. Soc.*, 2015, **137**, 13780–13783.

5 D. Deng, *et al.*, *Nat. Nanotechnol.*, 2016, **11**, 218–230.

6 D. H. Seo, *et al.*, *Nat. Chem.*, 2016, **8**, 692–697.

7 C. M. Simon, *et al.*, *Energy Environ. Sci.*, 2015, **8**, 1190–1199.

8 R. Gómez-Bombarelli, *et al.*, *Nat. Mater.*, 2016, **15**, 1120–1127.

9 L.-C. Lin, *et al.*, *Nat. Mater.*, 2012, **11**, 633.

10 P. Z. Moghadam, *et al.*, *Nat. Commun.*, 2018, **9**, 1378.

11 A. Singh, K. Mathew, H. Zhuang and R. Hennig, *J. Phys. Chem. Lett.*, 2015, **6**, 1087–1098.

12 C. E. Wilmer, *et al.*, *Nat. Chem.*, 2012, **4**, 83–89.

13 S. Chong, G. Thiele and J. Kim, *Nat. Commun.*, 2017, **8**, 1539.

14 Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436–444.

15 D. Silver, *et al.*, *Nature*, 2016, **529**, 484–489.

16 G. Carleo and M. Troyer, *Science*, 2017, **355**, 602–606.

17 C. W. Tom Kalil, http://www.whitehouse.gov/blog/2011/06/24/materials-genome-initiative-renaissance-american-manufacturing, 2011.

18 F. Gándara, H. Furukawa, S. Lee and O. M. Yaghi, *J. Am. Chem. Soc.*, 2014, **136**, 5271–5274.

19 Y. Peng, *et al.*, *J. Am. Chem. Soc.*, 2013, **135**, 11887–11894.

20 R. Gómez-Bombarelli, *et al.*, *ACS Cent. Sci.*, 2018, **4**, 268–276.

21 International Zeolites Association (IZA) Home Page, http://www.iza-structure.org/databases, (accessed April, 2018).

22 R. Pophale, P. A. Cheeseman and M. W. Deem, *Phys. Chem. Chem. Phys.*, 2011, **13**, 12407–12412.

23 D. J. Earl and M. W. Deem, *Ind. Eng. Chem. Res.*, 2006, **45**, 5449–5454.

24 J. Kim, R. L. Martin, O. Rübel, M. Haranczyk and B. Smit, *J. Chem. Theory Comput.*, 2012, **8**, 1684–1693.

25 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.

26 Y. Lee, *et al.*, *Nat. Commun.*, 2017, **8**, 15396.

27 D. P. Kingma and M. Welling, 2013, arxiv.1312.6114.

28 I. J. Goodfellow *et al.*, 2014, arxiv.1406.2661.

29 T. Salimans *et al.*, 2016, arxiv.1606.03498.

30 M. Arjovsky, S. Chintala and L. Bottou, 2017, arxiv.1701.07875.

31 D. Berthelot, T. Schumm and L. Metz, 2017, arxiv.1703.10717.

32 I. Goodfellow, 2016, arxiv.1701.00160.

33 W. S. Jeong, *et al.*, *Proc. Natl. Acad. Sci.*, 2017, **114**, 7923–7928.

34 Y. Colon and R. Snurr, *Chem. Soc. Rev.*, 2014, **43**, 5735–5749.