Published in final edited form as:
Nat Mater. 2022 December ; 21(12): 1419–1425. doi:10.1038/s41563-022-01374-3.

A Data-Science Approach to Predict the Heat Capacity of Nanoporous Materials

Seyed Mohamad Moosavi[1,2,*], **Balázs Álmos Novotny**[1], **Daniele Ongari**[1], **Elias Moubarak**[1], **Mehrdad Asgari**[1,3], **Özge Kadioglu**[1], **Charithea Charalambous**[4], **Andres Ortega-Guerrero**[1], **Amir H. Farmahini**[5], **Lev Sarkisov**[5], **Susana Garcia**[4], **Frank Noé**[2,6,7,8], **Berend Smit**[1,*]

[1]Laboratory of Molecular Simulation, Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Valais, Switzerland

[2]Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 12, 14195 Berlin, Germany

[3]Department of Chemical Engineering & Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK

[4]The Research Centre for Carbon Solutions (RCCS), School of Engineering and Physical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, United Kingdom

[5]Department of Chemical Engineering, School of Engineering, The University of Manchester, Manchester M13 9PL, United Kingdom

[6]Department of Physics, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany

[7]Department of Chemistry, Rice University, Houston TX, USA

[8]Microsoft Research, Cambridge, UK

Abstract

The heat capacity of a material is a fundamental property that is of great practical importance. For example, in a carbon capture process, the heat required to regenerate a solid sorbent is directly related to the heat capacity of the material. However, for most materials suitable for carbon capture applications the heat capacity is not known, and thus the standard procedure is to assume the same value for all materials. In this work, we developed a machine-learning approach, trained on density functional theory simulations, to accurately predict the heat capacity of these materials, i.e., zeolites, metal-organic frameworks, and covalent-organic frameworks. The

* seyedmohamad.moosavi@fu-berlin.de, berend.smit@epfl.ch.

Author contributions

S.M.M., A.H.F., C.C., L.S., S.G., and B.S. formulated the scientific problem and the engineering context. S.M.M. developed the machine learning framework with the help of F.N. S.M.M., D.O., Ö.K., and A.O.G. performed the heat capacity calculations. E.M., C.C., and S.G. performed process modeling. B.A.N. and M.A. synthesized the MOFs and measured the heat capacities. All authors contributed to the analysis and discussion of the results. S.M.M. and B.S. wrote the paper with the contribution of all authors.

Competing interest

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

accuracy of our prediction is confirmed with experimental data. Finally, for a temperature swing adsorption process that captures carbon from the flue gas of a coal-fired power plant, we show that for some materials the heat requirement is reduced by as much as a factor of two using the correct heat capacity.

Europe PMC Funders Author Manuscripts

Metal-organic frameworks (MOFs) and related nanoporous materials are promising candidates for a wide range of energy-related applications, including gas separation, gas storage, and catalysis.[1–3] In recent years, considerable effort has been focused on the understanding and tuning of the adsorption[4, 5] and catalytic[6] properties of these materials.

For many applications, the thermal properties of the materials also play a role in their performance.[7–9] For example, the energy penalty of a carbon capture process using a temperature swing adsorption process (TSA), where the adsorbent material is regenerated by heating the adsorption bed, is directly influenced by the heat capacity of the adsorbent.[8] For MOFs, it is generally assumed that the heat capacity is a constant for all these materials.[10–12] However, there is little foundation for this assumption, and it is rather a pragmatic simplification. Only a few experimental studies on the heat capacity of these materials exist,[13, 14] and their reported values have been used throughout the literature ever since for energy penalty calculations. Currently, we lack an understanding of how the heat capacity is related to the underlying crystal structure. Also, we do not have a methodology to accurately and efficiently evaluate the heat capacity at scale for the enormous number of available porous materials. Developing methods to accurately predict the heat capacity is essential to improve the accuracy of large-scale performance evaluation of porous materials.

In this work, we have used state-of-the-art quantum mechanical calculations to accurately predict the heat capacity of a set of representative nanoporous materials. However, as the cost of such calculation scales with the number of atoms per unit cell (N_a) as $\mathcal{O}\left(N_a^4\right)$, it can only be carried out for a subset of materials whose number of atoms is not too large (<200 atoms). We show that we can take advantage of the fact that the heat capacity can be estimated as a local property, and hence use this data as a training set for a machine-learning methodology that enables fast, accurate, reliable prediction of the heat capacity, even if the number of atoms far exceeds the limit of our quantum calculations. We evaluate the accuracy of this approach using newly measured experimental data. In addition, we show that for carbon capture applications, the assumption of a constant heat capacity of these materials has resulted in a substantial overestimation of the energy requirements.

Theoretical aspects

The heat capacity measures how much energy it takes to raise the temperature of a solid one degree. Theoretically, the amount of energy is proportional to the change in the average energy of the collective vibration of atoms or phonons. Within the harmonic approximation, the heat capacity is formulated as a summation over the contribution of these lattice vibrations via:

Europe PMC Funders Author Manuscripts

$$C_v = \sum_{\omega} k_B \left(\frac{\hbar\omega_i}{k_B T}\right)^2 \frac{\exp\left(\frac{\hbar\omega_i}{k_B T}\right)}{\left(\exp\left(\frac{\hbar\omega_i}{k_B T}\right) - 1\right)^2}, \quad (1)$$

where the summation is over the vibrational frequencies (ω), and ω_i is a vibrational frequency, T the temperature, and $\hbar$ and k_B Planck's and Boltzmann constants, respectively. See the method section for the details of phonon frequency calculations.

This formulation of the heat capacity follows the experimental observation of how the heat capacity of solids changes with respect to the temperature. The heat capacity tends to zero at low temperatures, where only low-frequency modes are occupied, and then converges to a constant (3R) at high temperatures, where all vibrational modes are active. This behavior is shown for MOF-74 in Figure 1, where we show the density functional theory (DFT) computed heat capacity of the materials with two different metals, Cobalt and Zinc (Co-MOF-74 and Zn-MOF-74, respectively). In this figure, we also compare the DFT results with experimental data. For these experiments, it was essential to ensure proper activation of the material. The details of the synthesis and experimental procedure can be found in the method section and SI. We observed a good agreement between the computed values and the experimental measurements. These results provide confidence for our computational methodology and support the conclusion of the previous works[15, 16] that the harmonic approximation provides sufficiently reliable data for the heat capacity of an empty framework.

In this study, we focus on predicting the heat capacity of an empty framework. For carbon capture applications one also needs to take into account the effect of the adsorbed molecules. Kapil et al.[15] have shown that the effect of the adsorbed CO_2 molecules on the lattice vibrations of the MOF atoms is small, and hence does not influence the heat capacity of the material. As a consequence, to obtain the heat capacity of the total system, we can simply add the contributions of the gas molecules. In SI sections 4.2 and 4.3, we show that the contributions of these gas molecules are relatively small for carbon capture applications.

Understanding structure - heat capacity relationships

From a reticular chemistry point of view, a MOF structure is an assembly of metal nodes and organic linkers on a topology network. Therefore, it is important to understand how the variation of each of these factors influences heat capacity.

Isoreticular MOF-74 structures are an ideal case for understanding the role of metal on the heat capacity as they have been made with a wide range of metals, including Mg, Mn, Fe, Ni, Co, and Zn.[17, 18] Figure 2a shows that the computed heat capacity depends on the type of metal, for which Mg-MOF-74 has the lowest molar heat capacity. To explain this observation, we inspect the lattice vibrational frequencies of these systems. As the only difference between these structures are their metals, we inspect the lattice vibrational frequencies of the metal centers. Figure 2d shows the histograms of the projected lattice vibrational frequencies on metal centers for Mg and Zn MOF-74, where we see

that Zn has a shift in the vibrational frequencies to lower frequencies compared to Mg. Theoretically, different frequencies contribute to the heat capacity differently: At low and intermediate temperatures, the low-frequency modes are dominant in contributing to the heat capacity, and only at high temperatures the high-frequency modes also participate in the heat capacity (see Figure 2b for a quantitative picture). Substituting Mg with Zn shifts the vibrational frequencies to lower frequencies, and hence, leading to a higher heat capacity in Zn-MOF-74. One can explain the effect of changing the mass of the metal node if we envision this metal to be connected with harmonic springs to its neighbors. The vibrational frequency of this metal is proportional to $\sqrt{k/m_{\text{eff}}}$, where k is the spring stiffness and m_{eff} is the effective mass. If change of the mass would be the only contribution, one would expect monotonically increasing heat capacity with increasing the molecular mass of the metal. Figure 2a shows that differences in mass only partly explains the data; when the masses are similar, the details of the metal-linker interactions are also important.

