COMMUNICATION

Check for updates

A data-driven approach for the modeling of a ball-milled dispersion of BaTiO₃ nanoparticles

Takumi Ono,[a] Tarojiro Matsumura,[b] Kiwamu Sue[a] and Satoru Takeshita [ID] *[a]

Wet-state ball milling of ceramic nanoparticles is analyzed by machine learning and machine-learning-assisted model formulation. A linear model formula is constructed from the high-impact input features revealed in the machine learning. The formula explains the relation between the ball-milling conditions and hydrodynamic size with less precision but better analytical processability compared to the original machine learning.

New concepts

The current work provides a pioneering study on how to adapt machine learning to wet-state ball milling of ceramic nanoparticles. While machine learning has been becoming commoditized in chemistry fields, its less analytical nature has remained as a major issue, particularly in unsophisticated, industry-oriented processes including ball milling and nanoparticle dispersion–aggregation control. Several attempts have recently been reported regarding the use of machine learning in ball milling, but they have been limited in the range of simple regression or classification without analytical insights. We would like to take these attempts one step forward: utilizing machine learning to assist analytical model formulation. A linear formula is constructed based on the importance of the input ball-milling parameters as revealed by machine learning. The obtained formula explains the relation between the ball-milling conditions and hydrodynamic size with better analytical processability compared to the original machine learning. The current approach will open up a new methodology of connecting data-driven approaches and analytical modeling for the potential production of a wide range of ceramic devices.

Introduction

After the recent commoditization of AI-related technologies, machine learning has become a major tool in chemistry.[1] Machine-learning methodology can represent mathematical relationships between inputs and outputs even when underlying phenomena are unclear. Such methodology can also evaluate the contributions of input features to predicted outputs, whereas it is not good at making models for actual phenomena due to its less analytical nature. Applications of machine learning are expected to provide new insights from experimental data as a data-driven approach. Successful data-driven approaches in chemistry have focused on issues with large databases and specific features, such as the exploration of new material compositions,[2] catalyst optimization,[3] macromolecular structure prediction,[4] and property regression/prediction of functional materials.[5] Besides, industrial processes have many unclear "dirty" features, for which many challenges still remain in order to introduce a data-driven approach. In particular, the manufacturing of ceramic devices is full of such know-how-based, unsophisticated features.

Ball milling is a popular process in solid device fabrication in both industry and academia. Impulsive forces of impact with balls are utilized for many purposes: physical mixing, physical pulverization and downsizing,[6] disintegration of aggregates/agglomerates,[7] mechanical alloying,[8] and various mechanochemical reactions.[9] In the production of ceramic devices, wet-state ball milling is commonly used to prepare well-dispersed colloids of ingredient nanoparticles before molding, coating, or sintering treatments.[10] Several theoretical models have been established for aggregation–dispersion of colloidal nanoparticles focusing on aqueous systems with electrostatic repulsion, *e.g.*, DLVO theory.[11] On the other hand, nonaqueous systems with steric repulsion, which are typical in practical ball-milling processes, still lack general theoretical models. The wide variety of industrial processes also makes general modeling difficult, where we suggest that a data-driven approach has certain room for contribution.

In fact, several reports have applied machine learning and statistical methods to ball milling[12] and aggregation–dispersion of particles,[13] but they have been limited to simple regression or classification without analytical insights. In this communication, we would like to contribute a pioneering study

[a] *Research Institute for Chemical Process Technology, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 5, 1-1-1 Higashi, 3058565 Tsukuba, Japan. E-mail: s.takeshita@aist.go.jp*

[b] *Materials DX Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 2, 1-1-1 Umezono, 3058568 Tsukuba, Japan*

on how to adapt machine learning to hydrodynamic size data of ball-milled ceramic nanoparticles. To take this work one step forward from a simple case study, we suggest an additional role of machine learning as an assistant in instant model formulation. Our approach consists of three simple steps: First, machine-learning regression is employed on the hydrodynamic size dataset. Second, some data-processing tools visualize the influence of input variables on the machine-learning prediction. Third, a linear formula is constructed based on the important inputs revealed in the data processing. The comparison between machine learning and formulation reveals their characteristics and applicability.

