



Melting temperature prediction using a graph neural network model: From ancient minerals to new materials

Qi-Jun Hong^{a,1}, Sergey V. Ushakov^b, Axel van de Walle^c, and Alexandra Navrotsky^{b,1}

Contributed by Alexandra Navrotsky; received June 5, 2022; accepted August 4, 2022; reviewed by Benjamin Burton and Alexandra Khvan

The melting point is a fundamental property that is time-consuming to measure or compute, thus hindering high-throughput analyses of melting relations and phase diagrams over large sets of candidate compounds. To address this, we build a machine learning model, trained on a database of ~10,000 compounds, that can predict the melting temperature in a fraction of a second. The model, made publicly available online, features graph neural network and residual neural network architectures. We demonstrate the model's usefulness in diverse applications. For the purpose of materials design and discovery, we show that it can quickly discover novel multicomponent materials with high melting points. These predictions are confirmed by density functional theory calculations and experimentally validated. In an application to planetary science and geology, we employ the model to analyze the melting temperatures of ~4,800 minerals to uncover correlations relevant to the study of mineral evolution.

machine learning | melting temperature | mineral evolution

Melting points play an important role in a wide variety of disciplines. High-performance refractory materials (1–5) have applications ranging from gas turbines to heat shields for hypersonic vehicles. In this context, high melting points correlate with desirable mechanical properties (e.g., high-temperature materials strength as well as good ablation and creep resistance). In geology and planetary science, knowledge of the melting points of minerals provides insight into their formation and evolution in addition to shedding light into the structure of exoplanets.

In these examples, the melting points tend to be high, which considerably complicates their experimental measurement (due to containment and calibration issues). For minerals, these issues are compounded but the fact that many new species are identified from small grains and are thus accessible in limited amounts. Complex phase equilibria and incongruent melting lead to further complications. As a result, the melting temperature is known for less than 10% of the more than 200,000 inorganic substances with known crystal structures.

In light of these limitations, it would be natural to turn to computational methods. Unfortunately, the calculation and prediction of melting temperatures is also an expensive and complex procedure because it involves sampling a large number of configurations.

Numerous efficient methods have been devised to capture melting temperatures from computations (6). Using empirical potentials is relatively inexpensive, but it depends on availability and reliability of such potentials. It is both complicated and time-consuming to build a new classical interatomic potential for every new material, not to mention the issue of reliability regarding accuracy. Density functional theory (DFT) calculations are clearly better in terms of generalizability and reliability. However, they remain notoriously expensive, despite increasing power and capability of our computers. The large-size coexistence method (7, 8), which is generally considered the gold standard and widely utilized as a benchmark, typically requires a system size too large for DFT simulations, rendering this approach prohibitively expensive in practice. The single-phase small-size “Z method” (9), which heats a solid until it melts, seeks to address this, but suffers from well-documented practical and conceptual problems (10). Alternatively, one can compute melting temperatures via the free energy method (7, 11), which locates the intersection of the free energy curves of the solid and the liquid. This approach requires highly accurate free energy calculation of the liquid phase, because the two curves cross at a very shallow angle and thus a small free energy shift will result in a large error in melting temperature. Unfortunately, all methods for liquid state free energy computation, such as thermodynamic integration (12), the particle insertion method (13, 14), and the two-phase thermodynamics method (15), are expensive and challenging. In earlier work, we proposed the accurate and relatively more efficient small-size coexistence method (6, 16), and developed the solid and liquid in ultra-small coexistence with hovering interfaces

Significance

High temperature materials properties, such as melting temperature, are generally challenging to measure or compute rapidly. As a result, melting points are known for less than 10% of the ~200,000 known inorganic compounds. Here, we employ machine learning methods to fill this gap by building a rapid and accurate mapping from chemical formula to melting temperature. The model, which we have made publicly available, will facilitate large-scale data analysis involving melting temperature in a wide range of areas, including the discovery of refractory materials, the design of novel extractive metallurgy processes, the modeling of mineral formation and evolution over geological time and the prediction of exoplanet structure.