A more affordable way to describe the interactions between the atoms of a MOF is to use a classical force field. For example, the universal force field (UFF) can predict the mechanical properties of a MOF with similar accuracy as DFT.[19, 20] However, Figure 2a shows that this force field systematically underestimates the heat capacity. If we softened the spring constants between metals and linkers, we can obtain better agreement with the DFT results. However, one would need a metal-linker dependent correction factor to correctly capture the subtleties of the chemistry, which limits the usefulness of the UFF force field to predict the heat capacity of MOFs. Further analysis (see section 1.2 in SI) shows that the deficiency of the UFF force field is mainly due to the metal-linker interactions, whereas the linker contributions are well described.

Another important variation in MOF structures is the modification of linkers and the underlying network topology. To better understand these variations, we look at a set of zeolitic imidazolate framework (ZIF) structures with different organic linkers. These materials are ideal for this investigation as they can be synthesized with the same topology but varying linkers, as well as the same linkers but varying topology. However, the number of atoms and the dimensions of the unit cell of these ZIF structures are too large for DFT calculations, and therefore we used the UFF force field. As the metal is not changing among the ZIFs, one can expect that we underestimate the experimental value, but this underestimation is equal for all ZIFs.

We analyze the heat capacity of 200 hypothetical ZIF structures assembled using four unique linkers and 50 topologies.[21] Figure 2c shows the heat capacity of the ZIFs. We observe a large dependence of the heat capacity on the linker functionalization (dots with different colors) and a very small influence by the framework topology (dots with the same color). Comparing structures with the dichloroimidazolate (dcIM) with imidazolate (IM) linkers is instructive to further understand the role of linker functionalization, as the only difference between these structures is the replacement of hydrogen atoms on the linker with chlorine. Clearly, replacing the hydrogens with heavier atoms (i.e., chlorine) increases the heat capacity of the material, similar to the simple spring-mass model described above, where the heavy chlorine atoms shift the vibrational frequencies to smaller values, leading to a higher heat capacity per atom.

A machine learning approach

A machine learning model exploits the similarities between structures in the training set to predict the target properties of new structures in the test set without performing the actual computations or experimental measurements on those.[22, 23] The similarity between different materials and structures can be explicitly encoded by chemically motivated descriptors[24–26] or learned as part of a deep learning framework.[27] For example, machine learning models based on the chemical formula[25, 28] or property-labeled materials fragment descriptor[24] were found to be effective in predicting heat capacity of solids and semiconductors.[29] These approaches are based on a global description of the material. However, these approaches based on a global description are of limited use for MOFs as their chemical diversity is so large that one needs a large data set to be able to train a predictive model. As we have to rely on DFT calculations, we can only generate data for a small subset of MOFs.

A machine learning method that follows the relevant physics will typically require less training data than a naive model. Recall that the heat capacity is given by the summation over atomic vibrations (equation 1). Therefore, we devise an approach based on the summation of the local contribution of each atom to the total heat capacity. Particularly, our simulations on ZIFs show that changes in the topology have a minor effect on the heat capacity, indicating that the relevant chemical environment is relatively short-ranged. Hence, our features only need to account for the local environment around each atom (see Figure 3a). Compared to the previous machine learning approaches, the main advantage of our methodology is that each MOF contains many different chemical environments for each atom. Hence, in the selection of the training set, we only need to generate data for an adequately diverse set of atoms in different chemical environments. These environments include different elements and coordination environments, as well as different sizes of crystals. As we can obtain such a diverse set using MOFs with a relatively small number of atoms or with high symmetry, we have a training set that we also can use to predict the heat capacities of those MOFs that are too large for DFT calculations. In essence, our machine learning approach is staying as close as possible to the physics of the problem. Once trained, it predicts the contribution of each atom to the heat capacity, depending on the local chemical environment.

To featurise local atomic environments, we include the atom identity and descriptors to capture both chemical and geometric similarities (Figure 3a). To capture the local chemistry, we use Voronoi-tessellation based descriptors,[30, 31] which give statistics of chemical heuristics for the neighboring atoms. Furthermore, we use symmetry functions[32] and AGNI fingerprints[33] to encode geometric similarities. These descriptors are expressive enough to provide adequate flexibility for the machine learning model to capture the similarities between the chemical environments to predict heat capacity.

We selected ~230 structures with diverse chemical environments from the experimental structures in the CoRE-MOF database,[34] experimental covalent organic frameworks in CURATED-COFs,[35] and the experimental all-silica zeolites in IZA database.[36] The DFT predicted heat capacities of these structures are shown in Figure 3b. Remarkably, Figure 3b shows that all silica zeolites have very similar heat capacities, which justifies the use

of a constant heat capacity in the performance evaluation of these materials in chemical processes. In contrast, MOFs and COFs show a wide range of heat capacities, ranging from 0.4 to 1.1 $J \cdot g^{-1} \cdot K^{-1}$. While the general correlation between the heat capacity and average atomic mass exists, to make quantitative predictions, we need to capture the subtle chemical differences of these materials in our machine learning approach.

We use 120 of the structures to train machine learning models to predict the heat capacity and keep the remaining 110 structures for evaluating the models. The performance of machine learning models on the test set is shown in Figure 3c. The model reaches an excellent accuracy (~ 3% relative error). Achieving such high accuracy with only using 120 structures for training the model demonstrates the advantage of incorporating the physics of the problem in the machine learning model. In addition, we estimate the uncertainty of the predictions using the variance in predictions of an ensemble of machine learning models trained on different training sets (see method section for details). Figure 3c shows that this approach effectively assigns high uncertainty to erroneous predictions. This enables us to use our machine learning models to predict the heat capacity of materials in the porous material databases with a controlled error and confidence.

Using our uncertainty-aware machine learning models, we predict the heat capacity of the enormous materials available from multiple nanoporous materials databases. In Figure 4a and 4b, we show the predicted heat capacity at room temperature for all the experimental structures reported in CoRE-MOF,[34] CURATED-COFs,[35] and IZA[36] zeolites databases. We note that for many of these materials, quantum calculations are not feasible due to the excessive computational costs.

It is interesting to compare our machine learning predictions with experimental measurements. In addition to the few experimental values reported in the literature,[37] we used Differential Scanning Calorimetry (DSC) to measure the heat capacity of some additional MOFs. Figure 4c and 4d shows that except for a few cases, we have a good match between the predicted and measured heat capacities. By carefully inspecting the experiment protocols, we realize that the materials with a large discrepancy appear to be not (fully) activated. Solvent molecules that remain in the pores of the material have a large effect on the heat capacity. In our experimental procedure, we ensure that the MOFs are fully activated (see methods and SI section 3 for further discussion). If we discard the materials that are not fully activated, these results show that the machine learning model can provide fairly accurate predictions for the heat capacity of fully activated MOFs.