Experimental

The target experimental system is aggregation–dispersion control of $BaTiO_3$ nanoparticles, a typical ingredient of dielectric devices,[14] through wet-state ball milling (detailed experimental procedures are given in the SI). A conventional nonionic polyalkylene-glycol-derived dispersant[15] and alcohol solvents with different alkyl chains were selected. We varied several ball-milling conditions (Fig. 1). The input variables of machine learning were selected from the independent experimental conditions avoiding multicollinearity (details in Section S2 in the SI).

The output targets were the particle sizes of ball-milled dispersions measured by dynamic light scattering (DLS).[16] Two types of hydrodynamic size were used as machine-learning targets (see Fig. 1, bottom and Fig. S1 for details): (i) Cumulant diameter, a typical size from DLS under an assumption of monomodal normal size distribution.[17] Samples having multimodality tend to show large polydispersity indices. (ii) Median diameter, d_{50}, from integrated size distribution profiles obtained by the histogram method.[18] This method reflects

multimodality but is more sensitive to noise. The distribution width of the histogram was evaluated using SPAN, $(d_{90} - d_{10})/d_{50}$. Different bases were used for the vertical axis: scattering intensity-, volume-, and number-bases (see Fig. S2 for a typical example of an intensity-based profile).

Results and discussion

Among 2570 combinations of the input variables (the dataset csv file is in the SI), the minimum cumulant diameter and intensity-based d_{50} are ~ 35 nm. This value is close enough to single-nanoparticle dispersion, considering the intrinsic difference between the hydrodynamic and physical sizes of the primary nanoparticles (~ 20 nm in the number average from TEM, Fig. S3). The input-feature dependences of these hydrodynamic sizes are summarized in Fig. S4 and S5. There is no clear dependence on one single input variable; the sizes are determined by multiple contributions of the inputs. The relation between the cumulant diameter and intensity-based d_{50} (Fig. S6) shows a small gap at the size of $> \sim 10$ μm in d_{50}, which is almost out of the measurable range of DLS, showing the multimodality in some samples and a large influence of noise in the histogram method.

We used random forest for machine-learning regression of the output hydrodynamic sizes.[19] The obtained results targeting the cumulant diameter and polydispersity index (Fig. 2a) show the best predictions ($R^2 = 0.913$ and 0.841, respectively, Table S3 in the SI) compared to those on d_{50} and SPANs (Fig. 2b). Predictions of volume- and number-based d_{50} are particularly not successful mainly because of large noise in the $< \sim 1$–10 nm region. We therefore focus on the cumulant diameter and intensity-based d_{50} for further analysis. As a supplementary calculation, we also conducted a different machine-learning method, XGBoost, to compare with the random forest prediction (details in Section S2 and Fig. S9 in the SI).

Random forest has the advantage of being easy to connect with data processing modules that can interpret feature importances: the influences of input features on prediction precision (details and additional evaluations are in Section S2 in the SI). According to Gini importances in Fig. 2, milling ball size, dispersant mass, milling ball mass, solvent type (represented by its viscosity in the calculation), and total milling time have certain influences on the predicted hydrodynamic sizes and polydispersity/SPANs, whereas the nominal $BaTiO_3$ particle mass does not have a large influence on these items.

For more interpretability, SHapley Additive exPlanations (SHAP)[20] was employed on the cumulant diameter prediction (Fig. 3). This method visualizes both importances and their directions of input features. The horizontal axis in SHAP corresponds to the impact of input items on target values (negative-to-positive from left-to-right) and the plot color of red/blue corresponds to large/small input values. In the present case, high-impact items in SHAP are roughly consistent with those in Gini importance. Milling ball size is the most important with red plots on the left, meaning that large balls decrease
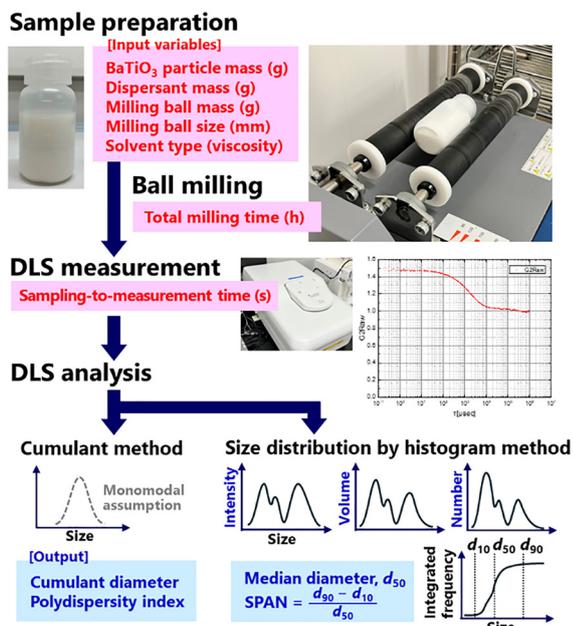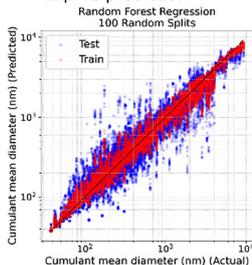