Author affiliations: ^aSchool for Engineering of Transport, Energy and Matter, Arizona State University, Tempe, AZ 85287; ^bCenter for Materials of the University, School of Molecular Sciences, Arizona State University, Tempe, AZ 85287; and ^cSchool of Engineering, Brown University, Providence, RI 02912

Author contributions: Q.-J.H. and A.N. designed research; Q.-J.H. and S.V.U. performed research; Q.-J.H., S.V.U., A.v.d.W., and A.N. analyzed data; and Q.-J.H., S.V.U., A.v.d.W., and A.N. wrote the paper.

Reviewers: B.B., NIST; and A.V.K., Nacional'nyj issledovatel'skij tehnologiceskij universitet MISiS.

The authors declare no competing interest.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: qhong7@asu.edu or alexnav@asu.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2209630119/-/DCSupplemental>.

Published August 31, 2022.

(SLUSCHI) package (17) to automate the computation process. We have utilized this method to calculate melting temperatures of hundreds of materials, at the cost of several days of computations per compound. Despite such progress, these efforts still fall short of providing a comprehensive view of the melting point landscape.

To overcome these limitations, we turn to machine learning (ML) methods, which are increasingly used for the prediction of materials properties and missing thermodynamic data (18, 19). We build a ML model to predict melting temperature, with an ultimate goal of integrating DFT and ML, which complement each other in terms of speed and accuracy. The ML model allows us to rapidly estimate melting temperatures, at a speed on the order of milliseconds per material, while the DFT calculation provides robustness and accuracy at a much higher cost on the order of several days of computations per material.

To provide training data for our ML procedure, we first build a melting temperature database via web crawling. Melting temperatures are collected and included in our database mostly from Ref (20). DFT melting temperature calculations are included in the database as well. Our current melting temperature database contains 9,375 materials, out of which 982 compounds are high-melting-temperature materials with melting points above 2,000 K. The database consists of chemical compositions (i.e., elements and concentrations) or equivalently chemical formula, of the materials, and their corresponding melting temperatures.

To illustrate the usefulness of our software tool, two application examples are described: (i) the prediction of melting temperatures for 4828 mineral species, and (ii) the prediction of compositions with melting temperatures above 3,500 K.

Architecture of Neural Network Model for Melting Temperature Prediction

The majority of entries for melting temperatures used for machine learning were collected by parsing data for ~26,000 single-phase compounds from a ten-volume compilation of thermodynamic constants of substances (20). The values are based on experimental data from ~51,500 publications before 1982. Only congruent melting temperatures were included in the current version of the database used for machine learning. The data were complemented by results obtained from ab initio molecular dynamic calculations within the SLUSCHI framework (17).

In our machine learning model, we seek to build a mapping from chemical formula to melting temperature. The minimal requirement of input, only chemical formula here, should facilitate the model's extensive application in the future: no additional materials properties are required as input, and thus neither computational nor experiment data are needed. In our view, relying on input feature would reduce the broad applicability of the method, especially in high-throughput exploration or screening applications, where chemistry is typically the only a priori known input. The machine learning model combines the Graph Neural Network (GNN) (21) and residual neural network (ResNet) (22) architectures within the Tensorflow (23) framework (Fig. 1). The GNN architecture is designed to impose permutation invariance (e.g., ZrO_2 and O_2Zr are the same material), which drastically reduces model complexity and thus improves efficiency. The ResNet architecture avoids the problem of vanishing gradient by skipping connections, which also effectively simplifies the network. When a material (i.e., its elements and composition) is fed to the neural network, each element is first converted to 14 features, such as atomic radius, atomic mass, electronegativity, core and valence electrons, ionization energy, electron affinity, density,