In Figure 2a, we presented the DFT values of the heat capacity for MOF-74 with different metals. An unexpected result is the heat capacity (per atom) of Mn-MOF-74, which is higher than the corresponding value for other similar mass MOFs, e.g., Ni-MOF-74. This is a puzzling result, as the only change is the metal, and if we increase the atomic mass we expect a monotonic increase of the heat capacity due to a shift of frequencies to lower values. A feature analysis of the machine learning model shows that 60% of the heat capacity prediction can be explained by knowing the atomic mass of all atoms in the material (see Figure 3d and method section for details). The remaining 40% is the local geometry and chemistry. In the case of Mn-MOF-74, the feature analysis shows that

the geometry is the decisive factor (see Figure S.11 in SI). Interestingly, if we take the Ni-MOF-74 geometry and simply replace the metal, the machine learning model gives similar values for all but Mn-MOF-74. For Mn-MOF-74, the feature analysis shows that the higher predicted value is caused by longer metal-linker bonds, where these weaker bonds shift the frequencies to lower values.

Discussion

From a practical point of view, for new MOFs or COFs crystal structures, we can now provide an estimate of the heat capacity of the material, together with an estimate of the reliability of this prediction. We provide these for the materials available from multiple nanoporous materials databases, including the MOFs in the CoRE-MOF[34] and QMOF[38] databases, zeolites in IZA,[36] and COFs in CURATED-COFs[35] (see section 2.2 in SI for the details of these databases). One can expect that, for new structures with chemical environments that are very different from the training set, our model will indicate that the predictions are unreliable. For these structures, we need to update the machine learning model by including additional DFT calculations.

One of the practical motivations of our work is to understand the importance of accurate knowledge of the heat capacity in the use of porous materials in carbon capture processes. In particular, in a temperature swing adsorption process (TSA), an objective for material design is to minimize the heat required to regenerate the material per kilogram of recovered CO_2. As the contributions from heating the adsorbed gases in the column are negligible for this application (see sections 4.2 and 4.3 in SI for discussion), this energy requirement depends mostly on the heat capacity of the material. Figure 4 shows that for zeolites, it is indeed reasonable to assume that the heat capacity does not differ considerably among different materials. However, this assumption does not hold for MOFs. As our machine learning model gives the heat capacity with high accuracy, we can quantify the impact of the differences in the heat capacity on the ranking of materials in a TSA process.

In Figure 5a, we compare the heat requirement using the actual values of the heat capacity with the assumption of a constant heat capacity (0.985 $J \cdot g^{-1} \cdot K^{-1}$).[11] For materials with a low heat capacity (red and yellow dots), the heat requirement reduces by as much as 50% and makes some of them top performings. In fact, using the actual values of the heat capacity changes the ranking completely compared to using a constant value (see Figure 5b). In particular, the ranking correlation is almost lost when we look at the materials with similar CO_2 uptake: the ones with lower heat capacity appear to be performing considerably better. As the heat capacity used in the previous studies[11] was considerably above the average, many of these studies have overestimated the energy requirements; for some materials as much as 50%. Since these energy requirements contribute substantially to the total cost of the capture process, our results can have profound practical impacts (see the discussion in SI). However, it is important to note that the total energy requirement for a carbon capture process depends on the specific details of the process design, and it will be also affected by factors such as the shaping of the materials and the type of adsorption column.

Europe PMC Funders Author Manuscripts

Our machine learning approach is not limited to porous materials and can be extended to other classes of materials, provided one can generate the relevant training set.

Methods

Computations of heat capacity

To compute the heat capacity, we use the harmonic approximation. Similar to our observation (see Figure 1), previous studies[15, 16] have shown that it gives sufficient accuracy for estimating the heat capacity of an empty framework. However, it is important to note that the harmonic approximation might fail to correctly describe materials with anharmonic effects, e.g., flexible materials or large host-guest interactions. Hence, it will be interesting to investigate the extent of validity of this approximation for estimating the heat capacity of these materials using path integral molecular dynamics[15, 16] simulations in the future.

Within the harmonic approximation, the lattice vibrational frequencies (phonons) were extracted from the Hessian matrix of the lattice energy. The Hessian matrix is the second-order partial derivative of the lattice energy with respect to atom displacements. To build the Hessian matrix, we use a finite difference approach where each atom of the crystal is displaced, and density functional theory (DFT) or a molecular mechanics force field is used to compute the force on all the atoms of the crystal upon this displacement. The eigenvalues and eigenvectors of this matrix correspond to the vibrational frequencies and vibrational modes of the system, respectively. Mathematically speaking, for a crystal with N_a atoms, there are $3N_a$ phonon modes (eigenvectors and eigenvalues of an N_a body system in 3D). However, three of these modes always have a negative frequency because of the translation degrees of freedom. The labels for training the machine learning models are the contribution of each atom to the total heat capacity of the material. We estimate these contributions by projecting the phonon density of state on each atom of the crystal.

The DFT calculations were performed within the generalized gradient approximation (GGA) level of theory using Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional with DFT-D3(BJ) dispersion corrections.[39, 40] We use GTH pseudopotentials,[41] and DZVP-MOLOPT-SR contracted Gaussian with an auxiliary plane wave basis set. Since the phonon calculation procedure relies on the assumption that the structures are at the minimum energy configurations that are consistent with the quantum or classical mechanical method used to describe the potential energy surface of the material. Therefore, in the DFT calculations, we use a tight optimization setting that is described in section 1.1 in SI to avoid negative frequencies.

All lattice displacement and post-processing for vibration calculations were performed using Phonopy.[42] We use CP2K[43] for DFT calculations and LAMMPS,[44] LAMMPS-interface,[19] and phono-LAMMPS[45] for the molecular mechanics calculations. The DFT calculation recipe is adapted from our previous work[35] (see section 1.1 in SI for details).

Measurements of heat capacity

The MOFs were synthetized based on the reported procedures in the literature, outlined in detail in SI. To obtain a reliable measurement of the heat capacity, we consider MOFs with

Europe PMC Funders Author Manuscripts

a robust solvothermal synthesis and stability of framework up to 200 °C. Furthermore, a well-defined state of sorption and clear stoichiometry at this temperature were requisites. Obtained crystalline phases were identified by matching powder x-ray diffraction (PXRD) patterns and thermogravimetric analysis (TGA) results to previous reports (See section 3 in SI).

The ground MOF samples were crimped in non-hermetically sealed aluminum pans. Heat capacity was determined against the sapphire standard by Differential Scanning Calorimetry (DSC), in the range of 20 °C to 200 °C. Samples were measured repeatedly in an argon atmosphere until three consecutive DSC traces were overplayed. The gas phase was purged after each measurement, and upon converging of traces, the lack of heat flow peaks, which may be associated with desorption phenomena or phase change, was confirmed. Heat flow was normalized to the mass of the activated samples. Sample mass reductions were normalized to initial sample masses and confirmed to correspond to expectations based on TGA results. Asymmetry of stray heat flow was determined from the 200 °C dwell phase of each evaluated DSC trace. This asymmetry was assumed to be proportional to the deviation of block temperature from ambient temperature and corrected for accordingly.

Machine learning

The chemical environment of each atom is described using a feature vector comprised of elemental properties as well as descriptors for chemical and geometric similarities. For the elemental properties, we include atomic number and mass, row and column on the periodic table, covalent radii, and Pauling electronegativity. To capture geometric and chemical similarities, we use AGNI fingerprint and Gaussian symmetry functions,[32, 33] in addition to the Voronoi tessellations based features on the statistics of atomic properties of neighbouring atoms.[30, 31] These features were computed for each atom of the crystals using Matminer and Pymatgen.[46, 47]

We use gradient boosted decision trees (GBDT) machine learning model as implemented in XGBoost[48] to map the feature vectors to the labels. To train the model, we first split our dataset into 120 structures for training and the remaining 112 structures for testing. The 120 training structures contain ~8500 atoms, that we use 90% of them for training, 10% for validation. The feature vectors were centered to zero and scaled using the mean and standard deviation for each training set, respectively. The hyperparameters of the GBDT model were optimized using 10-fold cross-validation.