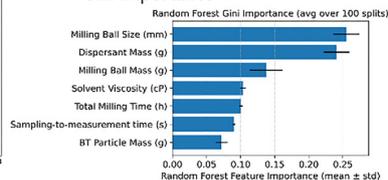**Fig. 1** Input variables and output hydrodynamic sizes.
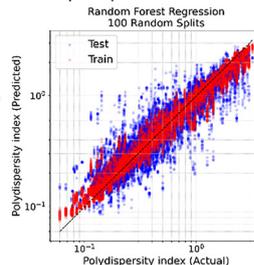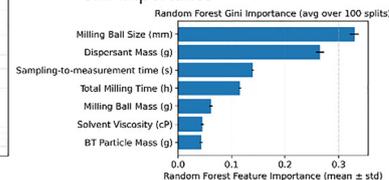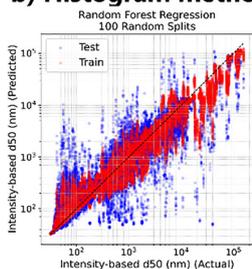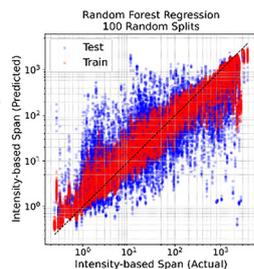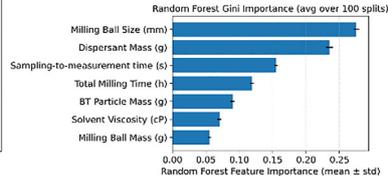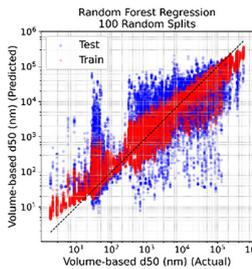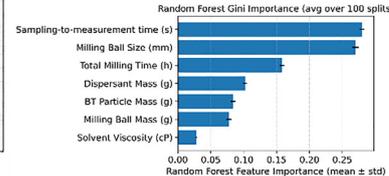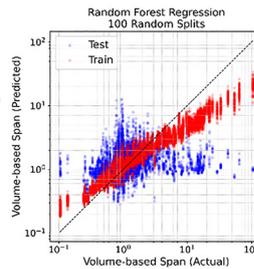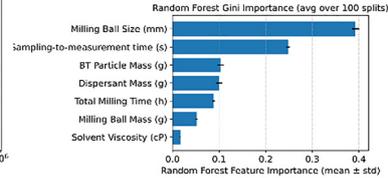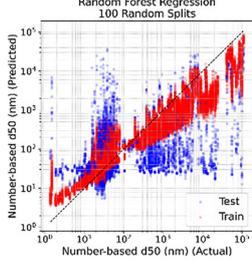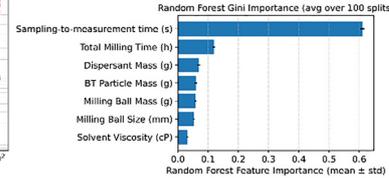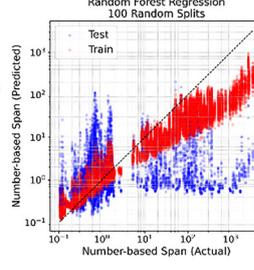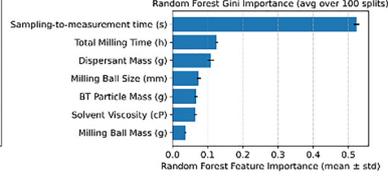
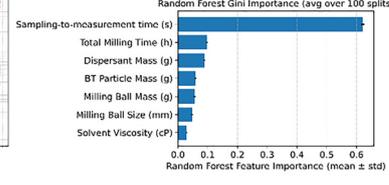**Fig. 2** Machine-learning regressions and their Gini importances using random forest targeting on (a) cumulant diameter and polydispersity index and (b) d_{50} and SPAN by the histogram method.
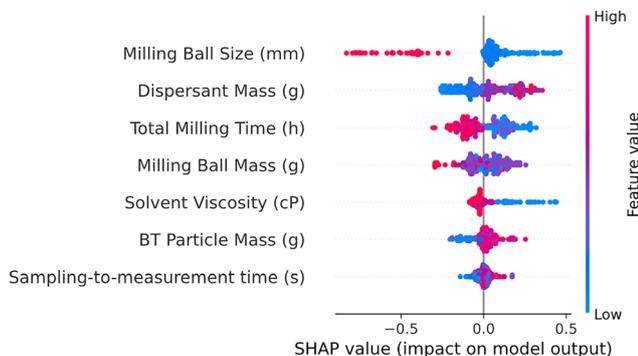