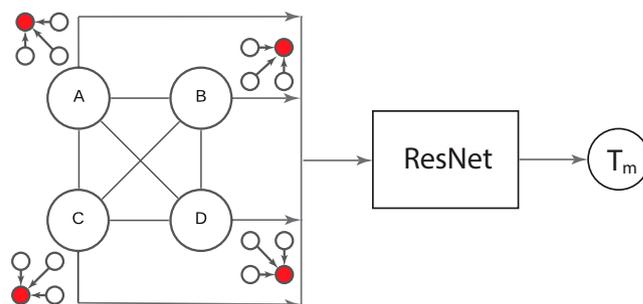


Fig. 1. Architecture of the GNN model for ML melting temperature prediction. Each circle represents an element and its composition in the material. Up to four elements are connected in this graph (denoted as A, B, C, and D). First, each element and composition are converted to 14 features, which are then encoded and fed to the ResNet input layer. The circles communicate with each other in order to account for higher order contributions. For example, each circle (element and composition) pulls information from other circles via the GNN. The outputs are then sent to the ResNet. The latter consists of four fully connected layers with skipping connections and leads to the regression analysis for melting temperature prediction.

and position in the periodic table. These features are encoded and passed to the next layer, which we expect to include the key determinants of the individual atomic contributions to the melting temperature. In addition, elemental features interact with each other via the GNN connections, thus leading to contributions from the binary, ternary, and higher-order combination of elements. These encoded contributions are passed to the next layer as well. This layer, consisting of unary, binary, and ternary interactions of the elements and compositions of the material of interest, is fed into a four-layer ResNet, which leads to the regression and the estimation of melting temperature. Currently the number of elements is limited to four, but this constraint can be removed if a larger combination of elements is needed. For instance, the limit is relaxed to five in the study of minerals later in this paper. More elements increase the complexity of the model and thus the risk of overfitting. The GNN architecture undergoes two iterations of communication among elements, as we find more rounds do not significantly improve performance of the model. Dropout layers are heavily employed in the architecture to avoid overfitting.

The 9,375 materials are randomly assigned to training and testing sets, with 8,635 materials in the training set, and 740 materials for testing. The training process takes 2,000–4,000 epochs of optimization. As shown in Fig. 2, the root-mean-square errors (rmse) of melting temperature are 110 and 160 K for the training and testing sets, respectively. These surprisingly small errors [DFT error is typically 100 K (6) due to imperfect density functionals] represent the method's accuracy over the ranges of composition that are included in the database. While the testing set is a holdout dataset and thus it provides an unbiased evaluation of the final model, we note that accuracy could degrade for prediction request in relatively poorly sampled regions of composition space, which would demand considerable extrapolation. The errors for materials of different types are shown in *SI Appendix, Fig. S1*.

We benchmark our GNN model with XGBoost, one of the most popular gradient boosting methods. Our model score (R₂: 0.933, rmse: 160K) outperforms that of XGBoost (R₂: 0.919, rmse: 183K). This observation is consistent with our expectation and understanding of the two methods. With only chemical formula and elemental features as input, data features are very limited in this work, and thus it favors the neural networks method, which is more capable of combining and generating features by itself. The moderately large size of our dataset also works well with the neural networks method.