To quantify the uncertainty of the model, we train an ensemble of 100 GBDT models and use the standard deviation of the prediction as a metric of uncertainty. The training set for each model in the ensemble was resampled with replacement from the original train set using bootstrapping as implemented in Scikit-learn.[49]

The final models, which are made available online, are trained using all structures to exploit the full dataset.

TSA process modelling and adsorption data

To demonstrate the effect of the specific heat capacity of the adsorbent in the process performance, we use an equilibrium shortcut temperature swing adsorption (TSA) model based on the model developed by Ajenifuja et al.[50] For this, we assumed a standard TSA process that consists of 3 steps: (i) adsorption, (ii) open heating, and (iii) open cooling. The full description of the model and model parameters are available in SI. Also, the details of adsorption data calculations, charges, and force field are outlined in section 4 in SI.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

The authors would like to acknowledge support from the ACT PrISMa Project (No 299659), which has received funding through the ACT programme (Accelerating CCS Technologies, Horizon2020 Project No 294766). Financial contributions made from BEIS together with extra funding from NERC and EPSRC, UK; RCN, Norway; SFOE, Switzerland and US-DOE, USA, are gratefully acknowledged. Additional financial support from TOTAL and Equinor, is also gratefully acknowledged. S.M.M. was supported by the Swiss National Science Foundation (SNSF) under Grant P2ELP2_195155. M.A. was supported by the Swiss National Science Foundation (SNSF) under Grant number P2ELP2_195134. C.C. is also supported by the UKRI ISCF Industrial Challenge within the UK Industrial Decarbonisation Research and Innovation Centre (IDRIC) Award number EP/V027050/1. F.N. acknowledges funding from the European Commission (ERC CoG 772230), the BMBF (Research center BIFOLD), and the Berlin Mathematics Center MATH+ (AA2-8). The calculations of this work were enabled by the Swiss National Supercomputing Centre (CSCS), under project ID s1019. S.M.M. and B.A.N. thank Miriam Jasmin Pougin for making experimental MOF structures computation ready. S.M.M. thanks Dr. Leopold Talirz and Dr. Mohsen Sadeghi for fruitful discussions.

Data availability

Supplementary Information, containing details of the theoretical aspects, machine learning methodology, experimental measurements, and process modeling, is available for this paper. In addition, the crystal structures of the training set together with the DFT-optimised geometries and the corresponding calculated thermal properties, tabulated heat capacity of MOFs in CoRE-MOF[34] and QMOF database,[38] zeolites in IZA,[36] COFs in CURATED-COFs,[35] at different temperatures, the data that were used and are needed to reproduce the results of this study, the thermogravimetric analysis, synthesis protocols, powder X-ray diffraction exported from electronic lab notebooks, and the codes to generate figures of the paper are deposited on the Materials Cloud[51] archive that can be accessed via https://doi.org/10.24435/materialscloud:p1-2y. Correspondence and requests for additional materials should be addressed to the corresponding authors.

Code availability

The featurisation, prediction, and trained models are available from https://github.com/SeyedMohamadMoosavi/tools-cp-porousmat.

References

1. Furukawa, Hiroyasu; Cordova, Kyle E; O'Keeffe, Michael; Yaghi, Omar M. The chemistry and applications of metal-organic frameworks. Science. 2013; 341 (6149) 1230444–1230456. [PubMed: 23990564]
2. Bavykina, Anastasiya; Kolobov, Nikita; Khan, Il Son; Bau, Jeremy A; Ramirez, Adrian; Gascon, Jorge. Metal–organic frameworks in heterogeneous catalysis: recent progress, new trends, and future perspectives. Chemical Reviews. 2020; 120 (16) 8468–8535. [PubMed: 32223183]
3. Rosen, Andrew S; Notestein, Justin M; Snurr, Randall Q. Realizing the data-driven, computational discovery of metal-organic framework catalysts. Current Opinion in Chemical Engineering. 2022; 35 100760
4. Boyd, Peter G; Chidambaram, Arunraj; García-Díez, Enrique; Ireland, Christopher P; Daff, Thomas D; Bounds, Richard; Gładysiak, Andrzej; Schouwink, Pascal; Moosavi, Seyed Mohamad; Mercedes Maroto-Valer, M; , et al. Data-driven design of metal–organic frameworks for wet flue gas co 2 capture. Nature. 2019; 576 (7786) 253–256. [PubMed: 31827290]
5. Chen, Zhijie; Li, Penghao; Anderson, Ryther; Wang, Xingjie; Zhang, Xuan; Robison, Lee; Redfern, Louis R; Moribe, Shinya; Islamoglu, Timur; Gómez-Gualdrón, Diego A; , et al. Balancing volumetric and gravimetric uptake in highly porous materials for clean energy. Science. 2020; 368 (6488) 297–303. [PubMed: 32299950]
6. Diercks, Christian S; Liu, Yuzhong; Cordova, Kyle E; Yaghi, Omar M. The role of reticular chemistry in the design of co 2 reduction catalysts. Nature materials. 2018; 17 (4) 301–307. [PubMed: 29483634]
7. Farmahini, Amir H; Friedrich, Daniel; Brandani, Stefano; Sarkisov, Lev. Exploring new sources of efficiency in process-driven materials screening for post-combustion carbon capture. Energy & Environmental Science. 2020; 13 (3) 1018–1037.
8. Mason, Jarad A; Sumida, Kenji; Herm, Zoey R; Krishna, Rajamani; Long, Jeffrey R. Evaluating metal-organic frameworks for post-combustion carbon dioxide capture via temperature swing adsorption. Energy & Environmental Science. 2011; 4 (8) 3030–3040.
9. Nandy, Aditya; Duan, Chenru; Kulik, Heather J. Using machine learning and data mining to leverage community knowledge for the engineering of stable metal–organic frameworks. Journal of the American Chemical Society. 2021; 143 (42) 17535–17547. [PubMed: 34643374]
10. Subraveti, Sai Gokul; Roussanaly, Simon; Anantharaman, Rahul; Riboldi, Luca; Rajendran, Arvind. How much can novel solid sorbents reduce the cost of postcombustion co2 capture? a techno-economic investigation on the cost limits of pressure–vacuum swing adsorption. Applied Energy. 2022; 306 117955
11. Huck, Johanna M; Lin, Li-Chiang; Berger, Adam H; Shahrak, Mahdi Niknam; Martin, Richard L; Bhown, Abhoyjit S; Haranczyk, Maciej; Reuter, Karsten; Smit, Berend. Evaluating different classes of porous materials for carbon capture. Energy & Environmental Science. 2014; 7 (12) 4132–4146.
12. Farmahini, Amir H; Krishnamurthy, Shreenath; Friedrich, Daniel; Brandani, Stefano; Sarkisov, Lev. Performance-based screening of porous materials for carbon capture. Chemical Reviews. 2021; 121 (17) 10666–10741. [PubMed: 34374527]
13. Mu, Bin; Walton, Krista S. Thermal analysis and heat capacity study of metal–organic frameworks. The Journal of Physical Chemistry C. 2011; 115 (46) 22748–22754.
14. Kloutse FA, Zacharia R, Cossement D, Chahine R. Specific heat capacities of mof-5, cu-btc, fe-btc, mof-177 and mil-53 (al) over wide temperature ranges: Measurements and application of empirical group contribution method. Microporous and Mesoporous Materials. 2015; 217: 1–5.
15. Kapil, Venkat; Wieme, Jelle; Vandenbrande, Steven; Lamaire, Aran; Van Speybroeck, Veronique; Ceriotti, Michele. Modeling the structural and thermal properties of loaded metal-organic frameworks. an interplay of quantum and anharmonic fluctuations. Journal of chemical theory and computation. 2019; 15 (5) 3237–3249. [PubMed: 31002500]
16. Wieme, Jelle; Vandenbrande, Steven; Lamaire, Aran; Kapil, Venkat; Vanduyfhuys, Louis; Van Speybroeck, Veronique. Thermal engineering of metal–organic frameworks for adsorption