Fig. 3 SHAP analysis on cumulant diameter.

the predicted cumulant diameter. Dispersant mass has a complex influence. Blue plots on the left show that a small dispersant amount makes the cumulant diameter small, whereas its influence is not fully monotonic as shown in the red/blue mixture on the right. Total milling time is roughly monotonic. A certain time is needed to make the diameter small, but overtreatment makes it large probably by breaking the crystals to form aggregates again (this will be discussed with XRD data later). Milling ball mass is not very straightforward with a weak negative tendency; a large mass makes the diameter small. Solvent viscosity has a monotonic influence; small-viscosity solvents make the diameter large. BaTiO$_3$ mass and sampling-to-measurement time have complex and relatively small influences compared to the above items.

Even though the above importance-oriented data processing can increase the interpretability, the weak point of machine

learning remains; regressions are nonlinear, less analytical, and hard to use to make a single approximation formula explaining dependences of input features. We suggest constructing a linear formula based on the feature importances of machine learning. The high-impact features are picked up in the order they appear in the SHAP, *i.e.*, milling ball size, dispersant mass, ... and simply multiplied in the following eqn (1) with exponent constants:

$$\phi = d_0 \left(d_{\text{ball}}^*\right)^a \left(m_{\text{dispersant}}^*\right)^b \left(t_{\text{total}}^*\right)^c \left(m_{\text{ball}}^*\right)^d$$
$$\times \left(\mu_{\text{solvent}}^*\right)^e \left(m_{\text{BT}}^*\right)^f \left(t_{\text{sampling}}^*\right)^g \quad (1)$$

where $d_{\text{ball}}^*, m_{\text{dispersant}}^*, t_{\text{total}}^*, m_{\text{ball}}^*, \mu_{\text{solvent}}^*, m_{\text{BT}}^*,$ and t_{sampling}^* are the milling ball size, dispersant mass, total milling time, milling ball mass, solvent viscosity, $BaTiO_3$ mass, and sampling-to-measurement time, respectively. Superscript $\ast$ means that each item is unitless through dividing by the respective unit, d_0 is a constant, a–g are the exponent constants, and ϕ is the predicted cumulant diameter by this formulation. The equation is converted into a linear polynomial for simplification:

$$\log(\phi) = \log(d_0) + a\log\left(d_{\text{ball}}^*\right) + b\log\left(m_{\text{dispersant}}^*\right)$$
$$+ c\log\left(t_{\text{total}}^*\right) + d\log\left(m_{\text{ball}}^*\right) + e\log\left(\mu_{\text{solvent}}^*\right) \quad (2)$$
$$+ f\log\left(m_{\text{BT}}^*\right) + g\log\left(t_{\text{sampling}}^*\right)$$

Approximation formulae are constructed by adding the terms in eqn (2) one by one, *i.e.*, the first approximation consists of $\log(d_0) + a\log(d_{\text{ball}}^*)$, the second one adds $b\log\left(m_{\text{dispersant}}^*\right)$ to the first one,..., and displayed in Fig. 4 with prediction R^2 values to the actual cumulant diameter. The first approximation cannot fit the actual diameter at all with $R^2 = 0.196$. The fitting gradually improves from the second to third approximations with increasing R^2 from 0.529 to 0.592. Then, the fourth and fifth approximations increase the fitting up to $R^2 = 0.709$. After that, the sixth and seventh approximations do not drastically improve the fitting. Thus, the cumulant diameter can be predicted roughly by a linear connection of the 1st–3rd important variables then more precisely by that of the 1st–5th ones. Fig. 4 also shows a tendency of polydispersity index (the plot color of prediction–actual diameter diagrams). The yellowish plots around the upper right in the diagrams indicate that some conditions that give smaller predicted diameters have large polydispersity.