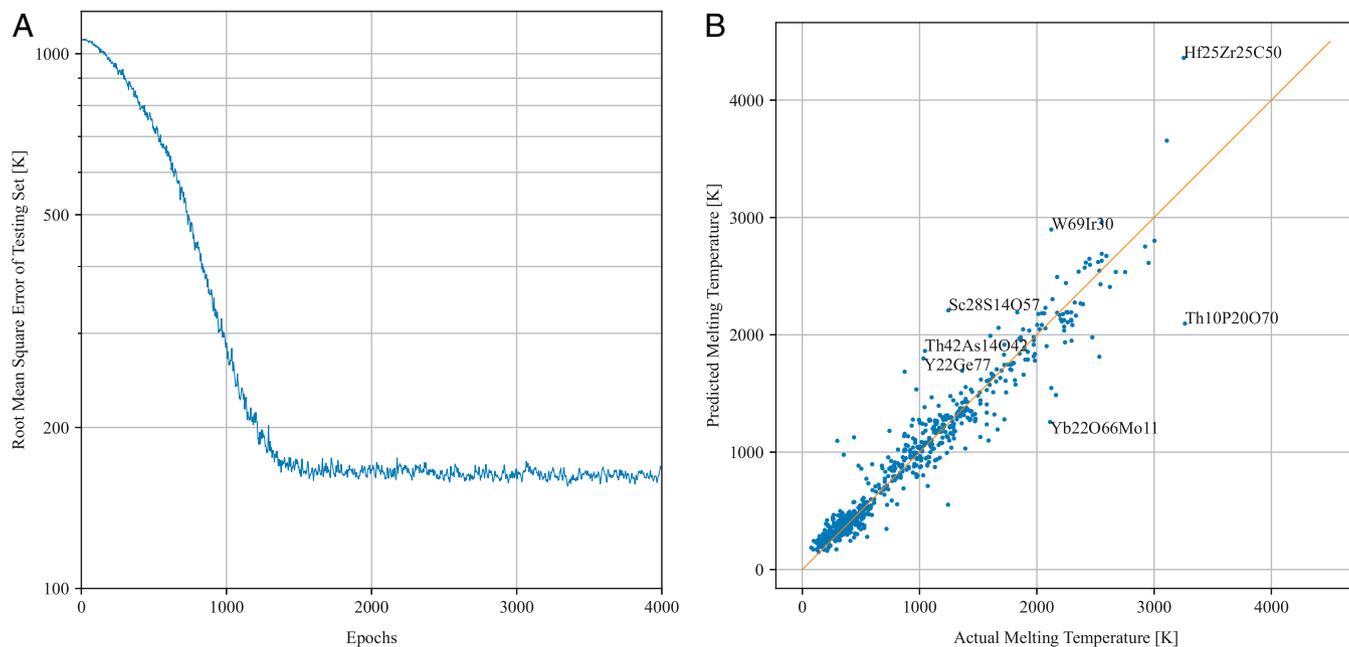


Fig. 2. (A) rmse during model training. The training process was completed after 2,000 epochs. (B) Predicted vs. actual melting temperatures in the testing dataset. The rmse are 110 and 160 K for the training and testing sets, respectively. Compounds with large errors are labeled for further investigation.

The model is currently hosted at the ASU Research Computing Facilities and available through a web page (24) and Application Programming Interface (API) interface. To use the model, a user needs to visit the webpage and input the chemical compositions of the material of interest. The model will respond with a predicted melting temperature in seconds, as well as the actual melting temperatures of the nearest neighbors (i.e., the most similar materials) in the database. Thus this model serves as not only a predictive model, but a handbook of melting temperature as well. A user may also run batch calculations via command line with much shorter latency, by sending an HTTP POST request to the API server and providing JSON data (elements and compositions of multiple materials) in the body of the POST message. Detailed instruction is available at the webpage.

Mineral Melting Temperatures: Structure and Deep Time Correlations

There are currently more than 5,700 approved minerals (25), all naturally occurring compounds which include phases identified exclusively in meteorites and formed during geological processes on Earth. While the composition and structure of all minerals are known (as required for approval of new mineral species), thermodynamic properties, such as melting or decomposition temperatures, are only available for a small fraction of them. The field of mineral evolution, pioneered by Hazen et al. (26), studies occurrence of new mineral species and their increasing chemical and structural complexity as a function of geologic time (26, 27). Recently, the oldest known ages were assigned for 4,828 mineral species based on more than 190,000 dated mineral locality occurrences (28, 29). The dataset is constantly updated and made openly available to promote data-driven discovery in mineralogy (30, 31).

We employed our melting temperature database and model (based on 9,375 compounds) to analyze the mineral dataset, as well as a subset of 412 minerals containing rare earth elements (lanthanides, Y and Sc). Approximately 6% of the minerals in the dataset have direct matches in our melting temperature database. For them, experimentally measured values were used

in the analysis. For the rest of the minerals, melting temperature (ML T_m) was predicted based on our ML model. Since the majority of minerals are ternary and higher order compounds, containing structural water and carbonate groups, they are unlikely to melt congruently and in this case ML T_m correlates with their decomposition temperature. We interpret ML T_m as the upper boundary of decomposition temperature, as these materials decompose before melting. The dataset is included in the *SI Appendix*.