Europe PMC Funders Author Manuscripts

applications: A molecular simulation perspective. ACS applied materials & interfaces. 2019; 11 (42) 38697–38707. [PubMed: 31556593]

17. Caskey, Stephen R; Wong-Foy, Antek G; Matzger, Adam J. Dramatic tuning of carbon dioxide uptake via metal substitution in a coordination polymer with cylindrical pores. Journal of the American Chemical Society. 2008; 130 (33) 10870–10871. [PubMed: 18661979]
18. Queen, Wendy L; Hudson, Matthew R; Bloch, Eric D; Mason, Jarad A; Gonzalez, Miguel I; Lee, Jason S; Gygi, David; Howe, Joshua D; Lee, Kyuho; Darwish, Tamim A; , et al. Comprehensive study of carbon dioxide adsorption in the metal–organic frameworks m 2 (dobdc)(m= mg, mn, fe, co, ni, cu, zn). Chemical Science. 2014; 5 (12) 4569–4581.
19. Boyd, Peter G; Moosavi, Seyed Mohamad; Witman, Matthew; Smit, Berend. Forcefield prediction of materials properties in metal-organic frameworks. The journal of physical chemistry letters. 2017; 8 (2) 357–363. [PubMed: 28008758]
20. Moghadam, Peyman Z; Rogge, Sven MJ; Li, Aurelia; Chow, Chun-Man; Wieme, Jelle; Moharrami, Noushin; Aragones-Anglada, Marta; Conduit, Gareth; Gomez-Gualdron, Diego A; Van Speybroeck, Veronique; , et al. Structure-mechanical stability relations of metal-organic frameworks via machine learning. Matter. 2019; 1 (1) 219–234.
21. Moosavi, Seyed Mohamad; Boyd, Peter G; Sarkisov, Lev; Smit, Berend. Improving the mechanical stability of metal–organic frameworks using chemical caryatids. ACS central science. 2018; 4 (7) 832–839. [PubMed: 30062111]
22. Jablonka, Kevin Maik; Ongari, Daniele; Moosavi, Seyed Mohamad; Smit, Berend. Bigdata science in porous materials: materials genomics and machine learning. Chemical reviews. 2020; 120 (16) 8066–8129. [PubMed: 32520531]
23. Rahimi, Mohammad; Moosavi, Seyed Mohamad; Smit, Berend; Alan Hatton, T. Toward smart carbon capture with machine learning. Cell Reports Physical Science. 2021; 2 (4) 100396
24. Tawfik, Sherif Abdulkader; Isayev, Olexandr; Spencer, Michelle JS; Winkler, David A. Predicting thermal properties of crystals using machine learning. Advanced Theory and Simulations. 2020; 3 (2) 1900208
25. Legrain, Fleur; Carrete, Jesús; van Roekeghem, Ambroise; Curtarolo, Stefano; Mingo, Natalio. How chemical composition alone can predict vibrational free energies and entropies of solids. Chemistry of Materials. 2017; 29 (15) 6220–6227.
26. Ducamp, Maxime; Coudert, François-Xavier. Prediction of thermal properties of zeolites through machine learning. The Journal of Physical Chemistry C. 2022; 126 (3) 1651–1660.
27. Chen, Zhantao; Andrejevic, Nina; Smidt, Tess; Ding, Zhiwei; Xu, Qian; Chi, Yen-Ting; Nguyen, Quynh T; Alatas, Ahmet; Kong, Jing; Li, Mingda. Direct prediction of phonon density of states with euclidean neural networks. Advanced Science. 2021. 2004214 [PubMed: 34165895]
28. Kauwe, Steven K; Graser, Jake; Vazquez, Antonio; Sparks, Taylor D. Machine learning prediction of heat capacity for solid inorganics. Integrating Materials and Manufacturing Innovation. 2018; 7 (2) 43–51.
29. Petretto, Guido; Dwaraknath, Shyam; Miranda, Henrique PC; Winston, Donald; Giantomassi, Matteo; Van Setten, Michiel J; Gonze, Xavier; Persson, Kristin A; Hautier, Geoffroy; Rignanese, Gian-Marco. High-throughput density-functional perturbation theory phonons for inorganic materials. Scientific data. 2018; 5 (1) 1–12. [PubMed: 30482902]
30. Ward, Logan; Liu, Ruoqian; Krishna, Amar; Hegde, Vinay I; Agrawal, Ankit; Choudhary, Alok; Wolverton, Chris. Including crystal structure attributes in machine learning models of formation energies via voronoi tessellations. Physical Review B. 2017; 96 (2) 024104
31. Jablonka, Kevin Maik; Ongari, Daniele; Moosavi, Seyed Mohamad; Smit, Berend. Using collective knowledge to assign oxidation states of metal cations in metal–organic frameworks. Nature Chemistry. 2021; 13 (8) 771–777.
32. Behler, Jörg; Parrinello, Michele. Generalized neural-network representation of highdimensional potential-energy surfaces. Physical review letters. 2007; 98 (14) 146401 [PubMed: 17501293]
33. Botu, Venkatesh; Batra, Rohit; Chapman, James; Ramprasad, Rampi. Machine learning force fields: construction, validation, and outlook. The Journal of Physical Chemistry C. 2017; 121 (1) 511–522.

34. Chung, Yongchul G; Haldoupis, Emmanuel; Bucior, Benjamin J; Haranczyk, Maciej; Lee, Seulchan; Zhang, Hongda; Vogiatzis, Konstantinos D; Milisavljevic, Marija; Ling, Sanliang; Camp, Jeffrey S; , et al. Advances, updates, and analytics for the computation-ready, experimental metal-organic framework database: Core mof 2019. Journal of Chemical & Engineering Data. 2019; 64 (12) 5985–5998.
35. Ongari, Daniele; Yakutovich, Aliaksandr V; Talirz, Leopold; Smit, Berend. Building a consistent and reproducible database for adsorption evaluation in covalent-organic frameworks. ACS central science. 2019; 5 (10) 1663–1675. [PubMed: 31681834]
36. Baerlocher, Ch; McCusker, Ch. Iza: Database of zeolite structures. http://www.iza-structure.org/databases/
37. Ming, Yang; Purewal, Justin; Sudik, Andrea; Xu, Chunchuan; Yang, Jun; Veenstra, Mike; Rhodes, Kevin; Soltis, Richard; Warner, James; Gaab, Manuela; , et al. Thermophysical properties of mof-5 powders. Microporous and mesoporous materials. 2014; 185: 235–244.
38. Rosen, Andrew S; Iyer, Shaelyn M; Ray, Debmalya; Yao, Zhenpeng; Aspuru-Guzik, Alan; Gagliardi, Laura; Notestein, Justin M; Snurr, Randall Q. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. Matter. 2021; 4 (5) 1578–1597.
39. Perdew, John P; Burke, Kieron; Ernzerhof, Matthias. Generalized gradient approximation made simple. Physical review letters. 1996; 77 (18) 3865. [PubMed: 10062328]
40. Grimme, Stefan; Ehrlich, Stephan; Goerigk, Lars. Effect of the damping function in dispersion corrected density functional theory. Journal of computational chemistry. 2011; 32 (7) 1456–1465. [PubMed: 21370243]
41. Goedecker, Stefan; Teter, Michael; Hutter, Jürg. Separable dual-space gaussian pseudopotentials. Physical Review B. 1996; 54 (3) 1703.
42. Togo, Atsushi; Tanaka, Isao. First principles phonon calculations in materials science. Scripta Materialia. 2015; 108: 1–5.
43. Hutter, Jürg; Iannuzzi, Marcella; Schiffmann, Florian; VandeVondele, Joost. cp2k: atomistic simulations of condensed matter systems. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2014; 4 (1) 15–25.
44. Plimpton, Steve. Fast parallel algorithms for short-range molecular dynamics. Journal of computational physics. 1995; 117 (1) 1–19.
45. Carreras, Abel. phonoLAMMPS: A python interface for LAMMPS phonon calculations using phonopy, July 2020.
46. Ward, Logan; Dunn, Alexander; Faghaninia, Alireza; Zimmermann, Nils ER; Bajaj, Saurabh; Wang, Qi; Montoya, Joseph; Chen, Jiming; Bystrom, Kyle; Dylla, Maxwell; , et al. Matminer: An open source toolkit for materials data mining. Computational Materials Science. 2018; 152: 60–69.
47. Ong, Shyue Ping; Richards, William Davidson; Jain, Anubhav; Hautier, Geoffroy; Kocher, Michael; Cholia, Shreyas; Gunter, Dan; Chevrier, Vincent L; Persson, Kristin A; Ceder, Gerbrand. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. Computational Materials Science. 2013; 68: 314–319.
48. Chen, Tianqi; Guestrin, Carlos. Xgboost: A scalable tree boosting system; Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016. 785–794.
49. Pedregosa, Fabian; Varoquaux, Gaël; Gramfort, Alexandre; Michel, Vincent; Thirion, Bertrand; Grisel, Olivier; Blondel, Mathieu; Prettenhofer, Peter; Weiss, Ron; Dubourg, Vincent; Vanderplas, Jake; , et al. Scikit-learn: Machine learning in python. Journal of machine learning research. 2011; Oct. 12: 2825–2830.
50. Ajenifuja, Abdulmalik; Joss, Lisa; Jobson, Megan. A new equilibrium shortcut temperature swing adsorption model for fast adsorbent screening. Industrial & Engineering Chemistry Research. 2020; 59 (8) 3485–3497.
51. Talirz, Leopold; Kumbhar, Snehal; Passaro, Elsa; Yakutovich, Aliaksandr V; Granata, Valeria; Gargiulo, Fernando; Borelli, Marco; Uhrin, Martin; Huber, Sebastiaan P; Zoupanos, Spyros; , et al. Materials cloud, a platform for open computational science. Scientific data. 2020; 7 (1) 1–12. [PubMed: 31896794]