For further comparison, we also tried screening 35 different combinations of 3 input variables without considering the feature importance order, *e.g.*, 1st + 2nd + 7th terms. As shown in Fig. S10, the R^2 values varied from ~0.6 to ~0.08 depending on the combinations. While the original combination in Fig. 4, 1st + 2nd + 3rd terms ($R^2 = 0.592$), is not the best fitting, it still ranks among the top three with an identical R^2 range of ~0.6. This result means that the feature-importance-assisted modeling is a fast and convenient way to avoid time-consuming comprehensive trials and errors of input variable selections.

The exponent constants, a–g (the bar graphs in Fig. 4), show their influence directions in the formula; those of the 1st–5th terms are negative, positive, negative, negative, and negative. The SHAP shows the same trend, *i.e.*, increases in milling ball size (1st), total milling time (3rd), milling ball mass (4th), and solvent viscosity (5th) roughly decrease the cumulant diameter, while that in dispersant mass (2nd) roughly increases the diameter. A decisive characteristic of the model formula is its linearity. The input features in the random forest have certain degrees of correlations (Fig. S7), *e.g.*, solvent viscosity and dispersant mass have a correlation coefficient of -0.4. On the other hand, the formula assumes perfectly independent input variables as it is a simple linear polynomial. This nature would decrease the prediction precision of the formula ($R^2 = 0.711$ at the maximum) compared to the original random forest ($R^2 = 0.913$). Also, each term in the formula describes only a monotonic increase or decrease. This item would not be a drastic disadvantage in the current case, since the 1st–5th items in the SHAP show roughly monotonic behaviors.

To discuss the applicability of the linear model formulation not only for the current ball-milling system but also for other potential systems, we note its advantages and limitations. First, a linear formula would be useful for analytical investigation, *e.g.*, grasping a rough quantitative influence of each input variable with the exponent constant. This would be beneficial for on-site modulation/tuning of input conditions in actual lab/industrial processes. In fact, several previous studies, including ball milling of alumina particles,[21] have already used linear formulae to describe industrial-oriented phenomena. The current methodology would help construct such linear models without comprehensive trials and errors of input variable selections.

On the other hand, the linear formula has two drawbacks: the absence of physical meaning and less precision due to the lack of nonlinear components. Regarding the former item, the formula reflects the correlation of inputs–outputs in the dataset, but does not provide causation or explanation for physical phenomena behind the data. For example, a theoretical description of ball milling needs more physicochemical information, such as microscopic solvent–dispersant–nanoparticle interactions, specific motion trajectory of the balls, and time dependence of rheological indices. Besides, industrial phenomena often lack physical models. Researchers usually start with a linear model in such cases thanks to the simplicity of linear formulae. Although the model does not contain physical meanings, it could be an initial model to derive an empirical formula for the phenomenon.

Regarding the latter item, the current machine-learning-assisted methodology is not limited to linear formulation. After understanding input-item dependences with a linear formula, we can then expand/modify the formula to make it nonlinear. Actually, we already started a preliminary trial for nonlinear modeling (Fig. S11), where the fitting precision was slightly better than that in the linear formula. Since the possible forms of nonlinear formulae have infinite variations, we suggest that a meaningful nonlinear model should be accompanied by some physical justification based on the above-mentioned physicochemical information.

For the final section, we also point out more practical items regarding the ball-milling process. Impact with balls can

Fig. 4 Linear model formulation of cumulant diameter prediction using high-impact features in SHAP analysis.

disintegrate the aggregates but also introduce some defects into the $BaTiO_3$ crystals. We measured the crystallite size of (110) as an indicator of crystallinity change through ball milling. Visualization of the crystallite size in the machine-learning prediction diagram (Fig. S14 left) unveils that the samples with both smaller and larger cumulant diameters tend to show smaller crystallites. The former means that certain damage occurs in the crystals as a result of the ball milling. The latter indicates that the damaged particles would make aggregates again through overtreatment.