Fig. 3A shows the average ML T_m versus oldest known age, grouped with an interval of 250 million years. As expected, the oldest minerals, interstellar and solar nebula condensates predating Earth formation 4.5 billion years ago, are the most refractory, with average and median melting temperatures around 1700 K. The gradual overall decrease in ML T_m of minerals formed during Earth history is interrupted with two anomalies, which are distinctly pronounced in average and medium ML T_m using 250 or 500 Ma binning. The spike at 3.75 Ga correlates to the proposed timing of late heavy bombardment, hypothesized exclusively from dating of lunar samples and currently debated (32, 33). The dip at 1.75 Ga is related to the first known occurrences of a large number of hydrous minerals and correlates with the Huronian glaciation (34), the longest ice age and thought to be the first time Earth was completely covered in ice.

The rise in average ML melting temperature with increase in symmetry from triclinic to cubic structures (Fig. 3B) is consistent with the observations of the predominant stability of high symmetry structures at high temperatures, as suggested by established experimental phase diagrams. Low symmetry minerals with complex composition typically do not melt congruently, but instead decompose, often to phases with higher symmetry. Analysis of melting temperatures of 412 rare earth containing minerals gives a higher average ML T_m compared to the overall mineral dataset (1,296 vs. 1,005 K, respectively). This is expected due to high melting temperatures of rare earth oxides. Consistent with the full dataset, cubic and tetragonal rare earth minerals represent the smallest fraction, and the highest average melting temperatures, but there is not a clear sequence among monoclinic, hexagonal, and orthorhombic minerals (*SI Appendix*, Fig. S2).

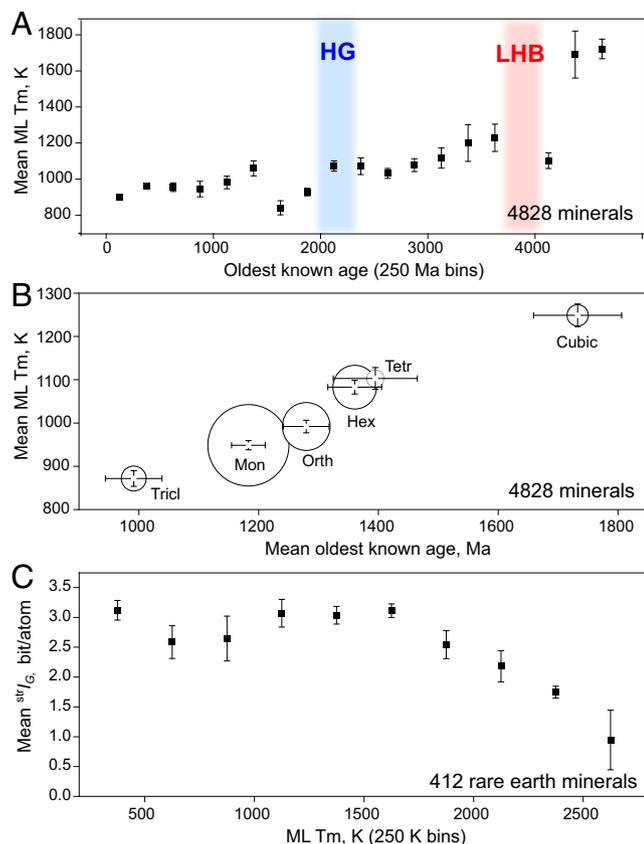


Fig. 3. Machine learned melting temperature (ML Tm) for minerals from RRUFF dataset. (A) Mean ML Tm vs. oldest known age. The highlights correspond to the timing of hypothesized Late Heavy Bombardment (LHB) and Huronian glaciation (HG) events. (B) ML Tm vs. oldest known age for all minerals grouped by crystal systems. The size of the circles scaled with the number of minerals: cubic: 500; tetragonal: 383; hexagonal: 855; orthorhombic: 932; monoclinic: 1600; triclinic: 509. (C) Mean structural complexity index vs. predicted melting temperature for 412 rare earth containing minerals. The error bars correspond to SEs.

The approach based on Shannon's entropy (35) is increasingly used in thermodynamic modeling (36, 37). Krivovichev's index (38) provides a quantitative evaluation of structural complexity in bits of Shannon's information per atom. In addition to symmetry, it accounts for the size of the unit cell and chemical diversity. The higher the bit/atom values, the lower the entropy of the structure (39). For rare earth containing minerals, predicted melting temperature shows a strong negative correlation with the structure complexity index (Fig. 3C). The rare earth minerals with index below 2.5 bits/atom have predicted melting temperatures above 1,500 K.