Europe PMC Funders Author Manuscripts

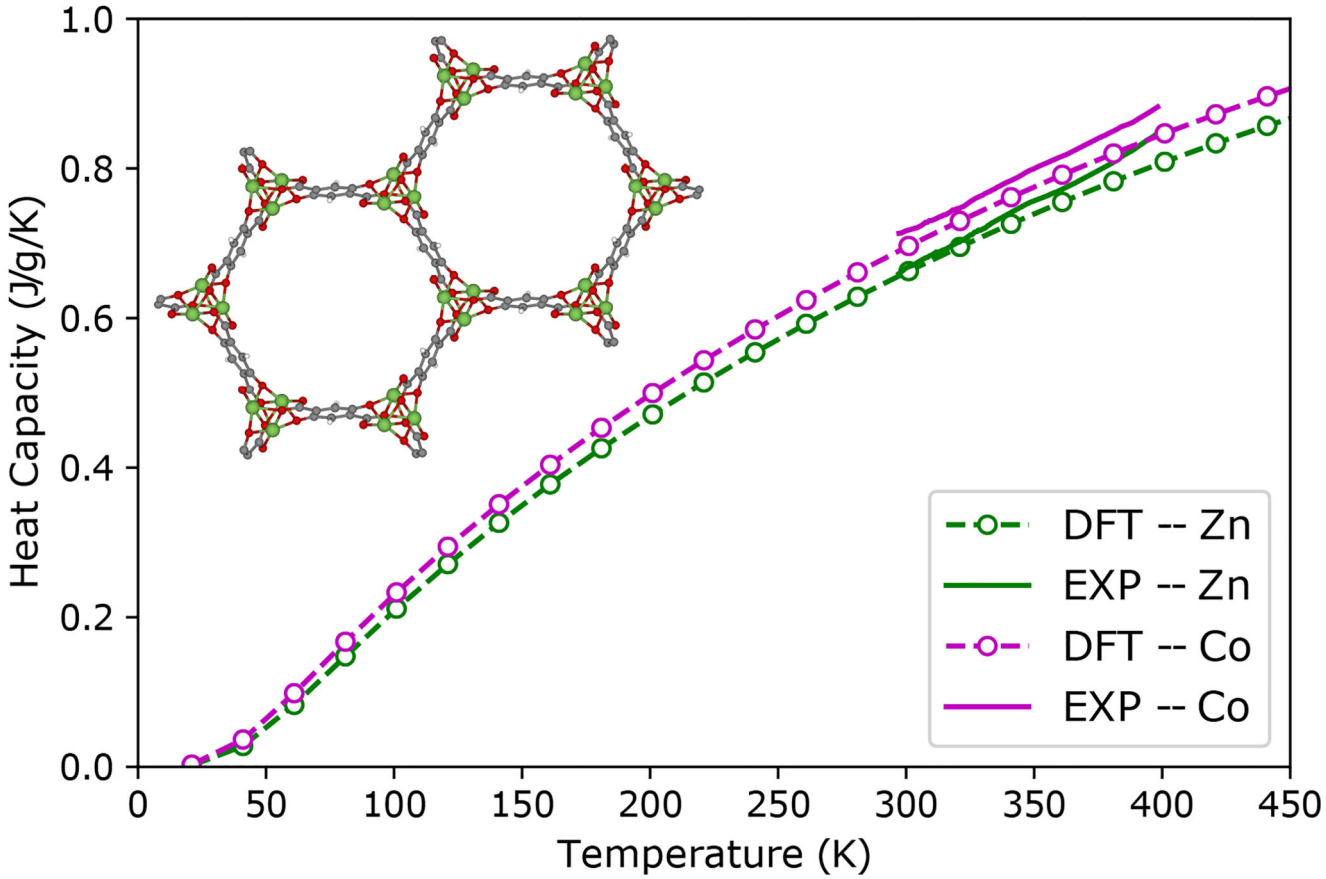

Figure 1. Temperature dependence of the heat capacity of MOF-74.

The DFT predictions of the heat capacity of Zn-MOF-74 and Co-MOF-74 are compared with the experimental (EXP) measurements obtained in this work. Ball and stick model of the hexagonal porous structure of MOF-74 is shown in the inset. Green, red, grey, and white spheres represent the metal (Zn or Co), oxygen, carbon, and hydrogen respectively.