In the next practical step, we need an optimization of the particle size and crystallinity for subsequent processes, e.g., forming, coating, and/or sintering, to design the production process of the desired ceramic devices.

Conclusions

We suggested a new methodology utilizing machine learning for instant model formulation of ball-milled ceramic nanoparticles.

After making the hydrodynamic size prediction with random forest, feature importance and SHAP analyses were employed on the prediction to extract the important input features. Then, the high-impact features revealed in the SHAP analysis were utilized for constructing a rough, linear model formula. The obtained formula explained the input-feature dependence of the hydrodynamic size with better interpretability and less precision compared to the original machine learning.

As concluding remarks, we suggest two potential future directions and their challenges toward a wide range of industrial issues: (i) machine-learning-assisted empirical formulae, typically starting with a linear formula but not limited to this, for various industrial phenomena. This direction would require microscopic physicochemical analyses to connect the formulae with actual physical phenomena. (ii) Closed-loop frameworks of experiments and computational predictions for faster discoveries of materials/processes. This direction would require automated experimental/characterization systems, which are becoming a hot topic in materials chemistry.[22] For both directions, we should continue case studies connecting machine learning and analytical formulation to expand the applicable range of the data-driven approach toward a wide range of industrial issues.

Author contributions

T. Ono: data curation, formal analysis, investigation, and writing – review and editing. T. Matsumura: validation and writing – review and editing. K. Sue: conceptualization, data curation, project administration, and writing – review and editing. S. Takeshita: conceptualization, formal analysis, visualization, supervision, and writing – original draft.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article, including the dataset of input conditions–output DLS data, are available as part of the supplementary information (SI) (dataset_MAIN_v02.csv). Supplementary information: experimental details and additional data, machine-learning codes, and a dataset for the ball-milling conditions and hydrodynamic size. See DOI: https://doi.org/10.1039/d5mh02452h.

Acknowledgements

Notes and references

1 (a) Z. J. Baum, X. Yu, P. Y. Ayala, Y. Zhao, S. P. Watkins and Q. Zhou, *J. Chem. Inf. Model.*, 2021, **61**, 3197; (b) C. M. Biship, *Pattern recognition and machine learning (information science and statistics)*, Springer, 2007; (c) S. Raschka and V. Mirjalili, *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*, Packt publishing Ltd., 2019.

2 (a) P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73; (b) A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80; (c) N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng and G. Ceder, *Nature*, 2023, **624**, 86.

3 (a) Y. Bai, L. Wilbraham, B. J. Slater, M. A. Zwijnenburg, R. S. Sprick and A. I. Cooper, *J. Am. Chem. Soc.*, 2019, **141**, 9063; (b) K. Takahashi, L. Takahashi, S. D. Le, T. Kinoshita, S. Nishimura and J. Ohyama, *J. Am. Chem. Soc.*, 2022, **144**, 15735.

4 (a) A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu and D. Hassabis, *Nature*, 2020, **577**, 706; (b) J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583.

5 (a) R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D.-G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams and A. Aspuru-Guzik, *Nat. Mater.*, 2016, **15**, 1120; (b) T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301; (c) S. Takeshita, S. Mine and T. Ono, *Angew. Chem., Int. Ed.*, 2025, **64**, e202504250.

6 (a) J. D. Fougere, M. Lynch, J. Zhao, Y. Zheng and K. Li, *Energy Fuels*, 2014, **28**, 2645; (b) L. Protesescu, S. Yakunin, O. Nazarenko, D. N. Dirin and M. V. Kovalenko, *ACS Appl. Nano Mater.*, 2018, **1**, 1300; (c) Z. Yang, M. Jadidi, K. Chen, S. Zhou, Y. Zou, A. Dolatabadi and T. W. Coyle, *J. Eur. Ceram. Soc.*, 2026, **46**, 117828.

7 (a) F. Müller, W. Peukert, R. Polke and F. Stenger, *Int. J. Miner. Process.*, 2004, **74**, S31; (b) T. Ogi, R. Zulhijah, T. Iwaki and K. Okuyama, *Kona Powder Part J.*, 2017, **34**, 3; (c) B. H. Watson III, J. Schliesser, C. D. Eadie, M. Michie,

M. A. Fanton and R. J. Meyer Jr., *J. Appl. Ceram. Technol.*, 2025, **22**, e70013.

8 (*a*) C. C. Koch, O. B. Cavin, C. G. McKamey and J. O. Scarbrough, *Appl. Phys. Lett.*, 1983, **43**, 1017; (*b*) L. L. Hsiung, M. J. Fluss, S. J. Tumey, B. W. Choi, Y. Serruys, F. Willaime and A. Kimura, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **82**, 184103.