Discovery of New High Temperature Materials

Based on this moderately accurate but extremely rapid model, we run simulations to showcase its possible applications. Here, we present one potential application in the design and discovery of high-melting-point materials. We run Monte Carlo (MC) simulations to generate a list of ternary compounds, which are predicted by the model as top candidates for high-melting-point materials. Since the model takes inputs in the form of elements and compositions, there are only five variables, three elements, and two compositions (the sum of mole fractions must equal 1), to describe a ternary compound. Any combination of any element and composition is allowed in our simulation (i.e., any element in the periodic table). The Metropolis algorithm and

simulated annealing technique are employed to maximize melting temperature. The simulation explores the surface of melting temperature and searches for the global maximum of the surface, which is defined by the elements and composition and estimated by the GNN model. The initial MC temperature is set to 100 K, sufficiently high to allow the exploration to escape local minima. The temperature is linearly decreased to 0 K over 10,000 MC steps. We allow changes in both elements and compositions. After each MC run, we obtain a candidate, which presumably is the global maximum of the melting temperature surface as long as the MC trajectory is sufficiently long. In order to generate a list of top high-melting-temperature materials, we run a series of MC simulations in sequence, in which we exclude materials already found from the search in the next iterations, i.e., in the $(n + 1)$ th iteration, the top n material systems already discovered in the previous n iterations are excluded from the search, in order to encourage the exploration of new materials. After this series of MC simulations, we generate a list of top candidates, ranked by their melting temperatures.

As illustrated in *SI Appendix, Fig. S3*, the top 20 candidates are overwhelmingly carbides and nitrides (i.e., two metallic elements plus C or N), with the only exception being the Hf-C-N system, exactly the carbonitride we previously predicted as the material with the world's highest melting temperature, based on DFT molecular dynamics (MD) calculations using the SLUSCHI package (23). This discovery was also later confirmed independently from experiment (24–26). We note that the DFT melting points of the Hf-C-N system were deliberately excluded from the database and the GNN model, in order to challenge the model and examine its predictive capability. The outcome, that the model accurately discovers the Hf-C-N system, is promising and this approach arguably outperforms our human intuition: when we found the Hf-C-N system from DFT, we searched among possible combinations of five elements, Hf, Ta, B, C, and N, and discovered Hf-C-N as the best compound. In contrast, the model correctly predicts Hf-C-N as the most promising candidates, which could have saved us significant time and effort in DFT MD simulations. This capability suggests the model's potential application in materials design and discovery. The melting temperature predicted by the GNN model is at least 500 K lower than that from DFT, which is not surprising since there is no explicit information of the new material system in the dataset and this melting temperature must be pieced together from other similar material systems.

In the next step, we will include DFT melting temperature of the Hf-C-N system, retrain and improve the ML model, and repeat the MC simulations to search for even higher melting temperatures. As summarized in Fig. 4, the new list of top candidates now suggests possibilities with other nonmetals in addition to C and N. We plan to calculate DFT melting temperatures for these materials, which will not only corroborate the discovery of high melting temperature, if indeed favorable, but also further improve our melting temperature database when we include the new DFT results. This work is beyond the scope of this paper but it will be carried out in the near future.

Summary and Future Directions

We built a melting temperature database and an ML GNN model to predict melting temperature from chemical formula. We demonstrated the utility of prediction of melting temperature by providing correlations for mineral evolution and directions for further experimental and computational search for new high temperature materials. The model is openly available through web interface and will be updated as new data for neural network