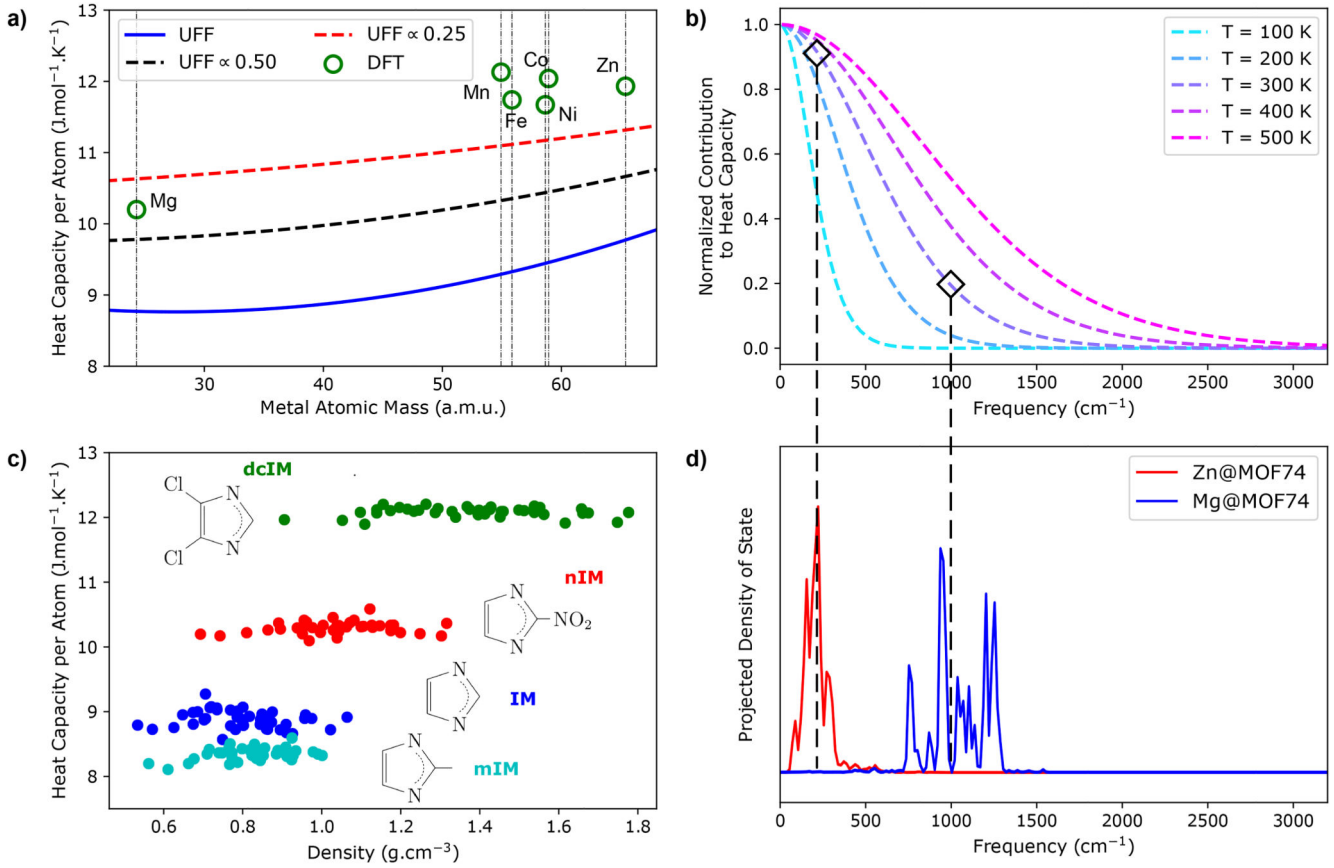

Figure 2. Structure–heat capacity relationships.

a) The heat capacity of MOF-74 with different metals, computed using DFT (open circles) and UFF (for simplicity we show fitted lines to the data points – see Figure S.3 in SI for the data points). The dashed lines show the results when we scaled the metal-linker force constants with a factor of $\propto 0.25$ or $\propto 0.5$. b and d) The differences in the vibrational frequencies of metal centers in Mg and Zn MOF-74, and their contributions to the heat capacity at different temperatures. c) The heat capacity of ZIF structures, color-coded with their linker type that are shown as inset (IM = imidazolate, mIM = 2-methylimidazolate, dcIM = dichloroimidazolate, and nIM = 2-nitroimidazolate). While from an engineering point of view the heat capacity is often reported per gram of materials, it is easier to explain structure-heat capacity relationships when we look at the heat capacity per atom.

Europe PMC Funders Author Manuscripts

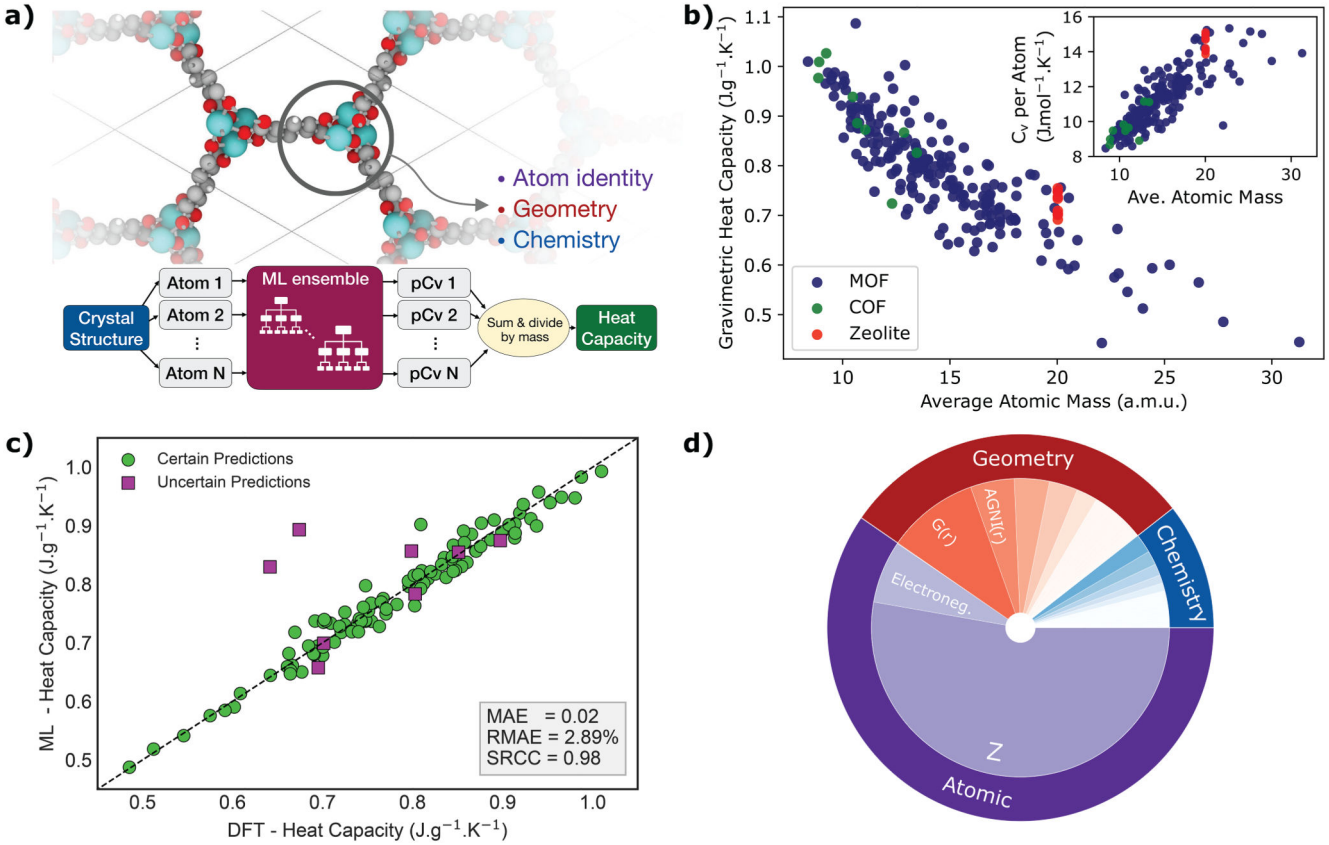

Figure 3. DFT and machine learning predictions of the heat capacity.

(a) The gravimetric and molar heat capacity of the structures at room temperature are shown with respect to the average atomic mass. The inset gives the molar heat capacity per atom. (b) The sketch of a local environment that is featurised in terms of atomic identity, local geometry and chemistry. For each atom within the unit cell of a crystal structure, our machine learning model predicts the contribution to the heat capacity (pCv), which is summed and normalized by the total mass of the crystal. (c) The correlation plot of the predictions of machine learning models compared to the DFT reference results. The statistics of the predictions for which the machine learning model was certain are shown in the inset, where MAE is mean absolute error, RMAE relative mean absolute error, and SRCC the Spearman rank correlation coefficient. (d) The pie chart shows the importance of different descriptors in the prediction of the heat capacity estimated using SHapley Additive exPlanations (SHAP) values (see method section for details). The descriptors are categorized into three groups of atomic identity, geometry, and chemistry (see section 2.1 in SI for details).