9 (*a*) S. Hutsch, A. Leonard, S. Grätz, M. V. Höfler, T. Gutmann and L. Borchardt, *Angew. Chem., Int. Ed.*, 2024, **63**, e202403649; (*b*) O. F. Jafter, S. Lee, J. Park, C. Cabanetos and D. Lungerich, *Angew. Chem., Int. Ed.*, 2024, **63**, e202409731; (*c*) S. Suginome, K. Murota, A. Yamamoto, H. Yoshida and Y. Nishibayashi, *Nat. Synth.*, 2025, **4**, 243.

10 (*a*) I. U. Idehenre, Y. A. Barnakov, S. A. Basun and D. R. Evans, *J. Appl. Phys.*, 2018, **124**, 165501; (*b*) J. Lee, J.-H. Ha, H. J. Lee, I.-H. Song and S. Z. A. Bukhari, *J. Eur. Ceram. Soc.*, 2025, **45**, 117523.

11 J. N. Israelachvili, Chapter 14 – Electrostatic Forces between Surfaces in Liquids, in *Intermolecular and Surface Forces*, Academic Press, 3rd edn, 2011, pp. 291–340.

12 (*a*) A. Tohry, S. C. Chelgani, S. S. Matin and M. Noormohammadi, *Adv. Powder Technol.*, 2020, **31**, 967; (*b*) Y. Li, J. Bao, T. Chen, A. Yu and R. Yang, *Powder Technol.*, 2022, **403**, 117409.

13 (*a*) C. Glaubitz, B. Rothen-Rutishauser, M. Lattuada, S. Balog and A. Petri-Fink, *Nanoscale*, 2022, **14**, 12940; (*b*) F. Torabi and S. G. Sabouri, *Measurement*, 2025, **256**, 118586.

14 M. Acosta, N. Novak, V. Rojas, S. Patel, R. Vaish, J. Koruza, G. A. Rossetti Jr and J. Rödel, *Appl. Phys. Lett.*, 2017, **4**, 041305.

15 Y. Liao, S. Takezawa, H. Takeda, T. Tsurumi and T. Hoshina, *Jpn. J. Appl. Phys.*, 2019, **58**, SLLA07.

16 (*a*) D. Sidebottom, Dynamic Light Scattering, in *Characterization of Materials*, ed. E. N. Kaufmann, John Wiley & Sons, Inc., 2012, DOI: 10.1002/0471266965.com111; (*b*) J. Stetefeld, S. A. McKenna and T. R. Patel, *Biophys. Rev.*, 2016, **8**, 409.

17 (*a*) D. E. Koppel, *J. Chem. Phys.*, 1972, **57**, 4814; (*b*) B. J. Frisken, *Appl. Opt.*, 2001, **40**, 4087.

18 (*a*) M. X. J. J. Fernandes, N. C. Santos and M. A. R. B. Castanho, *J. Biochem. Biophys. Methods*, 1998, **36**, 101; (*b*) A. Scotti, W. Liu, J. S. Hyatt, E. S. Herman, H. S. Choi, J. W. Kim, L. A. Lyon, U. Gasser and A. Fernandez-Nieves, *J. Chem. Phys.*, 2015, **142**, 234905.

19 (*a*) L. Breiman, *Mach. Learn.*, 2001, **45**, 5; (*b*) R. Genuer, J.-M. Poggi and C. Tuleau-Malot, *Pattern Recognit. Lett.*, 2010, **31**, 2225; (*c*) G. Biau, *J. Mach. Learn. Res.*, 2012, **13**, 1063.

20 S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim and S.-I. Lee, *Nat. Biomed. Eng.*, 2018, **2**, 749.

21 J. Yu, S.-H. Jin, K. Raju, Y. Lee and H.-K. Lee, *Ceram. Int.*, 2021, **47**, 31202.

22 T. Dai, S. Vijayakrishnan, F. T. Szczypiński, J.-F. Ayme, E. Simaei, T. Fellowes, R. Clowes, L. Kotopanov, C. E. Shields, Z. Zhou, J. W. Ward and A. I. Cooper, *Nature*, 2024, **635**, 890.