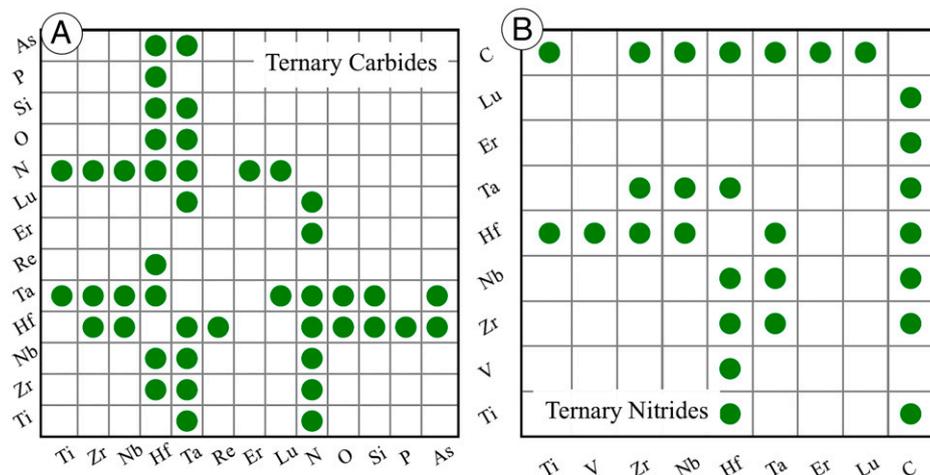


Fig. 4. Constituent elements in ternary carbides (A) and nitrides (B) with predicted melting temperatures above 3,500 K. Each dot represents one ternary compound, M1-M2-C, or M1-M2-N with M1 and M2 being the corresponding elements on the axes.

training will be collected. We have built a next version of the model, which further improves the performance and deployed it online at our webpage. The model is an ensemble model of 30 GNN models based on reshuffling training and testing datasets using bootstrap, which further reduces overfitting.

Data Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This research was funded by the US National Science Foundation under Collaborative Research Awards DMR-2015852, 2209026 (Arizona State University), and DMR-1835939, 2209027 (Brown University) with use of Research Computing at Arizona State University and the Extreme Science and Engineering Discovery Environment (XSEDE), supported by the National Science Foundation (ACI-1548562). S.V.U. would like to thank Sergey Krivovichev for discussions. Q.-J.H. would like to thank Yang Jiao for discussions.