Europe PMC Funders Author Manuscripts

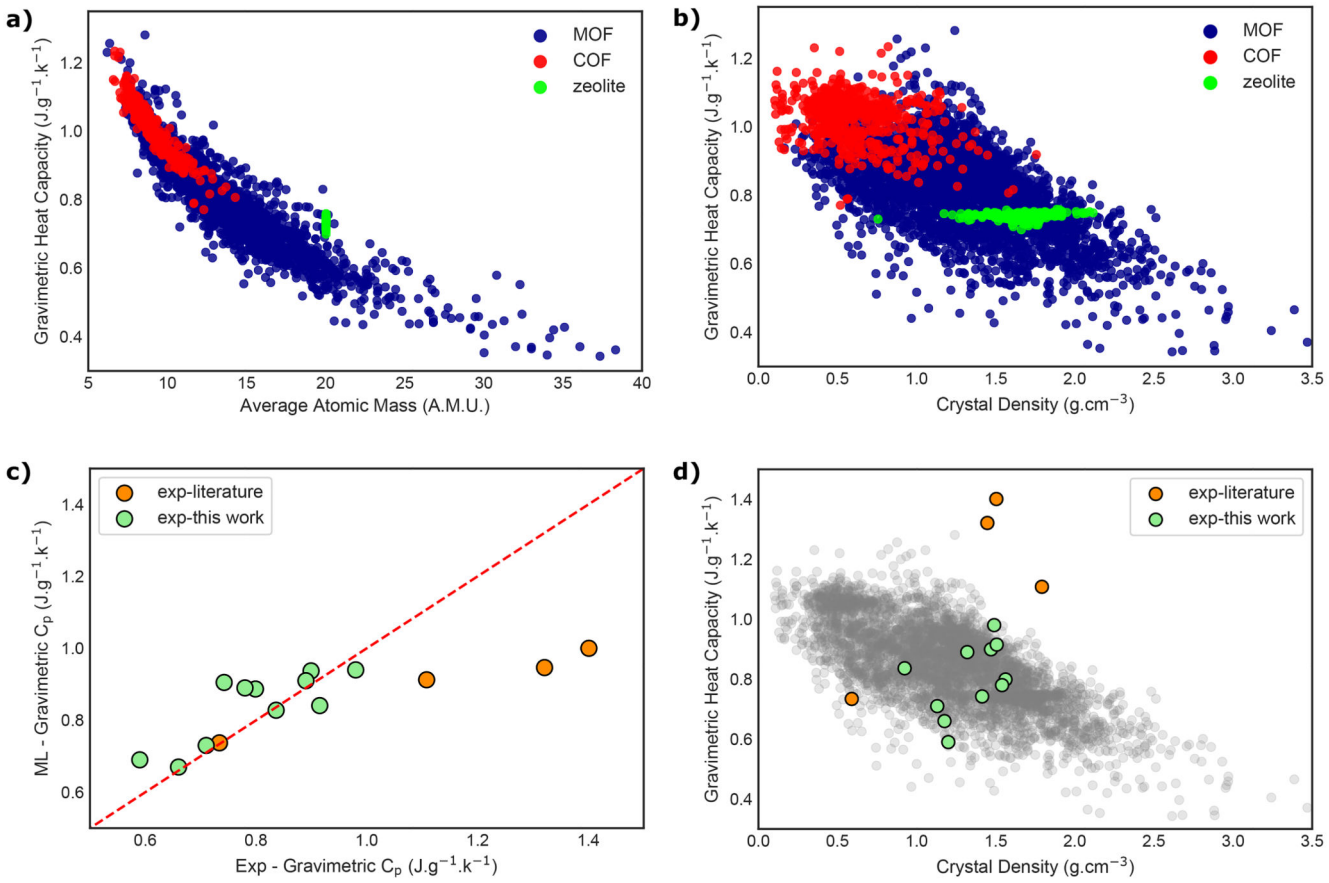

Figure 4. Mapping the heat capacity of nanoporous materials.

(a-b) Predictions of the heat capacity at 300 K for the experimental structures in the CoRE-MOF,[34] CURATED-COFs,[35] and IZA zeolites.[36] (c-d) Comparison of the machine learning predictions of the heat capacity at 300K with the experimental values reported in the literature[37] or measured in this work. The gray dots in (d) show the observed range of the heat capacity of materials in (b).

Europe PMC Funders Author Manuscripts

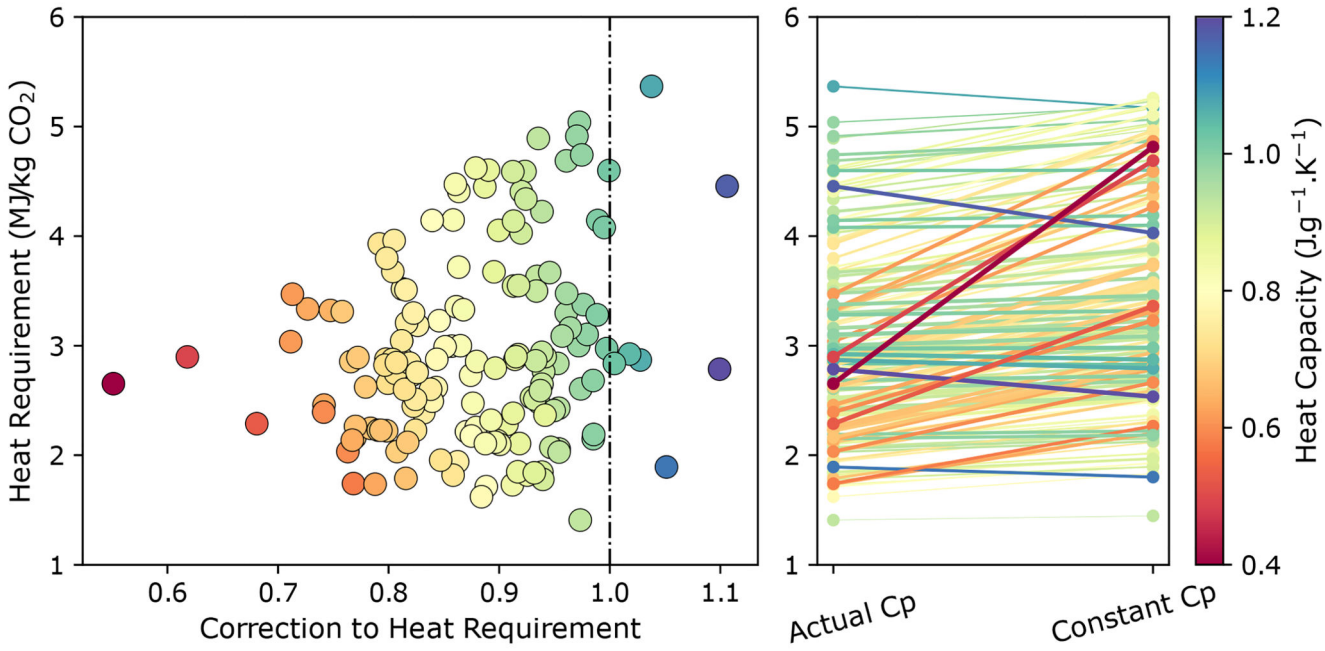

Figure 5. The role of heat capacity on the performance and ranking of porous materials in carbon capture.

In these figures, the heat requirement gives the energy requirement per kilogram of CO_2 for regenerating materials in a temperature swing adsorption process to capture CO_2 from a coal-fired power plant. The color coding indicates the value of the heat capacity. (Left) The vertical axis shows the heat requirement when the actual values of heat capacity of the materials are used. The horizontal axes give the correction factor, which is the ratio of this heat requirement to when the heat requirement is computed assuming all materials have the same heat capacity of 0.985 $J \cdot g^{-1} \cdot K^{-1}$. (Right) The change in the ranking of the materials: The left side gives the new ranking with the heat capacity from our machine learning model, and the right side the old ranking assuming a constant heat capacity for all materials. The thickness of the lines connecting these two points is a measure of the change in the ranking.

Europe PMC Funders Author Manuscripts