- N. P. Padture, M. Gell, E. H. Jordan, Thermal barrier coatings for gas-turbine engine applications. *Science* **296**, 280–284 (2002).
- E. Wuchina, E. Opila, M. Opeka, W. Fahrenholtz, I. Talmy, UHTCs: Ultra-high temperature ceramic materials for extreme environment applications. *Electrochem. Soc. Interface* **16**, 30–36 (2007).
- J. H. Perepezko, Materials science. The hotter the engine, the better. *Science* **326**, 1068–1069 (2009).
- K. Lu, Materials science. The future of metals. *Science* **328**, 319–320 (2010).
- G. Liu *et al.*, Nanostructured high-strength molybdenum alloys with unprecedented tensile ductility. *Nat. Mater.* **12**, 344–350 (2013).
- Q.-J. Hong, "Methods for melting temperature calculation," Ph.D. thesis, California Institute of Technology (2015).
- J. Mei, J. W. Davenport, Free-energy calculations and the melting point of Al. *Phys. Rev. B Condens. Matter* **46**, 21–25 (1992).
- J. R. Morris, C. Z. Wang, K. M. Ho, C. T. Chan, Melting line of aluminum from simulations of coexisting phases. *Phys. Rev. B Condens. Matter* **49**, 3109–3115 (1994).
- A. B. Belonoshko, N. V. Skorodumova, A. Rosengren, B. Johansson, Melting and critical superheating. *Phys. Rev. B Condens. Matter Mater. Phys.* **73**, 1–3 (2006).
- D. Alfè, C. Cazorla, M. J. Gillan, The kinetics of homogeneous melting beyond the limit of superheating. *J. Chem. Phys.* **135**, 024102 (2011).
- O. Sugino, R. Car, Ab initio molecular dynamics study of first-order phase transitions: melting of silicon. *Phys. Rev. Lett.* **74**, 1823–1826 (1995).
- G. de Wijs, G. Kresse, M. Gillan, First-order phase transitions by first-principles free-energy calculations: The melting of Al. *Phys. Rev. B* **57**, 8223 (1998).
- B. Widom, Potential-distribution theory and the statistical mechanics of fluids. *J. Phys. Chem.* **86**, 869–872 (1982).
- Q. J. Hong, A. van de Walle, Direct first-principles chemical potential calculations of liquids. *J. Chem. Phys.* **137**, 094114 (2012).
- S. T. Lin, M. Blanco, W. A. Goddard, The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: Validation for the phase diagram of Lennard-Jones fluids. *J. Chem. Phys.* **119**, 11792–11805 (2003).
- Q.-J. Hong, A. van de Walle, Solid-liquid coexistence in small systems: A statistical method to calculate melting temperatures. *J. Chem. Phys.* **139**, 094114 (2013).
- Q.-J. Hong, A. van de Walle, A user guide for SLUSCHI: Solid and liquid in ultra small coexistence with hovering interfaces. *Calphad: Comput. Coupling Ph. Diagr. Thermochem.* **52**, 88–97 (2016).
- A. M. Krajewski, J. W. Siegel, J. Xu, Z.-K. Liu, Extensible structure-informed prediction of formation energy with improved accuracy and usability employing neural networks. *Comput. Mater. Sci.* **208**, 111254 (2022).
- D. Jha *et al.*, ElemNet: Deep learning the chemistry of materials from only elemental composition. *Sci. Rep.* **8**, 17593 (2018).
- V. P. Glushko, Ed., *Thermodynamic Properties of Individual Substances, Vol. 1–10* (VINITI, Moscow, 1965–1982).
- F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model. *IEEE Trans. Neural Netw.* **20**, 61–80 (2009).
- K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition" in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.
- M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning" in *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation (USENIX Association, Savannah, GA, USA, 2016)*, pp. 265–283.
- Q. J. Hong, Melting temperature predictor based on machine learning. <https://faculty.engineering.asu.edu/hong/melting-temperature-predictor/>. Accessed 6 February 2022.
- International Mineralogical Association database of mineral properties. <https://rruff.info/ima/>. Accessed 6 February 2022.
- R. M. Hazen *et al.*, Mineral evolution. *Am. Mineral.* **93**, 1693–1720 (2008).
- S. V. Krivovichev, V. G. Krivovichev, R. M. Hazen, Structural and chemical complexity of minerals: correlations and time evolution. *Eur. J. Mineral.* **30**, 231–236 (2018).
- J. J. Golden, "Mineral evolution database: Data model for mineral age associations," M.S. thesis, University of Arizona, Tucson, AZ (2019).
- J. J. Golden, R. T. Downs, R. M. Hazen, A. J. Pires, J. Rlph, "Data-driven age assignment, how does a mineral get an age?" in *Abstracts of Geological Society of America (GSA) meeting* (Phoenix, Arizona, 2019).
- R. M. Hazen *et al.*, Data-driven discovery in mineralogy: Recent advances in data resources, analysis, and visualization. *Engineering (Beijing, China)* **5**, 397–405 (2019).
- A. Prabhu *et al.*, Global earth mineral inventory: A data legacy. *Geosci. Data J.* **8**, 74–89 (2021).
- P. Boehnke, T. M. Harrison, Illusory late heavy bombardments. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10802–10806 (2016).
- A. Mann, Bashing holes in the tale of Earth's troubled youth. *Nature* **553**, 393–395 (2018).
- H. Tang, Y. Chen, Global glaciations and atmospheric change at ca. 2.3 Ga. *Geoscience Frontiers* **4**, 583–596 (2013).
- C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
- M. Pfleger, T. Wallek, A. Pfennig, Constraints of compound systems: Prerequisites for thermodynamic modeling based on Shannon entropy. *Entropy (Basel)* **16**, 2990–3008 (2014).
- V. Vanchurin, Y. I. Wolf, E. V. Koonin, M. I. Katsnelson, Thermodynamics of evolution and the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2120042119 (2022).
- S. Krivovichev, Topological complexity of crystal structures: Quantitative approach. *Acta Crystallogr. A* **68**, 393–398 (2012).
- S. V. Krivovichev, Structural complexity and configurational entropy of crystals. *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* **72**, 274–276 (2016).